International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 7, No. 3, 2024

Research on the Assistive Strategy of Medical Image Segmentation Technology based on Medical Transformer Model in Intelligent Diagnosis of Tuberculosis in Xinjiang Region

Hui Sun, Yuan Fang

Xinjiang College Of Science & Technology, Korla Xinjiang, 841000, China

Abstract: The aim of this study is to explore the application of Medical Transformer model-based medical image segmentation technology in the intelligent diagnosis of tuberculosis in Xinjiang. Tuberculosis (TB), as a serious infectious disease, is still a major public health problem in Xinjiang and other regions where medical resources are relatively scarce. Traditional tuberculosis diagnosis methods suffer from low efficiency and insufficient accuracy, while deep learning technology provides a new solution for automatic medical image segmentation of tuberculosis. In this paper, a Medical Transformer model based on Gated Axial-Attention is proposed and combined with a local-global training strategy (LoGo) to effectively improve the segmentation accuracy of CT images of pulmonary tuberculosis. The problem of data scarcity in Xinjiang region was alleviated by strategies such as data enhancement, migration learning and cross-region data sharing. The results show that the model not only reduces the computational complexity but also performs well with small datasets, which is suitable for the environment with limited medical conditions in Xinjiang and is expected to improve the efficiency of early screening and diagnosis of TB. In addition, future research can further optimize the performance of the model through model lightweighting and self-supervised learning to enhance its potential application in practical diagnosis

Keywords: Medical Transformer; Medical Image Segmentation; Tuberculosis; Deep Learning; Computer Aided Diagnosis CAD; Gated Axial Attention; Local-Global Training Strategy; Data Augmentation; Migration Learning.

1. Introduction

Tuberculosis (TB), as a highly contagious disease, remains an important threat to public health, especially in areas with relatively scarce medical resources such as Xinjiang. As an important means of diagnosing TB, CT imaging of the lungs can provide precise lesion localization and lesion analysis. However, due to the limited professional ability of doctors, the traditional manual diagnosis method suffers from low efficiency and lack of accuracy. Therefore, automatic segmentation and analysis of medical images using deep learning technology to assist doctors in diagnosis and treatment has become a current research hotspot in the field of medical image processing.

In recent years, Transformer models have demonstrated powerful feature extraction and modeling capabilities in the fields of natural language processing and computer vision, and have been widely used in the processing of various types of complex data. However, for the medical image segmentation task, the existing Convolutional Neural Network (CNN) is difficult to capture global information due to the limitation of its fixed receptive field, and the Transformer model has bottlenecks in terms of computational complexity and training data volume. Therefore, this paper proposes an innovative segmentation model based on Medical Transformer, the Gated Axial-Attention model, which is able to achieve high-precision medical image segmentation with less training data by introducing a gating mechanism into the self-attention mechanism. This model can realize high accuracy medical image segmentation with less training data. Meanwhile, this paper also adopts the Local-Global Training (LoGo) strategy to further improve the performance of the model in TB image segmentation.

The purpose of this study is to improve the efficiency and accuracy of tuberculosis diagnosis by constructing an accurate and efficient medical image segmentation model to assist the relatively backward medical care in Xinjiang, thus providing effective support for the early screening and treatment of tuberculosis in the region. By combining deep learning with medical image segmentation, we provide new solutions for the development and application of Computer-Aided Diagnosis (CAD) systems.

2. National Distribution of Conjunctivitis and Trends in Incidence

Data on the distribution of TB in Xinjiang were obtained from several authoritative statistical organizations and public health reports. The study was mainly based on TB incidence and demographic data from the China Statistical Yearbook and Xinjiang Statistical Yearbook between 2005 and 2020.Monthly incidence report data from January 2005 to 2017 were obtained from the Public Health Science Data Center, while data from January 2018 to 2020 August 2018 were taken from the official releases of the Xinjiang Uygur Autonomous Region Health and Health Commission. The seasonal characteristics and epidemiological trends of TB incidence in Xinjiang were systematically analyzed by constructing a connecting-point regression model (JPR) to compare and analyze the incidence trends of TB between the whole country and Xinjiang, and by combining the Prophet model with the seasonal autoregressive moving average model (ARIMA). The results showed that the incidence rate of TB in Xinjiang was significantly higher than the national

average, and exhibited significant seasonal fluctuations, especially during the winter-spring period when the incidence rate increased sharply. These data provide an important scientific basis and reference support for the development of accurate TB prevention and control strategies [1].

