Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Smart Medicine
DOI: 10.54254/2753-8818/2026.HZ37412

Challenges and Emerging Solutions for Clinical Deployment of Artificial Intelligence Medical Imaging

Siwen Wang

Shenzhen College of International Education, Shenzhen, China
S24657.wang@stu.scie.com.cn

Abstract. Artificial intelligence (AI) is now a very promising technology for improving the efficiency of medical image analysis in CT, MRI, X-ray and other kinds of images. With the development of CNNs into Vision Transformers (ViTs) and foundation models, AI power has spread across many fields; thus, high-precision image classification and segmentation are now being carried out, and improvements in the efficiency and accuracy of medical diagnosis can be achieved. Although research on AI has achieved good results, only a small number of AI tools have been applied in hospitals to date. This paper will introduce the current situation of development for AI in medical imaging and discuss possible remedies for existing deficiencies and application scenarios of AI in diagnosis. At present, the problems that are faced by developers include the lack of sufficient amounts of annotated data, weak generalization of AI models, the issue of patients' privacy, and the problem of AI interpretability.

Keywords: Artificial intelligence, Medical imaging, Convolutional neural networks (CNNs), Vision Transformers (ViTs), Clinical deployment

1. Introduction

At present, artificial intelligence (AI) is being used in many places around the world to improve the speed and accuracy of diagnosis for diseases through medical images. To help diagnose and plan treatment for the disease, other ways are used to study the sick [1, 2]. With the development of healthcare systems in recent years, radiologists' work is increasingly dominated by analysing medical images. Advanced applications of artificial intelligence in medical imaging have reduced the problem. Artificial Intelligence (AI) in deep learning is now used to automate medical image analysis. Many studies have found that the diagnostic results of AI systems are comparable to those of experienced specialists in many areas of medical imaging. Thus, AI is now used to help diagnose diseases and make decisions about treatment based on medical images.

Although artificial intelligence technology for medical imaging is developing at a high speed, it has not been widely applied in hospitals yet. Although AI has achieved good research results, only a few systems have been applied in hospitals so far, and many practical obstacles need to be addressed before they can be used widely. Now that the main problem of model accuracy has been resolved, others include a lack of annotated data, poor generalisation across institutions, and a lack of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ37412

interpretability [3]. The above problems need to be solved to improve the stability and real-world applicability of AI systems for clinical use.

2. From CNNs to vision transformers and foundation models

2.1. CNN-based medical imaging

Convolutional Neural Networks (CNNs) have been applied to medical image analysis to some extent. CNNs have recently been used to automatically extract useful features from complex medical images to improve the performance of systems. At the bottom are several necessary layers of a CNN. As shown in Figure 1, the first few layers of a CNN generally consist of multiple convolutional layers, which use small filters (kernels) to slide over the input image and produce feature maps. These feature maps are the high-level features. After several convolutional layers, CNNs learn increasingly complex features and transition from simple-level learning to deep-level disease diagnosis.

