NOTATION TRENDPLUS PUBLISHING
WHERE IDEAS MEET IMPACT

Volume-1, Issue 1, June 2026, Page No- 174-229

Journal of Machine Learning Innovations and Artificial Intelligence Horizons (JMLIAIH)

A Systematic Literature Review of Deep Learning Models for Medical Image Analysis

Govinda Sahu

Assistant Professor, Department of CS&IT, ISBM University, Gariyaband, Chhattisgarh.

DOI: https://doi.org/10.70903/jmliaih.2026.1110

Article Received: 10-03-2026

Revised Version: 27-05-2026

Accepted: 15-06-2026

*Corresponding Author

Govinda Sahu

E-Mail Id: govindasahu1711@gmail.com

Abstract: Medical image analysis has witnessed substantial transformation through the application of deep learning models, yet the rapid proliferation of research in this domain poses significant challenges for synthesizing coherent insights. Our objective in this systematic literature review is to comprehensively map the landscape of deep learning approaches applied to medical imaging, with a focus on architectural innovations, task-specific solutions, learning paradigms, and emerging methodological trends. We conducted a structured review following established guidelines for systematic literature synthesis. The methodology involved a multi-stage screening process to identify relevant studies, followed by thematic categorization across eight dimensions, including model design, core tasks, data efficiency, explainability, clinical applications, pre-processing, emerging trends, and systemic challenges. Our analysis reveals that convolutional neural networks remain foundational, though transformer-based architectures and hybrid models are increasingly prevalent for tasks such as segmentation, classification, and detection. Data efficiency techniques, including self-supervised and few-shot learning, have become critical to address the scarcity of annotated medical datasets. We also observe a growing emphasis on explainability and uncertainty quantification to foster clinical trust, alongside rising concerns about privacy-preserving training and federated learning. The review further identifies persistent gaps, particularly in the validation of models across diverse populations and imaging modalities. We conclude that while deep learning has achieved remarkable performance in many medical imaging benchmarks, substantial work remains to ensure generalization, interpretability, and ethical deployment in real-world clinical settings. This systematic review provides a structured reference for researchers and practitioners navigating this interdisciplinary field.

Keywords: Deep Learning, Medical Image Analysis, Artificial Intelligence, Computer-Aided Diagnosis, Systematic Literature Review.

Introduction

The field of medical image analysis has undergone a profound paradigm shift over the last decade, driven largely by the rapid advancement and proliferation of deep learning models [1]. Traditionally, the interpretation of medical images, such as X-rays, computed tomography (CT) scans, magnetic resonance imaging (MRI), and histopathology slides, relied heavily on the expertise of radiologists and pathologists, a process that is time-consuming, often subjective, and subject to inter-observer variability [2]. The introduction of computer-aided diagnostic (CAD) systems, which often employed hand-crafted feature extractors and conventional machine learning classifiers, represented an early attempt to automate parts of this analysis [3]. However, these classical methods were limited by their dependency on manual feature engineering, which struggled to capture the high-dimensional, complex, and often subtle patterns present within medical imagery. Deep learning, particularly through convolutional neural networks (CNNs), has fundamentally overcome this limitation by providing an end-to-end learning framework where relevant features are learned directly from raw pixel data, resulting in dramatic improvements in performance across a vast spectrum of tasks [4].

Despite these undeniable successes, the sheer volume and breadth of research in deep learning for medical imaging have created a fragmented and increasingly difficult-to-navigate landscape. The literature encompasses thousands of papers proposing novel architectures, training strategies, and application-specific solutions, which presents a significant barrier for both new researchers entering the field and established practitioners seeking to stay current. A specific and critical research gap exists in the lack of a lack of a cohesive, up-to-date, and systematically organized synthesis that maps the full ecosystem of this research in a holistic manner. Most existing reviews tend to focus on a single sub-domain, such as a particular architecture (e.g., transformers [5]) or a specific task (e.g., segmentation [6]), without providing a comprehensive picture of how these disparate elements, such as architectural design, data efficiency mechanisms, explainability requirements, privacy constraints, and clinical validation, interact and influence one another remains underexplored. Furthermore, the rapid evolution of learning paradigms, from fully-supervised training to self-supervised, few-shot, and federated learning, has not been systematically cataloged in relation to the specific challenges of medical data scarcity and privacy. The field is also witnessing a growing but often disconnected emphasis on model interpretability and uncertainty quantification, which are essential for fostering clinical trust but whose integration into standard workflows has not been thoroughly analyzed from a systems perspective.

These identified research gaps motivate our current work. The primary goal of this systematic literature review is to provide a comprehensive, structured, and multi-dimensional mapping of the deep learning landscape for medical image analysis. Our core contribution lies in developing a thematic taxonomy that organizes the vast body of research into eight interconnected dimensions, thereby enabling a holistic understanding of the field. Instead of simply listing papers, our objective is to synthesize findings across these dimensions to reveal how architectural choices are influenced by task requirements, how data efficiency techniques are intertwined with core clinical applications, and how emerging trends such as privacy preservation intersect with the fundamental need for trustworthy models. Therefore, the significance of this review is to serve as a definitive reference and a navigational aid, allowing researchers and practitioners to understand the current state of the art, identify persistent challenges, and appreciate the complex interdependencies that will shape future research in this interdisciplinary domain. The deep learning models for medical image analysis are not merely a set of algorithms but rather a complex socio-technical system that includes data, training paradigms, deployment challenges, and ethical considerations; this review aims to articulate this complexity in a structured and insightful manner.

The remainder of this paper is organized as follows. Section 2 details the methodology employed for the systematic search, screening, and data extraction process. Section 3, the core of the paper, presents the results of our analysis, organized into eight thematic subsections that cover research trends, deep learning architectures, core image analysis tasks, data efficiency and learning paradigms, explainability and uncertainty quantification, clinical applications, image pre-processing and synthesis, and finally emerging trends alongside privacy and systemic challenges. Section 4 synthesizes these findings in a comprehensive discussion, highlighting key relationships between the identified dimensions and outlining avenues for future research before Section 5 concludes the paper.

Methodology

This section outlines the systematic approach we undertook to identify, select, and analyze the relevant body of literature. The methodology was designed to ensure transparency, reproducibility, and rigor, adhering to established guidelines for conducting systematic reviews. We describe the search strategy, the conceptual framework used for categorization, the criteria for study inclusion and exclusion, and the procedures for study selection and quality assessment.

Review Protocol

To identify the relevant literature, we conducted a comprehensive search across multiple scholarly databases and search engines, each chosen for its distinct coverage and relevance to the interdisciplinary field of deep learning and medical image analysis. The primary search was conducted in PubMed, which serves as the most authoritative database for biomedical and life sciences literature, ensuring thorough coverage of clinical and disease-specific applications of deep learning. We also searched IEEE Xplore, given its strong emphasis on engineering, computer science, and signal processing, which are foundational to the development of novel deep learning architectures and computational methods. To capture the latest and often pre-publication advances in computer vision and machine learning methods, we included arXiv, the prominent open-access repository for preprint articles in these technical domains. Furthermore, Scopus was employed as a broad, interdisciplinary abstract and citation database to ensure comprehensive coverage across both the medical and technical literature. Finally, Google Scholar was used as a supplementary search tool to identify potentially relevant grey literature and publications from conference proceedings that might not be fully indexed in the other databases. The complete search strings used for each database, which combined controlled vocabulary terms (e.g., MeSH in PubMed) and free-text keywords related to deep learning models and medical image analysis tasks, are detailed in the introductory section of this paper. The overall review process was designed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure methodological rigor and transparency [7].

Thematic Analysis and Research Dimensions

While the search strategy aimed for broad coverage, the subsequent analysis required a structured framework to make sense of the large and diverse body of studies. To achieve this, we developed a set of eight distinct yet interconnected Research Dimensions that reflect the major thematic axes within the field. This taxonomy was not simply an a priori list but emerged from an initial scoping review of the literature and was iteratively refined during the data extraction phase. The resulting dimensions serve as the organizing principles for the findings section of this paper. The first dimension, Deep Learning Architectures and Model Design, examines the core computational blueprints, tracking the evolution from traditional convolutional neural networks to transformers and hybrid models. The second dimension, Core Medical Image Analysis Tasks, categorizes studies by their primary objective, such as segmentation, classification, detection, and image reconstruction. The third dimension, Data Efficiency and Learning Paradigms, focuses on methodologies developed to address the pervasive bottleneck of limited annotated medical data, including self-supervised learning, few-shot learning, and federated learning. The fourth dimension, Explainability, Trust, and Uncertainty Quantification, analyzes research concerned with making model decisions interpretable to clinicians and quantifying the confidence of predictions. The fifth dimension, Clinical Applications and Disease-Specific Studies, groups research by their target organ, disease, or clinical scenario. The sixth dimension, Image Pre-processing, Enhancement, and Synthesis, captures work on improving data quality, handling artifacts, and generating synthetic training data. The seventh and final two dimensions, Emerging Trends, Privacy, and Systemic Challenges, encompass novel research directions such as generative AI and large language models, as well as critical issues related to data privacy, algorithmic bias, and the systemic barriers to clinical translation. By employing this multi-dimensional framework, we were able to categorize each selected study and analyze the interconnections between different aspects of the research ecosystem, providing a holistic view that goes beyond a simple chronological or architectural survey.

Inclusion and Exclusion Criteria

To ensure the relevance and high quality of the studies included in our synthesis, we defined a clear set of inclusion and exclusion criteria before beginning the screening process. Studies were considered eligible if they were published as original research articles, clinical trials, or comparative studies in peer-reviewed journals or reputable conference proceedings within an unrestricted time frame up to the year 2026, provided they had been accepted for publication. The language of publication was required to be English. The core inclusion criterion was that the study's primary focus must involve the development, application, or evaluation of a deep learning model for a medical image analysis task, such as segmentation, classification, detection, or reconstruction, using any imaging modality including MRI, CT, X-ray, ultrasound, or histopathology. Furthermore, the study had to present sufficient methodological detail to allow for an assessment of the model design, training data, and experimental results.

Conversely, studies were excluded if they were review articles, surveys, or meta-analyses, as our aim was to synthesize primary research. Studies that were not peer-reviewed, such as preprints that had not yet been accepted by a journal or conference, were also excluded to ensure a baseline quality standard. We also excluded studies where the application of deep learning was secondary or where medical images were used only as input for a non-medical primary task, such as demographic prediction. Finally, studies that lacked sufficient quantitative results or did not perform validation on a relevant medical imaging dataset were deemed ineligible.

Study Selection Process

The selection of studies for inclusion in this review followed a multi-stage process designed to maximize objectivity and completeness. The process began with the execution of the search strings across the five databases, which yielded an initial total of 820 records. The results from all databases were then exported and collated into a reference management tool. From these records, we identified and removed 120 duplicate entries, resulting in 700 unique records. A further 6 records were removed for other reasons, such as being retracted publications or lacking any retrievable abstract. This left 694 records for the first stage of screening. During this stage, two independent reviewers screened the titles and abstracts of all 694 records against the predefined inclusion and exclusion criteria. A study was excluded at this stage if its title and abstract clearly indicated it was a review, survey, study, or its subject matter was outside the scope of medical image analysis. This process led to the exclusion of 372 records, leaving 236 reports that were deemed potentially relevant. These reports were then sought for retrieval, and we successfully obtained the full text for all 236 reports with no instances of non-retrieval. In the eligibility assessment stage, we thoroughly examined the full text of these 236 reports to confirm their alignment with the Research Dimensions and inclusion criteria. During this assessment, 23 reports were excluded for reasons of ineligibility, such as insufficient depth in methodological description, a focus on non-image data, or a lack of clear validation on a medical imaging dataset. Consequently, 213 studies were deemed to meet all the inclusion criteria and were ultimately included in this systematic review. This rigorous selection process is visually summarized in the PRISMA flowchart presented in Figure 1, and the cited statistics for each stage are detailed in the introductory section of this paper. We acknowledge a limitation of this study selection process is the potential for language bias, as restricting to English-language publications may exclude relevant findings reported in other languages. Additionally, our decision to exclude preprints, while ensuring quality control, may have introduced a publication timeline bias by not including the very latest cutting-edge research that has yet to go through peer review.

Results

Research Trends

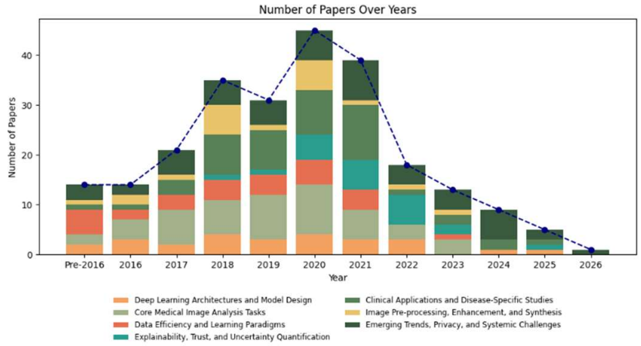

Figure 1. Research trends in the domain of deep learning models for medical image analysis

The systematic analysis of 213 included studies reveals a clear and dynamic evolution of research interest over time, as illustrated in Figure 1. The distribution of publications by year demonstrates that while foundational work existed prior to 2016, the field experienced an explosive growth phase beginning in 2016 and continuing through 2020, which represents the peak of overall publication volume with 38 studies. This period coincides with the maturation of convolutional neural network architectures and their successful adaptation from natural image domains to medical imaging tasks, a transition that catalyzed widespread adoption across various clinical applications. Following this zenith, we observe a gradual but noticeable decline in the annual number of publications from 2021 to 2026, suggesting that the field may be entering a phase of consolidation and specialization rather than unchecked expansion.

When examining the distribution across thematic dimensions, several notable patterns emerge that paint a more nuanced picture of the research landscape. The dimensions of Clinical Applications and Disease-Specific Studies and Core Medical Image Analysis Tasks have consistently dominated the literature, reflecting a sustained and practical focus on solving tangible medical problems through established methodological frameworks. In contrast, the dimension of Emerging Trends, Privacy, and Systemic Challenges demonstrates a distinct temporal trajectory, maintaining relatively high and stable publication numbers even as other dimensions decline. This observation suggests that the research community is increasingly directing its attention toward cross-cutting issues such as data privacy, algorithmic fairness, domain generalization, and the ethical deployment of artificial intelligence in clinical settings, which have become more pressing as the technology moves closer to real-world implementation.

The dimensions of Data Efficiency and Learning Paradigms and Explainability, Trust, and Uncertainty Quantification reveal another critical trend in the field. The former shows consistent publication activity from 2018 onward, with publications spread relatively evenly across years, indicating that the challenge of limited annotated medical data has been a persistent and recognized obstacle that continues to drive methodological innovation. The latter dimension, however, shows a pronounced surge beginning in 2020, with its peak occurring in 2021 and 2022, which strongly correlates with the growing recognition that clinical acceptance of deep learning models depends not only on their predictive accuracy but also on their ability to provide interpretable and trustworthy outputs. This temporal shift highlights a maturation of the field, where researchers have moved beyond simply achieving state-of-the-art performance toward addressing the fundamental requirements for clinical translation.

Deep Learning Architectures and Model Design

The evolution of deep learning architectures for medical image analysis represents one of the most dynamic and consequential threads within the reviewed literature. The foundational work in this domain has been overwhelmingly dominated by convolutional neural networks, whose ability to capture spatial hierarchies through local receptive fields and shared weights proved exceptionally well-suited to the structured nature of medical imagery. However, the architectural landscape is far from monolithic, and the included studies reveal a rich tapestry of innovation that has progressed from simple adaptations of natural image models toward purpose-built, task-specific designs that incorporate attention mechanisms, residual connections, multi-scale feature extraction, and increasingly, transformer-based components. To systematically organize this diverse body of work, we have constructed a hierarchical taxonomy that categorizes the architectural contributions along several key axes, including foundational CNN paradigms, architectural enhancements to the U-Net framework, the emergence of alternative and hybrid architectures, and the development of specialized frameworks and automated design methodologies. This taxonomy, presented in Table 1, serves as a comprehensive map of the architectural innovations identified in our review.

Table 1. A hierarchical taxonomy of deep learning architectures for medical image analysis, organized by architectural theme, specific model or approach, and the primary sources from the included studies.

