Proceedings of CONF-MLA 2026 Symposium: Learning and Decision Making in Multi Agent Software Systems
DOI: 10.54254/2755-2721/2026.BA38765

Progress and Trends of Deep Learning in CT Image Lesion Area Detection

Jiangxu Gao, Chun Ho Wong

Utahloy International School Guangzhou, Dongguan, China
13415979161@163.com

Abstract. Deep learning has achieved some good results in the field of medical image analysis to date, but most existing models for CT lesion detection have been developed for a single organ and lack a cross-site universal framework; in addition, their detection performance is highly dependent on specific clinical conditions. The following is an overview of the current stage, development and application cases of deep learning detection methods for four types of CT images: pulmonary nodules, brain lesions, liver lesions and bone lesions. The problems in Model Design and how to solve them are presented in this paper. Based on the comparison of the quantitative results and technical attributes of the main models, it can be observed that the general direction of development is moving towards three-dimensional representations, multiple tasks, integration of clinical knowledge, and so on. At the same time, the main bottlenecks in current research have also been identified in the three areas of data, clinical diagnosis and algorithms; high annotation costs, poor interpretability, insufficient use of three-dimensional information and weak generalisation ability are among them. This paper points out that there are deficiencies in assessing the strengths and weaknesses of methods that are not tied to particular situations. According to the kind of anatomical place and clinical conditions, a specific detection plan will be chosen. In the future, semi-supervised learning, multi-task integrated models and interpretable AI will be employed to increase the accuracy and clinical applicability of this field.

Keywords: Deep Learning, CT Images, Lesion Detection, Computer-Aided Diagnosis, Multi-Site Detection.

1. Introduction

The first problem to be addressed by intelligent analysis of medical images is the detection of lesions. There are many commonalities in the methodology of feature extraction and model design in the existing literature, but the detection performance is generally limited by the clinical application scenario. The shape, sharpness of the boundary and colour of tissues are all different in the lesions of various organs such as the lungs, brain, liver and bones. For instance, the isolated circular dense shadows in lung nodules and the diffuse low-density areas in liver lesions directly reduce the cross-site generalisation ability of the algorithm. However, most of the current research is still concentrated on the development of single-organ-specific models, and systematic exploration 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: 10.54254/2755-2721/2026.BA38765

multi-site universal detection frameworks and technology adaptation rules in various scenarios has not been achieved. Therefore, the goal of this paper is to build an adaptive detection strategy system that can meet all the different requirements for feature representation and detection logic in various parts of the anatomy, thereby improving the clinical application and robustness of Computer-Aided Diagnosis (CAD) systems. Therefore, the primary index for the classification logic in this review is application scenarios, and representative detection methods are organised according to the four main anatomical regions of the lungs, brain, liver and bones to examine the rationality and limitations of technical solutions under the specific conditions of each type of scenario.

In terms of the research value of this review, two levels of references will be provided for researchers in this field. First, by comparing the main technology routes in different scenarios horizontally, the paper can identify the achievements and common problems that current deep learning models have faced in CT image lesion detection; second, after systematically listing the shortcomings and deficiencies in the dimensions of data, clinical practice and algorithms, some directions for future research can be put forward. The current performance of deep learning models is highly dependent on large amounts of high-quality annotated data, but obtaining and annotating medical imaging data is expensive and subject to ethical restrictions. At the same time, the generalisation ability of models in cross-device and cross-population scenarios is not strong enough, and there is still a gap between this and the demands of clinical practice.

The following is the outline of the current article. In the second part, representative technical routes are introduced based on the four application scenarios of lung nodule detection, brain lesion detection, liver lesion detection and bone lesion detection. The third part discusses the shortcomings and limitations of the current research from the three levels of data, clinical diagnosis and algorithm. Finally, Part 4 summarizes the contents of this paper and indicates the future direction of research.