Table 1. Incidence rates of tuberculosis in the country and Xinjiang, 2005- 2020

particular year	National incidence (1/100,000)	Incidence in Xinjiang (1/100,000)
2005	96.31	205.63
2006	86.23	192.60
2007	88.55	199.99
2008	88.52	202.93
2009	81.09	186.35
2010	74.27	164.46
2011	71.09	157.83
2012	70.62	181.17
2013	66.80	172.73
2014	65.63	175.99
2015	63.42	184.53
2016	61.00	185.66
2017	60.53	202.59
2018	59.27	304.94
2019	55.55	169.05
2020	47.76	118.35

3. Current Technological Shortcomings in the Adjunctive Diagnosis of Tuberculosis

CT image segmentation techniques for tuberculosis have made significant progress in improving diagnostic accuracy, but existing methods still face many challenges and limitations. Traditional image segmentation methods (e.g., threshold segmentation, region growing, and edge detection) perform poorly when dealing with complex lesions. Threshold segmentation relies on image grayscale contrast and is difficult to deal with lesions with fuzzy boundaries or similar densities; the region growing method is sensitive to the initial seed points, which can easily lead to over- or under-segmentation; and the edge detection is sensitive to noise and is unable to accurately identify lesions with complex structures or fuzzy boundaries.

Deep learning models such as U-Net, FCN, and DeepLab series, while significantly improving segmentation accuracy, also have shortcomings. U-Net model requires a large amount of labeled data and is prone to overfitting for small sample datasets, while its high computational complexity and long training time make it difficult to adapt to resource-limited clinical environments. FCN performs poorly when dealing with small and complex lesions, and lacks the capability of processing multi scale features, resulting in insufficient segmentation accuracy for complex backgrounds and small foci. DeepLab, although utilizing null convolution to expand the sensory field, is weak in handling small foci and fuzzy boundaries, and the complex model increases computational cost.

Attention U-Net introduces the attention mechanism, which improves the ability to focus on key foci, but at the same time increases the amount of computation and the demand for hardware resources, and the segmentation effect may be unstable in low-quality data or abnormal situations. In addition, all deep learning models rely on large-scale, high-quality labeled data, and the model performance decreases significantly when the labeling is insufficient or the data quality is low. [2]

In lung CT diagnosis in Xinjiang's tertiary hospitals, CAD (computer-aided diagnosis) systems have improved the overall diagnostic accuracy, but there are still some significant challenges and drawbacks. First, deep learning models (e.g., CNN, U-Net, ResNet, etc.) are widely used in CAD systems, but they are highly dependent on a large amount of high-quality labeled data. In Xinjiang, especially in remote areas, medical data resources are relatively scarce, leading to unstable performance and insufficient generalization ability of these models in the absence of sufficient data. [3]

Second, these models have high hardware requirements, especially the 3D CNN and Transformer models for processing 3D CT images, which require high-performance computing devices. However, the limited hardware resources in some tertiary hospitals in Xinjiang limit the efficient operation of these complex models, which affects the ability of real-time diagnosis in the clinic.

In addition, CAD systems do not perform well enough in recognizing small lesions (e.g., lung nodules) at early stages. Common models such as U-Net and FCN are prone to miss or misclassify lesions when confronted with fuzzy boundaries or small-sized lesions, especially in complex lung backgrounds, which makes early detection of the disease challenging.

Finally, due to limited data, the model may suffer from overfitting phenomenon, which makes the model perform poorly when facing different CT data in practical applications. Therefore, in order to further enhance the effectiveness of the CAD system, Xinjiang Hospital needs to strengthen data collection, optimize hardware resources, and improve the existing model. [4]

4. Medical Transformer's Proposal and its Advantages

4.1. Introduction of the Transformer Model

The Transformer model was initially used in Natural Language Processing (NLP) tasks, where it was able to capture the global relationships between positions in a sequence through the Self-Attention Mechanism, and excelled in understanding long-range dependencies. Later, Transformer models were gradually introduced into computer vision tasks, such as image classification and target detection. However, directly applying Transformer to medical image segmentation tasks faces the following challenges:

High computational complexity: Medical images are usually of high resolution, and applying Transformer directly will bring huge computational overhead.