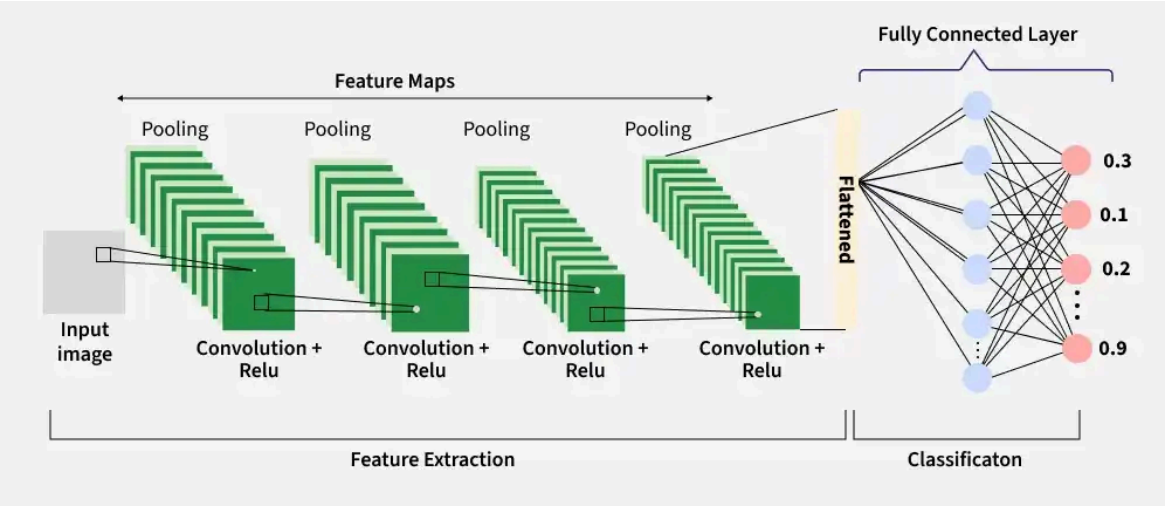

Figure 1. Convolution in CNN can be used to obtain features of different levels from the input image. Pooling reduces the size of the image and keeps the important information. The resulting feature maps are the features extracted by the CNN. Finally, a fully connected layer is used with the extracted features for the final prediction

CNNs are used in medical image analysis, and several layers of a neural network are combined to extract relevant features to help diagnose diseases in these images. A convolutional layer is followed by a ReLU activation function to add non-linearity to the network. The above simple operations can be used to learn complex, non-linear relationships in medical images more efficiently. Then, a pooling layer is used to reduce the size of the feature map through downsampling, thereby keeping essential features and lowering the amount of computation. Finally, the fully connected layer combines the extracted features and outputs the final prediction; this prediction can then be used as a diagnostic reference. Together, these layers help CNNs extract and learn from various levels of features in medical images to perform various tasks.

The Development of CNN-based models has addressed several problems in medical image analysis. For example, in image classification, a CNN can distinguish between normal and abnormal images to determine if a chest X-ray shows signs of lung cancer [4, 5]. The new C-Nets have improved significantly for use in medicine.

DOI: 10.54254/2753-8818/2026.HZ37412

CNNs have certain deficiencies that are not suitable for clinical application. First, CNNs generally need a large number of high-quality labelled medical data for training, and obtaining such data is difficult due to the considerable experience required [6]. Second, CNNs focus on local features through convolution operations, so they are not well-suited for learning long-distance dependencies among different areas in an image. This deficiency will impact on the performance of far-off structure interaction and reduce interpretation efficiency.

2.2. Vision transformers in medical imaging

Due to some deficiencies in CNNs, Vision Transformers (ViTs) have recently been applied to address these issues for medical image analysis. ViTs divide an image into multiple small patches and treat these patches as a sequence of visual tokens, rather than using the local convolution operation in CNNs. Then, using self-attention, the model learns the relationships among different areas in the image [7].

The self-attention mechanism can determine how much attention to pay to other patches in an image by ViTs. It can gather global information and learn long-range dependencies in the data better than traditional CNNs. Instead of first focusing on a small area, as CNNs do, ViT can learn relationships among various parts of the body to find a cause of disease.

ViTs are also more scalable than the previous models. Given a large amount of data, ViTs can be made more powerful by increasing their scale. Therefore, they can learn from experience and suitable for large-scale medical imaging applications that have abundant data.

There are some defects in ViTs as well. CNNs have built-in image-specific biases, so ViTs need a larger amount of data to learn these features effectively, and thus are difficult to train. ViTs are relatively computationally expensive; therefore, their implementation and widespread application are more difficult.

2.3. Foundation models for medical imaging

Based on the development of CNNs and ViTs, foundation models are now the next generation of AI systems for medical images. Foundation models are large-scale models that have been trained on a vast number of varied data to learn general representations applicable to multiple tasks.

Foundation models in medical imaging have reduced the demand for large amounts of labelled data to some extent. Through self-supervised learning, these models can learn useful visual representations from large amounts of unlabelled medical images and thus ensure that there are plenty of medical images.

Foundation Models also have their own problems. Large-scale training is relatively resource-intensive and demands high-quality data. In addition, the models must also be reliable, interpretable and safe before they can be used widely in healthcare.

In short, the development of CNNs, Vision Transformers and foundation models has been advancing medical AI at a high speed. CNNs first provided the foundation for automatic image analysis by learning features, and then ViTs improved this with self-attention. Foundation Models expand the scope of AI to provide flexible applications. Together, these optimisations offer some paths to address the current obstacles and realise clinical applications of medical AI.

DOI: 10.54254/2753-8818/2026.HZ37412

3. Solutions for clinical deployment of AI

Although CNNs have performed well, ViTs and foundation models for medical image analysis still have some deficiencies that limit their wide-scale application in clinical practice. Therefore, the solutions will be to reduce reliance on labelled data, improve model generalisation, and protect patient privacy.