2. Technical path and methodological analysis of multi-site CT spot detection

2.1. Pulmonary nodule detection

To solve the problem of a high false-positive rate caused by the difficulty in distinguishing lung cancer from benign nodules in low-dose CT screening, deep learning has been applied to the area of lung nodule detection because it has good feature discrimination ability. The first CAD system required people to manually draw all sorts of shapes, such as roundness, lobes, and spicules, and it was not very good at detection or generalisation [1, 2]. With the development of Convolutional Neural Networks (CNNs), an end-to-end automatic detection framework based on CNNs has gradually appeared. A good trade-off between sensitivity and specificity has been achieved in nodule detection using a two-stage framework; to improve the accuracy of benign and malignant classification, multi-scale feature fusion networks can be employed to obtain local details and global context, as shown in [3], and an attention mechanism is added to increase the response to boundary features that have the characteristics of blurred edges and low contrast in ground-glass nodules. Recently, 3D CNNs have been used to make full use of all the spatial information in the inter-layers of CT sequences to improve detection continuity [4], and the Transformer architecture has also introduced global modeling to address the short-range dependency problem of traditional CNNs [5]. In short, generally speaking, the trend in this field has been to develop it in three dimensions, shift from single-task operation to multi-task operation, and start to incorporate existing clinical knowledge with data analysis.

DOI: 10.54254/2755-2721/2026.BA38765

2.2. Brain lesion detection

The problem of detecting brain CT lesions is difficult due to the many kinds of lesions, complex anatomy and the urgent need for rapid diagnosis [6]. The old way is neither quick nor accurate enough. At present, the development direction of deep learning methods in this area is to achieve all the above goals simultaneously. To solve the problem of detecting intracranial haemorrhage, a CNN-based segmentation model can accurately determine the range and size of the hematoma, and the problem of irregular hematomas in previous studies has also been resolved [7]. The joint framework of the object detection and segmentation network can be used to find fracture lines in skull fracture detection and map them to the standard divisions of the skull for a positional reference in surgery. Diffuse axonal injury is characterised by a small number of widely distributed haemorrhagic foci, so it is difficult to detect. To keep the detailed information in the network, U-Net models generally have skip connections [8]. It should be pointed out that the new comprehensive diagnostic model has been added to cover all the previously mentioned serious injuries, such as intracranial hemorrhage, skull fractures, midline shift and space-occupying lesions, etc. Its diagnostic sensitivity is that of senior doctors, and it has been used in the emergency triage of hospitals and other resource-limited places [9].

2.3. Liver lesion detection

Given that the density difference between the tumour and the liver tissue is small, there are fuzzy boundaries, and the infiltration and growth are diffuse, traditional single-mode models may fail to detect the tumour properly and thus have a reduced contrast sensitivity and spatial perception capability. Microvascular infiltration is a factor that affects the recurrence and prognosis after surgery, but it is not easily observed in traditional examinations and is therefore difficult to detect. One research direction in deep learning at present is multimodal image fusion [10]. A differential diagnosis system based on multi-phase enhanced CT (arterial phase, portal phase, delayed phase) and multiple-sequence MRI can effectively integrate dynamic enhancement information in the time domain, and it has been shown to be in good agreement with pathological results in the diagnosis of liver cancer. For the "invisible" problem of microvascular invasion, the paper can indirectly infer heterogeneity in the texture of the tumour and changes in the surrounding tumour environment, and models that integrate clinical data and radiomic features have also shown good results in predicting postoperative recurrence. Generally speaking, the technical characteristic of liver lesion detection is deep fusion of deep learning with multi-phase and multi-modal images, and the main problem has changed from "determining whether there is a lesion" to "forecasting the biological behaviour of a lesion".