Data-dependent: Transformer models usually require a large amount of labeled data for training, while medical imaging datasets are relatively scarce.

4.2. Improvements to Medical Transformer

The Medical Transformer (MedT) model demonstrates significant advantages in medical image segmentation, especially in the task of segmenting CT images of tuberculosis. It effectively solves the problems of high computational complexity and data dependency of traditional deep learning models in dealing with complex lesions by introducing the Gated Axial Attention module and the local-global training strategy (LoGo). Traditional models, such as U-Net, FCN, and DeepLab, have limited performance in dealing with complex backgrounds and multiple foci, and especially have obvious deficiencies in long-range dependency modeling and capturing global features. In contrast, the MedT model introduces a gating mechanism into the self-attention mechanism through the Gated Axial Attention module, which is able to compute the attention weights along the rows and columns of the image separately, thus reducing the computational complexity while still maintaining a good capture of long-range dependencies of the image. This enables MedT not only to deal with complex lesions such as tuberculosis, but also to cope with background noise in the image, reducing the possibility of misdiagnosis and missed diagnosis.

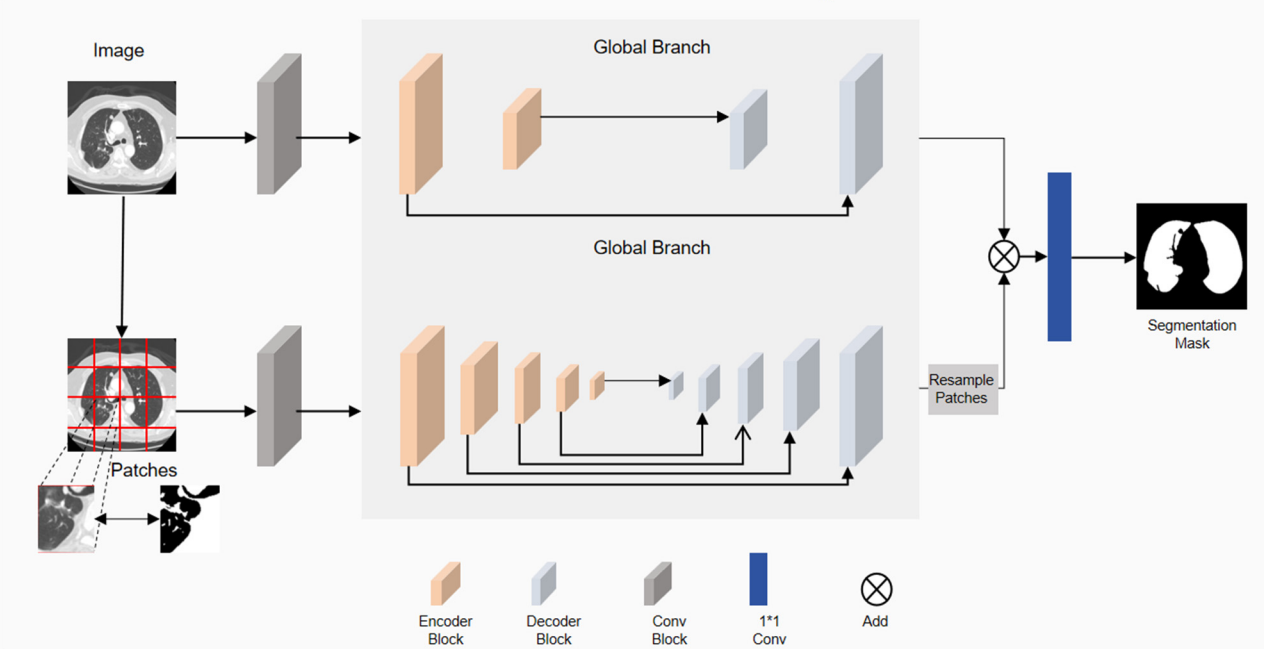

Fig 1. Medical Transformer (MedT) architecture

The local-global training strategy (LoGo) is another major innovation of MedT, which combines local and global feature extraction to enable the model to learn both local detail features and global contextual information of an image during training. Conventional models often perform poorly in the segmentation of multi-scale lesions, and it is difficult to take into account the simultaneous segmentation of small nodules and large-scale lesions, while the LoGo strategy enables the MedT model to ensure detail accuracy while maintaining the accuracy of the overall structure. As a result, MedT is able to more comprehensively detect early fine lesions and extensive lesion areas in TB image segmentation, which greatly improves the diagnostic accuracy.