3.1. Self-supervised learning and semi-supervised learning

The first problem of medical AI is a shortage of high-quality labelled data. Medical images are not natural pictures, so annotating them requires experts and is relatively expensive and time-consuming [8]. Therefore, self-supervised learning has been gradually applied to automate the annotation of images by learning.

Self-supervised learning can be used to learn good representations of medical images without labels for AI models. Without human annotations, self-supervised learning generates training tasks directly from the data to help the model learn good visual features and structures. Thus, a smaller amount of annotation is needed, and the generalisation ability of AI models across different medical imaging tasks will be enhanced [9, 10].

Semi-supervised learning is another way that uses a small amount of labelled data in conjunction with a large volume of unlabelled data. Both types of information can be used to enhance the performance of the model and reduce annotation costs, thus achieving a high-cost-effectiveness.

Together, self-supervised learning and semi-supervised learning have been successful in addressing the problem of limited data and have promoted new developments in artificial intelligence for medicine.

3.2. Generative AI

Generative Artificial Intelligence has recently been developed to address the problem of insufficient medical data in another way. Generative AI can create new fake medical images by means other than those of traditional AI models.

GANs and diffusion models are typical examples of generative models. The two networks in a GAN are a generator and a discriminator, and together they generate realistic fake data. Diffusion models generate images by gradually denoising random noise to produce meaningful medical images through a learning process [11].

Several applications of generative AI in medical imaging include generating synthetic medical images for model training, expanding the pool of data for rare diseases with limited samples, and improving dataset balance by increasing the representation of underrepresented cases.

Generative AI has new problems as well. Before using the synthetic images, check if they are good and accurate; otherwise, the model may not work well with fake data. The data generated may also contain the same biases present in the training set.

4. Future outlook: multimodal generative AI

The future development of AI in medicine will likely be multimodal intelligent models, moving beyond the current focus on image analysis. All kinds of data are to be collected here, such as CT images, MRI scans, pathological images and electronic health records (EHRs).

DOI: 10.54254/2753-8818/2026.HZ37412

Multimodal foundation models can use all these kinds of different information to construct a full-featured model of a person's health. The future system will be able to analyse medical images and, at the same time, integrate the results of analysis with laboratory information for personalised recommendations.

Some achievements have been made to date; however, many problems still need to be solved before multimodal AI can be applied broadly in healthcare. Among them are the requirement for regulatory approval, clinical validation and ethics.

Therefore, in the future, the development of medical AI will require not only new technologies but also evaluations and applications of these technologies, as well as cooperation among users. AI researchers, clinicians and other institutions in the field of health.

5. Conclusion

Artificial intelligence has greatly advanced medical imaging, and now it is based on vision transformers and foundation models that learn more sophisticated representations than CNNs. However, it can be seen that there are still many problems to be solved before the clinical application of machine learning can be successfully realised, such as the lack of sufficient training data, problems of generalisation, privacy concerns and a lack of interpretability. Now, self-supervised learning and generative artificial intelligence are relatively advanced approaches that can be used. In the future, multimodal generative AI will use the available clinical data to provide more support for healthcare.

References

[1] Obuchowicz, R., Strzelecki, M., & Piórkowski, A. (2024). Clinical applications of artificial intelligence in medical imaging and image processing: A review. *Cancers, 16*(10), 1870.

[2] Wenderott, K., Krups, J., Zaruchas, F., & Weigl, M. (2024). Effects of artificial intelligence implementation on efficiency in medical imaging—a systematic literature review and meta-analysis. *NPJ Digital Medicine, 7*(1), 265.

[3] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine, 17*(1), 195.

[4] Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., ... & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health, 1*(6), e271-e297.

[5] Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine, 25*(6), 954–961.

[6] Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., ... & Hamamoto, R. (2024). Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. *Journal of Medical Systems, 48*(1), 84.

[7] Li, J., Chen, J., Tang, Y., Wang, C., Landman, B. A., & Zhou, S. K. (2023). Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives. *Medical Image Analysis, 85*, 102762. https://doi.org/10.1016/j.media.2023.102762

[8] Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., ... & Norouzi, M. (2021, October). Big self-supervised models advance medical image classification. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 3458-3468). IEEE.

[9] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning* (pp. 1597-1607). PMLR.

[10] Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis, 58*, 101552.

[11] Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature, 616*(7956), 259-265.