

Review

# Large scale models in radiology: revolutionizing the future of medical imaging

Yilin Li<sup>a,b,1</sup>, Jiaojiao Liu<sup>c,1</sup>, Yanyan Zhang<sup>c,1</sup>, Jingwei Wei<sup>b,d,1</sup>, Yang Luo<sup>b,d,e</sup>, Xue Yang<sup>f</sup>, Yanbin Shi<sup>g</sup>, Lingling Zhao<sup>g</sup>, Wanshui Yang<sup>h,\*</sup>, Hongjun Li<sup>c,\*</sup>, Jie Tian<sup>b,d,i,j,\*</sup>

<sup>a</sup>School of Public Health, Anhui Medical University, Hefei, Anhui, China

<sup>b</sup>Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

<sup>c</sup>Department of Radiology, Beijing Youan Hospital, Capital Medical University, Beijing, China

<sup>d</sup>Beijing Key Laboratory of Molecular Imaging, Beijing 100190, China

<sup>e</sup>School of Life Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China

<sup>f</sup>The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

<sup>g</sup>Department of Radiology, The Sixth Peoples Hospital of Zhengzhou, Henan, China

<sup>h</sup>Department of Nutrition, School of Public Health, Anhui Medical University, Hefei, Anhui, China

<sup>i</sup>Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing 100191, China

<sup>j</sup>Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, Shaanxi 710126, China

<sup>1</sup>These authors contributed to the work equally and should be regarded as co-first authors.

\*Correspondence: [wanshuiyang@gmail.com](mailto:wanshuiyang@gmail.com) (W. Yang); [lihongjun00113@126.com](mailto:lihongjun00113@126.com) (H. Li); [tian@ieee.org](mailto:tian@ieee.org) (J. Tian)

Received: 22 November 2023; Revised: 12 January 2024; Accepted: 17 January 2024

Published online: 3 February 2024

DOI 10.15212/RADSCI-2023-0018

## Abstract

In the domain of medical image analysis, there is a burgeoning recognition and adoption of large models distinguished by their extensive parameter count and intricate neural network architecture that is predominantly due to their outstanding performance. This review article seeks to concisely explore the historical evolution, specific applications, and training methodologies associated with these large models considering their current prominence in medical image analysis. Moreover, we delve into the prevailing challenges and prospective opportunities related to the utilization of large models in the context of medical image analysis. Through a comprehensive analysis of these substantial models, this study aspires to provide valuable insights and guidance to researchers in the field of radiology, fostering further advances and optimizations in their incorporation into medical image analysis practices, in accordance with the submission requirements.

**Keywords:** Large scale model, Neural network architecture, Medical image analysis, Artificial intelligence, Radiomics

## 1. INTRODUCTION

In recent years the rapid advancement of artificial intelligence (AI) technology has sparked a growing interest in the integration of large scale models within the field of medical imaging. Large models, often denoting neural network models with a profusion of parameters, intricate architecture, and abundant neurons, have gained prominence in the realm of deep learning. Large models consistently exhibit exceptional performance and robust generalization capabilities, rendering them versatile tools in various domains, including medical image analysis and natural language processing [NLP] [1]. The roots of large models can be traced back to the foundational concept of neuron modeling that

was initially proposed by Warren McCulloch and Walter Pitts in 1943. However, the field of deep learning, where these expansive models reside, grappled with technical constraints for an extended period [2]. It was not until 2012, when Alex Krizhevsky and colleagues introduced the AlexNet model, that a crucial turning point was reached. Their victory in the ImageNet image classification competition underscored the considerable advantages of large deep models in computer vision tasks, marking the dawn of an era where substantial models flourished [3]. Subsequent developments introduced models, such as VGG, GoogLeNet, and the Residual Neural Network (ResNet), all contributed significantly to enhanced model performance [4-6]. The subsequent advent of the Transformer model pioneered the concept

of self-attention, providing the foundation for large-scale language modeling [7]. Building upon this foundation, researchers unveiled the Bidirectional Encoder Representations from Transformers (BERT) model, bringing about substantial improvements in the performance of large models in NLP tasks [8]. Today, the OpenAI's Generative Pre-trained Transformer (GPT) series of models, boasting billions of parameters, exemplify remarkable capabilities [9]. The evolution of large models, as witnessed today, owes much to the contributions of computational resources. In this context, computational resources encompass the hardware devices employed for training deep learning models, the computational time required for this training, and the energy consumption necessary to sustain these devices. Since the release of the AlexNet model in 2012, there has been an exponential surge in the utilization of computational resources by researchers for model training. The deployment of large-scale computational resources has significantly enhanced model performance [10]. However, the extensive use of large-scale computational resources has also given rise to a set of challenges, including substantial economic investments, heightened energy demands, increased carbon emissions, and concerns related to research inequality [11].

Despite the persistent nature of these challenges, the pivotal role played by large models in the realm of medical image analysis remains undisputed. The hierarchical architecture inherent in deep neural networks within large models facilitates a systematic process for the identification and accentuation of crucial features within input images, while concurrently eliminating superfluous elements. This process reveals the intrinsic characteristics latent within the original images [12]. This remarkable capability empowers large models to conduct medical image analysis with enhanced efficiency and precision. Moreover, the integration of large models in the field of medical imaging has catalyzed innovative research directions in medical image analysis, encompassing automated image segmentation and the automated generation of comprehensive medical image analysis reports [13].

