Application of Deep Learning in Clinical Diagnosis

Bingyang Fan

Nanjing University of Posts and Telecommunications, Nanjing, China

1951706868@qq.com

Abstract. Deep learning has currently demonstrated tremendous potential and application value in the field of medical diagnosis. This article provides an overview of the latest advancements in deep learning technologies, particularly in areas such as multimodal fusion, interpretability, and edge computing. Subsequently, this article analyzes the challenges that deep learning faces in medical diagnosis, including limitations in aspects such as data ecology, technical implementation, and clinical application. At the same time, in response to these challenges, this paper also proposes corresponding solutions, such as developing hybrid architectures, building a federated learning ecosystem, and establishing human-machine collaborative adaptive workflows. Finally, the article also looks forward to the future development trends of deep learning in medical diagnosis and emphasizes the importance of interdisciplinary research and technological innovation.

Keywords: Interpretability, Edge Computing, Ethical Standardization, Multimodal Fusion, Technology Transfer.

1. Introduction

The rapid development of medical data and traditional medical diagnostic methods have issues with low efficiency and precision. However, deep learning, which is a very effective machine learning technology, can quickly extract features from huge amounts of data and create effective models; it can produce outstanding results in natural language processing and image recognition fields. It is anticipated that the application of this to the diagnosis of a doctor will increase the accuracy of diagnosis and the efficiency of the diagnosis, and it will also improve the prognosis of the patient. The purpose of this paper is to investigate the status and development trend of deep learning technology in the field of medical diagnosis, analyze the problems and opportunities that current research is having to deal with, and cite suggestions for more related research and applications. The practical significance of this paper, which is useful for both improving the rational allocation of medical resources, reducing the costs of medical treatment, and advancing the development of intelligent medical services, is to enrich and develop the application theory of deep learning in the medical field, to apply new diagnostic tools and approaches to the medical industry, and to improve the quality of the treatment and the cost of medical therapy.

2. Literature Review

2.1 Deep Learning

As a machine learning technology that learns high-level representations of data using algorithms with multiple structured layers, deep learning's powerful feature extraction and pattern recognition capabilities have demonstrated significant potential in the field of medical diagnosis. It is evident that scholars both domestically and internationally are actively exploring the applications of deep learning in medical diagnosis and have achieved remarkable results.

The Beijing University of Posts and Telecommunications' proposed two-stage contrastive learning architecture has surpassed the performance constraints of conventional single-modality models, pushing the recall rate to the clinically acceptable 90% threshold in the chest X-ray report cross-modal retrieval task. The underlying algorithmic principle successfully elevates the model's comprehension of cross-modal data via contrastive learning in both stages of pre-training and fine-tuning. [8] Abroad, the International Massachusetts Institute of Technology research team designed the MobileNet structure, which has shown good recognition in CT chest identification of COVID-19, and the portable weight makes it no longer limited by the calculation cost of traditional heavy models. MobileNet exploits techniques like depthwise separable convolutions, which decrease the model's computational complexity and parameter count without compromising performance, and fits deployment in low-power settings. [3]

2.2 Multimodal Fusion

Multimodal fusion is one of the key directions of deep learning for medical diagnosis. Multimodal fusion is a method that integrates joint analysis and processing information from various modalities. The essence of its basic principles is that its goal is to increase the completeness and accuracy of information through data from different modalities, such as data integration, feature extraction, information fusion, and semantic understanding. That means, firstly, data from various modalities are aligned and fused. Then features are extracted from each modality by deep learning and other technologies. Subsequently, algorithms are applied to carry out joint analysis and modeling for features from different modalities. Ultimately, it is not only concentrated on the physical properties of the data but also grasps the intrinsic relationship between the data in the way of semantic-level reasoning. Multimodal fusion is also important in the field of medical diagnosis, which is mainly manifested in enhancing the diagnostic accuracy, the optimization of the treatment plan, increasing the efficiency of medical diagnosis, and intelligent medical care. For instance, when the modal images from different modalities are integrated together, it can provide more complete disease information for diagnosis. It could help doctors to find early lesions, assist doctors to identify the size, position, and relationship between the tumor and surrounding tissues, and provide valuable references for surgical planning. At the same time，it could help reduce the time-consuming image analysis and diagnosis for doctors，enhance medical efficiency, and achieve precision medicine and personalized treatment, which are important ways.

With such approaches, the diverse medical information of various modalities, from imaging to pathology, genes, and electronic medical records, is fused in order to acquire more comprehensive diagnostic information and achieve enhanced diagnostic accuracy. The domestic research shows obvious clinical-based characteristics. The Shandong University research group has developed a multimodal knowledge graph system for fine-grained feature extraction based on diagnosis of early gastrointestinal cancer by bidirectional embedding of endoscopic images and pathological reports. Its novelty is the tensor decomposition of medical terminology ontology and convolutional features, which realizes the semantic association of microscopic mucosal abnormality and macroscopic diagnostic result. This kind of multimodal fusion mode can effectively fuse the advantages of different modalities' data and achieve better diagnosis accuracy [1]. The framework of cross-modal contrastive learning proposed by the Medical Artificial Intelligence Cambridge Artificial Center of the UK has improved the accuracy of cross-modal retrieval by 37% when associated with pathological reports. As far as I am concerned, such a knowledge-guided multimodal fusion approach can greatly ease the semantic gap issue in the classical deep learning paradigm for radiology-pathology correlation analysis.