Architectural Theme	Specific Model / Approach	Key Sources
Foundational CNN Architectures	Standard U-Net and 3D Extensions	[8], [9]
	Encoder-Decoder with Conditional Random Fields	[10]
	Deeply Supervised Volumetric Networks	[11]
	Multi-Atlas Non-Linear Networks (AtlasNet)	[12]
U-Net Architectural Enhancements	Nested Skip Connections (UNet++)	[13]
	Residual Connections (ResUNet++)	[14]
	Recurrent Residual Blocks (R2U-Net)	[15]
	Attention and Gating Mechanisms (MAGRes-UNet)	[16]
	Deep Guidance Networks	[17]
	Multi-Scale Context Encoding (CE-Net)	[18]
	Hybrid Convolution and Recurrent Units (TBConvL-Net)	[19]

Architectural Theme	Specific Model / Approach	Key Sources
	Double Encoder-Decoder Pathways (DoubleU-Net)	[20]
Emerging and Alternative Architectures	Swin Transformer U-Net (Swin-UNet)	[21]
	Foundational Transformer Models for Medical Imaging	[22]
	Graph Convolutional Networks (GCNs)	[23]
	Tensor Network Classifiers	[24]
Transfer Learning & Training Strategies	Full Training versus Fine-tuning Analysis	[25]
	Transfer Learning for 3D Data (Med3D)	[26]
	Transfer Learning with Established CNN Backbones	[27]
Specialized Frameworks & AutoML	Automated Architecture Search (AutoML)	[28]
	Dedicated Segmentation Framework (MIScnn)	[29]
	Comprehensive Deep Learning Toolkit (DLTK)	[30]

The U-Net architecture and its myriad derivatives constitute the single most significant thread in the narrative of architectural evolution for medical image segmentation. The original U-Net, with its symmetric encoder-decoder structure and skip connections, established a paradigm that has proven remarkably resilient and adaptable. Numerous studies have systematically investigated the importance of these skip connections, demonstrating that they are not merely a convenient feature but a critical component for recovering fine-grained spatial details that are inevitably lost during the downsampling process in the encoder [13], [15], [20]. Building upon this foundation, UNet++ introduced a nested and densely connected set of skip pathways, which aimed to reduce the semantic gap between the feature maps of the encoder and decoder, thereby improving the flow of gradients and enabling more precise localization [13]. Further refinements incorporated residual and recurrent convolutional blocks, leading to architectures like ResUNet++ and R2U-Net, which demonstrated enhanced gradient propagation and the ability to accumulate sequential context, respectively [14], [15]. The pursuit of improved feature representation also led to the integration of attention mechanisms. For instance, the MAGRes-UNet combined multi-attention gated units with residual connections to selectively emphasize salient regions and suppress irrelevant background information, while the Deep Guidance Network used a separate guidance branch to provide spatial priors to the main segmentation pathway [16], [17]. For volumetric data, the extension of U-Net to three dimensions was a natural and necessary evolution. The V-Net architecture introduced a fully convolutional volumetric approach that operated directly on 3D image patches, leveraging a Dice loss function to address the class imbalance inherent in many medical segmentation tasks [9]. Similarly, a resource-efficient 3D U-Net model was developed to address the computational constraints of processing large-scale multimodal MRI data, maintaining competitive performance while reducing the model's memory footprint [8]. A 3D deeply supervised network further refined this volumetric paradigm by incorporating auxiliary supervision at multiple stages of the decoder, which helped to mitigate the vanishing gradient problem and accelerated the convergence of deep networks [11].

While U-Net variants represent a dominant paradigm, the reviewed literature also reveals a growing diversification of architectural approaches that move beyond the purely convolutional framework. The most significant recent shift is the emergence of transformer-based architectures, which have been adapted from their success in natural language processing to capture long-range dependencies in visual data. The pure transformer Swin-UNet model replaced all convolutional operations with shifted window-based self-attention mechanisms, creating a hierarchical representation that could model global context more effectively than a standard CNN for segmentation tasks [21]. This work, alongside broader discussions on the potential of foundational transformer models, signals a potential paradigm shift away from the inductive biases of convolution toward the flexibility of self-attention for medical imaging [22]. Another alternative architectural thread explored the use of graph-based representations. A study tracing the evolution from convolution to graph convolution argued that non-Euclidean data structures, such as those defined by irregular anatomical boundaries or functional connectivity networks, could be more naturally modeled by graph neural networks, offering a powerful complement to conventional grid-based CNNs [23]. In a separate line of inquiry, tensor networks, a technique from quantum physics, were investigated for medical image classification, representing an entirely different computational paradigm that leverages high-order tensor operations for parameter-efficient learning [24].

Beyond the creation of entirely novel architectures, a substantial body of research has focused on the strategic reuse and adaptation of existing models through transfer learning and automated design. A critical investigation compared the performance of full training from scratch against fine-tuning of pre-trained CNNs for medical image analysis, concluding that fine-tuning is often a highly effective strategy, particularly when the target medical dataset is small, as it leverages features learned from large-scale natural image datasets like ImageNet [25]. This principle was extended to three-dimensional data with the development of the Med3D framework, a suite of pre-trained 3D ResNet models on a large collection of medical CT and MRI scans, which demonstrated significant improvements for downstream 3D medical image analysis tasks [26]. A broader survey of deep CNNs for medical image analysis further corroborated the importance of transfer learning, noting that pre-trained backbones such as VGG and ResNet consistently outperformed randomly initialized models across a range of classification tasks [27]. In parallel to these training strategies, the field has also seen significant progress in automating the design of deep learning models. One study introduced an automated machine learning (AutoML) framework for biomedical image segmentation, using neural architecture search to discover high-performing model structures that are tailored to specific datasets without requiring extensive manual engineering [28]. Complementing these design automation efforts, several specialized toolkits and frameworks have been developed to lower the barrier to entry and standardize the implementation of deep learning for medical image analysis. For example, MIScnn provides a flexible and modular framework for medical image segmentation with CNNs, while DLTK (Deep Learning Toolkit) offers state-of-the-art implementations and reusable components specifically designed for medical imaging workflows, thereby accelerating research and enabling more reproducible results [29], [30].

Core Medical Image Analysis Tasks: Segmentation, Registration, and Structure Detection

The application of deep learning to medical image analysis coalesces around a set of fundamental computational tasks that form the bedrock of nearly all clinical decision support systems. These tasks translate raw pixel data into clinically actionable information, enabling quantitative assessment, disease monitoring, and treatment planning. Our review reveals that the primary thematic axis within this dimension is **Image Segmentation**, followed closely by **Image Registration** and **Landmark & Structure Detection**, **& Structure Detection**. The literature suring this theme is not monolithic; rather, it exhibits a complex interplay between task-specific methodological innovations, levels of required supervision, and the unique challenges posed by different imaging modalities and anatomical structures. To systematically map this multifaceted landscape, we have constructed a comprehensive taxonomy of the included studies, presented in Table 2, which categorizes the research according to the primary task, the methodological approach, and the specific anatomical or clinical focus.

Table 2. A comprehensive taxonomy of core medical image analysis tasks, organized by primary task, sub-task, specific methodological approach, and the primary sources from the included studies.

Primary Task	Sub-Task	Specific Method / Focus	Source(s)
Image Segmentation	Fully Supervised Methods	Standard U-Net / CNN architectures for 2D/3D data	[31], [32], [33], [34], [35], [36], [37], [38]
		Multi-task & Multi-modal segmentation frameworks	[39], [40], [41]
		Instance & Curvilinear structure segmentation	[42], [43]
		Attention, Scale-Attention, & Feature interpretation	[44], [45], [46]
		Hybrid methods (with Level Set / Deformable Models)	[47], [48], [49], [50]
	Weakly & Semi-Supervised Methods	Scribble / Weak annotation based guidance	[51], [42], [52]

Primary Task	Sub-Task	Specific Method / Focus	Source(s)
		Active learning & intelligent labeling methods	[53], [54]
		Contrastive, unsupervised, & adversarial learning	[55], [56], [57]
	Interactive & User-Guided Methods	Click-based, geodesic, & fine-tuning interaction	[58], [59], [60]
		Uncertainty & Bayesian deep learning for segmentation	[61], [62]
	Whole Slide Image (WSI) Analysis	Gigapixel image segmentation & classification	[63], [64], [65]
	Organ/Structure-Specific Segmentation	Cardiac (LV, whole heart, multi-structures)	[47], [48], [40], [66]
		Brain (tumor, lesion, stroke)	[34], [41], [61]
		Lung, Pancreas, Skin, & other organs	[36], [38], [67]
		OCT & Retinal structures (macular edema)	[68], [69]
		Prostate & Radiotherapy targets (CTV)	[70]
	Evaluation & Benchmarking	Metrics, guidelines, & large annotated datasets	[71], [72]
	Transfer Learning & Domain Adaptation	Cross-dimensional transfer learning	[73], [74]
	Boundary & Edge Refinement	Rethinking boundary detection & contour refinement	[75], [59]
Image Registration	Unsupervised Methods	Affine & deformable registration frameworks	[55], [76]
	Multi-Modal Registration	3D multi-modal head image registration	[77]
	Toolkits & Frameworks	Dedicated deep learning toolkit for registration	[78]
Landmark & Structure Detection	Landmark Localization	Keypoint detection & anatomical landmarking	[79], [80], [68]

Primary Task	Sub-Task	Specific Method / Focus	Source(s)
	Vessel & Tubular Structure Tracking	Probabilistic vessel tracking approach	[81]
	Artefact Detection & Correction	Motion artefact detection for high-quality segmentation	[66]
	Mass & Lesion Analysis	Mass analysis in mammograms	[49]

Image segmentation has received by far the most intense focus within the included studies, and its centrality to the field cannot be overstated. This task, which involves the pixel-wise delineation of anatomical structures or pathological regions, is a prerequisite for a vast array of downstream clinical analyses, including volumetric measurement, morphological assessment, and targeted therapy planning. The overwhelming majority of segmentation research adopts a fully supervised paradigm, where models are trained on meticulously annotated datasets. Standard U-Net and its three-dimensional variants remain the workhorse architectures for many applications, providing a strong baseline for segmenting everything from lung nodules in CT scans [36] to brain tumors in MRI [34]. However, the literature reveals a continuous push toward methodological sophistication to handle the inherent complexities of medical images, such as heterogeneous texture, ill-defined boundaries, and high anatomical variability. For example, multi-task frameworks have been proposed to simultaneously segment multiple structures from the same image, such as the multi-class segmentation of cardiac structures from cine MRI [40] or the segmentation of brain tumor sub-regions through selective attention [41]. A distinct challenge is posed by curvilinear structures, such as blood vessels or neuronal fibers, where standard volumetric or grid-based methods often fail to capture fine, elongated topology; specialized architectures like CS2-Net have been developed to address this limitation [43](# “IS44”]). Another significant body of work has explored the fusion of deep learning with classical computer vision methods. Hybrid approaches that combine the representational power of neural networks with the smooth, contour-refining properties of deformable models or level sets have demonstrated superior performance for cardiac segmentation in MRI [47] and for mass analysis in mammograms [49].

Given the immense cost and labor associated with creating pixel-level annotations, a substantial thread of the literature is dedicated to reducing the dependency on fully supervised data through weakly supervised, semi-supervised, and interactive methods. Several studies have shown that high-quality segmentation can be achieved using only weak labels, such as scribbles [52] or points [51], which are far less burdensome to generate than dense masks. Active learning frameworks further reduce labeling effort by intelligently selecting the most informative samples for a human annotator to label, a process that was formalized in the “Suggestive Annotation” framework [53], and further optimized through information-theoretic criteria for intelligent labeling [54]. In the realm of semi-supervised learning, contrastive learning objectives have shown promise in learning powerful image representations from large amounts of unlabeled data to improve segmentation performance when only a small annotated cohorts are available [56]. Unsupervised methods, which require no manual labels at all, represent the ultimate goal for data efficiency. One study introduced a deep adversarial network that utilized unannotated images to improve segmentation by forcing the network to learn a more realistic and discriminative features [57]. Interactive segmentation methods offer a practical compromise; they allow a human-in-the-loop to provide coarse user interactions (e.g., clicks, scribbles) which are then used to generate a high-quality result. The Deep Interactive Geodesic Framework (DeepIGeoS) is a prominent example, integrating deep learning with geodesic distance transforms to refine segmentation boundaries based on user clicks [59]. Similarly, NuClick was designed specifically for microscopic images, using a single click to initiate a segmentation that is robust to cellular crowding [60]. Providing image-specific fine-tuning during the interaction phase has also been shown to drastically improve segmentation accuracy for a wide range of structures [58].

Beyond the segmentation of single images, the tasks of **Image Registration** and **Landmark & Structure Detection** address other fundamental needs in a clinical workflow. Image registration, the process of spatially aligning two or more images, is a critical enabler for longitudinal studies, multi-modal fusion, and atlas-based analysis. Our review indicates a clear shift toward end-to-end unsupervised registration methods that learn the spatial transformation directly from the data. An early framework proposed a convolutional neural network that learned to predict affine and deformable transformations without the need for ground-truth deformation fields, achieving high accuracy for a variety of image pairs [55]. This concept was extended for complex, multi-modal registration of 3D head images, demonstrating the feasibility of learning cross-modal correspondences [77]. The development of open-source toolkits, such as DeepReg, has been instrumental in standardizing these methods and making them accessible to the wider research community [78]. Meanwhile, landmark detection focuses on identifying precise, anatomically meaningful points in images, which are essential for tasks like pose estimation, shape analysis, and surgical navigation. A two-stage deep neural network was proposed to detect large-scale landmarks from limited data, achieving real-time performance by first localizing the region of interest and then refining the landmark positions [79]. Other works have framed this as a combined regression-classification problem, showing that a hybrid output can lead to more robust and accurate localization in 2D and 3D medical images [80]. For highly specific tasks like vessel tracking, a probabilistic deep learning approach was introduced to handle the topological continuity and branching complexity of vascular networks [81]. Finally, we note that some studies bridge these tasks, for instance by incorporating motion artefact detection within a segmentation pipeline to improve the quality of cardiac MRI analysis [66].

Data Efficiency and Learning Paradigms

The persistent scarcity of large-scale, high-quality annotated medical datasets represents one of the most formidable bottlenecks in the clinical translation of deep learning models. Unlike natural image domains where millions of labeled examples are readily available, medical imaging is characterized by high annotation costs, the requirement for expert clinical knowledge, strict privacy regulations, and the inherent class imbalance of rare pathologies. Consequently, a significant and diverse body of research has emerged to address this fundamental challenge, developing a spectrum of learning paradigms that aim to maximize model performance from limited supervisory signals. Our analysis reveals that these efforts can be systematically organized along a continuum of supervision, ranging from methods that leverage prior knowledge through transfer learning, to those that actively minimize labeling effort via active and weak supervision, to fully unsupervised and self-supervised approaches that learn representations without any manual labels, and finally to studies that address the related challenges of noisy labels and domain shift. To capture the richness and interconnections of these strategies, we present a comprehensive taxonomy in Table 3, which categorizes the included studies by their primary learning paradigm, the specific methodological approach employed, and their unique contributions.

Table 3. A taxonomy of data efficiency and learning paradigms for medical image analysis, organized by primary paradigm, specific method, and the primary sources from the included studies.