The applications of large models in the field of medical imaging encompass several distinct areas. The first of these is image classification and segmentation, a critical task in medical image analysis that finds wide utility in assisted diagnosis and lesion localization. Large models can autonomously discern salient features within original medical images, offering precise classification outcomes and the ability to delineate and segment various tissues and organs within the image with exceptional accuracy [14, 15]. The second area focuses on the detection and prediction of anomalies within medical images. Large models exhibit the capability to identify diverse pathologic anomalies, including viral infections, exemplified by the capacity to detect early-stage COVID-19 infections through medical images [16]. Furthermore, large models can effectively forecast disease onset or future

progression, as evidenced by the accurate predictions in the context of glaucoma onset and progression [17]. The third domain pertains to multimodal medical image analysis, which addresses the multifaceted nature of contemporary medical image data. Large models adeptly combine multiple types of medical images, extracting common features across all modalities to effectively analyze target images [18]. Lastly, large models play an invaluable complementary role for radiologists. Large model applications, whether focused on image segmentation or anomaly detection, significantly streamline the work of radiologists, enhancing the accuracy and efficiency of their tasks within the realm of medical image analysis.

## 2. METHODS

In this section the fundamental architecture of large models, along with the training strategies and optimization techniques, will be introduced. The goal is to foster a deeper comprehension of large models.

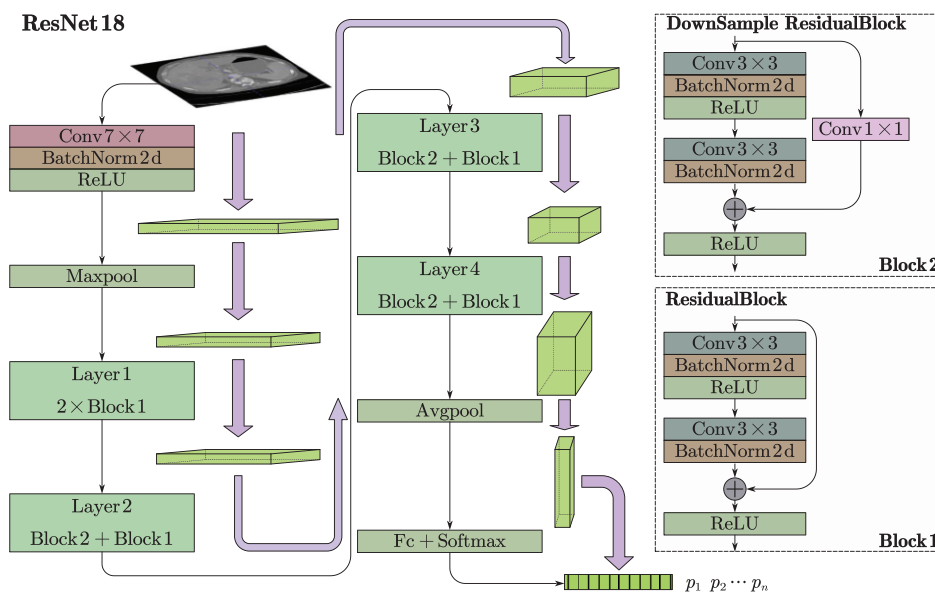
### 2.1 Basic architecture

**2.1.1 Categories of models.** The landscape of large models has evolved significantly over time. Currently, these models can be broadly classified into the following three groups based on their foundational architectural structures:

#### (1) Convolutional Neural Networks (CNNs)

CNNs typically comprise three distinct types of layers (convolutional, pooling, and fully connected). Within the convolutional layer, a pivotal element emerges—the convolutional kernel, which is often referred to as the filter. The presence of these filters empowers CNNs to adeptly discern salient features within input images, enhancing the efficiency of image processing. However, traditional CNNs grapple with certain limitations, including the challenge of gradient vanishing, which hinders the capacity of the model to grasp intricate features [19]. In response, researchers have devised innovative models building upon the foundation of CNN architecture to surmount these challenges. Noteworthy exemplars encompass AlexNet, VGGNet, and ResNet models. AlexNet leverages multiple convolutional layers, ReLU activation functions, maximum pooling, and normalization to optimize model accuracy. VGGNet enhances AlexNet by introducing a sequence of convolutional layers characterized by smaller convolutional kernels, thereby enhancing feature recognition. The ResNet model (Figure 1) introduces the concept of residual learning, effectively addressing the gradient vanishing problem and enabling the model to grasp more intricate features [20]. In the realm of medical image analysis, these models predominantly find application in image classification and segmentation. A multitude of studies have unequivocally demonstrated the adeptness of this class of models in accurately classifying and segmenting medical images [21-23].

