Review

A Review on Deep Learning Techniques for Medical Image Segmentation and Classification

Seyedeh Sara Hosseini *

Posted Date: 28 May 2025

doi: 10.20944/preprints202505.2151.v1

Keywords: AI; Deep Learning; Medical Imaging; Classification; Neural Network

 doi:10.20944/preprints202505.2151.v1

Review

A Review on Deep Learning Techniques for Medical Image Segmentation and Classification

Seyedeh Sara Hosseini

Affiliation; syedehsarahosseini@gmail.com

Abstract: The intersection of artificial intelligence (AI) with medical imaging has advanced at an unprecedented rate over the last decade, enabling the diagnosis of disease, the planning of treatment, and monitoring of patients. Deep learning, specifically convolutional neural networks and their derivatives, has been employed to interpret medical images with specialized methods and proven high accuracy in working with complex forms of medical imaging. This review summarizes the most well-known of deep learning architectures in the implementation of medical image analysis, including CNNs, U-Net, ResNet, DenseNet, GANs, and transformer architecture-based models. We discuss the most common applications (e.g., tumor detection, organ segmentation of medical images and image enhancement) associated with each of the architectures. Even with successful applications of AI analyses in medical imaging, specific challenges remain that limit adoption of AI tools by the clinical community that range from lack of annotated data to questions of interpretability and ultimately clinical applicability. We highlight the current work and future directions intended to mitigate main challenges, thus maximizing the potential of AI in clinical practice. This article serves as a detailed overview of AI and medical imaging, and will serve as a deep resource to the reader that wishes to engage in the discussion of AI and its applications in medical imaging.

Keywords: AI; Deep Learning; Medical Imaging; Classification; Neural Network

1. Introduction

In the last several years, the expanding intersection of medicine and artificial intelligence (AI) has brought on unprecedented advancements in medical imaging interpretation and processing. Deep learning—a series of algorithms based on the structure and operation of the human brain—is perhaps the most impactful AI extension in this regard. Deep learning has been an impressive agent for the interpretation of complex visual data, sometimes rivalling or even bettering human experts at certain diagnostics [1,2].

Medical imaging is an essential component of modern medicine, enabling physicians to diagnose disease, plan treatment, and monitor recovery. Modalities like MRI, CT, ultrasound, X-rays, and PET scans generate enormous volumes of data to be read. Traditionally, radiologists read these images using experience and visual inspection—a time-consuming, error-prone, and unscaleable process [3,4].

Deep learning reshapes this landscape by allowing computers to learn autonomously patterns and features from large sets of medical images. Instead of writing rules by hand or extracting features manually, models such as convolutional neural networks (CNNs) learn from data directly, occasionally enhancing the accuracy as well as the efficiency. Such models have been applied to tumor detection, organ segmentation, anomaly classification, and even image enhancement [5,6].

While their growing use, deep learning models remain beset by several challenges, such as the existence of limited amounts of labeled data, model interpretability, and integration into clinical workflows. In this work, we try to provide an overview understandable to the field of how medical image analysis is being transformed by deep learning. We describe the fundamental architectures of these models, summarize their uses, and outline the main challenges that remain below [7,8].

 doi:10.20944/preprints202505.2151.v1

2. Medical Image Analysis Architectures in Deep Learning

Deep learning has offered a range of neural network architectures, each applied for a specific type of task in image analysis. While many were originally developed for natural images, scientists have adapted them successfully to suit the particular challenges in deep learning for medical imaging—low data availability, 3D imaging, and pixel-level accuracy [9].

2.1. Convolutional Neural Networks (CNNs)

CNNs are the backbone of most image-based deep models. They are employed for spatial pattern recognition through convolutional layers and have been broadly used for classification, feature detection, and also medical image registration [10].

2.2. U-Net and Its Variants

One of the most influential architectures for medical image segmentation is the U-Net. It uses an encoder-decoder architecture with skip connections, which help preserve fine-grained spatial information. U-Net is very effective when there is scarce labeled data, a common situation in medical applications. Versions like 3D U-Net and Attention U-Net extended its usage to volumetric data and focusing attention on the target areas [5,6,11].