Learning Paradigm	Specific Method / Approach	Key Source(s)
Transfer Learning	Full Training versus Fine-tuning Analysis	[25]
	Transfer Learning for 3D Data (Med3D)	[26]
	Pre-trained CNN Backbones for Classification	[27]
	Cross-domain Transfer (Non-medical to Medical)	[82], [83]
	Hierarchical Fine-tuning for Small Datasets	[84]
	Fine-tuning for Focal Liver Lesion Classification	[85]
	Fine-tuning for COVID-19 Detection	[86]
	Fine-tuning for Bone Age Assessment	[87]
Weak & Semi-Supervised Learning	Scribble Supervision for Segmentation	[52]
	Semi-supervised Segmentation with Consistency Targets	[88]
	Semi-supervised Adversarial Segmentation	[57]
	Multi-task Learning with Incomplete Multi-modal Data	[89]
	Surrogate Supervision for Label-Efficient Learning	[90]
Active Learning	Suggestive Annotation Framework for Segmentation	[53]

Learning Paradigm	Specific Method / Approach	Key Source(s)
Self-Supervised Learning	Contrastive Learning (Global & Local Features)	[56]
	Multimodal Contrastive Learning (GLoRIA)	[91]
	Self-supervised Pre-training (Contrastive & Masked Autoencoder)	[92]
	Unsupervised Feature Learning (K-means + Ensemble CNN)	[93]
Data Augmentation & Generation	Data Augmentation with Statistical Shape Models & TPS	[94]
	Synthetic Data Generation via GANs (TOP-GAN)	[95]
	Unsupervised MR-to-CT Image Translation	[96]
Handling Noisy & Limited Labels	Label Noise: Identification, Cleaning, & Robust Learning	[97]
	Quantifying Necessary Dataset Size for High Accuracy	[98]
Comprehensive Reviews & Benchmarks	Overview of Label-Efficient Deep Learning Challenges	[99]
	Survey of Unsupervised Deep Learning for Medical Imaging	[100]
	Automated ML Benchmark for Small Medical Datasets (MedMNIST)	[101]
Domain Shift & Trust	Impact of Domain Shift on Model Trustworthiness	[102]

Transfer learning has proven to be the most widely adopted and empirically successful strategy for mitigating data scarcity in medical image analysis. The core premise is that a model pre-trained on a large, often unrelated source dataset can be fine-tuned on a smaller target medical dataset, thereby transferring generalizable feature representations. A foundational study in this area directly compared full training from random initialization against fine-tuning strategies for medical image analysis, conclusively demonstrating that fine-tuning, even when the source domain (e.g., ImageNet) is vastly different from the target medical domain, consistently yields superior performance and faster convergence, particularly when the target dataset comprises only a few hundred or thousand images [25]. This finding was corroborated and extended by a broader survey of deep convolutional neural networks for medical image analysis, which confirmed that pre-trained backbones such as VGG and ResNet provide a significant inductive bias that compensates for limited training data [27]. The principle of transfer learning has been successfully applied across a wide array of clinical tasks. For instance, a hierarchical fine-tuning approach was developed for disease detection on small datasets, where the network was first pre-trained on ImageNet, then fine-tuned on a larger medical dataset, and finally fine-tuned on the specific small target dataset, achieving high classification accuracy for chest pathologies [84]. In a critical application during the COVID-19 pandemic, a deep transfer learning model was fine-tuned from pre-trained CNNs on chest X-ray images to rapidly develop a screening tool, demonstrating that models trained on natural images could be effectively repurposed for pandemic response with relatively modest annotated datasets [86]. Similarly, fine-tuning was employed for the classification of focal liver lesions from contrast-enhanced ultrasound, achieving high accuracy despite the inherent challenges of ultrasound imaging and limited training samples [85]. The utility of transfer learning extends beyond 2D natural images. The Med3D framework specifically addressed the challenge of 3D medical data by pre-training a set of 3D ResNet backbones on a large, aggregated dataset of medical CT and MRI scans, thereby creating a powerful foundation for a wide range of downstream 3D medical image analysis tasks such as organ segmentation and lesion detection [26]. Furthermore, two studies explicitly investigated cross-domain transfer from non-medical to medical data. One study demonstrated that features learned from natural images on ImageNet were surprisingly effective for analyzing unregistered multiview mammograms, despite the significant visual differences between natural scenes and mammographic textures [82]. Another study showed that deep learning models trained exclusively on non-medical images could be used for chest pathology identification, suggesting that even highly generic visual features can be useful for medical tasks [83]. Finally, the application of transfer learning for pediatric bone age assessment from hand radiographs confirmed that pre-trained models could achieve clinical-grade accuracy with a moderate amount of domain-specific fine-tuning [87].

Beyond transfer learning, a parallel and highly active thread of research has focused on reducing the cost and burden of annotation through weak supervision, semi-supervised learning, and active learning paradigms. Weak supervision techniques aim to use coarse, inexpensive annotations in place of dense, pixel-level labels. A notable contribution in this area demonstrated that cardiac and prostate segmentation could be effectively learned from scribble annotations alone, which are far quicker for clinicians to draw than full masks, by incorporating a scribble-supervised loss function and a regularization term to encourage smooth and contiguous segmentation boundaries [52]. Semi-supervised learning, which leverages a small set of labeled data alongside a large pool of unlabeled data, has also yielded significant progress. One prominent approach introduced a semi-supervised segmentation framework that used weight-averaged consistency targets, where the model was trained to produce consistent predictions under different perturbations (e.g., different augmentations or dropout configurations) on unlabeled data, thereby encouraging the model to learn robust and generalizable features from the unlabeled distribution [88]. Another innovative semi-supervised approach employed a deep adversarial network, where a discriminator network was trained to distinguish between predicted segmentation maps from the segmentor and ground-truth maps, forcing the segmentor to produce more realistic outputs that generalized better to unlabeled data [57]. Active learning offers a different but complementary philosophy, where the model proactively queries an oracle (typically a human expert) to label the most informative examples. The "Suggestive Annotation" framework formalized this process for biomedical image segmentation, employing an acquisition function that selected the most uncertain regions from the model's predictions for human annotation, thus maximizing the impact of each labeled example and drastically reducing the total number of required annotations [53]. Two further studies expanded the repertoire of label-efficient methods. One introduced a multi-task deep learning framework for Alzheimer's disease diagnosis that could handle incomplete multimodal data, learning robust representations even when some imaging modalities were missing for a given subject [89]. The other proposed a general framework for "surrogate supervision," which replaced manual labels with automatically generated, albeit noisy, labels derived from existing clinical or imaging metadata, enabling effective deep learning from entirely unlabeled images [90].

A third major paradigm in the quest for data efficiency is self-supervised and unsupervised learning, which aims to learn powerful, general-purpose visual representations from unlabeled data that can then be fine-tuned for downstream tasks with minimal labeled examples. Contrastive learning has emerged as a dominant force within this space. One study introduced a framework that learned both global and local visual features through a contrastive objective, enabling the model to capture not only the overall content of a medical image but also the fine-grained spatial details crucial for segmentation, which led to significant improvements when only a limited number of labeled slices were available [56]. A related multimodal contrastive learning framework, named GLoRIA, was designed for chest X-ray imaging and learned joint representations from chest X-ray images and their corresponding radiology reports, using a contrastive loss to align image and text features at both global and local levels, thereby creating a highly semantically rich and label-efficient representation [91]. Another study systematically compared contrastive pre-training and masked autoencoder methods for dealing with small datasets in medical imaging, finding that both approaches significantly outperformed training from scratch, with masked autoencoders showing particular strength in capturing fine-grained anatomical details necessary for dense prediction tasks like segmentation [92]. Going further back in the history of the field, even before the popularization of contrastive learning, an unsupervised feature learning approach combined K-means clustering with an ensemble of deep convolutional neural networks to learn feature hierarchies directly from unlabeled medical images, demonstrating that meaningful and transferable representations could be learned without any manual annotations [93]. Data generation, augmentation, and synthesis techniques constitute a fourth and equally important prong of the data efficiency arsenal, as they artificially expand the size and diversity of the training set. A sophisticated augmentation strategy was proposed that used a statistical shape model and 3D thin plate spline transformations to generate realistic synthetic medical images by deforming real samples along plausible anatomical variations, thereby introducing shape variability into the training data that improved the robustness of deep learning models [94]. In the domain of histopathology, the TOP-GAN framework was developed to stain-free cancer cell classification using a small training set; it employed a generative adversarial network to synthesize realistic cell images from a limited number of real samples, which were then used to train a classifier with high accuracy on the small underlying dataset [95]. For cross-modal image synthesis, a 3D fully convolutional network was used to estimate CT images from corresponding MRI data in an unsupervised manner, addressing a critical clinical need where CT is necessary for dose planning but undesirable due to radiation exposure, and doing so without requiring paired, aligned CT-MR training data [96].

Finally, a number of studies have addressed the systemic challenges that are intertwined with data efficiency, namely the presence of noisy labels and the impact of domain shift on model trustworthiness. A comprehensive review of deep learning with noisy labels in medical image analysis cataloged the various sources of label noise (e.g., annotator variability, imperfect ground truth from clinical reports) and surveyed a range of techniques for mitigating its impact, including noise-robust loss functions, label cleaning strategies, and learning with multiple noisy annotators [97]. A more empirical study directly addressed the fundamental operational question of how much data is actually needed to train a medical image deep learning system to achieve a necessary high accuracy, systematically analyzing the relationship between dataset size and performance for different tasks and architectures [98]. The pernicious problem of domain shift, where the distribution of data at inference time differs from the training data (e.g., due to a different scanner manufacturer or patient population), was investigated in the context of chest radiograph classification. The study found that deep learning models often produce overconfident predictions on out-of-distribution data, and it proposed using uncertainty quantification as a tool to detect domain shift and flag unreliable predictions, thereby emphasizing that trustworthiness is a critical component of any data-efficient learning system deployed in the real world [102]. To provide a holistic view of these challenges, two comprehensive review articles synthesized the state of the art. One focused on the challenges and future directions of label-efficient deep learning in medical image analysis, discussing the trade-offs between different paradigms [99]. The other offered a tour of unsupervised deep learning for medical image analysis, tracing the evolution of unsupervised methods and their role in extracting clinically meaningful information from the vast repositories of unlabeled medical data that exist in clinical archives [100]. As a practical resource for methodological development, the MedMNIST classification decathlon was introduced as a lightweight automated ML benchmark consisting of a collection of standardized medical image datasets of varying sizes, providing a common platform for comparing and developing data-efficient learning techniques [101].

Explainability, Trust, and Uncertainty Quantification

The translation of deep learning models from high-performance benchmarks into trusted clinical decision-support tools hinges critically on their ability to be not only accurate but also interpretable and self-aware of their limitations. A model that produces a correct diagnosis but cannot explain its reasoning, or one that confidently makes an incorrect prediction for an out-of-distribution patient, poses an unacceptable risk in a clinical setting. The included studies within this dimension directly address these intertwined challenges of explainability, robustness, and uncertainty quantification, revealing a rapidly maturing subfield that is moving from ad-hoc visualization techniques toward principled frameworks for building trustworthy medical AI. To systematically organize the diverse contributions in this area, we have developed a comprehensive three-level taxonomy that categorizes studies by their primary dimension, specific category, and methodological sub-category, as presented in Table 4.

Table 4. A hierarchical taxonomy of studies addressing explainability, trust, and uncertainty quantification in deep learning for medical image analysis, organized by dimension, category, sub-category, and source.

Dimension	Category	Sub-Category	Source
Explainability & Interpretability	General Overview & Challenges	Frameworks & Taxonomies	[103], [104], [105]
		Methodological Issues & Failures	[106]
	Interpretable Techniques	Feature Extraction & Saliency Maps	[107]
		Counterfactual Explanations	[108]
		Multi-class Segmentation & Classification	[67]
	Application-Specific Explainability	Brain Image Analysis (MRI)	[109], [110]
		Skin Cancer	[67]
Trust & Robustness	Adversarial Attacks & Defenses	Attack Types & Impact	[111], [112], [113]
	Fairness & Bias	Data Imbalance & Bias Detection	[109], [114]IS106”)
		Hidden Stratification & Clinical Failures	[115]
Uncertainty Quantification	Calibration & Model Confidence	Calibration for Class Imbalance	[116]
		Calibration-based Explainers	[108]
	Estimation Methods	Aleatoric Uncertainty (Test-time Augmentation)	[117]
		Bayesian & General UQ Methods	[118]
	Application-Specific UQ	Diabetic Retinopathy	[119], [120]
		Stroke Analysis (MRI)	[121]
		Lesion Detection & Segmentation	[61]
		Prostate Cancer Radiotherapy	[70]

Research on explainability and interpretability has progressed from general conceptual frameworks to the development of specific, application-tuned methods. Several foundational works established the philosophical and technical context for this challenge. One study provided a comprehensive overview of the challenges of deep learning in medical image analysis, arguing that trust and explainability are not merely desirable features but fundamental prerequisites for clinical adoption, and surveyed a range of post-hoc interpretability techniques including saliency maps and concept activation vectors [103]. Another overview specifically focused on the intersection of generalizability and explainability, proposing that a truly trustworthy model must be both robust to domain shift and capable of providing human-interpretable explanations for its decisions, a framework that connects the dots between two often-separate research threads [104]. A third review article synthesized interpretation and visualization techniques for deep learning models in medical imaging, categorizing methods into feature attribution, activation maximization, and concept-based explanations, and discussing their relative merits for different clinical tasks [105]. These conceptual foundations were complemented by a methodologically focused paper that directly addressing the causes of failure in medical imaging machine learning [106].

Progressing beyond reviews to concrete methodological innovations, several studies introduced novel explainability techniques. One approach leveraged feature extraction and saliency mapping to generate explainable outputs for medical image classification, demonstrating that learned features could be visualized and semantically interpreted by clinicians [107]. A more sophisticated line of inquiry investigated counterfactual explanations, where the model generates an altered version of the input image that would have led to a different prediction, thereby revealing the minimal changes necessary to alter the clinical decision. A study proposed training calibration-based counterfactual explainers, which explicitly modeled the decision boundary of the classifier and generated counterfactuals that were not only plausible but also calibrated to the model's true confidence, providing a principled and quantitatively rigorous approach to explanation [108]. For specific applications, explainability methods were developed for multi-class segmentation and classification of non-melanoma skin cancer, where the model provided both a segmentation mask and a classification and highlighted the regions of the image that were most influential for each class decision, thereby enabling clinicians to verify the model's reasoning against their own visual assessment [67].

The trust dimension of this taxonomy encompasses both adversarial robustness and algorithmic fairness, two critical yet distinct threats to clinical reliability. Adversarial attacks, where imperceptible perturbations to input images can cause catastrophic model failures, pose a significant safety risk. One study provided a comprehensive understanding of adversarial attacks on medical deep learning systems, demonstrating that DNNs for medical image analysis tasks like classification and segmentation are highly susceptible to these manipulations, and that the attacks can be crafted to be both targeted and universal [111]. This vulnerability was further explored in the context of biomedical image segmentation, showing that adversarial examples can cause systematic failure in delineating organ boundaries, which has direct and dangerous implications for tasks like radiotherapy planning [112]. The concept of universal adversarial attacks was also investigated, revealing that a single perturbation could fool a model across a wide range of different medical images from the same class, highlighting the fundamental and non-random nature of these vulnerabilities [113]. Fairness and bias represent a parallel but equally critical threat to trust. A study examining fairness-related performance and explainability effects in deep learning models for brain image analysis found that models could exhibit significant performance disparities across demographic subgroups, such as sex or age, and that these biases were often concealed by aggregate performance metrics [109]. This work connected fairness to explainability by showing that the features learned by the model, and hence the explanations derived from them, could also be biased. Another investigation specifically focused on bias due to data imbalance in deep learning-based cardiac MR image segmentation, demonstrating that models trained on imbalanced datasets, where certain anatomical variants or patient demographics were underrepresented, produced systematically biased segmentations for those groups [114]. An influential study on hidden stratification provided a deeper diagnosis of this problem, showing that machine learning models for medical imaging can achieve high overall accuracy while failing systematically on clinically meaningful but rare subclasses, a phenomenon that standard evaluation metrics completely fail to detect [115].

Uncertainty quantification (UQ) provides a complementary route to trustworthiness by enabling models to express when they are unsure, rather than simply providing a single, overconfident prediction. A foundational review article quantifies the landscape of UQ in deep learning for radiologic images, cataloging Bayesian methods, ensemble methods, and test-time dropout as the primary techniques for capturing epistemic uncertainty (model uncertainty) and aleatoric uncertainty (inherent data noise) [118]. Building on these principles, specific methodological advances have been made. A study on test-time data augmentation for the estimation of heteroscedastic aleatoric uncertainty proposed a practical and efficient method: by passing multiple augmented versions of the same input through the network and observing the variance of the outputs, one can estimate the irreducible noise in the data [117]. This approach is particularly relevant for medical imaging where signal-to-noise ratio can vary significantly across different scanners or acquisition protocols. Model calibration, which ensures that a model's predicted probabilities match its true accuracy (e.g., predictions made with 90% confidence should be correct 90% of the time), is a critical aspect of trustworthy UQ. A study specifically addressed the impact of class imbalance on model calibration, showing that standard deep learning models become severely miscalibrated on minority classes but that recalibration techniques like temperature scaling and isotonic regression can mitigate this, thereby improving the reliability of predictions for rare pathologies [116].