2.4. Bone disease detection

Skeletal system CT lesions can be caused by various diseases, including fractures, bone tumours and osteoporosis. The three problems of the traditional empirical interpretation are strong subjectivity, late discovery and lack of quantification [11]. The shape of the bones is fixed in anatomy, so they can be automatically recognised and segmented. The density of the tumour is not uniform, and there are other problems, such as bone destruction; thus, feature learning is difficult. At present, the technology roadmap for "classification segmentation detection" is divided into three stages, and development has been carried out in phases. First, Convolutional Neural Networks (CNNs) are employed to classify images as either normal or abnormal. Second, a U-Net variant is used to

DOI: 10.54254/2755-2721/2026.BA38765

perform fine-grained segmentation of abnormalities, such as vertebral bodies, and at this time, quantitative analysis is also carried out. Finally, YOLO single-stage detection algorithms are employed to locate fracture lines and lesions in real time for emergency response [12]. It should be pointed out that deep learning has advanced the direction of diagnosis from qualitative description to quantitative analysis, and an objective basis for evaluating the disease has been established by measuring parameters of segmentation structures. Interpretability research can tell us why a model made a certain decision in a comprehensible manner, for example, Class Activation Maps, to address the problem of "black box" models for clinical physicians and promote multidisciplinary trust and clinical application.

3. Comparison of experimental results

Table 1. Comparison of technical routes and quantitative results of deep learning models for pulmonary nodule detection

Deep learning models	test object	Technical Features	quantitative results	Main Results
Faster R-CNN	Pulmonary nodule	Two stage object detection network with high target localization capability	Average FROC (CPM) = 0.891	Improve the accuracy of pulmonary nodule localization and reduce the rate of missed detections
LUNA16 Benchmark	Pulmonary nodule	International standard evaluation system, using FROC and CPM to evaluate detection performance	Benchmark: Sensitivity>95% @<1 FP/Scan	Has good detection effect on small-sized pulmonary nodules
DeepSEED	Pulmonary nodule	Multi scale feature fusion enhances feature expression capability	Average FROC=86.2%	Improve the sensitivity of small nodule detection
3D CNN	Pulmonary nodule	Utilizing CT three-dimensional spatial information for feature learning	Sensitivity>95% @<1 FP/Scan	Improve spatial feature expression ability
Transformer-CNN	Pulmonary nodule	CNN and Transformer Fusion	Sensitivity=97.05% , FPs/Scan=4.09	Improve detection performance in complex backgrounds

As shown in Table 1, pulmonary nodules are one of the first signs of lung cancer. Due to their small size, many shapes and the influence of normal tissues such as blood vessels and bronchi, traditional image processing methods are likely to fail in obtaining some of the images and be mistaken for others. With the development of deep learning, lung nodule detection has gradually moved from traditional Convolutional Neural Networks (CNNs) to 3D CNNs and Transformer

DOI: 10.54254/2755-2721/2026.BA38765

hybrid networks, and now these models can learn the spatial structural information in CT images better.

With the progress of technology, two-stage object detection models such as Faster R-CNN have good localization accuracy [13] and are suitable for the first stage of lung nodule detection; later, DeepSEED and other models have been developed to add segmentation networks, multi-scale feature fusion and attention mechanisms to improve the detection ability of small nodules. In recent years, the fusion model of Transformer and CNN has further enhanced the model's ability to learn global information and is now one of the main directions for development in lung CT lesion detection.

Table 2. Comparison of technical routes and quantitative results of deep learning models for brain disease detection

Deep learning models	test object	Technical Features	quantitative results	Main Results
U-Net	Brian lesions	Encoder–Decoder architecture.	Detection Accuracy = 98.59%;Dice = 0.8033;IoU = 0.6919	It Can accurately segment brain lesion areas
Improved U-net	Cerebal hemorrhage	Uses DenseNet as the encoder to improve feature extraction capability	Segmentation Accuracy = 99.0%	Accurately segments brain lesions
transformer	Brian tumors	Global attention mechanism, improving the ability to identify complex lesions	No information found	Improves the ability to identify lesions in brain CT images
CNN	Brian lesions	Automatically learns CT image features and performs lesion classification	AUC = 0.974	Provides stable performance across multiple datasets
TransHarDNet	Brian lesions	Combines HarDNet with Transformer to improve overall feature extraction capability	Dice = 0.712;IoU = 0.597;30.78 FPS	Improves the accuracy of boundary segmentation for complex lesions

As shown in Table 2, most of the brain CT lesions have irregular borders and small grayscale differences; thus, the demand for image segmentation algorithms is high. U-Net is a typical segmentation network for medical images that can accurately locate areas of lesions in the decoder of its encoder, and it has been applied to the automatic detection of brain lesions.

