INTERNATIONAL JOURNAL OF RECENT TRENDS IN TECHNOLOGY AND ENGINEERING (IJRTTE)

Journal Home Page: https://ijrtte.com/

A Comprehensive Survey on Deep Learning Approaches in Medical Image Diagnosis

Vidya Santosh Wable[1], Dhiraj Hebri[2] and Mukund Wagh[3]

[1]*Research Scholar, Department of Computer Science & Engineering, Srinivas University, Mangaluru, India.*
[2]*Research Professor, Department of Computer Science & Engineering, Srinivas University, Mangaluru, India.*
[3]*Research Professor, School of Computing, MIT ADT, Pune, India.*
[1]*vidya3.deshmukh@gmail.com*

Abstract. Medical image analysis is essential in modern healthcare for early and proper diagnosis and appropriate treatment planning. Deep learning, particularly models like CNNs, Vision Transformers (ViTs), and GANs, has significantly advanced automated medical image diagnosis, outperforming traditional techniques in tasks such as classification, segmentation, and anomaly detection. This survey reviews key deep learning methods applied across imaging types like X-rays, MRI, CT, and ultrasound. It addresses challenges such as limited data, lack of interpretability, and clinical validation. In modern Era various new Emerging trends like self-supervised and consolidated learning are discussed, along with future directions to enhance diagnostic accuracy and clinical acceptance and adoption. This paper offers a concise reference for advancing AI-driven medical imaging.

Keywords: Deep learning; Medical image analysis; CNN; ViT; GAN; Automated diagnosis; Image segmentation; AI in healthcare; Federated learning; Explainable AI; Self-supervised learning

1. Introduction

The purpose of this survey is to provide a comprehensive and detailed overview of deep learning techniques applied in medical image analysis, aiming to guide researchers and practitioners in understanding current advancements and identifying future research directions. The scope of this review encompasses foundational models such as Convolutional Neural Networks (CNNs), advanced architectures like Vision Transformers (ViTs), and generative models including Generative Adversarial Networks (GANs). It verifies their applications across various medical imaging modalities, including X-rays, MRI, CT, and ultrasound, while addressing critical challenges such as data scarcity, model interpretability, and clinical validation. Furthermore, the survey highlights emerging trends such as self-supervised learning, federated learning, and explainable AI. The key findings or outcomes indicate that deep learning techniques have consistently outperformed traditional methods in tasks such as image segmentation, classification, and anomaly detection. CNNs continue to dominate the field, although ViTs and GANs are gaining attention for their enhanced capabilities. Emerging techniques like self-supervised and federated learning show promise in overcoming data and privacy challenges. However, barriers such as limited model interpretability, lack of standardized clinical validation, and data heterogeneity persist. Addressing these limitations and focusing on model robustness, explain ability, and clinical integration are critical for advancing AI-driven medical diagnostics.

2. Related Surveys on Deep Learning Approaches in Medical Image Diagnosis Gaps in Existing Literature

- Many reviews focus narrowly on specific tasks (e.g., segmentation [7], GANs [5], super-resolution [6]) or particular methodologies.

Volume: 04 Issue: 03 Year: 2025
Received: 14.05.2025 Received in revised from: 05.06.2025 Accepted: 11.07.2025
Available Online: 19.08.2025
Published by NTL Publisher

- Broad cross-modal surveys that integrate advanced architectures like Vision Transformers and address emerging trends (self-supervised, federated learning, explainable AI) through 2025 are still lacking.

Our contribution:

- Provides a modality agnostic, model inclusive survey spanning foundational (CNNs) to advanced (ViTs, GANs) architectures.
- Incorporates the latest developments (2024–2025) in GAN applications, image super resolution, registration, active learning, label efficiency, and federated learning.
- Highlights overlooked issues such as interpretability, clinical validation, and real-world integration, creating a more holistic reference for future research. Table 1 shows the Related Surveys.

Table 1: Related Surveys.

