Advances in Artificial Intelligence and Machine Learning;Review 5 (4) 4461-4483 Received 08-08-2025; Accepted 29-10-2025; Published online 05-11-2025

Deep Learning Algorithms in Medical Image Processing: A Critical and Comprehensive Review

Mohammed Ahmed Alharbi mahalharbe@kau.edu.sa
Faculty of Computing & Information Technology,
King Abdulaziz University, Saudi Arabia

Morched Derbali mderbali@kau.edu.sa
Faculty of Computing & Information Technology,
King Abdulaziz University, Saudi Arabia

Rayed Alakhtar ralakhtar@kau.edu.sa
Faculty of Computing & Information Technology,
King Abdulaziz University, Saudi Arabia

Mutasem Jarrah m_jarrah@asu.edu.jo
Faculty of Information Technology,
Applied Science Private University (ASU),
Jordan

Corresponding Author: Mohammed Ahmed Alharbi

Copyright © 2025 Mohammed Ahmed Alharbi, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Deep learning has soon taken medical image processing to the cutting-edge, with state-of-the-art performance in classification, segmentation, and anomaly detection. Convolutional neural networks (CNNs) led early breakthroughs, then generative adversarial networks (GANs) for data augmentation and super-resolution, and vision transformers (ViTs) and self-supervised learning (SSL) more recently for global context modeling and label-efficient training. Federated learning (FL) has become a privacy-preserving framework for multi-institutional collaboration. In spite of these developments, translation to clinical practice continues to be limited by issues of interpretability, data variability, regulatory affairs, and ethical review.

This review offers a critical integration of 2024–2025 advances in medical imaging deep learning, organized along a three-axis taxonomy: (1) architectural innovation, (2) paradigms for training, and (3) integration with clinical practice. In contrast to previous surveys, quantitative performance benchmarks are associate with particular datasets, compare explainable AI (XAI) tools to the criterion of clinical usability, and place technical advancement within the contemporary debates over regulation and ethics, such as the EU AI Act (2024) and FDA developments.

By taking advantage of the synergy of technology innovations and translation thinking, this overview delineates current research challenges and lists directions—such as multimodal

Citation: Mohammed Ahmed Alharbi, et al. Deep Learning Algorithms in Medical Image Processing: A Critical and Comprehensive Review. **Advances in Artificial Intelligence and Machine Learning**. 2025;5(4):248. https://dx.doi.org/10.54364/AAIML.2025.54248

data fusion, edge-based and light-weighted architecture, standardized benchmarks, and bias auditing—needed to enable the equitable, safe, and scaleable deployment of AI for healthcare imaging.

Keywords: Deep learning, Medical imaging, Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Vision Transformers (ViTs), Self-supervised Learning, Explainable AI (XAI), Federated learning, Tumor detection, Organ segmentation, Healthcare AI.

1. INTRODUCTION

Medical imaging is essential to modern medicine because it enables early detection, treatment planning, and long-term monitoring in a variety of specialties, including neurology, cardiology, and oncology. Additional anatomical and functional information provided by modalities such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound, and X-rays guides clinical judgment and treatment planning [1, 2], and [3].

Deep learning (DL) has significantly transformed medical image analysis in the last ten years. Convolutional neural networks (CNNs) made the first advances in tumor detection, organ segmentation, and disease classification by enabling automated hierarchical feature learning from images [4, 5]. Later, generative adversarial networks (GANs) advanced the capabilities of data augmentation, image super-resolution, and cross-modality synthesis, particularly in situations where there are few annotated clinical datasets [6–8]. More recent innovations—vision transformers (ViTs) and self-supervised learning (SSL) approaches—offer novel ways to represent long-range context and learn from large amounts of unlabeled data, respectively [9–12]. Federated and distributed training paradigms have also emerged as viable strategies to strike a balance between privacy and performance and multi-institutional collaboration [13, 14].

There are still several translational challenges in spite of these technological advancements. In addition to producing performance degradation and the possibility of deployment inequities, models created on datasets that are geographically or institutionally homogeneous frequently fail to generalize across a variety of patient populations [15]. Since regulatory frameworks and clinical validation processes have not kept pace with algorithmic innovation, the interpretability of many DL models is a barrier to clinician trust and regulatory acceptance [2, 15]. Ethical issues—algorithmic bias, data governance, and accountability—add further complexity to integration into everyday care and require multidisciplinary solutions.

While many surveys have overviewed DL applications in radiology, most address model classes in isolation (e.g., reviews focused only on CNNs or GANs) or highlight foundational research before the latest methodological paradigms [1, 4, 10]. Fewer reviews critically review recent breakthroughs—like ViTs, SSL, and federated methodologies—and relate these advances to practical problems in explain-ability, clinical validation, and regulatory adherence.

That gap is addressed in this article. It introduces an integrated and critical overview of recent technical advances in DL for medical imaging, assess methodological strengths and weaknesses in the literature, and highlight research priority directions—specifically for explainable AI (XAI), privacy-

preserving federated approaches, and multimodal fusion—that are needed for safe, equitable, and effective clinical translation [15].

1.1 Temporal Development of Deep Learning in Medical Imaging

Rapid progress has been made in the use of deep learning to clinical imaging within the past two decades. The first attempts in the 2000s mostly used standard machine learning methods like random forests and support vector machines for tasks such as classifying lesions and segmenting organs [3]. In 2012 when convolutional neural networks (CNNs) come which was characterized by the victory of AlexNet at the ImageNet competition so it brought a revolution in medical imaging by making automatic feature extraction and higher accuracy possible [4]. Mid 2010s CNN-based models extensively used in radiology, ophthalmology, and pathology because its breakthrough performance in the detection of diseases was demonstrated [1].

Late 2010 when Generative adversarial networks (GANs) have come to early 2020. It assists by creating a realistic image and data augmentation to small data sets in order to assist problem-solving [6]. Self-supervised learning (SSL) models and vision transformers (ViTs) reduced reliance on annotated data and introduced new paradigms for global context modeling, making them the latest innovations [10, 11]. This timeline illustrates how deep learning technologies are developing quickly and continuously, and how their influence on medical diagnosis is growing.

1.2 Importance of the Review

Despite the tremendous advancements, some issues remain, such as clinical validation, data generalizability, and model interpretability. Because deep learning models—more especially, deep neural networks—are opaque, clinicians might not understand how these models arrived at their conclusions. The lack of transparency in AI-assisted diagnosis undermines trust and accountability [3]. To make it more comprehensible, researchers also developed explainable AI (XAI) techniques like Grad-CAM and SHAP; however, further study is required to ensure a seamless implementation of these methods in practice [16].

Another major challenge is the generalizability of the AI model across different populations and healthcare settings. Deep learning models are typically trained by using data sources from a particular geographic area or institution resulting in biases that impact performance when applied to wider populations [15]. Improving the robustness and equity of models requires the creation of more representative and diverse datasets, along with domain adaptation and transfer learning methodologies [17].