In addition, MedT still has strong generalization ability under small datasets, avoiding the over-reliance on large amounts of labeled data by traditional deep learning models. Optimized by the LoGo strategy, MedT can effectively reduce overfitting and maintain stable segmentation performance on different types of datasets. This makes it still have efficient segmentation capability in scenarios with limited medical data such as tuberculosis.

Despite the high computational complexity of the MedT model, it dramatically reduces the demand for hardware resources and improves computational efficiency through the optimization of the Gated Axial Attention module. It is well adapted to computer-aided diagnosis (CAD) systems, and is capable of handling large-scale medical imaging data for real-time diagnosis and large-scale screening. This efficient and accurate segmentation capability makes MedT have a broad application prospect in the field of medical image

segmentation, which is expected to promote the early detection and accurate treatment of tuberculosis and other diseases. [5]

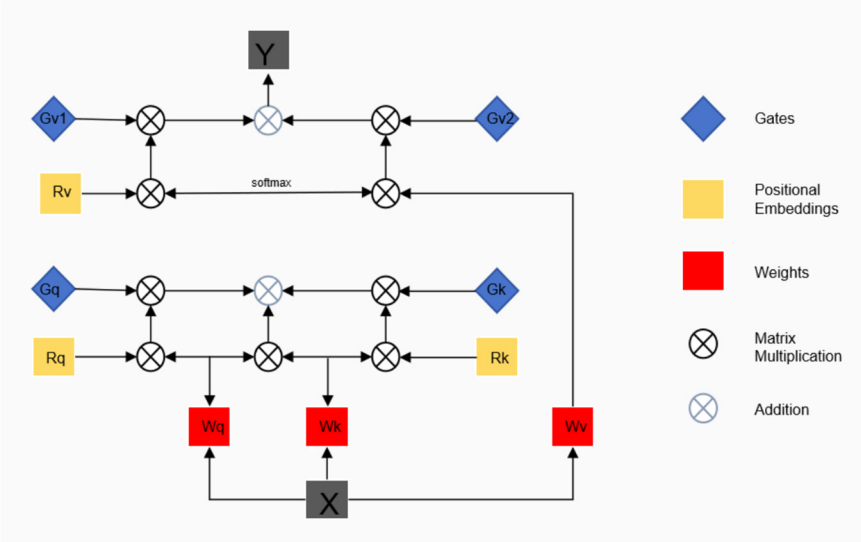

Fig 2. Gated Axial Attention Module

5. Methods to Address the Lack of Indigenous Datasets in the Xinjiang Region

Migration learning, data augmentation, and cross-region data sharing are three important strategies in solving the problem of scarcity of medical imaging data in Xinjiang region.

Migration learning utilizes models pre-trained on large-scale datasets to migrate their knowledge to local data in Xinjiang. This approach, by fine-tuning the pre-trained models, can reduce the need for large-scale labeled data and improve the performance on small sample data. For example, applying a model trained on a widely used global tuberculosis dataset to the Xinjiang region with localized fine-tuning helps the model better adapt to the specific lesion characteristics of Xinjiang.

Data augmentation techniques effectively expand the diversity of training data by applying multiple transformations (e.g., rotation, scaling, cropping, etc.) to existing data. In this way, the model is able to learn under different lesion representations and imaging conditions, enhancing its robustness and generalization ability, and making segmentation and diagnosis more consistently, especially when facing complex lesions or low-quality images.

Cross-region data sharing can enhance model training by uniting datasets from multiple regions. The Xinjiang region can cooperate with medical organizations in other provinces or countries to share TB imaging data and construct a larger training dataset. This not only enriches the training samples of the model, but also improves the adaptability of the model in different regions and lesions, which helps to improve its generalization performance. Through these three strategies, the problem of data scarcity in Xinjiang can be effectively alleviated, and the accuracy and usefulness of the intelligent diagnosis system can be improved.