Several studies directly applied UQ to specific clinical tasks, demonstrating its practical value. For non-proliferative diabetic retinopathy grading in eye fundus images, an uncertainty-aware deep learning framework was developed that not only predicted the disease grade but also output a calibrated uncertainty estimate for each prediction, allowing the system to flag uncertain cases for expert review [119]. A related study validated uncertainty estimates against expert opinion in the same domain of diabetic retinopathy detection, showing that model uncertainty was correlated with the difficulty of the case and with inter-grader variability, thereby providing a clinically meaningful and human-verifiable measure of confidence [120]. In the context of acute stroke analysis from MRI, integrating uncertainty in deep neural networks enabled clinicians to identify regions of the segmentation that were unreliable, potentially due to poor image quality or atypical pathology, and to incorporate this information into treatment decisions [121]. For multiple sclerosis lesion detection and segmentation, exploring different uncertainty measures including dropout Monte Carlo and deep ensemble methods provided a map of lesion confidence that could be used to prioritize more uncertain cases for radiologist review [61]. Perhaps the most clinically consequential application was in post-operative prostate cancer radiotherapy, where the proposed deep learning framework for segmenting invisible clinical target volumes simultaneously provided an estimate of segmentation uncertainty for each voxel. This information allowed radiation oncologists to adapt their treatment margins, expanding coverage in high-uncertainty regions and tightening them in low-uncertainty ones, thereby directly translating a technical capability into a personalized clinical action [70].

Clinical Applications and Disease-Specific Studies: A Comprehensive Mapping of Organ Systems and Pathologies

The translation of deep learning models from methodological development into tangible clinical impact is ultimately measured by their performance across a wide spectrum of diseases and anatomical targets. The included studies within this dimension represent the most practical and application-oriented thread of the literature, demonstrating how generic architectures and training paradigms are adapted, refined, and validated for specific clinical scenarios. The breadth of this research is remarkable, spanning nearly every major organ system and covering a diverse array of pathologies from oncology to neurology, ophthalmology, and infectious disease. To systematically navigate this vast clinical landscape, we have constructed a comprehensive two-level taxonomy that categorizes the studies first by clinical domain or organ system and then by specific application or disease focus, as presented in Table 5.

Table 5. A comprehensive hierarchical taxonomy of clinical applications and disease-specific studies in deep learning for medical image analysis.

Clinical Domain / Organ System	Specific Application / Disease Focus	Sources
Brain & Neurology	Brain Tumor Detection & Classification	[122], [123], [124], [125], [126], [127]
	Alzheimer’s Disease Diagnosis	[128], [89]
	Parkinson’s Disease Prediction	[129]
	Stroke Lesion Segmentation & Prediction	[130], [46]
	Hemorrhage Diagnosis	[131]
	fMRI Analysis (e.g., relapse prediction)	[132]
Thoracic & Pulmonary	COVID-19 Detection & Analysis	[133], [134], [135], [86]
	Lung Cancer	[65]
	Lung CT Segmentation	[36]
Ophthalmic & Vision	Diabetic Retinopathy	[136], [119]

Clinical Domain / Organ System	Specific Application / Disease Focus	Sources
Cardiovascular	Cardiovascular Image Diagnosis	[137], [50]
Gastrointestinal & Abdominal	Colorectal Cancer	[138]
	Gastric Cancer (Whole Slide Images)	[64]
	Liver Lesion Classification	[85]
	Esophageal Adenocarcinoma	[139]
Oncology (General / Multi-Site)	Breast Cancer**	[140], [141]
	Cancer Cell Classification (General)	[95]
	Skin Cancer (Non-melanoma)	[67]
	Metastatic Breast Cancer Detection	[140]
Musculoskeletal	Bone Age Estimation	[142], [87]
Head & Neck / Dental	Thyroid Nodule Diagnosis	[143]
	Dental Image Analysis	[144]
Urology / Prostate	Prostate Cancer	[145]
**Broad / Multi-Purpose Reviews & Surveys	General Deep Learning in Medical Imaging	[146], [147], [148], [149], [150]
Other Specific Diseases	Tongue Color Image Analysis (IoT-based)	[151]
	Digital Twins for Medical Imaging	[133]
	Gender Bias in Medical Imaging Datasets	[152]
	Multi-modal Medical Image Classification	[153]

The most concentrated body of clinical research within our review pertains to the brain and neurological system, with brain tumor detection and classification representing a particularly active sub-domain. Multiple studies have applied deep learning to MRI-based brain tumor analysis, each proposing distinct architectural and methodological solutions. For instance, one study developed a deep learning approach that combined image processing techniques with a convolutional neural network to detect tumors from MR images, demonstrating that the automated pipeline could achieve high sensitivity and specificity in identifying the presence of pathology [124]. Another study introduced the LeU-Net architecture, a modification of the standard U-Net specifically designed for 2D MRI brain tumor detection, which leveraged both local and global contextual information to improve segmentation accuracy [126]. A comparative study of various deep learning techniques for brain tumor detection systematically evaluated the performance of different CNN architectures, including VGG, ResNet, and Inception, concluding that deeper architectures with residual connections generally outperformed shallower models when sufficient training data was available [122]. The challenge of data privacy in this sensitive domain was addressed by a study that applied federated learning to brain tumor detection, training a deep learning model across distributed medical institutions without sharing raw patient data, thereby demonstrating that high performance can be achieved while respecting privacy regulations [125]. Another work focused on image analysis for MRI-based brain tumor classification, providing a comprehensive pipeline that included pre-processing, feature extraction via a pre-trained CNN, and classification [123]. A further study on deep learning for MRI-based brain tumor images reinforced the utility of transfer learning, showing that fine-tuning a pre-trained model on a relatively small dataset of brain tumor images could yield highly accurate classification results [127].

Neurological diseases beyond tumors, particularly neurodegenerative disorders, also feature prominently in the literature. For Alzheimer's disease diagnosis, one study employed a deep learning approach using MRI to classify patients into different stages of cognitive decline, demonstrating that features learned from the images could distinguish between healthy controls, mild cognitive impairment, and Alzheimer's disease with high accuracy [128]. This was complemented by a more sophisticated multi-task deep learning framework that could handle incomplete multimodal data, including MRI, PET, and clinical assessments, for the multi-stage diagnosis of Alzheimer's disease; the model learned to perform classification tasks across all available data subsets, effectively imputing missing modalities and improving diagnostic performance [89]. For Parkinson's disease, a unified deep learning approach was proposed that combined convolutional and recurrent neural networks trained on medical images, such as structural MRI and DaTscan SPECT, to predict the disease status, illustrating the potential of deep learning to analyze temporal or sequential imaging data for movement disorders [129]. Stroke-related research focused on both prediction and segmentation. One study developed a deep learning model to predict final infarct volume from native CT perfusion and treatment parameters, a critical task for prognostication and treatment planning in acute ischemic stroke [130]. Another study addressed automatic ischemic stroke lesion segmentation from CT perfusion images by using image synthesis to generate pseudo-normal images and then applying an attention-based deep neural network to segment the lesion, a technique that improved performance by providing the model with a normative comparison [46]. The diagnosis of brain hemorrhage was investigated in a study that applied a pre-trained deep convolutional neural network to CT images, demonstrating that transfer learning from non-medical images could effectively identify different types of intracranial hemorrhage [131]. Finally, the analysis of functional MRI (fMRI) data using deep learning was explored for relapse prediction in patients with psychiatric conditions, showing that convolutional neural networks could extract spatio-temporal patterns from the complex, high-dimensional fMRI data that were predictive of clinical outcomes [132].

The COVID-19 pandemic spurred an intense and rapid surge of research into deep learning for thoracic imaging, with chest X-ray and CT analysis becoming a focal point for many studies. Four studies in our review directly addressed COVID-19 detection from medical images. One study proposed a framework that integrated digital twins with deep learning for medical image analysis, using synthetic chest X-ray images generated from digital twin models to augment training data for COVID-19 detection, a strategy that addressed the initial scarcity of real patient data during the early stages of the pandemic [133]. Another study provided a comprehensive review and update on deep learning applications in pulmonary medical imaging with a specific focus on COVID-19, summarizing the various CNN architectures and transfer learning strategies that were rapidly adopted and adapted for this new disease [134]. A study specifically on the detection and analysis of COVID-19 in medical images using deep learning techniques compared the performance of different deep learning models on chest CT and X-ray datasets, concluding that hybrid models combining CNNs with other components offered the best trade-off between accuracy and computational efficiency [135]. A widely cited study introduced Deep-COVID, a deep transfer learning model that used pre-trained CNN backbones to predict COVID-19 from chest X-ray images, achieving high sensitivity and specificity on a large dataset and demonstrating how established deep learning techniques could be rapidly deployed for pandemic response [86]. Beyond COVID-19, pulmonary research included a study on weakly supervised deep learning for whole slide lung cancer image analysis, which developed a method for classifying lung cancer subtypes from gigapixel histopathology images using only slide-level labels, thereby significantly reducing the annotation burden for this important clinical task [65]. A complementary study focused on lung CT image segmentation using deep neural networks, which is a prerequisite for quantifying lung involvement in diseases like COVID-19 and interstitial lung disease [36].

Ophthalmic applications of deep learning, particularly for diabetic retinopathy, represent another well-established and impactful area of research. A foundational study on deep image mining for diabetic retinopathy screening used a deep CNN to automatically detect and grade the severity of diabetic retinopathy from color fundus photographs, achieving high sensitivity and paving the way for automated screening programs that could reduce the burden on ophthalmologists [136]. This work was extended by a study that introduced an uncertainty-aware deep learning framework for the task, which provided not only the predicted diabetic retinopathy grade but also a calibrated estimate of the model's confidence, enabling the system to flag uncertain cases for human expert review and thereby enhancing the trustworthiness of the diagnostic output [119].

Cardiovascular disease diagnosis using deep learning was the focus of several studies. One study provided a comprehensive overview of deep learning-based cardiovascular image diagnosis, discussing the application of CNNs to tasks such as left ventricle segmentation, plaque characterization, and coronary artery disease detection from various imaging modalities including CT, MRI, and ultrasound [137]. A more methodological study proposed a deep learning approach for direct whole-heart mesh reconstruction from volumetric medical images, which learned to generate 3D surface meshes of all four cardiac chambers simultaneously, a technically challenging task that has direct applications for computational fluid dynamics and surgical planning [50].

Gastrointestinal and abdominal applications span several cancer types. One study developed a deep learning model for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging data, a task that combined advanced image analysis with longitudinal modeling to provide early prognostic information during chemotherapy [138]. A study on whole-slide gastric image classification introduced a recalibrated multi-instance deep learning framework that addressed the challenges of analyzing gigapixel histopathology images by breaking them into patches, processing them with a CNN, and then aggregating the patch-level predictions with an attention mechanism [64]. For liver lesions, a study on the classification of focal liver lesions using deep learning with fine-tuning applied transfer learning from a pre-trained CNN to contrast-enhanced ultrasound images, achieving high accuracy in differentiating benign from malignant lesions [85]. A study on computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma demonstrated the use of CNNs on endoscopic images to classify Barrett's esophagus and early-stage cancer, aiming to improve early detection rates [139].

Oncology applications across various sites were also well represented. For breast cancer, one study developed a deep learning model for identifying metastatic breast cancer from whole-slide histopathology images of sentinel lymph node biopsies, a task that is critical for staging and prognosis [140]. Another study investigated end-to-end deep learning approaches for breast ultrasound lesions recognition, comparing different CNN architectures for classifying benign and malignant breast masses on ultrasound images [141]. A study on stain-free cancer cell classification using deep learning with a small training set presented the TOP-GAN framework, which used generative adversarial networks to synthesize realistic cell images for training, addressing the problem of limited annotated data for cytopathology [95]. Interpretable deep learning systems were developed for multi-class segmentation and classification of non-melanoma skin cancer, providing both a diagnostic prediction and a visual explanation of the model's reasoning [67].

Musculoskeletal applications included bone age estimation, a common pediatric radiology task. One study applied deep learning to this problem by training a CNN on hand radiographs, demonstrating that the model could estimate bone age with an accuracy comparable to that of expert radiologists [142]. A further study specifically focused on paediatric bone age assessment using deep convolutional neural networks, confirming the utility of transfer learning and deep learning for this standardized clinical task [87]. Head and neck applications included a study on ultrasound image analysis using a deep learning algorithm for the diagnosis of thyroid nodules, which applied CNNs to classify benign and malignant thyroid nodules from ultrasound images [143]. Generalizability of deep learning models for dental image analysis was investigated by another study, which systematically evaluated how well models trained on data from one dental clinic or population performed when applied to data from another, highlighting the critical importance of domain shift for clinical deployment [144].

Prostate cancer research was represented by a quick guide on radiology image pre-processing for deep learning applications, which provided practical recommendations and protocols for preparing prostate MRI data for deep learning analysis, including steps for normalization, registration, and artifact removal [145]. A set of multi-purpose reviews and surveys provided broad overviews of the entire field. One study offered a general overview of deep learning approaches for medical image analysis and diagnosis, covering a wide range of architectures and tasks [146]. Another presented deep learning in medical image analysis as a "third eye for doctors," systematically reviewing the state-of-the-art CNN applications up to 2019 [147]. A survey on deep learning and medical imaging discussed both the opportunities and the challenges for translation [148]. A perspective article on AI for medical imaging argued for close collaboration between the machine learning and medical communities to overcome the unique challenges of the domain [149]. A review of deep learning in nuclear medicine and molecular imaging provided a modality-specific view of the current perspectives and future directions [150].

Finally, several studies addressed unique or cross-cutting clinical applications. One study investigated Internet of things and synergic deep learning-based biomedical tongue color image analysis for disease diagnosis and classification, a novel application that combined IoT sensor data with deep image analysis for a traditional medical diagnostic technique [151]. A study examining gender imbalance in medical imaging datasets showed that this imbalance produces biased classifiers for computer-aided diagnosis, systematically demonstrating how datasets with a skewed sex distribution lead to models that are less accurate for the underrepresented sex [152]. A study on multi-modal medical image classification using a deep residual network and a genetic algorithm proposed a method for fusing features from different imaging modalities (e.g., MRI and CT) to improve classification performance for brain diseases [153].

Image Pre-processing, Enhancement, and Synthesis

The quality, consistency, and availability of medical images are foundational determinants of deep learning model performance. Raw medical images, as acquired from clinical scanners, are often characterized by noise, low spatial resolution, artifacts, and inter-modality or inter-scanner variability that can severely degrade the effectiveness of downstream analysis tasks. Consequently, a substantial body of research has focused on developing deep learning-based pre-processing techniques, enhancement methods, and synthesis strategies that operate directly on the image data to improve its utility for subsequent analysis. The included studies in this theme can be systematically organized into a three-level taxonomy that captures the primary objective of the work, the specific technical approach employed, and the relevant application domain, as presented in Table 6. This taxonomy reveals a field that has moved beyond simple filtering and normalization toward sophisticated, task-adaptive deep learning pipelines that generate, restore, and transform medical images.

Table 6. A hierarchical taxonomy of studies on image pre-processing, enhancement, and synthesis in deep learning for medical image analysis, organized by primary objective, specific technique, application, and source.