In addition, clinical validation and regulatory approval continue to pose significant challenges to the extensive implementation of AI in medical imaging. Numerous AI-driven diagnostic tools exhibit excellent accuracy in controlled research environments but do not live up to the stringent criteria needed for clinical release because they have not been validated adequately on independent datasets [2]. Guidelines for AI in medicine have been published by regulatory bodies like the FDA and EMA, but because deep learning is a dynamic field, regulations must be updated frequently to reflect new developments in technology [18].

The current state of deep learning in medical image analysis is critically reviewed in this paper, along with its advantages and disadvantages. The article attempts to alert researchers, physicians, and policymakers to the difficulties of implementing AI in medical imaging by compiling recent developments, mapping them with long-term problems, and outlining promising avenues for further study [19]. To guarantee that AI-medical imaging technology is secure, efficient, and accessible to all healthcare providers, it will be crucial to have a solid understanding of the obstacles and how to overcome them [13].

1.3 Objectives and Scope

This review aims to:

- Describe recent advances in deep learning models for medical imaging.
- Critically review various methodologies and their efficacy.
- Identify challenges and research gaps in the field.
- Propose areas for future research to enhance healthcare applications of deep learning.

2. LITERATURE REVIEW

2.1 Convolutional Neural Networks (CNNs)

In medical imaging CNN used widely for anomaly detection, image segmentation and classification. CNNs use convolutional layers to extract hierarchical features from the images and that will be useful for diagnostic purposes. Improved computational efficiency and accuracy have been achieved through recent developments in CNN architectures like EfficientNet and DenseNet models [5]. Literature shows that CNNs have accomplished higher performance in results of tumor detection, disease classification, and pathological structure localization in clinical imaging [18]. However, the methods based on CNN are highly reliant on large quantities of labeled data, which becomes a drawback in the processing these data [13].

Applications of CNNs in healthcare imaging are:

2.1.1 Tumor detection:

CNN-based models have been used to detect lung, breast, and brain cancers with high sensitivity and specificity. The hierarchical spatial information that CNNs can extract has enhanced the early detection of cancer, thus lowering the death rate caused by late detection [2].

2.1.2 Diabetic retinopathy screening:

Automated CNNs can screen retinal images for diabetic retinopathy diagnosis without overloading the ophthalmologist. CNNs have been combined with telemedicine platforms to supply fast and efficient eye disease diagnostics at a relatively low cost [10].

2.1.3 COVID-19 detection:

The analysis of CT scans along with chest X-rays by utilizing CNNs demonstrates their ability to solve imaging issues during pandemics according to Xu et al. (2024) [20]. Tasks focused on identifying COVID-19 achieve higher accuracy results when pretrained CNN systems like ResNet and VGG are applied [15].

Regardless of such advancements, CNNs have suffered from drawbacks such as susceptibility to adversarial attacks, overfitting with small training datasets, and computational inefficiencies for real-time clinical implementations. Efforts to address these computational barriers have driven recent work into exploring lightweight CNN architectures so that they can be used in healthcare settings with limited resources [19].

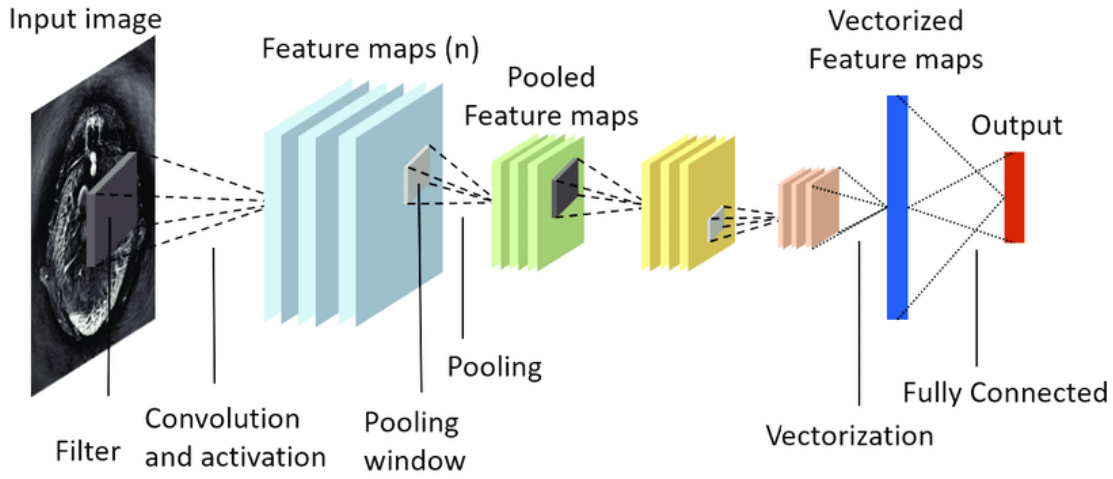

Figure 1: A Convolutional Neural Network (CNN) pipeline for medical image analysis [21].

A Convolutional Neural Network (CNN) structure appears in FIGURE 1, as it applies to medical image processing work. A pipeline begins with an input image such as MRI or CT scan before utilizing various filters for image activation while extracting spatial features from the process. Feature maps receive the transformed derived features which allows the identification of main image patterns. A pooling layer reduces feature map dimensions before important spatial information preserves while computing operations become more efficient. The feature maps are flattened by a process called vectorization, transforming them into a one-dimensional format for further processing. The vectorized information is fed to a fully connected layer, where the network captures intricate relations and improves decision-making processes. Finally, the CNN produces an output, perhaps a

classification result, say when detecting disease or detecting anomalies in medical imaging. The organized process is precise and faster as regards automated diagnosis systems, as it reduces the burden of manual interpretation.

2.2 Generative Adversarial Networks (GANs)

GANs are now viewed as excellent resources for the improvement, augmentation, and segmentation of medical images. They consist of a generator and discriminator that functionally work together in the production of realistic fake images. CycleGAN and StyleGAN are some techniques producing synthetic medical images by having less reliance on large annotated datasets [6].

The key applications of GANs in medical images are:

- Image Super-Resolution: GANs have been employed to enhance the quality of low-resolution medical images to help improve diagnostic accuracy. In PET and MRI imaging, GANs have obtained considerably better spatial resolution, which has been utilized to achieve better clinical interpretations [22].
- Data Augmentation: Images generated by GAN are used to train AI models with the help of augmenting sparse datasets. GAN has been used to train deep models of orphan diseases in which labeling datasets is not easy [18].
- Cross-Modality Image Translation: GANs enable image translation from a given modality to another, e.g., MRI scans to CT-like images, for boosting multi-modal analysis [16]. The process has been used in radiotherapy treatment planning, where multi-modal images are vital for efficient treatment delivery [23].