With the progress of deep learning technology, 3D convolutional networks can fully utilize the spatial information in CT volume data to improve the accuracy of brain lesion detection; in recent years, Transformers and attention mechanisms have also been gradually introduced in brain lesion

DOI: 10.54254/2755-2721/2026.BA38765

detection to enhance the recognition ability of complex lesion boundaries and have shown promising results in the development of intelligent brain CT detection. Table 3 shows the technical methods and quantitative results comparison of deep learning models for liver injury detection.

Table 3. Comparison of technical approaches and quantitative results of deep learning models for liver lesion detection

Deep Learning Models	Test Object	Technical Features	Quantitative Results	Main Results
FCN	Liver lesions	Fully Convolutional Network	Dice ≈ 0.94 (LiTS baseline)	Automatically segments liver lesion regions
U-Net	Liver lesions	Encoder–Decoder architecture	Dice ≈ 0.96 (LiTS)	Provides good lesion boundary recognition
H-DenseUNet	Liver tumors	2D and 3D fusion	Liver Dice = 0.963; Tumor Dice = 0.722	Improves liver tumor detection performance
ResUNet	Liver lesions	Incorporates residual learning	Dice > 0.97	Enhances segmentation of complex lesions
TransUNet	Liver lesions	Combines CNN with Transformer	Dice = 0.971	Improves segmentation accuracy of complex boundaries

Table 4. Comparison of technical approaches and quantitative results of deep learning models for bone lesion detection

Deep Learning Models	Test Object	Technical Features	Quantitative Results	Main Results
FracNet (3D U-Net)	Fracture detection	Uses 3D U-Net for fracture detection and segmentation, combining fracture detection with voxel-level segmentation	Detection Sensitivity = 92.9%; Dice = 71.5%; FPs/Scan = 5.27	Improves fracture localization accuracy
Deep CNN	Fracture detection	Deep convolutional neural network automatically extracts fracture texture features and improves detection efficiency	Sensitivity = 64.5%	Meets the requirements of real-time clinical detection
3D CNN	Vertebral disease	Uses 3D convolution to learn spatial features and automatically detects vertebral fractures	Patient-level AUC = 0.95; Vertebra-level AUC = 0.93	Effectively identifies abnormal vertebral regions

DOI: 10.54254/2755-2721/2026.BA38765

Table 4. (continued)

Automated CT Fracture Detection System	Bone CT	Automatically performs vertebral localization, fracture detection, and classification	Sensitivity = 95.7%; Specificity = 77.3%; Classification Accuracy = 95%	Improves the ability to identify complex bone lesions
nnDetection	Bone lesions	Automatic object detection framework suitable for lesion detection in 3D CT images	Average Sensitivity = 77.0%	Improves the detection performance of complex fractures

As shown in Table 4, the main applications of bone lesion detection are CT-assisted diagnosis of diseases such as rib fractures and vertebral fractures. With the development of deep learning, many studies have moved away from traditional CNNs and started exploring 3D convolutional networks and automated detection systems. FracNet is a 3D U-Net that has achieved a detection sensitivity of 92.9% in the RibFrac dataset, and it can use the three-dimensional spatial information of CT to improve the detection ability of fracture lesions; a 3D CNN has also attained an AUC of 0.95 for vertebral fracture detection and shown good classification results.

In short, all of them have advantages. Deep CNNs can improve the detection accuracy of fractures. The Sensitivity of the automated detection system for vertebral fractures is 95.7%, the classification accuracy is 95%, and nnDetection has advanced the automation of models. With the increase in accuracy and the development of automation and intelligence, deep learning technology is now a good way to find bone CT lesions.