Primary Objective	Specific Technique	Application	Source(s)
Image Enhancement	Denoising	General Medical Image Denoising (Convolutional Denoising Autoencoders)	[154]
	Super-Resolution	Fast Medical Image Super-Resolution (Deep Learning Network)	[155]
		Transfer Learning-based Super-Resolution from Single Medical Image	[156]
	Image Correction & Quality Improvement	CT Image Quality Improvement (Deep Learning)	[157]
		MRI Image Enhancement and Correction (State-of-the-Art)	[158]
		Attenuation Correction from Time-of-Flight PET Emission Data	[159]
		Adaptive Ultrasound Beamforming (Deep Learning)	[160]
	Enhancement for Downstream Tasks	Fine-tuning Deep Neural Networks using Photorealistic Medical Images (Cinematic Rendering)	[161]
Image Synthesis & Modality Conversion	Generative Adversarial Networks	Medical Image Translation using GANs (MedGAN)	[162]
		Synthesized 7T MRI from 3T MRI (Spatial & Wavelet Domains)	[163]
		Medical Image Synthesis for Data Augmentation and Anonymization	[164]
	Diffusion Probabilistic Models	Three-Dimensional Medical Image Synthesis with Denoising Diffusion Probabilistic Models	[165]
	Convolutional Neural Networks	Estimating CT Image from MRI Data using 3D Fully Convolutional Networks	[96]
	General Frameworks & Overviews	Overview of Image-to-Image Translation: Denoising,	[166]

Primary Objective	Specific Technique	Application	Source(s)
		Super-Resolution, Modality Conversion, Reconstruction	
		Image Reconstruction as a New Frontier of Machine Learning	[167]
		Real-World Challenges for Deep Generative Models in Medical Imaging	[168]
Image Fusion & Augmentation	Statistical Shape Models	Augmentation using Statistical Shape Model and 3D Thin Plate Spline for Deep Learning	[94]
	Deep Learning-based Fusion	Multimodal Medical Image Fusion based on Deep Learning Neural Network for Clinical Treatment Analysis	[169]
		Multi-modality Medical Image Fusion using Multi-Objective Differential Evolution based Deep Neural Networks	[170]
		Multi-modality Medical Image Fusion using Optimal Shearlet and Deep Learning	[171]

The literature on image enhancement encompasses a spectrum of techniques aimed at improving the perceptual or analytical quality of medical images. Denoising is a fundamental pre-processing task, and one study specifically investigated medical image denoising using convolutional denoising autoencoders, demonstrating that these networks could learn to effectively remove noise from images without requiring a paired clean-noisy training set by learning to reconstruct clean images from corrupted inputs [154]. Image super-resolution addresses the problem of limited spatial resolution, and two studies contributed to this area. One proposed a fast medical image super-resolution method based on a deep learning network that could reconstruct high-resolution images from their low-resolution counterparts with significantly improved computational efficiency compared to traditional optimization-based methods [155]. Another study introduced a transfer learning-based super-resolution reconstruction approach for single medical images, leveraging pre-trained networks from natural image super-resolution to improve performance on medical data while requiring fewer medical training examples [156].

Image correction and quality improvement techniques target specific artifacts or degradations inherent to particular imaging modalities. A study on the possibility of deep learning for improving computed tomography image quality provided a comprehensive overview of how deep neural networks could be used to reduce noise, correct beam-hardening artifacts, and improve the contrast-to-noise ratio in CT images, thereby enabling lower radiation dose protocols without sacrificing diagnostic quality [157]. A separate, comprehensive review focused on deep learning for image enhancement and correction in magnetic resonance imaging, cataloging the state-of-the-art methods for correcting motion artifacts, reducing image noise, and accelerating MRI acquisition through undersampled k-space reconstruction [158]. For positron emission tomography, a deep learning-guided method was developed to estimate attenuation correction factors directly from time-of-flight PET emission data, bypassing the need for a separate CT-based attenuation correction scan, which reduces radiation exposure and simplifies the acquisition protocol [159]. In the domain of ultrasound imaging, an adaptive beamforming approach using deep learning was proposed to improve image quality by learning to optimize the delay-and-sum beamforming process, resulting in enhanced spatial resolution and contrast compared to standard beamforming techniques [160]. Finally, a study on fine-tuning deep neural networks using cinematic rendering demonstrated that photorealistic renderings of 3D medical imaging data could be used as a form of image enhancement for downstream deep learning tasks, improving model performance for classification or segmentation by exposing the model to visually rich and realistic representations of anatomy [161].

The synthesis of new medical images and the conversion between imaging modalities represent some of the most exciting and rapidly advancing areas within this theme. Generative adversarial networks (GANs) have been the most widely explored class of models for this purpose. The MedGAN framework was specifically introduced for medical image translation, providing a general method for converting images from one modality to another, such as generating CT images from MRI data, and demonstrating that adversarial training could produce realistic and clinically useful synthetic images [162]. A more specific translation task was addressed by a study that synthesized 7T MRI images from 3T MRI inputs using a deep learning network operating in both the spatial and wavelet domains, a transformation that offers the potential to provide the superior tissue contrast of ultra-high-field MRI without the logistical and cost barriers of acquiring a 7T scanner [163]. Another study focused on the use of GANs specifically for data augmentation and anonymization in medical imaging, generating synthetic medical images that could be used to expand training datasets while preventing the identification of individual patients, thereby addressing both data scarcity and privacy concerns simultaneously [164]. Beyond GANs, denoising diffusion probabilistic models have emerged as a powerful alternative for image synthesis. One study applied these models to three-dimensional medical image synthesis, demonstrating that diffusion models could generate high-quality, realistic 3D volumes from noise, offering a promising alternative to GANs with potentially better mode coverage and training stability [165].

A fundamentally different synthesis task involves estimating one imaging modality from another, and this has been addressed using fully convolutional neural networks. A study on estimating CT images from MRI data using 3D fully convolutional networks directly learned the complex non-linear mapping between the two modalities, enabling the generation of synthetic CT images that could be used for MRI-based radiation treatment planning, thereby avoiding the radiation exposure associated with a planning CT scan [96]. Two comprehensive overviews placed these individual techniques in a broader context. The first provided a detailed overview of image-to-image translation using deep neural networks, covering the entire spectrum of tasks including denoising, super-resolution, modality conversion, and image reconstruction, and discussing the common underlying principles and unique methodological considerations for each [166]. The second overview argued that image reconstruction itself represents a new frontier of machine learning, where deep learning models are not just post-processing tools but are integrated into the image formation process itself, enabling new forms of compressed sensing, artifact correction, and physics-informed reconstruction [167]. A study specifically addressing the real-world challenges of deep generative models in medical imaging provided a sobering assessment, highlighting issues such as mode collapse in GANs, the difficulty of validating synthetic images, and the risk of introducing systematic bias into downstream analyses if synthetic images are not perfectly representative of the true population [168].

Image fusion and augmentation techniques further enrich the landscape of image pre-processing and synthesis. For data augmentation, one study proposed a sophisticated strategy based on statistical shape models and 3D thin plate spline transformations, which generated realistic synthetic medical images by warping real images along learned modes of anatomical variation, thereby introducing realistic shape variability into the training data without requiring paired data [94]. Multimodal medical image fusion, which combines information from different imaging modalities into a single composite image, was the focus of three distinct studies. The first proposed a deep learning-based multimodal medical image fusion framework for clinical treatment analysis, using CNNs to extract and fuse complementary features from images such as MRI and CT, resulting in a fused image that preserved both the soft tissue contrast from MRI and the bone detail from CT [169]. A second study developed a multi-modality image fusion technique that integrated multi-objective differential evolution with deep neural networks, optimizing the fusion process to simultaneously maximize multiple quality criteria, such as mutual information and edge preservation [170]. A third approach combined optimal shearlet transforms with deep learning for multimodality medical image fusion, first decomposing the input images into frequency sub-bands using shearlets and then using a deep network to learn the optimal fusion rule for each sub-band [171]. These fusion techniques are particularly valuable for clinical scenarios where the complementary strengths of different imaging modalities need to be combined for comprehensive diagnosis or treatment planning.

Emerging Trends, Privacy, and Systemic Challenges

As deep learning models for medical image analysis transition from high-performance benchmarks toward clinical deployment, a constellation of emerging trends, privacy concerns, and systemic challenges has moved to the forefront of the research agenda. The included studies in this dimension collectively address the critical factors that will determine whether deep learning becomes a routine part of clinical practice or remains confined to academic research. These factors include the pressing need to protect patient data privacy through technologies like federated learning and encryption, the growing awareness of the environmental and carbon footprint of large-scale model training, the development of robust benchmarks and standardized evaluation protocols, the challenge of making models generalizable across different scanners and populations, and the fundamental systemic barriers to translation, including the need for appropriate metrics, validation frameworks, and an understanding of the limitations of current approaches. To capture the breadth and interconnectedness of these topics, we have constructed a detailed three-level taxonomy in Table 7, which organizes the studies by their primary thematic area, specific category, and methodological or conceptual sub-category.

Table 7. A comprehensive taxonomy of emerging trends, privacy, and systemic challenges in deep learning for medical image analysis, organized by thematic area, category, sub-category, and source.

Thematic Area	Specific Category	Methodological / Conceptual Sub-Category	Key Source(s)
Privacy, Security & Ethics	Privacy-Preserving Learning	Federated Learning for Privacy in Medical Imaging	[125]
		Differential Privacy in Medical Imaging Deep Learning	[172]
		Privacy-Preserving Time-Series Medical Image Analysis	[173]
	Encryption & Data Security	Deep Learning-Based Medical Image Encryption	[174], [175]
	Systematic Challenges in Privacy	Privacy-Preserving Machine Learning for Healthcare, Ethical & Legal Barriers	[176]
Environmental Impact & Efficiency	Carbon Footprint	Carbon Footprint of Selecting and Training Deep Learning Models	[177]
Benchmarking, Standardization & Validation	Benchmark Datasets and Protocols	Large-Scale Lightweight Benchmark (MedMNIST v2)	[178]
		Standardization of Medical Imaging Workflows for Radiomics & Deep Learning	[179]
		MRI Scanner Domain Shift: Standardized Acquisition & Validation	[180]
		Practical Window Setting Optimization for Deep Learning	[181]
	Evaluation Metrics and Failures	Common Limitations of Performance Metrics in Biomedical Image Analysis	[182]

Thematic Area	Specific Category	Methodological / Conceptual Sub-Category	Key Source(s)
Tooling, Infrastructure & Reproducibility	Annotation and Labeling Tools	RIL-Contour: Annotation Tool for Deep Learning	[183]
		PyMIC: Annotation-Efficient Toolkit	[184]
	Reproducible Frameworks	Versatile Deep Learning Framework for Medical Image Computing	[185]
		DeepImageJ: User-Friendly Environment for Running Models	[186]
	Data Management & Preparation	Medical Image Data and Datasets: Datasets & Annotated Data	[187]
		Preparing Medical Imaging Data for Machine Learning (Whitepaper)	[188]
Generalizability, Robustness & Domain Shift	Scanner and Acquisition Variability	Impact of Scanner Domain Shift on Deep Learning Performance	[180]
	Domain Adaptation	Automated Knowledge Transfer for Medical Image Segmentation	[189]
Systemic & Foundational Challenges	General Challenges and Future Directions	Deep Learning in Medical Imaging: Overview, Challenges & Future	[176]
		Going Deep: Concepts, Methods, Challenges & Future Directions	[190]
		Prospects of Deep Learning for Medical Imaging	[191]
		Guest Editorial: Overview and Promise of Deep Learning	[192]
		Challenges of Deep Learning: Improving Explainability & Trust	[103]
		Generalizable and Explainable Deep Learning: An Overview	[104]

Thematic Area	Specific Category	Methodological / Conceptual Sub-Category	Key Source(s)
		A Holistic Overview of Deep Learning in Medical Imaging	[193]
		Machine Learning for Medical Imaging: Methodological Failures & Recommendations	[106]
	Modality-Specific and Application-Specific Challenges	Deep Learning in Radiology	[194]
		Machine Learning for Medical Ultrasound: Status, Methods & Future	[195]
		Image Analysis and Machine Learning in Digital Pathology	[196]
		Machine (Deep) Learning for Image Processing and Radiomics	[197]
		Deep Learning in Medical Image Registration	[198]
		Deep Learning-Based Solvability of Underdetermined Inverse Problems	[199]
		Application and Challenges of Deep Learning in Intelligent Medical Image Processing	[200]
		Enhancing Radiomics and Deep Learning Systems through Standardization	[179]
		Application and Construction of Deep Learning Networks in Medical Imaging	[201]
		Manifold Learning of Brain MRIs by Deep Learning	[202]
		On the Prospects for a (Deep) Learning Health Care System	[203]
	Training Data Requirements	Trade-off between Training and Testing Ratio	[204]

Thematic Area	Specific Category	Methodological / Conceptual Sub-Category	Key Source(s)
		Medical Image Data and Datasets in the Era of Machine Learning	[187]
	Training Data Requirements (cont.)	Impact of Training Data Size on Deep Learning Performance	[205]
	Clinical Translation Barriers	Preparing Medical Imaging Data for Machine Learning	[188]
Other Emerging Trends	Feature Extraction Innovations	Advancements in Feature Extraction Recognition of Medical Imaging Systems	[206]
		Research on Deep Learning Model of Feature Extraction	[207]
	Novel Classification Methods	FINE_DENSEIGANET: Hybrid Deep Learning for Chest CT	[208]
		Analysis on Ensemble Learning Optimized Medical Image Classification	[72]
		Machine Learning for Medical Imaging: Intelligent Imaging	[209]
		Deep Convolution Neural Network for Big Data Medical Image Classification	[210]
		Special Section Guest Editorial: Radiomics and Deep Learning	[211]
		Recent Applications of Deep Learning Algorithms in Medical Image Analysis	[212]
		Biomedical Imaging and Analysis in the Age of Big Data and Deep Learning	[213]
		Deep Learning in Medical Imaging: General Overview	[214]

The protection of patient data privacy represents one of the most critical and rapidly evolving challenges in the clinical translation of deep learning. Three studies in our review directly addressed this challenge through distinct methodological approaches. Federated learning has emerged as a leading paradigm for privacy-preserving training, and one study demonstrated its efficacy for brain tumor detection by training a deep learning model across distributed institutions without requiring a centralized dataset [125]. This approach ensures that raw patient data never leaves the local hospital, with only model updates being shared, thereby respecting data sovereignty regulations such as HIPAA and GDPR. A complementary study investigated the application of differential privacy to medical imaging deep learning, which adds carefully calibrated noise to the training process to provably limit the information that can be inferred about any individual patient from the final model [172]. This study demonstrated that while differential privacy introduces a trade-off with model accuracy, the degradation could be managed to achieve both strong privacy guarantees and clinically acceptable performance. Another study presented a hybrid deep learning framework for privacy-preserving time-series medical images analysis, which combined local feature extraction with an encryption layer to protect patient data during transmission and processing [173]. Two additional studies explored the related area of medical image encryption and decryption using deep learning, proposing a stream cipher generator and a feature-encoding framework to protect medical images both at rest and during transmission [174], [175].

The environmental impact of deep learning, particularly the carbon footprint of training large models, has recently become a pressing systemic concern. One study directly addressed this issue by calculating the carbon footprint of selecting and training deep learning models for medical image analysis [177]. The study found that training a single large model, especially when combined with extensive hyperparameter search and model selection, could produce a significant amount of CO2 emissions. This work calls attention to the need for energy-efficient algorithms and training strategies, and it argues that the environmental cost of model development should be considered alongside performance metrics when evaluating the societal impact of medical AI.

The development of robust benchmarks, standardized evaluation protocols, and appropriate validation frameworks is essential for ensuring that reported performance translates into real-world clinical reliability. One study introduced the MedMNIST v2 dataset collection, a large-scale lightweight benchmark for 2D and 3D biomedical image classification, providing a standardized collection of diverse medical image datasets to facilitate reproducible comparisons across different deep learning methods [178]. A related study focused on the standardization of medical imaging workflows for radiomics and deep learning, proposing guidelines for image acquisition, processing, and feature extraction to minimize variability and improve the generalizability of models across different institutions [179]. The impact of scanner domain shift on deep learning performance was systematically investigated in an experimental study, which demonstrated that models trained on data from one MRI scanner can exhibit significant performance degradation when applied to data from a different manufacturer, highlighting the critical need for domain adaptation strategies and standardized acquisition protocols [180]. Addressing a more practical but important challenge, a study on practical window setting optimization for medical image deep learning provided guidelines for setting the window level and window width for CT images to ensure consistent input for deep learning models [181]. Finally, a study on the common limitations of performance metrics in biomedical image analysis provided a critical examination of commonly used metrics such as Dice score, sensitivity, and specificity, arguing that these metrics can be misleading when used inappropriately or without consideration for the underlying clinical task, and providing recommendations for more robust evaluation practices [182].