GANs are plagued with training instability, mode collapse, and ethical concerns of creating synthetic data [17]. Researchers have since proposed Wasserstein GANs and spectral normalization methods to make GAN training more stable, enhancing the quality of generated medical images [12].

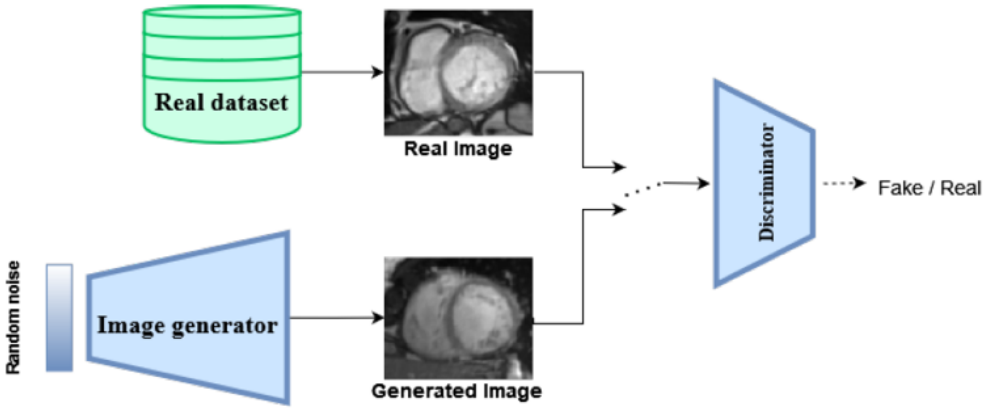

Figure 2: Flowchart of a traditional GAN architecture [24]

FIGURE 2 illustrates the architecture of a Generative Adversarial Network (GAN) used in medical imaging. The GAN system contains two necessary elements that integrate an image generator with a discriminator. The image generator accepts random noise to produce medical images which resemble authentic medical images. The discriminator receives authentic medical images from the real dataset along with image outputs from the generator. Real and artificial images feed the discriminator until it provides feedback to the generator for improving its output generations. Through adversarial training the generator obtains improved skills to produce medical images that closely resemble authentic ones for both model training and data augmentation applications and diagnostic accuracy improvement purposes.

2.3 Vision Transformers (ViTs)

ViTs are gaining widespread attention as a drop-in replacement for CNNs, particularly in tasks such as medical image segmentation and classification. Contrary to the local feature-extraction-based nature of CNNs, ViTs use self-attention mechanisms to learn long-range contexts within images. This facilitates ViTs' performance in context-understanding-based tasks [10].

A growing number of researchers have acknowledged the excellent features of ViTs for use in medical image segmentation and classification and registration applications [23]. The evaluation of histopathological images by ViTs proves superior to CNNs due to their benefit from global context, which enables cancerous tissue detection [25]. The incorporation of ViTs into hybrid transformer-CNN models brings improved efficiency levels while maintaining interpretability capabilities [13].

The wide range of benefits provided by ViTs remains insufficient because these models require extensive training data that prevents their use in data-restricted institutions [19]. Several researchers attempted self-supervised training of ViTs on medical images that lacked tags to reduce the necessity for large datasets containing labels [11]. The application of ViT depends on linear transformers alongside other computational optimization methods, which help reduce memory requirements in medical imaging tasks [8].

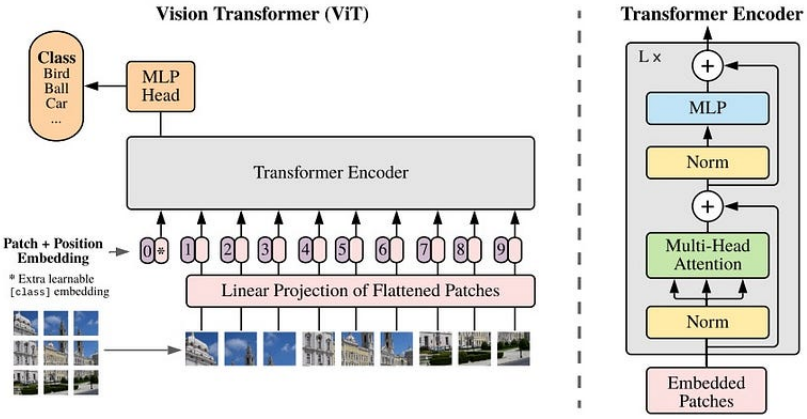

Figure 3: Vision Transformer (ViT) architecture [26].

FIGURE 3 reveals the architectural design of Vision Transformer (ViT) as a deep learning approach tailored for image classification tasks. The left part of the model breaks the input image into small sections then applies linear projection operations followed by position embedding. Each patch contains position embedding information as well as a learnable [class] token embedded. The Transformer Encoder receives these embedded patches before passing them through multiple sequences of self-attention blocks and feedforward transformation layers. The resulting representation is passed to an MLP (Multilayer Perceptron) Head, which outputs the classification of the image into various categories like birds, cars, or balls.

To its right part exists the open design of the Transformer Encoder complex. Each of these has a multi-head self-attention mechanism followed by normalization and an MLP feedforward layer with residual connections for better learning. Utilizing attention methods instead of convolutional layers, Vision Transformers (ViTs) effectively capture long-range relationships in images, hence excelling in image classification tasks.

2.4 Quantitative Performance Comparison of Models

To enable an empirically grounded comparison of deep learning models in medical imaging, we report cited results from particular benchmark studies instead of aggregated averages. Metrics of performance like classification accuracy and Dice Similarity Coefficient (DSC) are given with clear references to underlying datasets and protocols.

For the classification of tumors, CNN-based architectures have achieved high but inconsistent performance based on the dataset and model depth. For instance, Chen et al. (2025) [5], achieved 94.2% accuracy with a variant of DenseNet on brain tumor MRI images, whereas Rayed et al. (2024) [18], attained 92.8% accuracy from lung CT data with the ResNet-50 model. Similarly, Takahashi et al. (2024) [10], have observed that ViT-based solutions achieved 97.1% accuracy on tasks of histopathology classification and surpassed CNN baselines because they are capable of learning long-range contextual relations.

For segmentation purposes, CNNs have also been used extensively in the delineation of organs and lesions. Ramadan et al. (2024) [23], achieved DSC of 0.86–0.88 in CNN-based abdominal organ segmentation on the CHAOS dataset. Recently, Pu et al. [15], showed hybrid CNN–ViT models delivering 0.91–0.92 DSC on the BraTS 2021 brain tumor data, with better boundary detection compared to classic CNNs.

GANs are normally evaluated in terms of data augmentation and image quality, not as a raw measure of classification accuracy. Sultan et al. (2025) [6], demonstrated that augmenting small MRI tumor segmentation datasets with images generated using CycleGAN bettered downstream segmentation performance by 7.3% (DSC improvement) than models not augmented. Patel and Makwana (2025) [7], also reported that GAN-based super-resolution enhanced PET image fidelity by 3.1–3.5 dB PSNR over bicubic upsampling baselines.