4. Shortcomings and limitations

4.1. Deficiencies and limitations of the data level

High-quality annotations are relatively expensive, and the size of public datasets is generally small [14]. There are problems such as poor consistency in labelling, lack of uniform standards, the presence of noise, and difficulty in horizontal comparison; positive samples are scarce, especially for small and early-stage lesions, and the model has low sensitivity to minority classes. There are also considerable differences in the distribution of data among centers and devices, and the generalisation ability of the model is weak.

Research on semi-supervised learning, weak supervision and active learning is carried out to reduce the need for annotation; a common labelling standard is promoted; reweighting and synthetic minority oversampling techniques are used to tackle class imbalance; research on multi-center data joint modelling and domain adaptation alignment technology is strengthened [15].

4.2. Shortcomings and limitations of clinical diagnosis

The model is not very interpretable, so doctors are not convinced that it is useful. Variations in image quality and differences among various groups of people in real clinical practice will affect the performance of the model, and its validation cases do not match actual clinical situations. There are

DOI: 10.54254/2755-2721/2026.BA38765

also technical and procedural barriers to integrating the model with the current clinical information system and workflow. Finally, in the event of a misdiagnosis, the legal and ethical responsibilities are unclear, and people are reluctant to use it [16].

Introduce an interpretability mechanism so that doctors can review the results of the model, carry out multicenter prospective validation in near-real clinical situations, promote the standardization of DICOM/PACS interfaces and integration into the Design of clinical pathways, clarify the division of responsibilities among all parties, and establish an ethical and legal agreement for AI-assisted diagnosis.

4.3. Shortcomings and limitations of algorithms

The sensitivity for finding small lesions and low-contrast targets is relatively low, the false-positive rate is high, there is no use of CT three-dimensional spatial information, there is a loss of inter-layer continuity in two-dimensional slice processing, conventional models are prone to overfitting due to limited availability of annotated data, and small-sample learning strategies have poor generalisation ability; there are considerable morphological differences between lesions and other parts of the anatomy, and the algorithm is not robust; finally, a single-task model cannot meet all the diagnostic and treatment requirements in clinical practice.

Build an effective 3D network structure, promote the application of pretrained models and prompt-learning paradigms in the field of medical imaging, focus the model on the corresponding area based on prior knowledge of lesions, and construct a multi-task learning framework that combines detection, segmentation, quantification and prognosis prediction to cover all links in the clinical decision pathway.

5. Conclusion

Some research has been carried out on the application of deep learning for lesion detection in CT images. Based on the four main application scenarios of lung nodules, brain lesions, liver lesions and bone lesions, the technical features and development of representative methods in each scenario are introduced here. In terms of the structure of the Model, it has moved from traditional CNNs to 3D convolutional networks and Transformer fusion architectures; in terms of the type of problem, it is now moving away from a single detection or segmentation task towards all-encompassing multi-task diagnosis; and regarding the use of data, multimodal image fusion and the embedding of clinical prior knowledge have gained much attention. Different situations have various problems in detection; therefore, many efforts have been made to improve the accuracy of detection, reduce false alarm rates and enhance the generalisation ability of models. The main idea of this paper is that the choice of deep learning CT imaging lesion detection technology depends on various factors in actual clinical situations. The anatomy and lesion characteristics of the lungs, brain, liver and bones are all different, so the main problems in designing detection algorithms are also different. The lungs need to distinguish small nodules from the background; the brain must diagnose many diseases simultaneously; the liver is based on continuous integration of dynamic information; and the bones are supported by the prior registration of anatomical structures. Therefore, discussing the merits and demerits of the methods without context is one-sided. Only by considering the rationality and deficiencies of technical solutions in light of particular clinical conditions can people truly advance research in this area. At the same time, there are general problems such as lack of data, poor interpretability and difficulty in clinical application in all areas, and these are the basic bottlenecks preventing deep learning from the laboratory to clinical practice.