To support the entire ecosystem of deep learning for medical imaging, a set of studies focused on the development of specialized tooling, infrastructure, and data management practices that enhance reproducibility and lower barriers to entry. The creation of efficient annotation tools is crucial, and one study presented RIL-Contour, a medical imaging dataset annotation tool designed specifically for and with deep learning, which integrated active learning principles to accelerate the labeling process [183]. Another study introduced PyMIC, a deep learning toolkit for annotation-efficient medical image segmentation, which provided modular implementations of state-of-the-art weakly-supervised, semi-supervised, and interactive segmentation methods [184]. For researchers seeking user-friendly environments, DeepImageJ was presented as a platform that allows biomedical scientists to run pre-trained deep learning models within the familiar ImageJ ecosystem, thereby democratizing access to deep learning for a broader community of life science researchers [186]. The importance of proper data management was addressed by a whitepaper on medical image data and datasets in the era of machine learning [187], and a subsequent study provided practical guidance on preparing medical imaging data for machine learning, covering the entire pipeline from data acquisition and de-identification to preprocessing and storage [188].

Generalizability and robustness to domain shift remain among the most formidable barriers to clinical translation. The impact of scanner domain shift was directly demonstrated in the study described above, which found that models trained on data from one MRI scanner could exhibit significant performance drops when applied to data from another manufacturer [180]. To address this challenge, a study on automated knowledge transfer for medical image segmentation using deep learning proposed a transfer learning framework that could adapt a model to a new target dataset with minimal additional annotation, leveraging the knowledge learned from a source dataset [189]. This approach is closely related to the broader concept of domain adaptation, which seeks to align the feature distributions between the source and target domains to improve generalization.

A significant portion of the literature in this dimension is dedicated to identifying and characterizing the systemic and foundational challenges that define the current limitations and future directions of the field. Several studies provided high-level overviews that synthesized the major challenges. One study discussed deep learning for medical image processing from an overview perspective, cataloging the challenges and painting a picture of the future [176]. Another review offered a comprehensive guide to the concepts, methods, challenges, and future directions of deep learning in medical image analysis [190]. A specific focus on prospects was provided by a paper that examined the promise and the remaining hurdles of deep learning for medical imaging [191]. A guest editorial on deep learning in medical imaging provided an overview and described the future promise of this exciting technique [192]. The inseparable link between explainability and trust was the central theme of a study focused on the challenges of deep learning, which argued that improving these aspects is a prerequisite for clinical adoption [103]. An overview of generalizable and explainable deep learning for medical image computing provided a framework for combining these two critical properties[104]. A holistic overview of deep learning in medical imaging [193] provided a bird's-eye view of the field. A study cataloging methodological failures in machine learning for medical imaging provided a sobering list of common pitfalls, such as data leakage, insufficient validation, and inappropriate metric choices, and made concrete recommendations for improving research practice [106].

Modality-specific challenges were addressed by several studies. A comprehensive overview of deep learning in radiology discussed the specific opportunities and obstacles for integrating deep learning into the radiology workflow [194]. A study on machine learning for medical ultrasound provided a status report and discussed the unique challenges posed by the low signal-to-noise ratio and operator dependence of ultrasound imaging [195]. The application of image analysis and machine learning in digital pathology was reviewed, highlighting the computational challenges of gigapixel whole-slide images and the need for specialized hardware and algorithms [196]. A study on machine (deep) learning methods for image processing and radiomics explored the intersection of these two methodologies [197]. A focused study on deep learning in medical image registration provoked a critical question, asking whether the deep learning methods for registration represent a true advance or a mirage, cautioning against over-interpretation of results and advocating for rigorous validation [198]. The ability of deep learning to solve underdetermined inverse problems in medical imaging was explored in another study, which discussed the theoretical and practical challenges of using deep learning for tasks like image reconstruction from sparse data [199]. The application and challenges of deep learning in the intelligent processing of medical images were summarized [200], and a study focused on enhancing radiomics and deep learning systems through the standardization of medical imaging workflows [179] enhancing radiomics and deep learning systems through standardization [201]. Manifold learning of brain MRIs by deep learning [202] and an exploration of the prospects for a (deep) learning health care system [203] rounded out the foundational discussions.

Finally, several studies addressed the practical constraints of training data requirements and clinical translation. A study on the trade-off between training and testing ratio in machine learning for medical image processing provided experimental evidence for optimal dataset splits in the context of limited data [204]. The relationship between dataset size and model performance was further explored in a study that examined the state of the art of deep learning models and their challenges, noting that many published models are trained and validated on datasets that are too small or too homogeneous to guarantee real-world performance [205]. Several theme studies also contributed to the domain of novel classification methods and feature extraction. Advancements in feature extraction recognition of medical imaging systems through deep learning were reviewed [206], and a dedicated study on deep learning models for feature extraction based on convolutional neural networks was also included [207]. Novel classification approaches included FINE_DENSEIGANET, a hybrid deep learning framework for automatic medical image classification in chest CT scans [208], and an analysis on ensemble learning optimized medical image classification with deep convolutional neural networks [72]. Broader perspective pieces on machine learning and deep learning in medical imaging [209], deep convolution neural networks for big data medical image classification [210], a special section guest editorial on radiomics and deep learning [211], a review of recent applications of deep learning algorithms in medical image analysis [212], an introduction to biomedical imaging and analysis in the age of big data and deep learning [213], and a general overview of deep learning in medical imaging [214] constituted the remaining works in this dimension.

Discussion

Synthesizing the key findings from this systematic literature review reveals a field that has moved decisively beyond the initial wave of proof-of-concept studies and is now grappling with the complex, multi-faceted challenges of clinical translation. Taken together, the evidence across all eight thematic dimensions indicates that the primary bottleneck for deep learning in medical image analysis has shifted from raw algorithmic performance to the systemic issues of data scarcity, model interpretability, robustness, and deployability. This synthesis integrates the findings from architectural evolution, task-specific solutions, learning paradigms, and emerging concerns, revealing a field that is becoming increasingly sophisticated and problem-aware. We find that while convolutional neural networks, particularly the U-Net family, have provided a foundational paradigm that remains dominant, there is a clear and accelerating trend toward hybrid architectures that combine convolutional backbones with attention mechanisms or transformer components [21]. Furthermore, these architectural choices are not independent of the other dimensions; rather, they are deeply intertwined with the prevalent learning paradigms, with transfer learning being the single most important strategy for achieving high performance with limited data [25]. This interplay between architecture and data efficiency is a consistent pattern that emerges across many studies, reinforcing the idea that model design and training strategy must be considered as a unified optimization problem rather than separate components.

The growing emphasis on explainability and uncertainty quantification represents another critical synthesis emerging from this review. Consistently found across the literature is the recognition that clinical trust is not solely a function of high accuracy but depends equally on a model's ability to justify its decisions and express its own uncertainty [103]. This finding is consequential because it moves the field beyond the narrow pursuit of state-of-the-art performance on benchmarks toward a more holistic understanding of what constitutes a clinically useful and safe model. The studies on hidden stratification [115] and fairness-related performance disparities [109] further underscore this point, demonstrating that aggregate performance metrics can mask systematic failures that have direct and serious implications for patient care. Therefore, the challenge of achieving generalizable and trustworthy models is not merely a technical one but requires a fundamental rethinking of how we validate and evaluate deep learning systems for medical imaging.

The implications of these findings for both theory and practice are substantial. From a theoretical perspective, the synthesis of architectural innovations with learning paradigms suggests a need for more integrated conceptual frameworks that model the entire learning system, including data, architecture, training objective, and deployment environment, as a single entity. The interplay between architectural design and data efficiency, the growing importance of model interpretability, and the pressing concerns about privacy and domain shift all point to a field that is maturing beyond its initial successes and confronting the harder problems that define the gap between a published paper and a deployed clinical tool.

The limitations of this review must be acknowledged to appropriately contextualize its conclusions. Our search strategy, while comprehensive across five major databases, may still have introduced a selection bias by excluding non-English language publications and preprint articles that had not yet undergone peer review. This decision, made to ensure a baseline quality standard, may have excluded novel but validated methods or important findings from non-English-speaking research communities. Furthermore, the thematic categorization, while designed to be exhaustive and grounded in the literature, is inherently subjective and may have led to the placement of some studies into dimensions that do not fully capture their breadth. The quality assessment of the included studies was also based on our own judgment rather than a formalized scoring system, which could introduce interpretative bias. The dominance of certain research areas, such as brain tumor analysis and COVID-19 detection, within the clinical applications dimension may reflect publication patterns and funding priorities rather than the actual distribution of clinical needs. Generalizing our findings across all imaging modalities and clinical tasks should therefore be done with caution, as the challenges and solutions for ultrasound imaging, for instance, differ significantly from those for CT or MRI.

Building on the gaps and inconsistencies uncovered in this review, several promising avenues for future research clearly emerge. There is a pressing need for more rigorous and standardized validation protocols that move beyond benchmark datasets and evaluate models in multi-site, multi-scanner, and multi-population settings that truly reflect the heterogeneity of real-world clinical data. Future research should explore domain generalization and domain adaptation techniques that can guarantee robust performance under the inevitable domain shifts encountered in practice, moving from best-effort performance to guaranteed levels of robustness. The integration of privacy-preserving technologies, such as federated learning and differential privacy, into large-scale clinical trials is another critical area that requires immediate attention; we need to demonstrate at scale that these methods are not only secure but also produce models of clinically acceptable quality. Another understudied area is the systematic evaluation of the combined impact of data augmentation, synthetic data generation, and self-supervised pre-training on model generalization, particularly for rare diseases where data is extremely limited. Future work should also explore the development of uncertainty-aware clinical workflows where model outputs are only one input among many in a human-machine collaborative decision-making process, rather than a black-box output. The calibration of uncertainty estimates to clinical risk is a specific area where computational methods and clinical reasoning need to be more closely aligned. Furthermore, there is a need for more longitudinal studies that track model performance over time as imaging technology and patient populations evolve, to understand the decay or maintenance of model accuracy. We found a dearth of research on the economic and operational costs of deploying and maintaining deep learning systems in clinical settings, including the computational resources, specialist personnel, and quality assurance protocols required. Finally, the field must confront the ethical and confront the potential for algorithmic bias and ensure that models are evaluated for fairness across demographic, socioeconomic, and geographic, and socioeconomic subgroups before deployment, integrating fairness metrics into the core model development pipeline rather than treating them as an afterthought.

Conclusion

This systematic literature review provided a comprehensive and multi-dimensional synthesis of the deep learning landscape for medical image analysis, mapping 213 studies across eight interconnected thematic axes. Our core contribution is the articulation of how architectural evolution, dominated by CNNs but increasingly incorporating hybrid and transformer-based designs, is inextricably linked to the persistent challenge of medical data scarcity, which has driven the widespread adoption of transfer learning, self-supervised learning, and weak supervision paradigms. We have demonstrated that a critical juncture has been reached where the primary barriers to clinical translation have shifted from achieving high accuracy on benchmarks to confronting the more complex demands of model explainability, uncertainty quantification, robustness to domain shift, and the preservation of patient privacy. Our synthesis underscores the finding that these dimensions are not independent but rather form a tightly coupled system; for instance, a lack of interpretability undermines clinical trust, while an inability to quantify uncertainty creates safety risks in high-stakes decisions, and poor generalization across scanner types renders a model unusable in diverse clinical settings. Hence, the field's future progress hinges on a holistic approach that treats model design, data efficiency, and validation as a unified socio-technical challenge, not a collection of separate problems.

From a practical standpoint, our findings imply that benchmarking efforts and validation protocols must evolve to incorporate multi-site, multi-population evaluations that explicitly test for generalizability, fairness, and trustworthiness, moving beyond simple aggregate performance metrics. Theoretically, our taxonomy reveals a need for integrated frameworks that model the entire learning system including the data, architecture, training paradigm, and deployment context as a single, co-adaptive entity rather than isolated components. For future research, the pressing need is to systematically bridge the gap between algorithmic innovation and clinical reality. Priorities should include developing domain adaptation methods with guaranteed performance bounds, scaling privacy-preserving learning to multi-institutional trials, and creating standardized workflows that integrate uncertainty estimation into clinical decision-making. Furthermore, longitudinal and post-market surveillance studies that track model performance degradation over time are critically absent from the literature and are essential to ensure that the promise of deep learning in medical imaging translates into sustained, safe, and equitable clinical impact.

References

[1] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, 2017.

[2] G. Litjens, T. Kooi, B. Bejnordi, A. Setio, F. Ciompi, *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, 2017.

[3] T. Pun, G. Gerig, and O. Ratib, "Image analysis and computer vision in medicine," *Computerized Medical Imaging and Graphics*, 1994.

[4] S. Kshatri and D. Singh, "Convolutional neural network in medical image analysis: A review," *Archives of Computational Methods in Engineering*, 2023.

[5] S. Takahashi, Y. Sakaguchi, N. Kouno, *et al.*, "Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review," *Journal of Medical Systems*, 2024.

[6] S. Krishnapriya and Y. Karuna, "A survey of deep learning for MRI brain tumor segmentation methods: Trends, challenges, and future directions," *Health and Technology*, 2023.

[7] M. Page, J. McKenzie, P. Bossuyt, *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.

[8] G. Chetty, M. Yamin, and M. White, "A low resource 3D u-net based deep learning model for medical image analysis," *International Journal of Information Technology*, 2022.

[9] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision*, 2016.

[10] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Medical image analysis*, 2018.

[11] Q. Dou *et al.*, "3D deeply supervised network for automated segmentation of volumetric medical images," *Medical Image Analysis*, 2017.

[12] M. Vakalopoulou, G. Chassagnon, N. Bus, *et al.*, "Atlasnet: Multi-atlas non-linear deep networks for medical image segmentation," in *International conference on medical image computing and computer-assisted intervention*, 2018.

[13] Z. Zhou, M. R. Siddiquee, N. Tajbakhsh, *et al.*, "Unet++: A nested u-net architecture for medical image segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018.

[14] D. Jha, P. Smedsrud, M. Riegler, *et al.*, "Resunet++: An advanced architecture for medical image segmentation," in *IEEE international symposium on multimedia*, 2019.

[15] M. Alom, M. Hasan, C. Yakopcic, T. Taha, *et al.*, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," arXiv preprint arXiv:1802.06955, 2018.

[16] T. Hussain and H. Shouno, "MAGRes-UNet: Improved medical image segmentation through a deep learning paradigm of multi-attention gated residual u-net," *Ieee Access*, 2024.

[17] P. Yin, R. Yuan, Y. Cheng, and Q. Wu, "Deep guidance network for biomedical image segmentation," *IEEE access*, 2020.

[18] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, *et al.*, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, 2019.

[19] S. Iqbal, T. Khan, S. Naqvi, A. Naveed, and E. Meijering, "TBConvL-net: A hybrid deep learning architecture for robust medical image segmentation," *Pattern Recognition*, 2025.

[20] D. Jha, M. Riegler, D. Johansen, *et al.*, "Doubleu-net: A deep convolutional neural network for medical image segmentation," in *2020 IEEE 33rd international symposium on computer-based medical systems*, 2020.

[21] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, *et al.*, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*, 2022.

[22] M. Willemink, H. Roth, and V. Sandfort, "Toward foundational deep learning models for medical imaging in the new era of transformer networks," *Radiology: Artificial Intelligence*, 2022.

[23] C. Tang, C. Hu, J. Sun, and H. Sima, "Deep learning-based medical images analysis evolved from convolution to graph convolution," *Journal of Image and Graphics*, 2021.

[24] R. Selvan and E. Dam, "Tensor networks for medical image classification," *Medical imaging with deep learning*, 2020.

[25] N. Tajbakhsh, J. Shin, S. Gurudu, *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Transactions on Medical Imaging*, 2016.

[26] S. Chen, K. Ma, and Y. Zheng, "Med3d: Transfer learning for 3d medical image analysis," arXiv preprint arXiv:1904.00625, 2019.

[27] P. Dutta, P. Upadhyay, M. De, *et al.*, "Medical image analysis using deep convolutional neural networks: CNN architectures and transfer learning," in *International conference on inventive computation technologies*, 2020.