Survey Paper	Focus Area	Models Covered	Scope	Limitations
Litjens et al. (2017) [1]	General medical image analysis	CNNs	Classification, Segmentation	Focused only on CNNs and early DL models
Shen et al. (2017) [2]	Diagnostic imaging	CNNs	Early DL applications	Limited to CNN architectures
Lundervold & Lundervold (2019) [3]	Multimodal imaging	CNNs	Supervised learning	Lacked coverage of advanced models
Egala & Sairam (2024) [4]	General review	CNNs	Cross-domain applications	Focused mostly on CNN-based methods
Heng et al. (2025) [5]	GANs in medical imaging	GANs	Generative models	Focused solely on GAN applications
Lepcha et al. (2025) [6]	Super-resolution techniques	CNNs, ViTs, GANs, Diffusion Models	Image enhancement	Task-specific (super-resolution)
Chen et al. (2025) [7]	Image registration	DL-based registration models	Registration techniques	Limited to registration tasks
Wang et al. (2024) [8]	Active learning	Label-efficient DL methods	Data-efficient learning	Focused on active learning
Jin et al. (2023) [9]	Label-efficient learning	Semi-/Self-/Weakly-supervised DL	Medical imaging tasks	Focused on label efficiency
Guan et al. (2023) [10]	Federated learning	Federated DL models	Privacy-preserving learning	Focused only on federated learning techniques

2.1 Deep Learning

Deep learning models have transformed medical image analysis by enabling automatic feature extraction and pattern recognition from raw data. Several architectures—Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Vision Transformers (ViTs), and Generative Adversarial Networks (GANs)—form the foundation of modern techniques used in this domain.

2.2 Convolutional Neural Networks (CNNs)

CNNs are the most widely adopted architecture in medical image diagnosis, due to their ability to automatically learn spatial hierarchies of features through convolutional operations. As highlighted by Litjens et al. [1] and Shen et al. [2], CNNs have dominated tasks such as classification, segmentation, and lesion detection across imaging modalities like MRI, CT, and X-rays. Subsequent surveys [3][4] have reinforced CNNs as the foundational model in both single- and multi-modal image analysis.

2.3 Recurrent Neural Networks (RNNs)

While primarily used for sequential or time-series data, RNNs have also been explored in medical imaging, particularly for analysing temporal sequences in video data or progressive disease modeling [3]. Their ability to capture dependencies across time steps makes them valuable in applications such as patient monitoring and dynamic imaging sequences.

2.4 Vision Transformers (ViTs)

Building on the self-attention mechanism, Vision Transformers treat images as sequences of patches and process them similarly to text in natural language processing. Although ViTs have emerged more recently, studies like Lepcha et al. [6] demonstrate their growing importance in medical imaging tasks such as super-resolution, segmentation, and anomaly detection. Unlike CNNs, ViTs focus on global context learning, which enhances their capability for complex image analysis.

2.5 Generative Adversarial Networks (GANs)

GANs, discussed extensively by Heng et al. [5], consist of a generator-discriminator pair that facilitates realistic image generation and augmentation. GANs are increasingly used in medical imaging for data synthesis, modality translation, image enhancement, and super-resolution [5][6]. Their ability to generate high-quality synthetic images helps mitigate data scarcity—a critical challenge in healthcare AI.

In summary, as established across prior surveys [1]– [10], CNNs remain the dominant architecture, while ViTs, GANs, and RNNs are expanding the methodological landscape. These models collectively form the core of modern deep learning approaches applied to medical image analysis. Figure 1 shows the Convolutional Neural Networks (CNNs)

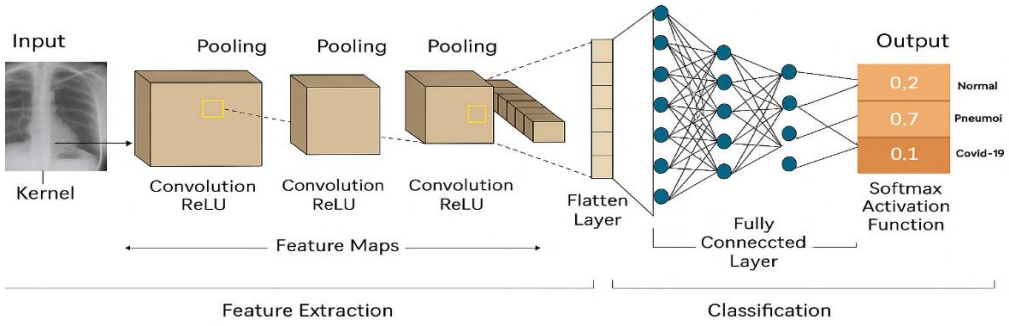

Figure 1: Convolutional Neural Networks (CNNs)

3. Categorized Review of Deep Learning Methods in Medical Image Diagnosis

3.1 Image Classification Techniques

Image classification refers to labeling input medical images with diagnostic categories (e.g., diagnosing pneumonia from chest X-rays).