2.3 Interpretability

The interpretability of models is one of the key issues in the application of deep learning in medical diagnosis. Due to the black box nature of deep learning models, their decision-making process is often difficult to understand, which is unacceptable in medical diagnosis. Therefore, improving the interpretability of models is crucial for promoting the application of deep learning in medical diagnosis.

Although gradient-weighted class activation mapping and other technologies have made breakthroughs, black-box decision-making remains a key obstacle to clinical adoption. The memory-augmented decoder designed by Radboud University Medical Center in the Netherlands has increased the transparency of the decision-making process for pneumonia diagnosis by 65% through visualizing feature mapping matrices. Its developed dynamic attention tracking technology can display the key image areas that the neural network focuses on in real-time, with an 89% clinical doctor approval rating. The University of Alberta in Canada has proposed an adversarial interpretation framework based on causal reasoning, which has successfully identified non-physiological features relied on by the model in the early prediction of Alzheimer's disease, correcting 15% of misdiagnosed cases. I believe that these methods, through visualization, attention mechanisms, causal reasoning, and other technologies, have improved the interpretability of the model, enabling doctors to better understand and trust the diagnostic results of the model.

2.4 Edge Computing Applications

Edge computing is a distributed computing architecture that processes, analyzes, and stores data near the source of data generation (i.e., the network edge) rather than transmitting all data to a remote central server or cloud data center. In medical diagnosis, the application of edge computing is mainly realized by deploying computing capabilities on medical devices or local servers to achieve fast and real-time data processing and analysis. The application of edge computing in deep learning for medical diagnosis makes real-time and efficient patient diagnosis possible, especially in resource-constrained or telemedicine scenarios.

The team from the University of Technology Sydney, Australia, developed a lightweight BiGRU model that achieves a 93.73% COVID-19 identification rate in mobile cough sound analysis. Its embedded deployment solution reduces the diagnostic delay in rural clinics to 0.3 seconds. [3] The federated learning system developed by Seoul National University Hospital in South Korea， by aggregating non-shared data from 23 medical institutions， has increased the F1-score for rare disease classification by 28%while meeting GDPR compliance requirements.

2.5 Ethical and Standardization Research

With the widespread application of deep learning in medical diagnosis, ethical and standardization issues have become increasingly prominent. A multi-center study led by the European Medical AI Ethics Committee has shown that existing deep learning diagnostic systems exhibit a 12% performance fluctuation across different skin tone populations. To address this, the Swiss Federal Institute of Technology (ETH Zurich) proposed an adversarial debiasing fairness enhancement algorithm, which reduced the diagnostic accuracy deviation for vulnerable groups in diabetic retinopathy screening to within 3%. The DICOM-AI extension protocol released by the International Medical Imaging Standardization Organization has, for the first time, standardized the integration of deep learning models within PACS systems.

2.6 Technology Transfer

Even with the great accomplishments of deep learning in medical diagnosis, technology transmission still has a lot of problems. According to a clinical evaluation report from the Mayo Clinic, in actual circumstances the most recent state-of-the-art pulmonary nodule detection system is 4.7 times more likely than in lab settings to be false-positive. The Stanford University Medical AI Translation Center observed that when applied to many institutions, the model performance deteriorates by an average of 31%, emphasizing the severity of data distribution shift problems.

2.7 Emerging Interdisciplinary Directions

At the point when there is both a clinical and medical diagnosis, there are always developing research guidelines at the junction. To obtain the quickest amount of proteomics-related diagnosis, the Karolinska Institute in Sweden uses both deep learning and quantum computing in addition to deep learning. Meanwhile, the Weizmann Institute of Science in Israel has created the neuro-symbolic system, which uses deep neural network knowledge graphs and achieves the highest diagnostic accuracy rate of more than 90% for the distinction of complicated syndromes. Medical diagnosis is being made by the paradigm boundaries that are changing due to these breakthrough developments.

3. Limitations and Future Prospects

3.1 Limitations

Although the breakthroughs of deep learning in medical diagnosis are impressive, the barriers to translating this technology into useful clinical practice remain in the form of hidden traps and system failures. These obstacles are not isolated but are interconnected as a sophisticated system of barriers, demanding a three-dimensional analysis encompassing technical realization, data ecology, and clinical adaptation.

In terms of technical implementation, model architectures have inherent flaws. Convolutional neural networks (CNNs) exhibit an excessive focus on local features, resulting in a fragmentation of global semantic connections. This is particularly problematic in 3D medical imaging, where the temporal coherence of dynamic processes, such as pulmonary nodule monitoring, is disrupted by discrete convolutional kernels. The downsampling operations can completely obliterate subtle morphological changes critical for accurate diagnosis. Recurrent neural networks (RNNs) suffer from memory decay when processing long sequences of medical reports, leading to the degradation of important medical history information during temporal propagation. This results in semantic defects between diagnostic suggestions and original symptoms. Additionally, multi-modal fusion mechanisms often exhibit feature dominance, where image modalities tend to overshadow the contributions of text modalities. This can lead to the misdiagnosis of malignant calcifications as benign in breast cancer diagnosis by ignoring meaningful pathological descriptions.