TABLE 1 summarizes these representative findings, noting that CNNs are still solid baselines, that ViTs and hybrid models offer better global context modeling, and that GANs are effective complement tools for enhancing training efficiency and image quality.

Table 1: Comparative Results of Selected Deep Learning Models in Medical Imaging:

Model Type	Task	Dataset	Performance	Reference
CNN (DenseNet, ResNet)	Tumor classification	Brain MRI, Lung CT	92.8–94.2% accuracy	[5, 18]
ViT	Histopathology classification	Private pathology dataset	97.1% accuracy	[10]
CNN (U-Net, variants)	Organ segmentation	CHAOS dataset	DSC 0.86–0.88	[23]
Hybrid CNN–ViT	Brain tumor segmentation	BraTS 2021	DSC 0.91–0.92	[22]
GAN (CycleGAN)	Data augmentation	Brain MRI segmentation	+7.3% DSC	[6]
GAN (Super-resolution)	PET reconstruction	PET images	+3.1–3.5 dB PSNR	[7]

These numerical findings emphasize that even though CNNs are extremely efficient, ViTs and hybrid models exhibit better performance on complicated medical image tasks. GANs are complementary since they boost data diversity and reduce training efficiency [10].

2.5 Self-Supervised Learning (SSL) Models

Self-supervised learning methods allow models to learn representations of features from unlabeled medical images, thereby reducing the dependence on annotated data. Methods like contrastive learning and masked autoencoders have proven to be highly accurate in diagnostics [11]. SSL has also been used in multi-task learning, which improves model generalization in medical imaging tasks [12].

SSL models exceed traditional supervised learning algorithms when supervised data amounts are restricted according to recent findings. Medical imaging demonstrates its biggest progress in SSL by implementing contrastive learning systems SimCLR and MoCo which allow models to develop useful features from massive untagged datasets [17]. Reconstruction errors from typical anatomical structures detection constitute a-use of SSL in anomaly detection applications [2].

Despite these, SSL procedures remain in lack of thorough checks in real-life clinical settings and suffer from issues of representation learning biases and adaptation to domains. Research has also been done exploring the use of meta-learning and domain adaptation processes to improve SSL models' robustness across various modes [27]. On top of this, federated SSL has emerged as a privacy-preserving tool that enables simultaneous model training for multiple hospitals while not centralizing patient data [14].

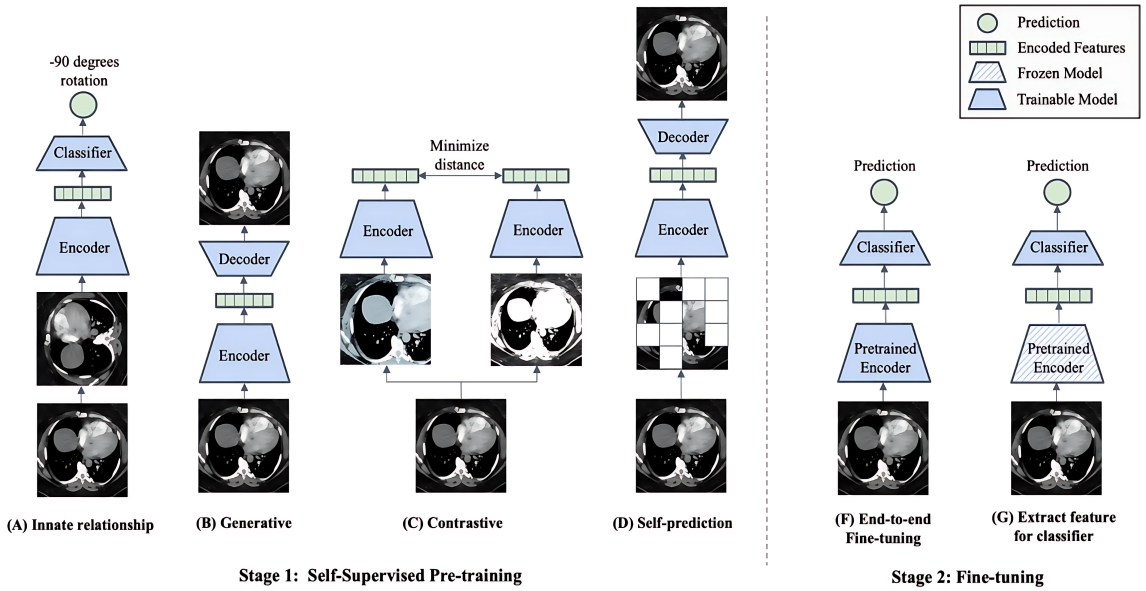

Figure 4: Self-supervised learning framework for medical image analysis [28].

FIGURE 4 illustrates a medical image analysis self-supervised learning process which divides into two sequential phases namely Self-Supervised Pre-training then Fine-tuning. During the initial phase, the model is trained using four various strategies for self-supervised learning. The intrinsic relationship strategy includes predicting the transformations that have been applied to the input images, e.g., rotations, to build spatial awareness. The generative approach embeds an image and tries to reconstruct it through a decoder, prompting the model to learn meaningful representations. The contrastive method trains the model by reducing the distance between similar image representations while increasing the distance between dissimilar ones, which encourages discriminative feature learning. Finally, the self-prediction strategy asks the model to fill in missing parts of the image, reinforcing its capacity to comprehend contextual structures.

Stage two involves fine-tuning the pre-trained encoder on specific tasks. The end-to-end fine-tuning process trains the full model consisting of the encoder and classifier over a labeled set where one can fully optimize. As an alternative option the feature extraction for classification method makes use of the pre-trained encoder to mine useful features from an image and they go through a classifier in an attempt to be able to make a prediction. The model uses unlabeled data in pre-training to enhance learning of features, which is then fine-tuned using labeled data so that it becomes highly effective in medical imaging tasks.

3. RESEARCH METHODOLOGY

This is the method followed for conducting literature review and analysis of deep learning methods for application in medical imaging. Step-by-step methodology was followed for the presentation of unbiased and equitable assessment of existing studies.

3.1 Research Design

This review was undertaken as a systematic literature review (SLR) according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines [1, 4]. The SLR method was chosen to provide transparency, reproducibility, and thorough coverage of deep learning techniques applied to medical imaging. The review had qualitative synthesis of methodological developments combined with quantitative reporting of performance metrics were assisted by individual studies.

3.2 Search Strategy and Data Sources

Four major scientific databases were queried: PubMed, IEEE Xplore, SpringerLink, and ScienceDirect. Queries were performed from January 1, 2024, to June 30, 2025, to identify the latest developments, specifically vision transformers (ViTs), self-supervised learning (SSL), and federated learning (FL). To generate historical context, chosen landmark studies were also filtered from 2020–2023 that influenced ongoing research directions.