DOI: 10.54254/2755-2721/2026.BA38765

Looking ahead, there will be several developments in this field in the future. First of all, at the level of data, semi-supervised learning, self-supervised pre-training and federated learning can be used to reduce the problem of a lack of annotated data. The construction of multi-centre large-scale datasets will also help to improve the generalisation ability of models; Secondly, at the level of algorithms, full use of three-dimensional space, deep fusion of Transformer and CNN, and robust learning for small samples are still research hot spots; Thirdly, in the area of clinical application, the progress of artificial intelligence and the construction of multi-task integrated diagnostic models will gradually meet the needs of clinical practice. With the continuous development of algorithms, data and computing power, it can be expected that the clinical application of deep learning in CT image lesion detection will continue to advance, and ultimately help to improve diagnostic efficiency, reduce the rate of misdiagnosis, optimize the allocation of medical resources, and so on.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Yang, H. W. (2010). *Research and system implementation on feature extraction and feature selection of pulmonary nodules*. Chongqing University, Master's Thesis.
[2] Zhou, X. T. (2022). *Application of deep learning based side scan sonar image matching and UUV assisted navigation*. Harbin Institute of Technology, Master's Thesis.
[3] Zhang, S. R. (2020). *Research on deep learning methods for lung nodule detection and classification*. Tianjin University, Master's Thesis.
[4] Chi, Z. W. (2023). *Research on CT image lung nodule detection method based on 3D convolutional network*. Hunan University of Science and Technology, Master's Thesis.
[5] Zhao, X. F. (2024). *Research on recognition method of diabetes retinopathy based on deep learning*. Shenyang University of Technology, Master's Thesis.
[6] Kang, D. W., Kim, J. H., Lee, S. H., Park, J. Y., & Choi, H. S. (2025). Deep learning-assisted detection of intracranial hemorrhage: Validation and impact on reader performance. *Neuroradiology, 67*, 1511–1519.
[7] Qin, H. J., Liu, Y. Y., Fu, E. H., Liu, Y. W., Tian, Z. L., Dong, H. W., Liu, T. A., Zou, D. H., Cheng, Y. B., & Liu, N. G. (2024). Intelligent recognition and segmentation of blunt traumatic brain injury CT images based on DeepLabV3+ model. *Journal of Forensic Medicine* [in Chinese].
[8] Zhang, Y. (2020). *In depth study on segmentation of small volume diffuse lesions in medical brain images*. University of Chinese Academy of Sciences (Shanghai Institute of Technical Physics, Chinese Academy of Sciences), Doctoral dissertation.
[9] Liang, H. L. (2025). *Research on retinal image segmentation algorithm based on deformable convolution and contrastive learning*. Jilin University, Master's thesis.
[10] Wu, Y. F., Zhang, L., Wang, H., Chen, X., & Li, Q. (2023). Phase difference network for efficient differentiation of hepatic tumors with multi-phase CT. *2023 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1-5.
[11] Elkohail, A., Alshammari, F., Alqahtani, S., & Alshammari, M. (2025). Artificial intelligence in bone fracture detection: A review of evidence, limitations, and clinical integration. *Cureus, 17*, e97674.
[12] Zhang, W. C. (2023). *Research on object detection and instance segmentation techniques based on deep learning*. Northeastern University, Doctoral dissertation.
[13] Su, Y., Li, D., & Chen, X. D. (2021). Lung nodule detection based on Faster R-CNN framework. *Computer Methods and Programs in Biomedicine, 200*, 105866.
[14] Felfeliyan, B., Jannati, A., Ghasemi, M., & Mohammadi, H. (2023). Self-supervised-RCNN for medical image segmentation with limited data annotation. *Computerized Medical Imaging and Graphics, 109*, 102297.
[15] Yuan, R. M. (2025). Research progress on non-contact physiological information perception technology of millimeter wave radar. *Computer Age* [in Chinese].
[16] Feng, J. F. (2022). *Legal regulation of algorithms in the digital age*. Henan University, Master's Thesis.