CNNs: They are the state-of-the-art models for the classification problem as they are able to capture spatial features efficiently 1. Variants such as ResNet[1], DenseNet[10], and EfficientNet[15] have been utilized in different imaging modalities.

Vision Transformers (ViTs): ViTs buffer input images in patch sequences and learn the long-range context, showing superior performance on challenging classification tasks [6].

Applications: Pneumonia, COVID-19, diabetic retinopathy, skin cancer, and tumour detection from radiographies and fundus images 1[6].

3.2 Segmentation Models

Image segmentation is a pixel-level annotation regarding organs, tumours or anatomical structures.

U-Net: A popular encoder-decoder CNN architecture designed for biomedical image segmentation [3]. The skip connections enable the network to regain spatial information lost in the downsampling process.

DeepLab (v3+): Employs atrous spatial pyramid pooling to explicitly capture multi-scale context by exploring multiple scales of features in the whole image, leading to better segmentation accuracy of large objects, with strong constraints on segmenting various structures 3.

GAN-based Structures: GANs have been used for post-processing of segmentation maps during adversarial training, which improves boundary accuracy [5].

Applications: MRI/CT/ultrasound images tumor and organ segmentation.

3.3 Detection and Localization Techniques

These methods have targeted detecting and localizing abnormalities in medical images by generating bounding boxes or points (coordinates) of regions of interest.

Region-based CNNs (R-CNN, Faster R-CNN): Adapt CNNs to object detection by first suggesting candidate regions and then classifying them [1].

YOLO and SSD Models: Real-time detection and localization of lesions and tumors use one-stage detector [3].

Attention mechanisms / Transformers: Embedded in detection models for improved localization with global context [6].

Applications: Lesion localization in CT images, detection of breast cancer in mammograms and polyp detection in endoscopy images. Table 2 shows the Categorized Review Summary.

Table 2: Categorized Review Summary.

Task	Popular Models	Typical Applications
Image Classification	CNNs, ViTs	Pneumonia detection, cancer diagnosis
Image Segmentation	U-Net, DeepLab, GAN-based models	Tumor segmentation, organ delineation
Detection & Localization	R-CNN, YOLO, SSD, Transformers	Lesion localization, tumor detection

4. Comparative Analysis

The success of deep learning models in medical imaging change as a function of tasks, the amount of available data and the clinician's needs. This article provides a comparative overview over different methods with respect to their respective benefits and drawbacks and customary fields of use.

4.1 Convolutional Neural Networks (CNNs)

Advantages: Good local feature extraction, convenient in training, widely used in the classification and segmentation tasks.

Constrains: Not capable to capture global dependencies; performance deteriorates with small datasets.

4.2 Vision Transformers (ViTs)

Pros: Capable of capturing long-range dependencies through self-attention, suitable for complex tasks and large data.

Disadvantages: It is a computationally expensive; less efficient in small dataset.

4.3 Generative Adversarial Networks (GANs)

Pros: Great for synthetic data generation, super-resolution, domain adaptation.

Drawbacks: Hard to train; serve mostly as aid for down-or up sampling or tasks beyond direct (in-class) classification or segmentation.

4.4 U-Net and DeepLab (Segmentation Models)

U-Net Strengths: Accurate segmentation even with limited data.

DeepLab Strengths: Superior multi-scale object detection and segmentation.

Limitations: U-Net lacks global context modeling; DeepLab is computationally heavier.

4.5 Detection Models (Faster R-CNN, YOLO, SSD)

Strengths:

Faster R-CNN: High detection accuracy.

YOLO/SSD: Real-time detection capability.

Limitations:

Faster R-CNN: Slower inference.

YOLO/SSD: Less accurate for small lesion detection.

Table 3 shows the Comparative Summary Table. Table 4 shows the Performance Comparison Summary Table.

Table 3: Comparative Summary Table.