6. Problems the Medical Transformer Model is Expected to Solve

The Medical Transformer (MedT) model provides an effective solution to some major shortcomings of the CAD system in lung CT diagnosis in Xinjiang's tertiary hospitals. First, the Gated Axial Attention module significantly reduces the computational complexity of the traditional Transformer model by introducing a gating mechanism into the self-attention mechanism. This innovation enables the CAD system to run complex models with limited hardware resources and improves the real-time processing capability of the system. Compared with the traditional CNN model, this optimization is particularly suitable for the problem of insufficient computational resources in Xinjiang's tertiary hospitals.

MedT combines local features and global feature extraction through the Local-Global Training Strategy (LoGo), which enables the model to effectively handle details and overall structure during training, and ensures accuracy even with a limited dataset. This has a significant mitigating effect on the problem of model overfitting in Xinjiang hospitals due to insufficient data resources. MedT excels in the detection of small lesions, and the Gated Axial Attention mechanism accurately captures long-range dependencies in complex backgrounds, reducing the problem of leakage and misdiagnosis of small lesions. Compared with the traditional model, it is more suitable for dealing with cases of blurred boundaries and early lesions, improving the accuracy of early diagnosis of diseases such as tuberculosis.

Overall, the MedT model effectively solves the major technical challenges faced by CAD systems in Xinjiang's tertiary hospitals by optimizing computational complexity, reducing data dependency, and improving lesion detection accuracy, which enhances its potential for application in resource-constrained environments.

7. Discussion of Future Directions for Improvement

Model lightweighting is particularly important in

tuberculosis diagnosis in Xinjiang region. Due to the limited hardware resources in the region, complex deep learning models are difficult to run efficiently in practical applications. Future research can reduce the computational complexity of the model through techniques such as model pruning, quantization, and knowledge distillation, thus reducing the hardware requirements without sacrificing segmentation accuracy. Such optimization can not only make the model applicable to real-time TB diagnosis in Xinjiang, but also better support large-scale screening and improve diagnostic efficiency.

Self-supervised learning has great potential in data-poor scenarios. In Xinjiang, especially in remote areas, obtaining a large number of high-quality medical image annotations is difficult. Self-supervised learning can enhance the feature extraction capability of the model through pre-training tasks on unlabeled data, such as image recovery or transform prediction. This will effectively reduce the dependence on labeled data and enable the model to achieve better generalization performance on smaller datasets, thus improving the detection ability under complex TB foci in Xinjiang.

Multi-task learning can further enhance the diagnostic ability of the model. Combining CT image segmentation of tuberculosis with tasks such as lesion classification and diagnosis prediction can improve the overall performance of the model by sharing feature representations. This joint training approach can reduce the model's misdiagnosis and omission in complex cases in Xinjiang and provide more accurate diagnostic results. In the future, by combining lightweighting, self-supervised learning and multi-task learning, the Medical Transformer model is expected to demonstrate more efficient performance with stronger generalization capabilities in resource-constrained medical scenarios such as tuberculosis in Xinjiang, significantly improving diagnostic quality and efficiency.

Acknowledgments

Project fund content:

This paper is funded by the Xinjiang College of Science & Technology (school-level) Student Innovation and Entrepreneurship Training Program project funds

Grant numbers: X202413561114.

References

[1] Buyiman-Bahani, Salavati Kharizbek, Nie Yanwu,et al. Comparison of seasonal epidemiologic characteristics and stage trends of tuberculosis in Xinjiang from 2005 to 2021[J]. Practical Preventive Medicine,2022,29(11):1285-1289.

[2] Shen Jiang. Research on medical image segmentation algorithm based on deep learning [D]. Shandong Institute of Commerce and Industry, 2023.DOI:10. 27903/d. cnki. gsdsg. 2023. 000137.

[3] Jinghua Z, Yue S, Hui C, et al. Cross Pyramid Transformer makes U-net stronger in medical image segmentation[J]. Biomedical Signal Processing and Control,2023,86(PC).

[4] Fay Zhao. A computer-aided diagnosis system for tuberculosis based on DR chest radiograph[J]. Electronic Technology and Software Engineering,2016, (05):97-98.

[5] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical Transformer: Gated Axial-Attention for Medical Image Segmentation," Johns Hopkins University and Rutgers University, 2021.