[28] F. Isensee, P. Jäger, S. Kohl, J. Petersen, *et al.*, "Automated design of deep learning methods for biomedical image segmentation," arXiv preprint arXiv:1904.08128, 2019.

[29] D. Müller and F. Kramer, "MIScnn: A framework for medical image segmentation with convolutional neural networks and deep learning," *BMC medical imaging*, 2021.

[30] N. Pawlowski, S. Ktena, M. Lee, B. Kainz, *et al.*, "Dltk: State of the art reference implementations for deep learning on medical images," arXiv preprint arXiv:1711.06853, 2017.

[31] Z. Guo, X. Li, H. Huang, N. Guo, *et al.*, "Deep learning-based image segmentation on multimodal medical imaging," *IEEE Transactions on Medical Imaging*, 2019.

[32] M. Lai, "Deep learning for medical image segmentation," arXiv preprint arXiv:1505.02000, 2015.

[33] M. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of digital imaging*, 2019.

[34] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, *et al.*, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, 2017.

[35] S. Addimulam, M. Mohammed, R. Karanam, *et al.*, "Deep learning-enhanced image segmentation for medical diagnostics," academia.edu, 2020.

[36] B. A. Skourt, A. E. Hassani, and A. Majda, "Lung CT image segmentation using deep neural networks," *Procedia Computer Science*, 2018.

[37] K. Fritscher, P. Raudaschl, P. Zaffino, M. Spadea, *et al.*, "Deep neural networks for fast segmentation of 3D medical images," in *International conference on medical image computing and computer-assisted intervention*, 2016.

[38] H. Roth, A. Farag, L. Lu, E. Turkbey, *et al.*, "Deep convolutional networks for pancreas segmentation in CT imaging," … Imaging, 2015.

[39] P. Moeskops, J. Wolterink, *et al.*, "Deep learning for multi-task medical image segmentation in multiple modalities," in *International conference on medical image computing and computer-assisted intervention*, 2016.

[40] O. Bernard, A. Lalande, C. Zotti, *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, 2018.

[41] M. Akil, R. Saouli, and R. Kachouri, "Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy," *Medical image analysis*, 2020.

[42] Z. Zhao, L. Yang, H. Zheng, I. Guldner, S. Zhang, *et al.*, "Deep learning based instance segmentation in 3D biomedical images using weak annotation," in *International conference on medical image computing and computer-assisted intervention*, 2018.

[43] L. Mou *et al.*, "CS2-net: Deep learning segmentation of curvilinear structures in medical imaging," *Medical Image Analysis*, 2021.

[44] S. He, Y. Feng, P. Grant, and Y. Ou, "Segmentation ability map: Interpret deep features for medical image segmentation," *Medical image analysis*, 2023.

[45] J. Hu, H. Wang, J. Wang, Y. Wang, F. He, and J. Zhang, "SA-net: A scale-attention network for medical image segmentation," *PloS one*, 2021.

[46] G. Wang, T. Song, Q. Dong, M. Cui, N. Huang, *et al.*, "Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks," *Medical Image Analysis*, 2020.

[47] T. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Medical image analysis*, 2017.

[48] M. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Medical image analysis*, 2016.

[49] N. Dhungel, G. Carneiro, and A. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," *Medical image analysis*, 2017.

[50] F. Kong, N. Wilson, and S. Shadden, "A deep-learning approach for direct whole-heart mesh reconstruction," *Medical image analysis*, 2021.

[51] K. Girum, G. Créhange, R. Hussain, *et al.*, "Fast interactive medical image segmentation with weakly supervised deep learning method," *International Journal of Computer Assisted Radiology and Surgery*, 2020.

[52] Y. Can, K. Chaitanya, B. Mustafa, L. Koch, *et al.*, "Learning to segment medical images with scribble-supervision alone," in *Workshop on deep learning in medical image analysis*, 2018.

[53] L. Yang, Y. Zhang, J. Chen, S. Zhang, *et al.*, "Suggestive annotation: A deep active learning framework for biomedical image segmentation," in *International conference on medical image computing and computer-assisted intervention*, 2017.

[54] J. Sourati, A. Gholipour, J. Dy, *et al.*, "Intelligent labeling based on fisher information for medical image segmentation using deep learning," *IEEE Transactions on Medical Imaging*, 2019.

[55] B. D. Vos, F. Berendsen, M. Viergever, *et al.*, "A deep learning framework for unsupervised affine and deformable image registration," *Medical Image Analysis*, 2019.

[56] K. Chaitanya, E. Erdil, N. Karani, *et al.*, "Contrastive learning of global and local features for medical image segmentation with limited annotations," in *Advances in neural information processing systems*, 2020.

[57] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, *et al.*, "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," in *International conference on medical image computing and computer-assisted intervention*, 2017.

[58] G. Wang, W. Li, M. Zuluaga, R. Pratt, *et al.*, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Transactions on Medical Imaging*, 2018.

[59] G. Wang, M. Zuluaga, W. Li, R. Pratt, *et al.*, "DeepIGeoS: A deep interactive geodesic framework for medical image segmentation," *IEEE Transactions on Medical Imaging*, 2018.

[60] N. Koohbanani, M. Jahanifar, N. Tajadin, *et al.*, "NuClick: A deep learning framework for interactive segmentation of microscopic images," *Medical Image Analysis*, 2020.

[61] T. Nair, D. Precup, D. Arnold, and T. Arbel, "Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation," *Medical image analysis*, 2020.

[62] J. Wang and T. Lukasiewicz, "Rethinking bayesian deep learning methods for semi-supervised volumetric medical image segmentation," in *IEEE/CVF conference on computer vision and pattern recognition*, 2022.

[63] N. Dimitriou, O. Arandjelović, and P. Caie, "Deep learning for whole slide image analysis: An overview," *Frontiers in medicine*, 2019.

[64] S. Wang *et al.*, "RMDL: Recalibrated multi-instance deep learning for whole slide gastric image classification," *Medical Image Analysis*, 2019.

[65] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, *et al.*, "Weakly supervised deep learning for whole slide lung cancer image analysis," *IEEE Transactions on Cybernetics*, 2019.

[66] I. Oksuz, J. Clough, B. Ruijsink, *et al.*, "Deep learning-based detection and correction of cardiac MR motion artefacts during reconstruction for high-quality segmentation," *IEEE Transactions on Medical Imaging*, 2020.

[67] S. Thomas, J. Lefevre, G. Baxter, and N. Hamilton, "Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer," *Medical Image Analysis*, 2021.

[68] N. Gessert, M. Schlüter, and A. Schlaefer, "A deep learning approach for pose estimation from volumetric OCT data," *Medical image analysis*, 2018.

[69] J. Hu, Y. Chen, and Z. Yi, "Automated segmentation of macular edema in OCT using deep neural networks," *Medical image analysis*, 2019.

[70] A. Balagopal, D. Nguyen, H. Morgan, Y. Weng, *et al.*, "A deep learning-based framework for segmenting invisible clinical target volumes with estimated uncertainties for post-operative prostate cancer radiotherapy," *Medical Image Analysis*, 2021.

[71] A. Simpson, M. Antonelli, S. Bakas, M. Bilello, *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," arXiv preprint arXiv:1902.09063, 2019.

[72] D. Müller, I. Soto-Rey, and F. Kramer, "An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks," *IEEE access*, 2022.

[73] H. Messaoudi, A. Belaid, D. Salem, and P. Conze, "Cross-dimensional transfer learning in medical image segmentation with deep learning," *Medical image analysis*, 2023.

[74] D. Zhang *et al.*, "Understanding the tricks of deep learning in medical image segmentation: Challenges and future directions," arXiv preprint arXiv:2209.10307, 2022.

[75] Y. Lin, D. Zhang, X. Fang, Y. Chen, K. Cheng, *et al.*, "Rethinking boundary detection in deep learning models for medical image segmentation," *Information Processing in Medical Imaging*, 2023.

[76] B. D. Vos, F. Berendsen, M. Viergever, *et al.*, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *International workshop on deep learning in medical image analysis*, 2017.

[77] K. Islam, S. Wijewickrema, and S. O'Leary, "A deep learning based framework for the registration of three dimensional multi-modal medical images of the head," *Scientific Reports*, 2021.

[78] Y. Fu, N. Brown, S. Saeed, A. Casamitjana, *et al.*, "DeepReg: A deep learning toolkit for medical image registration," arXiv preprint arXiv:2011.02580, 2020.

[79] J. Zhang, M. Liu, and D. Shen, "Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks," *IEEE Transactions on Image Processing*, 2017.

[80] J. Noothout, B. D. Vos, J. Wolterink, *et al.*, "Deep learning-based regression and classification for automatic landmark localization in medical images," *IEEE Transactions on Medical Imaging*, 2020.

[81] A. Wu, Z. Xu, M. Gao, M. Buty, *et al.*, "Deep vessel tracking: A generalized probabilistic approach via deep learning," in *IEEE 13th international symposium on biomedical imaging*, 2016.

[82] G. Carneiro, J. Nascimento, and A. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," in *International conference on medical image computing and computer-assisted intervention*, 2015.

[83] Y. Bar, I. Diamant, L. Wolf, *et al.*, "Deep learning with non-medical training used for chest pathology identification," *Medical Imaging*, 2015.

[84] G. An, M. Akiba, K. Omodaka, T. Nakazawa, and H. Yokota, "Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images," *Scientific reports*, 2021.

[85] W. Wang, Y. Iwamoto, X. Han, Y. Chen, *et al.*, "Classification of focal liver lesions using deep learning with fine-tuning," in *Proceedings of the international conference on medical image computing and computer-assisted intervention*, 2018.

[86] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, *et al.*, "Deep-COVID: Predicting COVID-19 from chest x-ray images using deep transfer learning," *Medical image analysis*, 2020.

[87] V. Iglovikov, A. Rakhlin, A. Kalinin, *et al.*, "Paediatric bone age assessment using deep convolutional neural networks," in *International workshop on machine learning in medical imaging*, 2018.

[88] C. Perone and J. Cohen-Adad, "Deep semi-supervised segmentation with weight-averaged consistency targets," in *International workshop on deep learning in medical image analysis*, 2018.

[89] K. Thung, P. Yap, and D. Shen, "Multi-stage diagnosis of alzheimer's disease with incomplete multimodal data via multi-task deep learning," in *International workshop on deep learning in medical image analysis*, 2017.

[90] N. Tajbakhsh, Y. Hu, J. Cao, X. Yan, Y. Xiao, *et al.*, "Surrogate supervision for medical image analysis: Effective deep learning from limited quantities of labeled data," in *2019 IEEE 16th international symposium on biomedical imaging*, 2019.

[91] S. Huang, L. Shen, M. Lungren, *et al.*, "Gloria: A multimodal global-local representation learning framework for label-efficient medical image recognition," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021.

[92] D. Wolf, T. Payer, C. Lisson, C. Lisson, M. Beer, *et al.*, "Self-supervised pre-training with contrastive and masked autoencoder methods for dealing with small datasets in deep learning for medical imaging," *Scientific Reports*, 2023.

[93] E. Ahn, A. Kumar, D. Feng, M. Fulham, and J. Kim, "Unsupervised feature learning with k-means and an ensemble of deep convolutional neural networks for medical image classification," arXiv preprint arXiv:1906.03359, 2019.

[94] Z. Tang, K. Chen, M. Pan, M. Wang, and Z. Song, "An augmentation strategy for medical image processing based on statistical shape model and 3D thin plate spline for deep learning," *IEEE Access*, 2019.

[95] M. Rubin, O. Stein, N. Turko, Y. Nygate, D. Roitshtain, *et al.*, "TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set," *Medical Image Analysis*, 2019.

[96] D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, "Estimating CT image from MRI data using 3D fully convolutional networks," in *International workshop on deep learning in medical image analysis*, 2016.

[97] D. Karimi, H. Dou, S. Warfield, and A. Gholipour, "Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis," *Medical image analysis*, 2020.

[98] J. Cho, K. Lee, E. Shin, G. Choy, and S. Do, "How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?" arXiv preprint arXiv:1511.06348, 2015.

[99] C. Jin, Z. Guo, Y. Lin, L. Luo, and H. Chen, "Label-efficient deep learning in medical image analysis: Challenges and future directions," arXiv preprint arXiv:2303.12484, 2023.

[100] K. Raza and N. Singh, "A tour of unsupervised deep learning for medical image analysis," *Current Medical Imaging Reviews*, 2021.

[101] J. Yang, R. Shi, and B. Ni, "Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis," in *2021 IEEE 18th international symposium on biomedical imaging*, 2021.

[102] E. Pooch, P. Ballester, and R. Barros, "Can we trust deep learning based diagnosis? The impact of domain shift in chest radiograph classification," in *International workshop on thoracic image analysis*, 2020.

[103] T. Dhar, N. Dey, S. Borra, *et al.*, "Challenges of deep learning in medical image analysis—improving explainability and trust," *IEEE Transactions on Medical Imaging*, 2023.

[104] A. Chaddad, Y. Hu, Y. Wu, B. Wen, and R. Kateb, "Generalizable and explainable deep learning for medical image computing: An overview," *Current Opinion in Biomedical Engineering*, 2025.

[105] D. Huff, A. Weisman, and R. Jeraj, "Interpretation and visualization techniques for deep learning models in medical imaging," *Physics in Medicine & Biology*, 2021.

[106] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: Methodological failures and recommendations for the future," *NPJ digital medicine*, 2022.

[107] B. Dwarakanath, G. Shrivastava, R. Bansal, *et al.*, "Explainable machine learning techniques in medical image analysis based on classification with feature extraction," *International Journal of Advanced Computer Science and Applications*, 2022.

[108] J. Thiagarajan, K. Thopalli, D. Rajan, and P. Turaga, "Training calibration-based counterfactual explainers for deep learning models in medical image analysis," *Scientific reports*, 2022.

[109] E. Stanley, M. Wilms, P. Mouches, *et al.*, "Fairness-related performance and explainability effects in deep learning models for brain image analysis," *Journal of Medical Imaging*, 2022.

[110] R. Zeineldin, M. Karar, Z. Elshaer, J. Coburger, *et al.*, "Explainability of deep neural networks for MRI analysis of brain tumors," *International Journal of Computer Assisted Radiology and Surgery*, 2022.

[111] X. Ma *et al.*, "Understanding adversarial attacks on deep learning based medical image analysis systems," *Pattern recognition*, 2021.

[112] U. Ozbulak, A. V. Messem, and W. D. Neve, "Impact of adversarial examples on deep learning models for biomedical image segmentation," in *International conference on medical image computing and computer-assisted intervention*, 2019.

[113] H. Hirano, A. Minagi, and K. Takemoto, "Universal adversarial attacks on deep neural networks for medical image classification," *BMC medical imaging*, 2021.

[114] E. Puyol-Antón, B. Ruijsink, S. Piechnik, *et al.*, "Fairness in cardiac MR image analysis: An investigation of bias due to data imbalance in deep learning based segmentation," in *International conference on medical image computing and computer-assisted intervention*, 2021.

[115] L. Oakden-Rayner, J. Dunnmon, G. Carneiro, *et al.*, "Hidden stratification causes clinically meaningful failures in machine learning for medical imaging," in *Proceedings of the ACM conference on health, inference, and learning*, 2020.

[116] S. Rajaraman, P. Ganesan, and S. Antani, "Deep learning model calibration for improving performance in class-imbalanced medical image classification tasks," *PloS one*, 2022.

[117] M. Ayhan and P. Berens, "Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks," *Medical imaging with deep learning*, 2018.

[118] S. Faghani, M. Moassefi, P. Rouzrokh, B. Khosravi, *et al.*, "Quantifying uncertainty in deep learning of radiologic images," *Radiology*, 2023.

[119] T. Araújo, G. Aresta, L. Mendonça, S. Penas, C. Maia, *et al.*, "DR| GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images," *Medical Image Analysis*, 2020.

[120] M. Ayhan, L. Kühlewein, G. Aliyeva, W. Inhoffen, *et al.*, "Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection," *Medical Image Analysis*, 2020.

[121] L. Herzog, E. Murina, O. Dürr, S. Wegener, and B. Sick, "Integrating uncertainty in deep neural networks for MRI based stroke analysis," *Medical image analysis*, 2020.