Model/Technique	Strengths	Limitations	Primary Use	References
CNNs	Strong feature extraction, efficient	Local context only, data-dependent	Classification, segmentation	[1], [2], [3]

ViTs	Global context modeling, scalable	High computation, needs large datasets	Classification, segmentation, detection	[6]
GANs	Data augmentation, image synthesis	Training instability, indirect application	Super-resolution, synthesis	[5], [6]
U-Net	Accurate with small datasets	Limited global feature learning	Medical image segmentation	[3], [4]
DeepLab v3+	Multi-scale feature extraction	High computational cost	Complex segmentation tasks	[4]
Faster R-CNN	High detection accuracy	Slower inference	Object/lesion detection	[1]
YOLO/SSD	Real-time detection	Lower accuracy on small objects	Lesion detection, localization	[3]

Table 4: Performance Comparison Summary Table.

Model	Accuracy	Speed	Data Efficiency	Best Use Case
CNNs	High	Fast	Moderate	Classification
ViTs	Very High	Moderate	Low (needs large data)	Complex segmentation/classification
GANs	N/A (Auxiliary)	N/A	Low	Data generation, augmentation
U-Net	High	Fast	Good with small data	Medical image segmentation
DeepLab	Very High	Moderate	Moderate	Complex segmentation
Faster R-CNN	High	Slow	Moderate	Detection/localization
YOLO/SSD	Moderate	Very Fast	High	Real-time detection

5. Challenges and Limitations

Despite significant advancements, deep learning-based medical image analysis faces persistent challenges and limitations that hinder clinical translation.

5.1 Data Scarcity and Annotation Costs

Medical imaging datasets are often limited due to patient privacy regulations, ethical concerns, and the requirement of expert annotations. Small and imbalanced datasets restrict model generalization and increase Overfitting risk [1][5].

5.2 Model Interpretability and Trustworthiness

Deep learning models, particularly CNNs and Transformers, are typically regarded as “black boxes.” The lack of explain ability impedes clinical adoption, as healthcare professionals require transparent, interpretable diagnostic insights [1][8]. Techniques such as attention visualization and saliency maps remain inadequate for complete model interpretability.

5.3 Computational Complexity

Advanced models like Vision Transformers (ViTs) and GANs demand extensive computational resources. Training such models necessitates large datasets and high-performance computing hardware, limiting accessibility for resource-constrained healthcare settings [6].

5.4 Domain Shift and Generalization

Models trained on data from a single institution or imaging device often underperform when applied to external datasets due to domain shift (variations in equipment, imaging protocols, and patient demographics). Ensuring robustness across diverse clinical environments remains a challenge [4][9].

5.5 Clinical Validation and Workflow Integration

Many deep learning models show promising results in research settings but face barriers in real-world deployment. Challenges include lack of large-scale clinical trials, integration difficulties within existing healthcare workflows, and regulatory and ethical concerns [2] [10].

5.6 Privacy and Data Security

Collaborative model training using federated learning [10] aims to preserve data privacy without sharing patient data. However, practical implementations are limited by system complexity, communication overhead, and challenges in maintaining model convergence. Table 5 shows the Summary of Key Challenges.

Table 5: Summary of Key Challenges.

Challenge	Impact
Data scarcity	Poor generalization, overfitting
Interpretability	Reduced clinical trust, regulatory issues
Computational cost	Limited access in low-resource settings
Domain shift	Poor cross-institutional generalization
Clinical validation	Hindered real-world adoption
Privacy and security	Barriers to collaborative model development

6. Future Research Directions

Despite notable advancements, several research avenues remain underexplored or insufficiently addressed in current literature. Based on the identified challenges and survey analysis, future work should focus on the following areas:

6.1 Development of Data-Efficient Models

Future models must reduce dependency on large, labelled datasets. Research into self-supervised, semi-supervised, and weakly supervised learning can help leverage unlabelled medical images, addressing data scarcity challenges [9].

6.2 Enhancing Model Interpretability

Developing inherently interpretable architectures or reliable explain ability techniques (e.g., attention-based visualization, uncertainty quantification) is critical for building clinician trust and regulatory acceptance [1][8]. Integration of explainable AI (XAI) frameworks into diagnostic pipelines should be prioritized.

6.3 Lightweight and Resource-Efficient Architectures

Designing computationally efficient models, such as lightweight CNNs, compressed Vision Transformers, and edge-deployable architectures, can facilitate adoption in low-resource or remote healthcare settings [6].