# Review



**Figure 1 | ResNet Architecture – Demonstrates how shortcut connections enable residual learning to address the vanishing gradient problem in deep neural networks.**

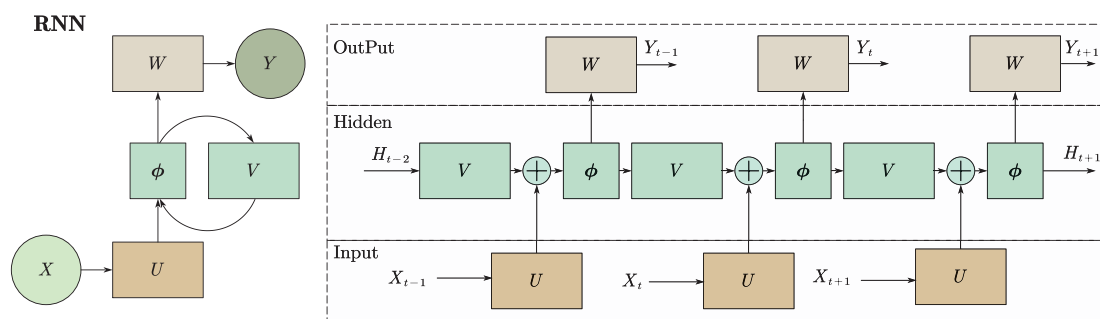
## (2) Recurrent Neural Network (RNNs)

RNNs (Figure 2) typically adopt a tree-like or directed acyclic graph structure. This architectural design facilitates recursive propagation and sharing of information across various network structures. This enables RNNs to demonstrate high efficiency when dealing with tree-structured and hierarchical data, particularly in natural language processing (NLP) or sequential data analysis [24]. By leveraging the functional characteristics of RNNs, the application in medical imaging extends to automating the generation of medical image reports and processing medical sequence data, such as time series images or video data [25, 26].

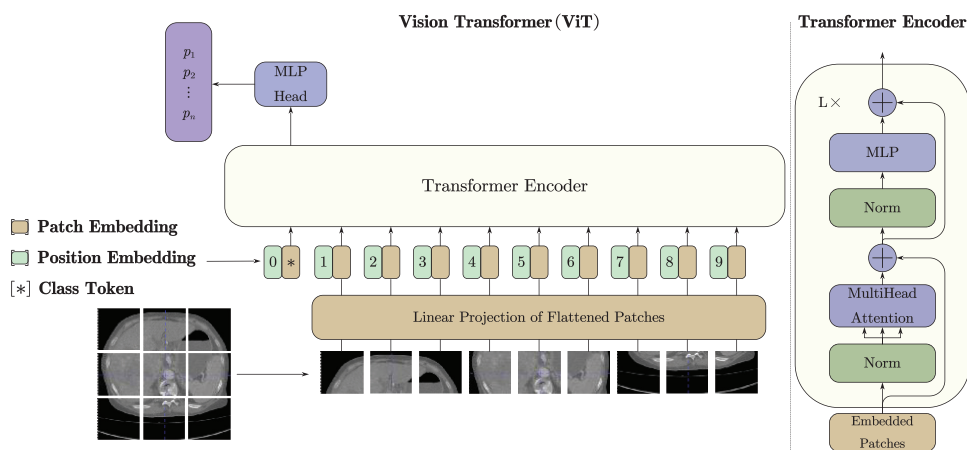
## (3) Transformer Model

The Transformer model (Figure 3) introduces a self-attention mechanism, dynamically adjusting the focus of the model on different segments of the

input based on task-specific features and input data characteristics, thereby enhancing model performance and resilience. This self-attention mechanism, when combined with feed-forward neural networks, enables global context modeling and maintains parallel processing capabilities, especially for extended sequences [7]. Building upon the foundations of the Transformer model, BERT inherits and expands its attributes. BERT introduces bidirectional training, addressing the unidirectional processing constraint of Transformers, while also featuring pre-training and fine-tuning capabilities [8]. The emergence of the GPT series models has elevated the application of neural language models based on the Transformer architecture. GPT-3, equipped with 175 billion parameters, boasts an innovative few-shot learning feature, allowing GPT-3 to proficiently handle various NLP tasks with minimal examples or task descriptions [9].



**Figure 2 | Recurrent Neural Network – Illustrates how RNNs use internal looping structures to handle sequential information, suitable for tasks like language processing and time-series analysis in medical imaging.**



**Figure 3 | Transformer Model – Showcases the self-attention mechanism, which dynamically tunes focus on input segments, thus enhancing performance and adaptability in processing sequential data.**

Hence, the significance of applying Transformer, BERT, and GPT-3 in the field of medical image analysis cannot be overstated. Numerous scholars have authored reviews highlighting the potential of Transformer models for medical image segmentation. BERT models have exhibited exceptional performance in automatic medical image report generation. The exploration of GPT series models for aiding clinical decision-making in the realm of radiology further underlines their promising utility [27-29].

**2.1.2 Model size and complexity.** Until now, all three categories of large models exhibit a considerable scale. Kaplan et al. conducted a comprehensive study to investigate the intricate relationship between model performance and model scale. Kaplan et al. reported that model performance is significantly contingent on the scale of the model in terms of parameters, the extent of the dataset, and the computational resources employed for training. It was evident that the judicious expansion of model size markedly enhances performance. Consequently, large models have demonstrated remarkable efficiency in the analysis of medical images, achieving commendable accuracy in tasks, such as tumor detection, image segmentation, and disease discrimination [30]. However, several challenges persist in the application of large models in the field of medical image analysis. A notable hurdle is the predominantly limited size of medical image datasets, which often fails to meet the demands of training large scale models [31].