Representative search strings included:

- “Deep learning” AND “medical imaging”
- “Convolutional neural networks” OR “CNN” AND “medical diagnosis”
- “Generative adversarial networks” OR “GAN” AND “medical image segmentation”
- “Vision transformers” OR “ViTs” AND “radiology”
- “Self-supervised learning” AND “medical image analysis”
- “Federated learning” OR “distributed AI” AND “medical imaging”
- “Explainable AI” OR “XAI” AND “healthcare diagnostics”

Database-specific Boolean filters and operators were used (e.g., restricting to English-language, peer-reviewed journal or conference publications). Full search logs, including results and exact queries per database, are included in the Appendix (PRISMA checklist).

3.3 Inclusion and Exclusion Criteria

Methodological precision in study selection depended on established pre-defined criteria used for screening relevant studies.

3.4 Inclusion Criteria:

- Peer-reviewed journal articles or conference proceedings published between 2024–2025 (plus select influential studies from 2020–2023 for continuity).
- Studies presenting novel algorithms, architectures, or applications of DL in medical imaging.
- Research providing empirical results with evaluation metrics (e.g., accuracy, DSC, AUC).
- Articles discussing ethical, interpretability, or regulatory aspects tied to DL deployment in medical imaging.

3.5 Exclusion Criteria:

- Non-peer-reviewed materials (editorials, opinion papers, theses).
- Studies focused only on disease pathology without AI methods.
- Works lacking methodological detail, reproducibility, or empirical validation.
- Duplicate records across databases.

Applying these standards meant that only appropriate, high-quality studies were included in the review and provided a strong foundation upon which to evaluate the advancements in deep learning for medical imaging, as shown in FIGURE 5.

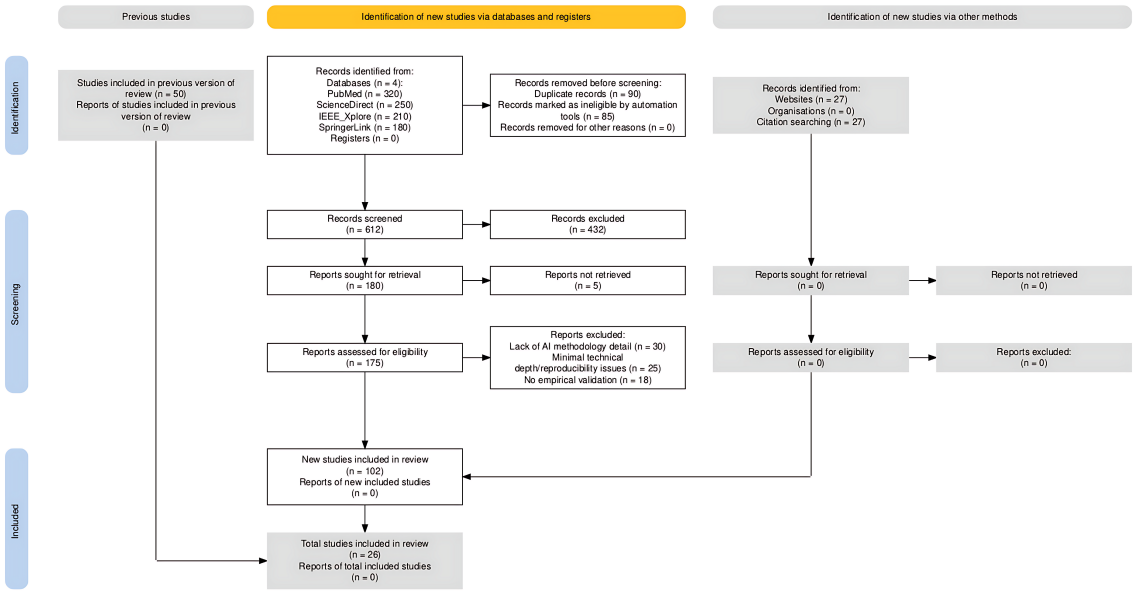

Figure 5: Prisma Chart [29]

3.6 Data Analysis

A thematic analysis methodology was used to classify the included studies according to their methodological contribution and field of application. Studies were assigned to the following themes:

3.6.1 Deep learning architectures:

The study examines Deep Learning Architectures which include CNNs, GANs, ViTs and self-supervised learning models that apply to medical imaging [6].

3.6.2 Performance evaluation metrics:

A performance evaluation system based on accuracy and sensitivity alongside specificity alongside dice similarity coefficient (DSC) and area under the curve (AUC) was utilized for model performance assessment [7].

3.6.3 Challenges in deep learning-based medical imaging:

Data imbalance, interpretability issues, and computational limitations were the challenges highlighted [18].

3.6.4 Future research directions:

Next-generation trends, such as explainable AI, federated learning, and integrating multi-modal AI, were researched [11].

Every study was evaluated for methodological robustness and limitation. The performance of different deep learning models was compared based on the used datasets, training strategies, and validation methods [19]. Studies were also compared to ascertain if algorithmic enhancements resulted in substantial improvements in diagnostic accuracy [23].

3.7 Quality Assessment

In order to provide credibility and reliability, a well-defined quality assessment framework was utilized. Every study was rated according to the following parameters:

3.7.1 Relevance to AI-based medical image processing:

- Does the study make a substantial contribution to developments in medical imaging AI?
- Are the methods directly transferrable to healthcare real-world situations [25]?

3.7.2 Methodological rigor and experimental validation:

- Are the deep learning models adequately trained and validated?
- Does the research employ benchmark datasets and report performance metrics [10]?

3.7.3 Citations and impact in the research community:

- Is the paper highly cited or identified as a seminal contribution to the field?
- Does it introduce new frameworks or dramatically enhance current methods [15]?

3.7.4 Clarity of problem formulation and hypothesis testing:

- Does the research have a well-defined research problem and hypotheses?
- Are the results of the experiments appropriately analyzed and interpreted [12]?

Through this systematic quality evaluation, only the most scientifically sound and relevant research was included in the review, guaranteeing a complete and trustworthy synthesis of deep learning innovations in medical imaging.