2.3. ResNet and DenseNet

Deeper models such as ResNet (Residual Networks) and DenseNet have been adopted for their ability to improve training stability as well as model accuracy. ResNet uses skip connections to reverse the vanishing gradient problem, while DenseNet allows feature reuse through dense connectivity. They have both been successful in classification tasks as well as feature extractors for hybrid systems [10,11].

2.4. V-Net

V-Net is a 3D U-Net specifically designed for volumetric segmentation. It uses 3D convolutions and is best suited for modalities like MRI and CT, which are inherently three-dimensional [9].

2.5. Generative Adversarial Networks (GANs)

GANs have opened up new opportunities for image synthesis and enhancement. In medical imaging, GANs have been used to reconstruct images, eliminate noise, and even generate synthetic data for augmenting training datasets. However, their training is often unstable, and their output is sometimes difficult to interpret for clinical application [12].

2.6. Transformer-based Architectures

As spurred by developments in natural language processing, transformers are increasingly making their presence known in medical image processing. Models like TransUNet combine the capacity of U-Net to localize with global contextual modeling via transformers. While promising, these models have a tendency to require vast amounts of data along with compute-intensive abilities [13].

 doi:10.20944/preprints202505.2151.v1

Table 1. Summary of the most prominent deep learning architectures used in medical image analysis, including their key characteristics, common applications, advantages, limitations, and relevant references.

Architecture	Key Characteristics	Common Applications	Strengths	Limitations	References
CNN (Classic)	Uses convolutional layers to extract spatial features	Classification, detection	Fast training, well-studied	Struggles with localization & spatial precision	[1,10]
U-Net	Encoder-decoder with skip connections	Segmentation (e.g., tumors, organs)	Excellent for pixel-level prediction with limited data	May overfit small datasets	[5,6]
3D U-Net	3D convolutions for volumetric data	3D segmentation (MRI, CT)	Preserves spatial context in 3D	Computationally expensive	[6,15]
ResNet	Deep architecture with residual blocks	Classification, feature extraction	Solves vanishing gradient problem	Complex for small datasets	[10]
DenseNet	Feature reuse via dense connections	Image classification, anomaly detection	Efficient in parameter usage	Slower training	[11]
V-Net	3D segmentation (volumetric U-Net variant)	Prostate, liver, brain segmentation	Works well with volumetric labels	High memory usage	[9,15]
GANs (e.g., Pix2Pix, CycleGAN)	Generative adversarial training	Image reconstruction, synthesis	High-quality image generation	Unstable training, hard to interpret	[12,16,17]
Transformer-based (e.g., TransUNet, Swin UNet)	Uses self-attention for spatial modeling	Segmentation, classification	Captures global context	Requires large datasets, high compute	[13,14,18]

3. Difficulties in Implementing Deep Learning in Medical Imaging

While deep learning in medical imaging is advancing tremendously, there remain a number of challenges that hinder their widespread implementation in clinical settings:

3.1. Lack of Labeled Data

High-quality labeled datasets in medical images are difficult to acquire, mainly because of privacy issues, the need for expert annotators to label the images and the costs. The scarcity of labeled data limits the ability of models to train and generalize, particularly because many medical imaging tasks use supervised training [7,14].

3.2. Model Explainability

Deep learning models, when appropriately constructed, are so-called "black boxes," meaning it is difficult for clinicians to understand how the decision was made. A lack of understanding hinders trust and buy-in in the clinical workflow [8,19].

3.3. Data Heterogeneity

Medical images, which can include regard to either modality of imaging, acquisition parameters, population demographics, and institution-specific influences vary immensely. Therefore, a model trained on a specific dataset may not generalize to other datasets [3,14].

3.4. Clinical Workflow

Integrating AI tools into existing hospital infrastructure remains a technical and operational difficulty given the need to collaborate with other clinicians [7].

 doi:10.20944/preprints202505.2151.v1

3.5. Regulatory and Ethical Issues

There exists numerous issues related to privacy of the patient, establishing regulatory approval and ethical issues related to bias and accountability related to Artificial Intelligence that create challenges regarding clinical deployment [19].

4. Future Directions and Opportunities

AI in medical imaging is swiftly changing and presents dynamic future possibilities.

4.1. Self-Supervised and Semi-Supervised Learning

Self-supervised and semi-supervised learning methods can harness vast amounts of unlabeled data to enhance the performance of machine-learning models [14], addressing the challenge of data scarcity.