**2.1.3 Pre-training and transfer learning.** The advent of pre-training and transfer learning methodologies has effectively addressed the quandary of limited data sizes. Pre-training involves the preliminary training of large models on extensive datasets, enabling the large models to glean generic features and structural knowledge from the data. Subsequently, transfer learning allows

these large models to apply the generalized features acquired during pre-training to specific datasets under scrutiny [32]. Notably, Hopson et al. [33] delved into the utilization of pre-trained CNN models for assessment of the quality of clinical PET images using transfer learning techniques. Hopson et al. [33] demonstrated that pre-training significantly enhances the performance of CNN models in the task of assessing the quality of clinical images, particularly in automating the prediction of PET images.

## 2.2 Training strategies

**2.2.1 Data preparation.** The quality of data significantly influences the performance of large models. It is imperative to undertake meticulous steps in preparing medical image data for model training. Commencing with a comprehensive summary of relevant medical image data for the study is crucial, followed by a rigorous assessment of data reliability. Subsequent steps involve data cleansing to eliminate non-compliant entries, standardization to ensure image consistency, and annotation tailored to the study requirements for effective model learning and comprehension [34].

**2.2.2 Data augmentation.** Beyond the above steps, the inclusion of data augmentation is imperative. Data augmentation involves the generation of new data from existing sources, incorporating techniques, such as rotation, translation, flipping, and cropping. This augmentation of the original dataset is vital to expanding its size and enhancing the generalization capabilities of the model. Given that collected data may fall short in meeting the demands of training a large model in practical scenarios, data augmentation becomes an essential facet of the preparation process [35].

**2.2.3 Loss functions and optimization objectives.** In the training regimen of large models, the meticulous

## Review

selection of appropriate loss functions emerges as a pivotal factor in augmenting model performance. Loss functions serve as metrics to gauge the deviation between the model predictions and actual values, elucidating how closely the model aligns with ground truth. The optimization objective is to minimize this deviation, with a smaller value of the loss function signifying superior model performance [36]. Commonly utilized loss functions in medical image segmentation tasks include Cross-Entropy Loss and Dice Loss. The judicious choice of a loss function in practical research hinges upon the specific data characteristics, research objectives, and the intended applications of large models [37].

### 2.3 Optimization techniques

**2.3.1 Training optimization algorithms.** Following data preprocessing, augmentation, and the selection of appropriate loss functions, a critical facet of training large models revolves around optimization algorithms. These algorithms aim to identify model parameters that minimize the loss function, thereby enabling optimal model performance. Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (Adam) stand out as commonly used optimization algorithms [38]. SGD serves as a foundational optimization algorithm that computes gradients for each training sample and updates model parameters accordingly. By selecting only one training sample at a time, this approach introduces randomness, aiding in escaping local optima and exploring the parameter space more comprehensively. Adam integrates SGD with a momentum-adaptive learning rate optimization algorithm, providing greater stability compared to SGD and facilitating faster convergence towards local optimal solutions [39].

**2.3.2 Regularization and overfitting control.** As model training progresses, the challenge of overfitting becomes increasingly pronounced. Overfitting manifests when the model excels on training data but performs poorly on unfamiliar data. To mitigate this issue, constraints or penalty terms can be incorporated into the loss function to reduce model complexity, an approach known as regularization. Typically, dropout and L2 Regularization serve as effective means for model regularization. Dropout entails the random deactivation of certain neurons during each training iteration, which prevents the model from relying too heavily on specific neurons and thereby enhancing generalization capabilities. In contrast, L2 Regularization introduces penalty terms to the loss function, encouraging optimization algorithms to favor smaller weight values during parameter selection and consequently diminishing the risk of overfitting [40].

**2.3.3 Model compression and acceleration.** Efficiently reducing the time and cost associated with training and inference for large models stands as a crucial facet in the training continuum. Presently, model compression and

acceleration primarily rely on methods, such as model pruning, quantization, and knowledge distillation. Model pruning involves judiciously trimming redundant weights, neurons, filters, and layers within large models based on CNN architecture. This process mitigates model storage requirements and expedites the inference phase. Quantization entails converting floating-point representations of model parameters and intermediate activation values into lower-precision integers or fixed-point numbers. Quantization not only reduces model size but also enhances inference efficiency. Knowledge distillation adopts the following two-step approach: initially training a large model, known as the teacher model; and subsequently constructing a smaller model, referred to as the student model, for the same task. Transferring knowledge from the teacher model to the student model results in a more streamlined architecture that demands fewer computational resources, thereby improving overall inference efficiency [41].

**2.3.4 Distributed and parallel training.** Additionally, distributed and parallel training assumes a pivotal role in expediting the training of large models and processing extensive datasets in medical imaging. Distributed training involves partitioning the parameters and training data of large models into multiple segments, each assigned to multiple computers or compute nodes. Independently computing updates to the model parameters on each node and sharing these updates facilitate simultaneous model training across multiple nodes, resulting in expedited training speeds [42]. In contrast, parallel training, distinct from distributed training (Table 1), necessitates a single computer or computer node. This method leverages multiple processing units within the computer to concurrently process various segments of the training task, thereby augmenting training speed. When applied to medical imaging data, the utilization of distributed and parallel training can notably accelerate the delivery of patient health information to healthcare professionals [43].