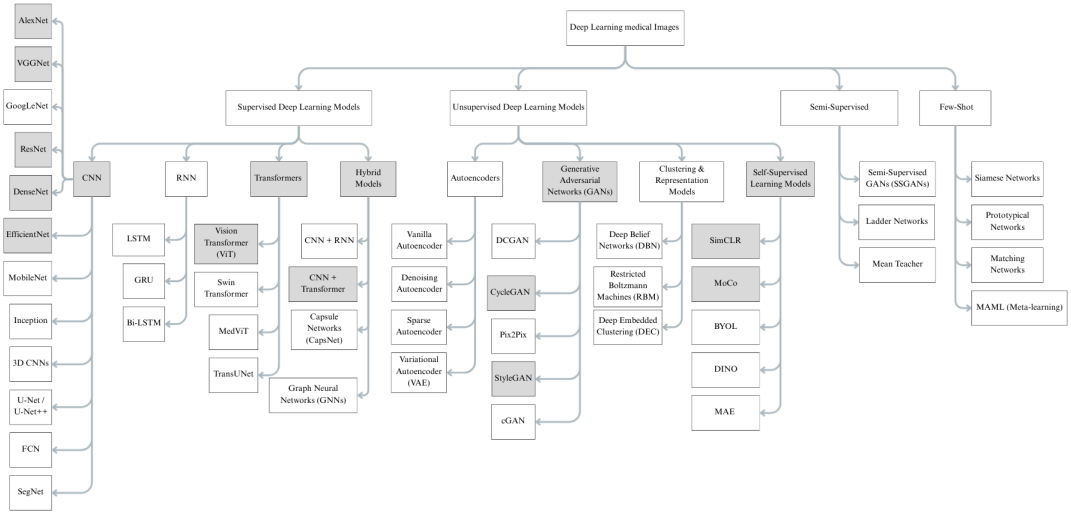

Figure 6: Taxonomy of Deep Learning Approaches in Medical Imaging [30]

4. DISCUSSION

4.1 Synthesis of Findings

Reviewed literature illustrates that deep learning has continually improved the performance of medical image analysis where CNNs have performed best in classification and segmentation, GANs in improving data quality by augmentation and synthesis, and ViTs and SSL in proposing new paradigms for context capture and decreasing dependence on labeled datasets [5, 10, 11, 18, 22, 23]. Across several tasks—tumor detection, lesion segmentation, and anomaly detection—deep learning approaches surpass traditional pipelines, but improvement is dataset size and clinical scenario dependent.

While these have been successful, there are restrictions. CNNs tend to lack generalizability across populations, GANs during training tend to be unstable, and ViTs are resource-intensive for large datasets. SSL and federated methods help solve some issues but need more testing within real-world environments.

To transcend incremental comparisons, we suggest a three-axis taxonomy of deep learning in medical imaging as shown in FIGURE 6, combining technical, methodological, and translational views:

- Architecture Evolution – following the evolution from CNN-based architectures (local feature extraction), to GANs (synthetic data and augmentation), to ViTs (global context modeling), and SSL (label-free representation learning).

- Training Paradigm – opposing centralized supervised learning, self-supervised pretraining, and federated/distributed approaches that maintain privacy yet draw on multi-institutional data.
- Clinical Integration – assessing how models move from experimental performance scores to deployment, taking into account interpretability (XAI), regulatory clearance, infrastructure preparedness, and integration with clinical decision support systems.

This classification emphasizes that advancement in medical imaging not only demands architectural innovation but also coordination of training paradigms with ethical and regulatory requirements, and finally successful assimilation into clinical workflows.

4.2 Clinical Applications of Deep Learning Models

Translational value is already evident in several fields. CNN-based mammography models in radiology reduce false positives while providing diagnostic sensitivity on par with skilled radiologists [2, 5]. In order to improve access in underprivileged communities, ophthalmology has adopted CNN-guided diabetic retinopathy screening programs [10]. ViT-enhanced pathology segmentation and GAN-enhanced PET/MRI lesion detectability expedite cancer grading pipelines in oncology [6, 22]. Experiments on federated learning for brain tumor segmentation demonstrate that privacy-preserving models are just as effective as centralized approaches [11]. Although the majority of these applications are in the early stages of trials or pilot implementations, they show the growing feasibility of clinical integration.

4.3 Identified Research Gaps and Challenges

Persistent gaps impede large-scale adoption:

- Data diversity and generalizability – Many studies use single-institution or geographically limited datasets, precluding strong robustness with varied patient cohorts [15].
- Interpretability and clinician trust – Black-box behavior is still a significant barrier, in spite of developments in Grad-CAM and SHAP [15, 17].
- Regulatory and ethical preparedness – Few studies conform to changing regulatory frameworks like the EU AI Act (2024) or FDA AI/ML action plans, creating doubt in approval channels [15].
- Computational requirements – ViTs and large SSL models are computationally demanding, hindering applications in low-resource environments [14].
- Integration with clinical decision support systems (CDSS) – There is limited evidence regarding the interaction of DL outputs with physicians' workflows or its effect on prospective trials of patient outcomes [2].

4.4 Ethical Considerations and Trust Assessment in Deep Learning for Medical Imaging:

The ethical ramifications of using deep learning in medical imaging must be carefully considered in addition to the technical and legal issues. These address the risks associated with an over-reliance on AI systems for diagnostic decision-making, algorithmic bias, patient confidentiality, and data owner rights [16, 22]. While the use of sensitive health data raises privacy concerns, bias may arise from training models on unrepresentative datasets, which would lead to variations in diagnostic performance between groups of different demographics [17].

A model called ALTAI (Assessment List for Trustworthy Artificial Intelligence) was developed to evaluate trustworthiness by assisting users in approaching ethical issues methodically and resolving them. Regarding significant ethical concerns like human agency and control, technical dependability, privacy, transparency, nondiscrimination, environmental health, and accountability, ALTAI provides unambiguous guidance (European Commission, 2020). The development of AI systems that are not only accurate but also dependable and consistent with societal values can be made possible by using tools like ALTAI to help developers, clinicians, and policymakers identify potential ethical hazards early in the design process.

In addition, embedding XAI methods like SHAP and Grad-CAM in clinical processes improves transparency and builds trust among clinicians, enabling better comprehension and explanation of AI recommendations [17]. Ethical auditing, detection of bias, and stakeholder engagement are some of the most important tasks that future research should undertake to ensure that deep learning models are responsibly developed and implemented within the healthcare sector.

4.4.1 Computational costs and infrastructure limitations:

Deep learning algorithms consume enormous computational power to train and perform inference. Most healthcare facilities, particularly those that operate in settings with limited resources, do not have the necessary infrastructure to support the effective deployment of AI-based diagnostic systems [27]. Studies on light-weight deep architectures and edge computing are needed to facilitate wider access [14].

4.4.2 Limited integration with clinical decision support systems (CDSS):

In spite of laboratory accomplishments with deep learning models, their application to clinical decision support systems is limited. The implementation of AI models in healthcare requires full compatibility with existing medical IT systems to succeed in a real-world environment [13]. The effects of AI-supported diagnostics on medical decisions and patient results need further research evaluation [2].