4.2. Explainable AI (XAI)

The development of ways to make AI models more transparent and interpretable will allay clinician worry about how clinical decision support can be operationalized, as well as acceptability by regulatory bodies to make the technology available to physicians [19].

4.3. Multi-Modal Learning

Multi-modal learning is based on the idea that we can combine imaging data with other related clinical data (e.g., genomic data, electronic health record data) in an informal way to provide additional context for diagnostic decisions, and improve diagnostic performance and accuracy [13].

4.4. Real-Time AI Assistance

The deployment of AI models that incorporate image acquisitions does so in real-time at the point of diagnosis or clinical aquantification will create fundamentally new workflows and impact patient outcomes positively [7].

4.5. Federated Learning

The utilization of federated-learning permits decentralized training of machine-learning models across multiple institutions while preserving patient data privacy and not copying patient data to other institutions [14].

5. Discussion

The application of deep learning for medical imaging has initiated a paradigm-shift in medical imaging, but there is still a long way to go for a seamless integration into routine clinical practice. The studies discussed here highlight some compelling architectures that successfully automated complex analyses of medical imaging data with an accuracy level that parallels or exceeds the level of confidence provided by human experts in select applications. Yet serious pitfalls still must be addressed.

Firstly, data quality and availability remain key bottlenecks. Contrary to a dataset of natural images, a medical image must be annotated by an expert, potentially creating a serious barrier because of the time and cost involved. Furthermore, the heterogeneity of imaging devices, imaging protocols, and patient populations create serious concerns with generalizability of a model. Identifying techniques such as transfer learning, data augmentation, and federated learning may provide encouraging alternatives but have not been fully developed through implementation.

Second, while the ability to trust a model is important in any decision-making process, the complexity of healthcare demands at least some level of model interpretability. Understanding why an AI tool is making a certain decision is important to clinicians trusting and using these tools. While

 doi:10.20944/preprints202505.2151.v1

the "black box" nature of machine learning-based models is very promising, this may present an impediment to acceptance because dilemmas may or may not have medical implications. There is an acute need for the development of explainable AI methods and we suggest the use of concordant reporting standards.

Lastly, creating an effective clinical pathway is challenging from both a technological and logistical standpoint. To avoid adding additional burdens on overworked team members, AI systems must convey and process data smoothly into the clinical workflow. One successful solution to this problem may come from collaborating with engineers and developers of AI, radiologists, and the IT teams within the hospitals to ensure a concerted effort is made.

6. Conclusions

Deep learning has transformed the reading of images for medical analysis to be both more accurate and efficient when interpreting complex imaging data. Improvements in architectures like CNNs, U-Net variants, GANs, and Transformers have helped enable applications from tumor detection to image enhancement. Some challenges remain as well: limited labeled data, interpretability, and clinical integration, for example. All of these must be resolved to realize the full potential of AI in medicine, and it will take the ongoing commitment of AI researchers, clinical delivery experts, and policy makers to transition these promising applications into everyday clinical practice.

References

1. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88.
2. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-118.
3. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol. 2017;10(3):257-273.
4. Suzuki K. Statistical machine learning in medical imaging. Springer; 2019.
5. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. MICCAI 2015.
6. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. MICCAI 2016.
7. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221-248.
8. Greenspan H, van Ginneken B, Summers RM. Deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Trans Med Imaging. 2016;35(5):1153-1159.
9. Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 3DV 2016.
10. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. CVPR 2016.
11. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. CVPR 2017.
12. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. NIPS 2014.
13. Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. MICCAI 2021.
14. Litjens G, Sánchez CI, Timofeeva N, et al. Deep learning as a tool for medical image analysis: overview, challenges and future promises. Med Image Anal. 2021;67:101815.
15. Shen Y, Dong J, Luo Y, et al. Transformer-based deep learning for medical image analysis: a review. Comput Methods Programs Biomed. 2023;230:107339.
16. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A nested U-Net architecture for medical image segmentation. IEEE Trans Med Imaging. 2019;39(6):1856-1867.
17. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal. 2019;58:101552.
18. Wang G, Li W, Ourselin S, Vercauteren T. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. MICCAI 2017; pp. 178-190.
19. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. ICLR 2021.
20. Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. MICCAI 2021.

 doi:10.20944/preprints202505.2151.v1