The foregoing information provides a foundational understanding of large model architecture, training methodologies, and optimization techniques. Subsequent sections will delve into the research advances and practical applications of large models within the domain of medical imaging.

## 3. EXPLORATION OF LARGE MODELS IN MEDICAL IMAGE ANALYSIS

### 3.1 Application examples

**3.1.1 Precision in image classification and segmentation.** The diligent efforts of researchers have yielded significant strides in the analysis of medical images through the integration of large models. Notably, Jin et al. introduced the RA-UNet model, a sophisticated architecture amalgamating CNNs, residual learning, and attention mechanisms. This model adeptly

**Table 1** | Comparison of distributed and parallel training.

Aspect	Distributed training	Parallel training
Concept	Utilizes a network of interconnected computers for distributed tasks.	Employs multiple processors within a single computer for concurrent tasks.
Primary Goal	To manage and expedite training with large datasets across several machines.	To optimize and expedite training within the constraints of a single machine.
Resource Requirements	Multiple interconnected computers or nodes; network bandwidth and latency are critical.	A computer with multi-core processors; dependent on the quality and number of cores.
Data Handling	Implements data or model parallelism across nodes, splitting tasks among multiple machines.	Executes simultaneous training on different parts or subsets of data or model within the same machine.
Communication Overhead	Higher due to the need for node synchronization and data exchange across the network.	Lower, as all processes occur within the same physical system, minimizing data exchange time.
Scalability Potential	Highly scalable with the ability to add more nodes; influenced by network architecture and data strategies.	Limited to the physical and technical specifications of the single computer; can be extended by upgrading hardware.
Operational Complexity	More complex due to coordination, network configuration, and data distribution across multiple machines.	Relatively simpler in setup but may require sophisticated parallel algorithms to fully utilize all cores efficiently.

achieves precise segmentation of the liver and tumors within three-dimensional computed tomography (CT) images. Leveraging datasets, such as Liver Tumor Segmentation Challenge (LiTS) and 3DIRCADb for model training and evaluation, the study used metrics, including the Dice coefficient and Jaccard index, to gauge segmentation quality. In liver segmentation, RA-UNet attained Dice coefficients of 0.961 and 0.977, along with Jaccard indices of 0.926 and 0.977 on the two datasets. Furthermore, RA-UNet demonstrated robust performance in tumor segmentation across both datasets. A noteworthy innovation in this study was the pioneering use of an attention-residual mechanism for tumor segmentation in three-dimensional medical images. The integration of residual modules within the model enables adaptive adjustments in attention-aware features, thereby amplifying overall model performance [44].

### 3.1.2 Advances in anomaly detection and prediction.

Large models have showcased remarkable progress in the domains of medical image anomaly detection and disease prediction. A notable example is the work of Brown et al., who harnessed deep CNNs for the automated diagnosis of “plus lesions” within retinal images of premature infants, a distinctive characteristic of retinopathy of prematurity (ROP). Given the critical importance of early plus lesion detection for effective ROP management, and considering the inherent low accuracy in clinical diagnosis, this research achieved remarkable precision and reproducibility in plus lesion diagnosis [45].

Furthermore, Jiang et al. introduced the “S-net,” a tailored deep neural network model designed for extracting image features from preoperative CT scans of gastric cancer patients to construct predictive models. These models not only forecast disease-free survival

and overall survival in gastric cancer patients but also identify individuals likely to benefit from postoperative adjuvant therapy. The study unveiled a unique image feature termed “DeLIS,” which enables accurate prognostication of patient outcomes when integrated with clinical factors [46].

### 3.1.3 Computer-aided diagnosis systems and automated report generation.

Moreover, the integration of large models has propelled the radiology field forward by assisting radiologists in disease diagnosis and automating the generation of medical imaging reports. Jiang et al. utilized a transformer-based image classification model employing optical coherence tomography (OCT) images to discern between age-related macular degeneration (AMD) and diabetic macular edema (DME), contributing significantly to the diagnosis of retinal diseases. The trained Transformer model achieved an impressive recognition accuracy of 90.9% when classifying normal, AMD, and DME OCT images, underscoring the potential of Transformer models in computer-aided diagnosis [47].

Furthermore, Yang et al. introduced an Adaptive Multimodal Attention network (AMAnet) designed for generating high-quality medical imaging reports, as evidenced by experiments conducted on a dataset of breast ultrasound images. The outcomes revealed that the AMAnet model autonomously produces semantically coherent and high-quality medical image reports, accurately portraying essential local features [48].

## 3.2 Technical challenges and solutions

### 3.2.1 Data scarcity and data bias.

While large models have demonstrated substantial advantages in medical image analysis, several challenges persist. Primarily, concerns arise regarding the availability and quality of datasets, specifically related to issues of data scarcity

## Review

and data bias. Large models demand considerable volumes of data for effective training, yet numerous research studies currently rely on medical imaging datasets that are relatively small in scale, falling short of the requirements for large model training. Additionally, some diseases exhibit an imbalanced data distribution, potentially leading to biased model outcomes. Furthermore, medical imaging data stems from diverse sources, posing challenges in ensuring data consistency. However, techniques, such as data augmentation, are presently used to alleviate these challenges, at least in part [49].