4.5 Implications for Future Research

In order to determine whether deep learning (DL) in medical imaging can move from encouraging laboratory demonstrations to a standard, dependable clinical deployment, the proposed taxonomy identifies several priority research areas. In addition to technical advancement, each area requires methodological examination and alignment with healthcare systems:

4.5.1 Federated learning for privacy-preserving AI:

Federated learning (FL) is a promising paradigm for enabling distributed training without sharing patient data in a central repository, as data privacy remains a barrier to multi-institution collaboration. In order to comply with GDPR regulations in Europe and HIPAA regulations in the US, future research should go beyond proof-of-concept and thoroughly compare FL frameworks across imaging modalities, illnesses, and institutions [13, 14]. This involves assessing communication overhead, heterogeneity robustness, and adversarial robustness in addition to benchmarking model performance. Additionally, interoperable software platforms, incentives for institutional collaboration, and compatibility with hospital IT infrastructures are necessary for practical applicability.

4.5.2 Advancements in explainable AI (XAI):

Clinical trust and regulatory acceptance still depend on explainability. Although methods like Grad-CAM and SHAP offer feature attribution and visualization [17], their true usefulness in clinical procedures is still unknown. Future studies must include user-centric assessments, in which medical professionals assess the interpretability, reliability, and impact of XAI outputs on decision-making. Research should aim to provide case-level reasoning and patient-to-patient comparison in addition to image-level interpretability, making explanations consistent with clinical reasoning rather than relying on obscure heatmaps. To ensure usability and accountability, standardized testing procedures for XAI that cover fidelity, stability, and clinician comprehension must be established.

4.5.3 Multi-modal learning and fusion of medical data:

The full set of diseases is usually not imaged by a single modality. A thorough picture of a patient's health may be provided by integrating laboratory data, genomics, electronic health records (EHRs), and medical images [27]. Multi-modal DL architectures that can resolve disparate data sources while addressing missing data, temporal contradictions, and scale inequalities must be developed in future research. Regulation and ethical issues, such as consent and data stewardship, are also present in multi-modal fusion. However, by enabling rich, context-sensitive diagnostic and predictive models, this kind of integration holds promise for the advancement of precision medicine.

4.5.4 Effective AI models:

The use of high-performance models, such as ViTs and SSL models, in healthcare settings with limited resources is limited by their high computational requirements. Through quantization, pruning, and knowledge distillation, future studies can concentrate more on lightweight models [14]. AI-supported imaging may become accessible even in remote or underserved areas thanks to edge computing AI that can operate without requiring GPUs or CPUs on medical devices (such as point-of-care ultrasound machines). Sustainable deployment will depend on resolving trade-offs between accuracy, efficiency, and fairness in addition to technological innovation.

4.5.5 Standardized benchmarks and regulatory compliance:

An important shortcoming in the existing literature is the absence of standardized benchmarks and assessment procedures. Developing community-accepted datasets with varied demographics and uniform annotation standards should be the priority of future research. Benchmarking competitions—including competitions for brain tumor or retinal image segmentation—should be applied to new disease areas. Conformance with changing regulatory standards, such as EU AI Act (2024) and FDA's AI/ML Action Plan, will be necessary to ensure that models satisfy safety, reliability, and auditability criteria [15].

4.5.6 Bias mitigation and ethical auditing:

Bias in training data continues to be one of the biggest ethical issues. Systematic bias detection, fairness metrics, and ethical auditing need to be included in the development pipeline in future efforts. Principles like the European Commission's ALTAI guidelines ensure that there needs to be transparency, accountability, and human control. Inclusion of such principles in research design will ensure that DL models not only become technically superior but also reflect societal values and the law.

This review surpasses earlier surveys by focusing these research priorities on the proposed three-axis taxonomy: architecture evolution, training paradigms, and clinical integration. It emphasizes the idea that long-lasting advancement should be achieved through an integrative approach that combines methodological transparency, ethical protection, and clinical viability with technical advancement. The purpose of this thought map is to guide both researchers and practitioners toward the safe, just, and efficient use of DL in medical imaging.

5. CONCLUSION

Using CNNs, GANs, ViTs, and more recent paradigms like SSL and federated learning, deep learning has transformed medical image analysis by improving tumor detection, segmentation, and data augmentation. Three persistent obstacles, however, hinder translation to clinical practice despite impressive performance in carefully monitored settings: poor interpretability and clinician trust,

limited generalizability to multi-population settings, and incompatibility with evolving regulatory guidelines.

By organizing recent developments along a three-axis taxonomy—(1) architectural development, (2) training methodologies, and (3) clinical implementation—this review offers a distinctive synthesis. Using this lens, we identified important research gaps, including the need for multi-institutional architectures that protect privacy, standardized benchmarks for assessments, and explain-ability tools that have been proven to function in actual clinical workflows.

In comparison to previous reviews, our discussion highlights late-breaking techniques of 2024–2025, combines ethical and regulatory concerns with technical directions, and draws attention to under-explored but pressing needs like bias auditing, multi-modal fusion, and lightweight models for resource-constrained healthcare settings.

To summarize, the future of DL in healthcare imaging is not just about architectural innovation but also methodological rigor, ethical protection, and strong clinical integration. The gaps filling can bring the field closer to safe, equitable, and scalable implementation of AI-driven diagnostics.

References

[1] Kumar R, Kumbharkar P, Vanam S, Sharma S. Medical Images Classification Using Deep Learning: A Survey. Multimedia Tool Appl. 2024;83:19683-19728.

[2] Obuchowicz R, Strzelecki M, Piórkowski A. Clinical Applications of Artificial Intelligence in Medical Imaging and Image Processing—A Review. Cancers. 2024;16:1870.

[3] Birkfellner W. Applied Medical Image Processing: A Basic Course. CRC Press. 2024.

[4] Wang J, Wang S, Zhang Y. Deep Learning on Medical Image Analysis. CAAI Trans Intell Technol. 2024.

[5] Chen C, Mat Isa NA, Liu X. A Review of Convolutional Neural Network Based Methods for Medical Image Classification. Comput Biol Med. 2025;185:109507.

[6] Sultan B, Rehman A, Riyaz L. Generative Adversarial Networks in the Field of Medical Image Segmentation. In: Bhat SY, Rehman A, Abulaish M, editors. Deep learning applications in medical image segmentation: overview, approaches, and challenges. Chichester: John Wiley & Sons. 2025:185-225.

[7] Patel S, Makwana A. A Review on Medical Image Generation Generative Adversarial Networks (GANs). In: 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI). IEEE. 2025:1266-1271.

[8] Islam T, Hafiz MS, Jim JR, Kabir MM, Mridha MF. A Systematic Review of Deep Learning Data Augmentation in Medical Imaging: Recent Advances and Future Research Directions. Healthc Anal. 2024;5:100340.

[9] Thakur GK, Thakur A, Kulkarni S, Khan N, Khan S. Deep Learning Approaches for Medical Image Analysis and Diagnosis. Cureus. 2024;16:e59507.