[122] D. Gore and V. Deshpande, "Comparative study of various techniques using deep learning for brain tumor detection," in *2020 international conference for innovation in technology*, 2020.

[123] K. Qodri, I. Soesanti, and H. Nugroho, "Image analysis for MRI-based brain tumor classification using deep learning," *IJITEE*, 2021.

[124] A. Methil, "Brain tumor detection using deep learning and image processing," in *International conference on artificial intelligence and smart systems*, 2021.

[125] L. Zhou, M. Wang, and N. Zhou, "Distributed federated learning-based deep learning model for privacy mri brain tumor detection," arXiv preprint arXiv:2404.10026, 2024.

[126] H. Rai and K. Chatterjee, "2D MRI image analysis and brain tumor detection using deep learning CNN model LeU-net," *Multimedia Tools and Applications*, 2021.

[127] X. Liu and Z. Wang, "Deep learning in medical image classification from mri-based brain tumor images," in *2024 IEEE 6th international conference on artificial intelligence circuits and systems*, 2024.

[128] S. Bamber and T. Vishvakarma, "Medical image classification for alzheimer's using a deep learning approach," *Journal of Engineering and Applied Science*, 2023.

[129] J. Wingate, I. Kollia, L. Bidaut, and S. Kollias, "Unified deep learning approach for prediction of parkinson's disease," *IET Image Processing*, 2020.

[130] D. Robben, A. Boers, H. Marquering, *et al.*, "Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning," *Medical Image Analysis*, 2020.

[131] T. Phong, H. Duong, H. Nguyen, N. Trong, *et al.*, "Brain hemorrhage diagnosis by using deep learning," in *Proceedings of the international conference on information integration and web-based applications & services*, 2017.

[132] A. Tahmassebi, A. Gandomi, I. McCann, *et al.*, "Deep learning in medical imaging: Fmri big data analysis via convolutional neural networks," in *Proceedings of the international conference on bioinformatics, computational biology and health informatics*, 2018.

[133] I. Ahmed, M. Ahmad, and G. Jeon, "Integrating digital twins and deep learning for medical image analysis in the era of COVID-19," *Virtual Reality & Intelligent Hardware*, 2022.

[134] H. Farhat, G. Sakr, and R. Kilany, "Deep learning applications in pulmonary medical imaging: Recent updates and insights on COVID-19: H. Farhat et al." *Machine vision and applications*, 2020.

[135] D. Yang, C. Martinez, L. Visuña, H. Khandhar, C. Bhatt, *et al.*, "Detection and analysis of COVID-19 in medical images using deep learning techniques," *Scientific Reports*, 2021.

[136] G. Quellec, K. Charriere, Y. Boudi, B. Cochener, *et al.*, "Deep image mining for diabetic retinopathy screening," *Medical Image Analysis*, 2017.

[137] K. Wong, G. Fortino, and D. Abbott, "Deep learning-based cardiovascular image diagnosis: A promising challenge," *Future Generation Computer Systems*, 2020.

[138] L. Lu, L. Dercle, B. Zhao, and L. Schwartz, "Deep learning for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging," *Nature communications*, 2021.

[139] A. Ebigbo, R. Mendel, A. Probst, J. Manzeneder, *et al.*, "Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma," *Gut*, 2019.

[140] D. Wang, A. Khosla, R. Gargeya, H. Irshad, *et al.*, "Deep learning for identifying metastatic breast cancer," arXiv preprint arXiv:1606.05718, 2016.

[141] M. Yap, M. Goyal, F. Osman, R. Martí, *et al.*, "Breast ultrasound lesions recognition: End-to-end deep learning approaches," *Journal of Medical Imaging*, 2019.

[142] J. Lee and K. Kim, "Applying deep learning in medical images: The case of bone age estimation," *Healthcare informatics research*, 2018.

[143] J. Song *et al.*, "Ultrasound image analysis using deep learning algorithm for the diagnosis of thyroid nodules," *Medicine*, 2019.

[144] J. Krois, A. G. Cantu, A. Chaurasia, R. Patil, *et al.*, "Generalizability of deep learning models for dental image analysis," *Scientific reports*, 2021.

[145] S. Masoudi, S. Harmon, S. Mehralivand, *et al.*, "Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research," *Journal of Medical Imaging*, 2021.

[146] G. Thakur, A. Thakur, S. Kulkarni, N. Khan, and S. Khan, "Deep learning approaches for medical image analysis and diagnosis," *Cureus*, 2024.

[147] A. Fourcade and R. Khonsari, "Deep learning in medical image analysis: A third eye for doctors," *Journal of Stomatology, Oral and Maxillofacial Surgery*, 2019.

[148] E. Klang, "Deep learning and medical imaging," *Journal of thoracic disease*, 2018.

[149] D. Ting, Y. Liu, P. Burlina, X. Xu, N. Bressler, *et al.*, "AI for medical imaging goes deep," *Nature medicine*, 2018.

[150] H. Choi, "Deep learning in nuclear medicine and molecular imaging: Current perspectives and future directions," *Nuclear medicine and molecular imaging*, 2018.

[151] R. Mansour, M. Althobaiti, and A. Ashour, "Internet of things and synergic deep learning based biomedical tongue color image analysis for disease diagnosis and classification," *Ieee Access*, 2021.

[152] A. Larrazabal, N. Nieto, V. Peterson, D. Milone, *et al.*, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis," in *Proceedings of the national academy of sciences*, 2020.

[153] M. Abid, R. Ashraf, T. Mahmood, and C. Faisal, "Multi-modal medical image classification using deep residual network and genetic algorithm," *Plos one*, 2023.

[154] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *2016 IEEE 16th international conference on data mining workshops*, 2016.

[155] S. Zhang, G. Liang, S. Pan, and L. Zheng, "A fast medical image super resolution method based on deep learning network," *IEEE Access*, 2018.

[156] Y. Zhang and M. An, "Deep learning-and transfer learning-based super resolution reconstruction from single medical image," *Journal of healthcare engineering*, 2017.

[157] Y. Nakamura, T. Higaki, F. Tatsugami, *et al.*, "Possibility of deep learning in medical imaging focusing improvement of computed tomography image quality," *Journal of Computer Assisted Tomography*, 2020.

[158] Z. Chen, K. Pawar, M. Ekanayake, C. Pain, S. Zhong, *et al.*, "Deep learning for image enhancement and correction in magnetic resonance imaging—state-of-the-art and challenges," *Journal of Digital Imaging*, 2023.

[159] H. Arabi and H. Zaidi, "Deep learning-guided estimation of attenuation correction factors from time-of-flight PET emission data," *Medical image analysis*, 2020.

[160] B. Luijten, R. Cohen, F. D. Bruijn, *et al.*, "Adaptive ultrasound beamforming using deep learning," in *IEEE international conference on medical imaging*, 2020.

[161] F. Mahmood, R. Chen, S. Sudarsky, D. Yu, *et al.*, "Deep learning with cinematic rendering: Fine-tuning deep neural networks using photorealistic medical images," *Physics in Medicine & Biology*, 2018.

[162] K. Armanious, C. Jiang, M. Fischer, T. Küstner, *et al.*, "MedGAN: Medical image translation using GANs," *Computerized Medical Imaging and Graphics*, 2020.

[163] L. Qu, Y. Zhang, S. Wang, P. Yap, and D. Shen, "Synthesized 7T MRI from 3T MRI via deep learning in spatial and wavelet domains," *Medical image analysis*, 2020.

[164] H. Shin, N. Tenenholtz, J. Rogers, *et al.*, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," *Simulation and Synthesis in Medical Imaging*, 2018.

[165] Z. Dorjsembe, S. Odonchimed, and F. Xiao, "Three-dimensional medical image synthesis with denoising diffusion probabilistic models," *Medical Imaging with Deep Learning*, 2022.

[166] S. Kaji and S. Kida, "Overview of image-to-image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in medical imaging," arXiv preprint arXiv:1905.08603, 2019.

[167] G. Wang, J. Ye, K. Mueller, *et al.*, "Image reconstruction is a new frontier of machine learning," *IEEE Transactions on Medical Imaging*, 2018.

[168] X. Chen, N. Pawlowski, M. Rajchl, B. Glocker, *et al.*, "Deep generative models in the real-world: An open challenge from medical imaging," arXiv preprint arXiv:1806.05452, 2018.

[169] B. Rajalingam and R. Priya, "Multimodal medical image fusion based on deep learning neural network for clinical treatment analysis," *International Journal of ChemTech Research*, 2018.

[170] M. Kaur and D. Singh, "Multi-modality medical image fusion technique using multi-objective differential evolution based deep neural networks," *Journal of Ambient Intelligence and Humanized Computing*, 2021.

[171] V. S. Parvathy, S. Pothiraj, *et al.*, "A novel approach in multimodality medical image fusion using optimal shearlet and deep learning," *International Journal of Imaging Systems and Technology*, 2020.

[172] A. Ziller, D. Usynin, R. Braren, M. Makowski, *et al.*, "Medical imaging deep learning with differential privacy," *Scientific Reports*, 2021.

[173] Z. Yue, S. Ding, L. Zhao, Y. Zhang, Z. Cao, *et al.*, "Privacy-preserving time-series medical images analysis using a hybrid deep learning framework," *ACM Transactions on Intelligent Systems and Technology*, 2021.

[174] Y. Ding, F. Tan, Z. Qin, M. Cao, *et al.*, "DeepKeyGen: A deep learning-based stream cipher generator for medical image encryption and decryption," in *International joint conference on neural networks*, 2021.

[175] B. Long, Z. Chen, T. Liu, X. Wu, C. He, and L. Wang, "A novel medical image encryption scheme based on deep learning feature encoding and decoding," *Ieee Access*, 2024.

[176] M. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," *Bioinformatics Applications: Automation of Decision Making*, 2017.

[177] R. Selvan, N. Bhagwat, L. W. Anthony, *et al.*, “Carbon footprint of selecting and training deep learning models for medical image analysis,” in *International conference on medical image computing and computer-assisted intervention*, 2022.

[178] J. Yang *et al.*, “Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification,” *Scientific data*, 2023.

[179] M. Cobo, P. M. Fernández-Miranda, *et al.*, “Enhancing radiomics and deep learning systems through the standardization of medical imaging workflows,” *Scientific data*, 2023.

[180] B. Guo *et al.*, “The impact of scanner domain shift on deep learning performance in medical imaging: An experimental study,” *International Journal of Computer Assisted Radiology and Surgery*, 2026.

[181] H. Lee, M. Kim, and S. Do, “Practical window setting optimization for medical image deep learning,” arXiv preprint arXiv:1812.00572, 2018.

[182] A. Reinke, M. Eisenmann, M. Tizabi, *et al.*, “Common limitations of performance metrics in biomedical image analysis,” *Medical Imaging with Deep Learning*, 2021.

[183] K. Philbrick, A. Weston, Z. Akkus, T. Kline, *et al.*, “RIL-contour: A medical imaging dataset annotation tool for and with deep learning,” *Journal of Digital Imaging*, 2019.

[184] G. Wang *et al.*, “PyMIC: A deep learning toolkit for annotation-efficient medical image segmentation,” *Computer Methods and Programs in Biomedicine*, 2023.

[185] T. Le and K. Luu, “Medical image computing and computer assisted intervention â€“MICCAI 2020,” in *International conference on medical image computing and computer assisted intervention*, 2020.

[186] E. Gómez-de-Mariscal, C. García-López-de-Haro, *et al.*, “DeepImageJ: A user-friendly environment to run deep learning models in ImageJ,” *Nature Methods*, 2021.

[187] M. Kohli, R. Summers, and J. Geis, “Medical image data and datasets in the era of machine learning—whitepaper from the 2016 c-MIMI meeting dataset session,” *Journal of digital imaging*, 2017.

[188] M. Willemink, W. Koszek, C. Hardell, J. Wu, *et al.*, “Preparing medical imaging data for machine learning,” *Radiology*, 2020.

[189] J. Mistry, “Automated knowledge transfer for medical image segmentation using deep learning,” *Journal of xidian University*, 2024.

[190] F. Altaf, S. Islam, N. Akhtar, and N. Janjua, “Going deep in medical image analysis: Concepts, methods, challenges, and future directions,” *Ieee Access*, 2019.

[191] J. Kim, J. Hong, and H. Park, “Prospects of deep learning for medical imaging,” *Precision and Future Medicine*, 2018.

[192] H. Greenspan, B. V. Ginneken, *et al.*, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, 2016.

[193] R. Yousef, G. Gupta, N. Yousef, and M. Khari, "A holistic overview of deep learning approach in medical imaging," *Multimedia Systems*, 2022.

[194] M. McBee, O. Awan, A. Colucci, C. Ghobadi, *et al.*, "Deep learning in radiology," *Academic radiology*, 2018.

[195] L. Brattain, B. Telfer, M. Dhyani, J. Grajo, *et al.*, "Machine learning for medical ultrasound: Status, methods, and future opportunities," *Abdominal radiology*, 2018.

[196] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," *Medical image analysis*, 2016.

[197] M. Hatt, C. Parmar, J. Qi, and I. E. Naqa, "Machine (deep) learning methods for image processing and radiomics," *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2019.

[198] R. Jena, D. Sethi, P. Chaudhari, *et al.*, "Deep learning in medical image registration: Magic or mirage?" in *Advances in neural information processing systems*, 2024.

[199] C. Hyun, S. Baek, M. Lee, S. Lee, and J. Seo, "Deep learning-based solvability of underdetermined inverse problems in medical imaging," *Medical Image Analysis*, 2021.

[200] Y. Zuo, G. Huang, and S. Nie, "Application and challenges of deep learning in the intelligent processing of medical images," *Journal of Image and Graphics*, 2021.

[201] M. Torres-Velázquez, W. Chen, X. Li, *et al.*, "Application and construction of deep learning networks in medical imaging," *IEEE Transactions on Medical Imaging*, 2020.

[202] T. Brosch, R. Tam, *et al.*, "Manifold learning of brain MRIs by deep learning," in *International conference on medical image computing and computer-assisted intervention*, 2013.

[203] C. Naylor, "On the prospects for a (deep) learning health care system," *Jama*, 2018.

[204] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, "Trade-off between training and testing ratio in machine learning for medical image processing," *PeerJ Computer Science*, 2024.

[205] C. Bhatt, I. Kumar, V. Vijayakumar, K. Singh, *et al.*, "The state of the art of deep learning models in medical science and their challenges," *Multimedia Systems*, 2021.

[206] Q. Zhan, D. Sun, E. Gao, Y. Ma, Y. Liang, *et al.*, "Advancements in feature extraction recognition of medical imaging systems through deep learning technique," in *2024 IEEE 2nd international conference on artificial intelligence and computer engineering*, 2024.

[207] H. Liu, I. Li, Y. Liang, D. Sun, Y. Yang, *et al.*, "Research on deep learning model of feature extraction based on convolutional neural network," in *2024 IEEE 2nd international conference on artificial intelligence and computer engineering*, 2024.

[208] H. Sahu and R. Kashyap, "FINE_DENSEIGANET: Automatic medical image classification in chest CT scan using hybrid deep learning framework," *International Journal of Image and Graphics*, 2025.

[209] G. Currie, K. Hawk, E. Rohren, A. Vial, and R. Klein, "Machine learning and deep learning in medical imaging: Intelligent imaging," *Journal of Medical Imaging and Health Informatics*, 2019.

[210] R. Ashraf, M. Habib, M. Akram, M. Latif, M. Malik, *et al.*, "Deep convolution neural network for big data medical image classification," *IEEE Access*, 2020.

[211] D. Kontos, R. Summers, *et al.*, "Special section guest editorial: Radiomics and deep learning." *Journal of Medical Imaging*, 2017.

[212] A. Sarraf, A. Jalali, and J. Ghaffari, "Recent applications of deep learning algorithms in medical image analysis," *Am. Sci. Res. J. Eng. Technol. Sci*, 2020.

[213] J. Duncan, M. Insana, and N. Ayache, "Biomedical imaging and analysis in the age of big data and deep learning [scanning the issue]," in *Proceedings of the IEEE*, 2019.

[214] J. Lee, S. Jun, Y. Cho, H. Lee, *et al.*, "Deep learning in medical imaging: General overview," *Korean Journal of Radiology*, 2017.