**3.2.2 Model interpretability.** Another crucial consideration is the interpretability of the model. The primary objective of using large models in medical image analysis is to support clinical decision-making, necessitating a transparent rationale behind every clinical decision. However, elucidating the decision-making process in large models is often challenging, potentially resulting in an inability to rectify errors, posing challenges for healthcare professionals and patients.

To address this concern, various techniques exist for model interpretation, such as Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations SHAP). LIME operates by using a set of perturbed data around the predictions of the original model, subsequently training an interpretable model to elucidate the decision-making process of the original model. SHAP utilizes game theory concepts to consider diverse combinations of features, calculating the contribution of each feature to the final prediction. This approach aids in comprehending how the model arrives at decisions. While both methods offer a degree of interpretability, rigorous research is imperative to ensure the accuracy and reliability of the results [50].

**3.2.3 Computational resources and efficiency.** Mitigating the demand for computational resources and minimizing energy consumption in large models constitutes a significant technical challenge. The training of large models necessitates substantial computational resources, and the escalating demand for computational resources concurrently amplifies the energy consumption associated with training large models.

Current techniques for model compression and acceleration, such as model pruning and quantization, can partially alleviate the strain on computational resources and energy consumption in large models. However, addressing this challenge comprehensively requires sustained research efforts [10].

## 4. FUTURE DIRECTIONS

Given the substantial potential of large models in the field of medical image analysis, ongoing research on the application in this domain is continually advancing. In this section, I will delineate the future directions of

large models in medical image analysis, encompassing, but not limited to, the following aspects:

### 4.1 Model performance optimization

As the trend toward increasing model scale persists, the complexity of large models rises, necessitating greater computational resources and energy. The training and deployment of large models encounter challenges related to inadequate computational resources and heightened energy consumption. Identifying model optimization and acceleration techniques that diminish computational resource requirements and energy consumption is imperative to propel the development and application of large models.

### 4.2 Enhancing model interpretability

While the application of large models in medical image analysis brings convenience to physicians and patients, the rigorous and specific nature of medical treatment mandates a clear rationale for treatment decisions. Enhancing the interpretability of large models is essential, enabling physicians and clinicians to comprehend the decision-making process of the models. This improvement provides a reliable foundation for large model-assisted clinical decision-making. These future directions underscore the importance of addressing challenges related to model scale, resource utilization, and interpretability to unlock the full potential of large models in advancing medical image analysis.

### 4.3 Multimodal medical image analysis

Medical imaging data is diverse, presenting in various formats, and large models exhibit the capability to seamlessly integrate information from multiple types of medical imaging data. This integration fosters information fusion and complementarity between distinct imaging modalities, ultimately enhancing diagnostic accuracy.

### 4.4 Self-supervised and few-shot learning

Advancing the application of self-supervised and few-shot learning in medical image analysis is crucial for mitigating the challenges posed by limited annotated data.

### 4.5 Automated medical report generation

Automated medical report generation remains a paramount focus. The ongoing evolution of large models will continue to propel the automation of medical image report generation, thereby alleviating the workload of radiologists.

### 4.6 Real-time medical image analysis

Real-time analysis and monitoring of medical images constitute a pivotal frontier for the future development of large models. Exploring the application of large models in real-time medical image analysis and monitoring is anticipated to provide substantial support to healthcare professionals and patients alike.

#### 4.7 Privacy protection

The widespread integration of large models in medical image analysis brings forth ethical and regulatory considerations that demand attention. The future trajectory of macro-modeling necessitates stringent privacy protection measures in accordance with ethical guidelines and regulatory requirements.

### 5. CONCLUSION

In conclusion, large models have demonstrated substantial advantages in the analysis of medical images, offering the potential to enhance the precision of disease diagnosis and introduce innovative possibilities to the field of medical image analysis. However, the application of these sophisticated models encounters several challenges, including insufficient data, interpretability of models, and computational resource demands. Researchers have proposed addressing these challenges through techniques, such as LIME explanatory modeling and model compression and acceleration, which mitigate these issues, at least in part.

The evolving landscape of large models continues to witness advances, with ongoing efforts focused on optimization, acceleration, and the augmentation of interpretability. Additionally, addressing challenges related to the analysis of multimodal medical image data, refining diagnostic accuracy, automating the generation of medical reports, and other dimensions signify the principal developmental trajectories for large models in the foreseeable future.

In summary, large models possess the robust capacity to conduct accurate and in-depth analysis of medical images, introducing unprecedented possibilities for their application. The existing challenges encountered by large models are serving as catalysts for their further refinement. Looking ahead, these models are poised to exhibit heightened performance in the realm of medical image analysis, experiencing deeper integration and continually charting new developmental pathways for the field.

#### ACKNOWLEDGEMENTS

This study has received funding from the National Key Research and Development Program of China under Grant Nos. 2021YFC2500402, 2017YFA0700401, 2022YFC2503700, and 2022YFC2503705, the Ministry of Science and Technology of China under Grant No. 2017YFA0205200, the National Natural Science Foundation of China under Grant Nos. 82001917, 81930053, 82090052, 62027901, 81227901, 92159202, U22A2023, U22A20343, and 82172039, and the Project of High-Level Talents Team Introduction in Zhuhai City (Zhuhai HLHPTP201703).