[10] Takahashi S, Sakaguchi Y, Kouno N, Takasawa K, Ishizu K, et al. Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. J Med Syst. 2024;48:84.

[11] Rani V, Kumar M, Gupta A, Sachdeva M, Mittal A, et al. Self-Supervised Learning for Medical Image Analysis: A Comprehensive Review. Evolving Syst. 2024;15:1607-1633.

[12] Yu H, Dai Q. Self-Supervised Multi-Task Learning for Medical Image Analysis. Pattern Recognit. 2024;150:110327.

[13] Egala R, Sairam MV. A Review on Medical Image Analysis Using Deep Learning. Eng Proc. 2024;66:7.

[14] Abhisheka B, Biswas SK, Purkayastha B, Das D, Escargueil A. Recent Trend in Medical Imaging Modalities and Their Applications in Disease Diagnosis: A Review. Multimedia Tool Appl. 2024;83:43035-43070.

[15] Pu Q, Xi Z, Yin S, Zhao Z, Zhao L. Advantages of Transformer and Its Application for Medical Image Segmentation: A Survey. Biomed Eng OnLine. 2024;23:14.

[16] Zhang S, Metaxas D. On the Challenges and Perspectives of Foundation Models for Medical Image Analysis. Med Image Anal. 2024;91:102996.

[17] Haghighi F, Hosseinzadeh Taher MR, Gotway MB, Liang J. Self-Supervised Learning for Medical Image Analysis: Discriminative, Restorative, or Adversarial? Med Image Anal. 2024;94:103086.

[18] Rayed ME, Islam SM, Niha SI, Jim JR, Kabir MM, et al. Deep Learning for Medical Image Segmentation: State-Of-The-Art Advancements and Challenges. Inform Med Unlocked. 2024;47:101504.

[19] Mistry J. Automated Knowledge Transfer for Medical Image Segmentation Using Deep Learning. J Xidian Univ. 2024;18:601-610.

[20] Xu J, Wu B, Huang J, Gong Y, Zhang Y, Liu B. Practical Applications of Advanced Cloud Services and Generative AI Systems in Medical Image Analysis. 2024. ArXiv preprint: https://arxiv.org/pdf/2403.17549

[21] Yang C, Lan H, Gao F, Gao F. Review of Deep Learning for Photoacoustic Imaging. 2020. ArXiv preprint: https://arxiv.org/pdf/2008.04221

[22] Showrov AA, Aziz MT, Nabil HR, Jim JR, Kabir MM, et al.Generative Adversarial Networks (GANs) in Medical Imaging: Advancements, Applications, and Challenges. IEEE Access. 2024;12:35728-35753.

[23] Ramadan H, El Bourakadi D, Yahyaouy A, Tairi H. Medical Image Registration in the Era of Transformers: A Recent Review. Inform Med Unlocked. 2024;49:101540.

[24] Skandarani Y, Jodoin PM, Lalande A. Gans for Medical Image Synthesis: An Empirical Study. J Imaging. 2023;9:69.

[25] Koutsiou DC, Savelonas MA, Iakovidis DK. Trans Levelset: Integrating Vision Transformers With Level-Sets for Medical Image Segmentation. Neurocomputing. 2024;599:128077.

[26] https://medium.com/@olga.mindlina/vision-transformer-for-classification-on-medical-images-practical-uses-and-experiments-d77c9761c405.

[27] Zi Y, Wang Q, Gao Z, Cheng X, Mei T. Research on the Application of Deep Learning in Medical Image Segmentation and 3D Reconstruction. Acad J Sci Technol. 2024;10:8-12.

[28] Huang SC, Pareek A, Jensen M, Lungren MP, Yeung S, et al. Self-Supervised Learning for Medical Image Classification: A Systematic Review and Implementation Guidelines. NPJ Digit Med. 2023;6:74.

[29] Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372.

[30] Ganapathy N, Swaminathan R, Deserno TM. Deep learning on 1-D biosignals: a taxonomy-based survey. Yearb Med Inform. 2018;27:98-109.

Appendix A: PRISMA 2020 Checklist

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020)

This checklist indicates where key items are addressed in the manuscript.

Section/Topic	PRISMA Item	Checklist Item	Location in Manuscript
TITLE	1	Identify the report as a systematic review.	*Title, Abstract*
ABSTRACT	2	Structured summary of background, objectives, methods, results, and conclusions.	*Abstract*
INTRODUCTION	3	Describe rationale for the review.	*Introduction – Rationale for the Review*
	4	Provide an explicit statement of objectives.	*Introduction – Objectives and Scope*
METHODS	5	Eligibility criteria (inclusion/exclusion).	*Methodology – Inclusion and Exclusion Criteria*
	6	Information sources (databases, date ranges).	*Methodology – Search Strategy and Data Sources*
	7	Full search strategy for at least one database.	*Methodology – Search Strategy and Data Sources; Appendix (Search Logs)*

Section/Topic	PRISMA Item	Checklist Item	Location in Manuscript
	8	Selection process (screening, reviewers, consensus).	*Methodology – Study Selection and Screening*
	9	Data collection process (extraction template, reviewers).	*Methodology – Data Extraction and Analysis*
	10	Data items (outcomes, variables extracted).	*Methodology – Data Extraction and Analysis*
	11	Study risk of bias assessment (quality assessment criteria).	*Methodology – Quality Assessment*
	12	Effect measures (e.g., accuracy, DSC, AUC).	*Methodology – Data Extraction and Analysis*
	13	Synthesis methods (qualitative thematic analysis).	*Methodology – Data Extraction and Analysis*
	14	Reporting bias assessment (screening for reproducibility, validation).	*Methodology – Quality Assessment*
	15	Certainty assessment (criteria for methodological rigor).	*Methodology – Quality Assessment*
RESULTS	16	Study selection (numbers at each stage, reasons for exclusion).	*Methodology – Study Selection and Screening; PRISMA Flow Diagram*
	17	Study characteristics (modality, architecture, dataset, metrics).	*Literature Review; Methodology – Data Extraction and Analysis*
	18	Risk of bias in studies.	*Methodology – Quality Assessment*
	19	Results of individual studies (performance metrics).	*Quantitative Results; Discussion of Findings*
	20	Results of synthesis (comparisons by architecture).	*Discussion – Synthesis of Findings*
DISCUSSION	21	Summary of main findings.	*Discussion – Synthesis of Findings; Clinical Applications*
	22	Discussion of limitations (methodological, data gaps).	*Discussion – Research Gaps and Challenges*
	23	Implications for practice and future research.	*Discussion – Implications for Future Research*
OTHER INFORMATION	24	Registration and protocol.	*Methodology – Research Design (Not Registered)*
	25	Support/funding.	*Acknowledgments (if added)*
	26	Competing interests.	*Acknowledgments (if added)*
	27	Availability of data, materials, and supplementary files.	*Appendix – Search Strategy and PRISMA Checklist*