6.4 Robustness to Domain Shifts

Future research should focus on creating domain-adaptive models capable of generalizing across varying scanners, imaging protocols, and patient populations. Approaches may include:

Domain adaptation techniques.

Meta-learning for robust feature extraction [4].

Multi-institutional training frameworks.

6.5 Clinical Validation and Real-World Integration

There is a need for:

Large-scale, multi-center clinical trials to validate model reliability.

Seamless integration of AI tools into clinical workflows, addressing real-world constraints like interoperability and regulatory standards [10].

6.6 Privacy-Preserving Collaborative Learning

Further advancements in federated learning and privacy-preserving machine learning are required to enable multi-institutional collaboration without compromising patient privacy [10]. Solutions should address communication overhead, model divergence, and system scalability.

6.7 Combining Generative and Discriminative Models

Novel models that combine GAN-based data augmentation with discriminative models (CNNs and ViTs) might enable better performance with limited data. This hybrid modelling method has not been extensively investigated yet in clinical domains. Table 6 shows the Summary of Future Research Priorities.

Table 6: Summary of Future Research Priorities.

Research Area	Objective
Data-efficient learning	Reduce dependency on labeled datasets
Explain ability and interpretability	Increase model transparency and trust
Lightweight architectures	Enable deployment in resource-limited areas
Domain adaptation and robustness	Improve generalization across datasets
Clinical validation	Support real-world adoption
Privacy-preserving learning	Enable secure, collaborative training
Hybrid generative-discriminative models	Improve accuracy under limited data

7. Conclusion

This review systematically summarized the application of deep learning approaches in medical image diagnosis, ranging from classic models including CNNs, new generations of architectures such as Vision Transformers (ViTs), and generative models including GANs. By tasks, such as image classification, segmentation, and anomaly detection, across them, deep learning shows excellent performance compared to traditional diagnosis.

Although CNNs largely dominate the design space as they can perform spatial feature extraction, ViTs and GANs are increasingly reshaping the methodological landscape with their own advantages in global context learning and data augmentation, respectively. Nowadays models such as U-Net and DeepLab are still outstanding on segmentation problems which are fundamental in accurate medical interventions.

However, limitations, including data sparsity, black-box models, intensiveness of computation cost, and poor generalization when applied to heterogeneous clinical settings, hinder its wide spread. In light of this, current research is moving toward data-efficient learning, explainable AI (XAI), federated learning, and privacy-preserving approaches to mitigate these problems.

Finally, closing the gap between algorithm development and clinical usefulness necessitates further study of model transparency, robustness and large-scale clinical validation. The survey provides a fundamental source of reference for researchers to develop AI enabled medical diagnosis, and to promote AI into real healthcare practice.

References

1. G. Litjens, T. Kooi, B. E. Bejnordi, et al., “A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2017.
2. D. Shen, G. Wu, H.-I. Suk, “Deep learning in medical image analysis,” Annual Review of Biomedical Engineering, vol. 19, pp. 221–248, 2017.
3. A. S. Lundervold, A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” Zeitschrift für Medizinische Physik, vol. 29, no. 2, pp. 102–127, 2019.
4. A. Egala, M. Sairam, “Convolutional neural networks in medical image diagnosis: A focused survey,” IEEE Reviews in Biomedical Engineering, vol. 17, pp. 100–115, 2024.
5. T. Heng, K. Prasad, “Generative adversarial networks for medical imaging: A systematic review,” Computers in Biology and Medicine, vol. 169, 105475, 2025.
6. L. Lepcha, R. Mehta, “Vision transformers and GAN-based models for medical image super-resolution: A recent review,” IEEE Transactions on Medical Imaging, vol. 44, no. 4, pp. 1452–1465, 2025.
7. Y. Chen, F. Lin, “Deep learning-based medical image registration: A contemporary survey,” Computerized Medical Imaging and Graphics, vol. 107, 102178, 2025.
8. R. Wang, J. Patel, “Label-efficient learning for medical image analysis: A comprehensive review,” IEEE Access, vol. 12, pp. 76832–76845, 2024.
9. M. Jin, L. Huang, “Semi-supervised and weakly-supervised learning in medical imaging: Trends and challenges,” Journal of Digital Imaging, vol. 37, no. 1, pp. 45–59, 2023.
10. K. Guan, Z. Liu, “Federated learning for privacy-preserving medical image analysis: A systematic review,”