#### CONFLICT OF INTEREST

There is no conflict to declare.

#### REFERENCES

- [1] Han K, Wang Y, Chen H, Chen X, Guo J, et al. A survey on vision transformer. *IEEE Trans Pattern Anal Mach Intell* 2023;45:87-110. [PMID: 35180075; DOI: 10.1109/TPAMI.2022.3152247]
- [2] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol* 1990, 52:99-115. [PMID: 2185863]
- [3] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90.[DOI: 10.1145/3065386]
- [4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [5] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016. Pp. 770-8.
- [6] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, et al. Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015. Pp. 1-9.
- [7] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al. Attention is all you need. *31st Annual Conference on Neural Information Processing Systems 30 (NIPS)*. 2017.
- [8] Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *Conference of the North-American-Chapter of the Association-for-Computational-Linguistics - Human Language Technologies (NAACL-HLT)*. 2019. Pp. 4171-86.
- [9] Brown TB, Mann B, Ryder N, Subbiah M, Kalpan J, et al. Language models are few-shot learners. *Adv Neural Inf Process Sys* 2020;33:1877-901.
- [10] Lohm AJ, Musser M. AI and compute. *Center for Security and Emerging Technology*. 2022.
- [11] Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. *57th Annual Meeting of the Association-for-Computational-Linguistics (ACL)*. 2019. Pp. 3645-50.
- [12] Chen X, Wang X, Zhang K, Fung KM, Thai TC, et al. Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal* 2022;79:102444. [PMID: 24006187; DOI: 10.1002/mrm.24937]
- [13] Shen D, Wu G, Suk HI. Deep Learning in medical image analysis. *Annu Rev Biomed Eng* 2017;19:221-48. [PMID: 28301734; DOI: 10.1146/annurev-bioeng-071516-044442]
- [14] Yousef R, Gupta G, Yousef N, Khari M. A holistic overview of deep learning approach in medical imaging. *Multimed Syst* 2022;28:881-914. [PMID: 35079207; DOI: 10.1007/s00530-021-00884-5]
- [15] Wang H, Minnema J, Batenburg KJ, Forouzanfar T, Hu FJ, et al. Multiclass CBCT image segmentation for orthodontics with deep learning. *J Dent Res* 2021;100:943-9. [PMID: 33783247; DOI: 10.1177/00220345211005338]
- [16] Zouch W, Sagga D, Echioui A, Khemakhem R, Ghorbel M, et al. Detection of COVID-19 from CT and chest X-ray images using deep learning models. *Ann Biomed Eng* 2022;50:825-35. [PMID: 35415768; DOI: 10.1007/s10439-022-02958-5]
- [17] Li F, Su Y, Lin F, Li Z, Song Y, et al. A deep-learning system predicts glaucoma incidence and progression using retinal photographs. *J Clin Invest* 2022;132:e157968. [PMID: 35642636; DOI: 10.1172/JCI157968]



## Review

- [18] Puyol-Anton E, Sidhu BS, Gould J, Porter B, Elliott MK, et al. A multimodal deep learning model for cardiac resynchronisation therapy response prediction. *Med Image Anal* 2022;79:102465. [PMID: 35487111; DOI: 10.1016/j.media.2022.102465]
- [19] Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;9:611-29. [PMID: 29934920; DOI: 10.1007/s13244-018-0639-9]
- [20] Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. *Comput Biol Med* 2021;128:104115. [PMID: 33227578; DOI: 10.1016/j.compbimed.2020.104115]
- [21] Chen J, Wan Z, Zhang J, Li W, Chen Y, et al. Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet. *Comput Methods Programs Biomed* 2021;200:105878. [PMID: 33308904; DOI: 10.1016/j.cmpb.2020.105878]
- [22] Hohn J, Krieghoff-Henning E, Jutzi TB, von Kalle C, Utikal JS, et al. Combining CNN-based histologic whole slide image analysis and patient data to improve skin cancer classification. *Eur J Cancer* 2021;149:94-101. [PMID: 33838393; DOI: 10.1016/j.ejca.2021.02.032]
- [23] Pei Y, Huang Y, Zou Q, Zhang X, Wang S. Effects of image degradation and degradation removal to CNN-based image classification. *IEEE Trans Pattern Anal Mach Intell* 2021;143:1239-53. [PMID: 31689183; DOI: 10.1109/TPAMI.2019.2950923]
- [24] Katte T. Recurrent neural network and its various. *Int. J. Res Sci Innov* 2018;5:124-9.
- [25] Hoogi A, Mishra A, Gimenez F, Dong J, Rubin D. Natural language generation model for mammography reports simulation. *IEEE J Biomed Health Inform* 2020;24:2711-7. [PMID: 32324577; DOI: 10.1109/JBHI.2020.2980118]
- [26] Yan W, Calhoun V, Song M, Cui Y, Yan H, et al. Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site fMRI data. *EBioMedicine* 2019;47:543-52. [PMID: 31420302; DOI: 10.1016/j.ebiom.2019.08.023]
- [27] Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, et al. Transformers in medical imaging: a survey. *Med Image Anal* 2023;88:102802. [PMID: 37315483; DOI: 10.1016/j.media.2023.102802]
- [28] Nakamura Y, Hanaoka S, Nomura Y, Nakao T, Miki S, et al. Automatic detection of actionable radiology reports using bidirectional encoder representations from transformers. *BMC Med Inform Decis Mak* 2021;21:262. [PMID: 34511100; DOI: 10.1186/s12911-021-01623-6]
- [29] Rao A, Kim J, Kamineni M, Pang M, Lie W, et al. Evaluating GPT as an adjunct for radiologic decision making: GPT-4 Versus GPT-3.5 in a breast imaging pilot. *J Am Coll Radiol* 2023;20:990-997. [PMID: 37356806; DOI: 10.1016/j.jacr.2023.05.003]
- [30] Sharma U, Kaplan J. Scaling laws from the data manifold dimension. *J Mac Learn Res* 2022;23:1-34.
- [31] Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: overview, challenges and the future. *Classification in BioApps*. 2018. Pp. 323-50.
- [32] Rogers A, Kovaleva O, Rumshisky A. A primer in BERTology: what we know about how BERT works. *Trans Assoc Comput Linguist* 2020;8:842-66. [DOI: 10.1162/tacl\_a\_00349]
- [33] Hopson JB, Neji R, Dunn JT, McGinnity CJ, Flaus A, et al. Pre-training via transfer learning and pretext learning a convolutional neural network for automated assessments of clinical PET image quality. *IEEE Trans Radiat Plasma Med Sci* 2023;7:372-81. [PMID: 37051163; DOI: 10.1109/TRPMS.2022.3231702]
- [34] Jeon SK, Lee JM, Joo I, Yoon JH, Lee G. Two-dimensional convolutional neural network using quantitative US for noninvasive assessment of hepatic steatosis in NAFLD. *Radiology* 2023;307:e221510. [PMID: 36594835; DOI: 10.1148/radiol.221510]
- [35] Kumar T, Mileo A, Brennan R, Bendechache M. Image data augmentation approaches: a comprehensive survey and future directions. *arXiv preprint arXiv:2301.02830*, 2023. [DOI: 10.48550/arXiv.2301.02830]
- [36] Nie FP, Hu ZX, Li XL. An investigation for loss functions widely used in machine learning. *Commun Inform Sys* 2018;18:37-52. [DOI: 10.4310/CIS.2018.v18.n1.a2]
- [37] El Jurdi R, Petitjean C, Honeine P, Cheplygina V, Abdallah F. High-level prior-based loss functions for medical image segmentation: a survey. *Comput Vis Image Underst* 2021;210: 103248. [DOI: 10.1016/j.cviu.2021.103248]
- [38] Sun RY. Optimization for deep learning: an overview. *J Oper Res Soc China* 2020;8:249-94. [DOI: 10.1007/s40305-020-00309-6]
- [39] Soydaner D. A comparison of optimization algorithms for deep learning. *Int J Pattern Recognit Artif Intell* 2020;34: 2052013. [DOI: 10.1142/S0218001420520138]
- [40] Phaisangittisagul E. Paper presented at the 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS). 2016.
- [41] Choudhary T, Mishra V, Goswami A, Sarangapani J. A comprehensive survey on model compression and acceleration. *Artif Intell Rev* 2020;53:5113-55. [DOI: 10.1007/s10462-020-09816-7]
- [42] Yan F, Ruwase O, He Y, Chilimbi T. Paper Presented at the Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015.
- [43] Farkas A, Kertesz G, Lovas R, Ieee. Parallel and distributed training of deep neural networks: a brief overview. 24th IEEE International Conference on Intelligent Engineering Systems (INES). 2020:165-70.
- [44] Jin Q, Meng Z, Sun C, Cui H, Su R. RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans. *Front Bioeng Biotechnol* 2020;8:605132. [PMID: 33425871; DOI: 10.3389/fbioe.2020.605132]
- [45] Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol* 2018;136:803-10. [PMID: 29801159; DOI: 10.1001/jamaophthalmol.2018.1934]
- [46] Jiang Y, Jin C, Yu H, Wu J, Chen C, et al. Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: a multicenter, retrospective study. *Ann Surg* 2021;274:e1153-e1161. [PMID: 31913871; DOI: 10.1097/SLA.0000000000003778]
- [47] Jiang Z, Niu J, Wang Y, Li Q, Wu J, et al. Computer-aided diagnosis of retinopathy based on vision transformer. *J Innov Opt Health Sci* 2022;15. [DOI: 10.1142/s1793545822500092]
- [48] Yang S, Niu J, Wu J, Wang Y, Liu X, et al. Automatic ultrasound image report generation with adaptive

- multimodal attention mechanism. *Neurocomputing* 2021;427:40-9. [DOI: 10.1016/j.neucom.2020.09.084]
- [49] Altaf F, Islam SMS, Akhtar N, Janjua NK. Going deep in medical image analysis: concepts, methods, challenges, and future directions. *IEEE Access* 2019;7:99540-72.
- [50] Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M. Explainable AI for retinoblastoma diagnosis: interpreting deep learning models with LIME and SHAP. *Diagnostics (Basel)* 2023;13:1932. [PMID: 37296784; DOI: 10.3390/diagnostics13111932]