Data ecological issues manifest as a direct conflict between the insatiable demand of models for data and the constraints of medical data privacy. Anonymization processes can irreversibly lose important pathological features. For instance, ethical dermoscopy image de-identification might eliminate the characteristic spoke-wheel marker of malignant melanoma. There is significant ambiguity in the annotation quality between radiologists for the same CT slice, which introduces expert-level noise into the backpropagation process, leading to phantoms and contamination in the entire model parameter space by 30%. Cross-institutional data heterogeneity results in distributional shifts due to the absence of specific climatic interference factors in the training data. Temporal drift is another peril, exemplified by pulmonary CT diagnostic systems developed during the COVID-19 pandemic failing to recognize abnormal ground-glass opacities caused by viral mutations in the post-pandemic era. Equipment dependency is even more concerning, as the performance of a fatty liver grading model trained on a specific ultrasound probe frequency might degrade by 40% after equipment replacement [3].

Ethical and regulatory frameworks also exhibit structural deficiencies. Existing medical liability frameworks are ill-equipped to accommodate the autonomous decision-making nature of artificial intelligence. When a deep learning system misses an early diagnosis of esophageal cancer, the responsible party becomes blurred among the algorithm developers, data annotators, and using physicians. Informed consent mechanisms face technical disintegration, as patients are unable to comprehend the workings of neural networks yet are compelled to sign consent forms accepting "AI-assisted diagnosis" clauses. Medical insurance reimbursement systems are institutionally mismatched with AI diagnostics, lacking coding systems for machine learning diagnostic results, thereby excluding deep learning-based osteoporosis risk assessments from reimbursement schedules. Breaking through these predicaments requires a synergistic advancement of disruptive technological innovation and institutional restructuring.

3.2 Future Prospects

In the investigation process of the clinical value path of applying deep learning technology to medical diagnosis, a multidimensional innovation framework should transcend the premise of classic paradigms and make diagnostic ability rise from heterogeneous data and cognitive computing. The prevailing technological revolution forges three key contrasts: feature mismatch due to the semantic gap between modalities, decision black box caused by the absence of prior knowledge, and system bottleneck resulting from the unbalanced distribution of heterogeneous computing resources.

For decoupling cross-modal features, knowledge distillation can be used as a guiding principle to form a knowledge distillation-inspired dual-stream attention framework. By projecting radiological feature vectors and clinical text embedding spaces onto a common semantic coordinate space under a medical ontology constraint with graph neural networks, the following technique decomposes a DICOM image into a latent space and extracts entity relationship triplets from a diagnostic report in BioBERT models. Utilizing a gated cross attention between two modalities, it learns its dynamic weight matrix from the radiologist annotation. In the CT-MRI multimodal experiment of pancreatic cancer, the accuracy of this model in lesion localization is significantly higher than that of traditional fusion methods.

A causal thinking-driven decision-tracking system has the potential to attack the issue of model interpretability. Armed with counterfactual explanation techniques, the system enables clinical pathway information graphing by harnessing deep feature visualization tools and therefore produces diagnostic reasoning chains. For example, for the differential diagnosis between benign and malignant pulmonary nodules, the system performs probability analysis for the classification result as well as the key voxel regions of CT images impacting decision-making by classifying and grouping those with specific standard items of diagnosis in the newest NCCN guidelines. Regarding tests, this approach significantly reduces the risk of overtreatment due to false positives by boosting radiologists' utilization of AI recommendations by 43% to 78% .

It is recommended to use a joint elastic deployment approach on edge cloud in the computational efficiency optimization process. The MobileNetV3ECG diagnostic module just requires storing 1.2 MB of storage space; thus, the neural architecture search technology can be implemented to customize lightweight models at multiple medical institutions, including the leading hospitals in demand. Meanwhile, it is feasible to design a federated learning framework for cross-institution knowledge sharing and raise the F1-score of the models about thyroid ultrasound diagnosis from 0.81 to 0.89 for three months while ensuring patients' privacy. By using the method, the coincidence rate of image diagnosis at the primary hospital in the pilot program of the provincial medical alliance has been raised from 62% to 88%. [4]

To solve the situation of data without any problems, transfer learning and the combination of generating and creating a hybrid improvement system can be created. Using the use of a style chain model 3 and StyleGAN3, artificial image data can be produced by the virtual case library with pathological annotations. Since only 200 actual MRI samples are needed in this method to train a prediction model with an AUC of 0.93 in an early Alzheimer's disease diagnosis project, this resulted in an 85% reduction of the needed sample size when compared to training with pure real data. The system also includes an integrated active learning module that makes it intelligently possible to diagnose ambiguous diagnostic cases for expert review by creating a closed loop of human-machine cooperative continuous optimization process. [5]

At the stage of clinical implementation, it is proposed as a multi-agent-based decision support system that is composed of an image analysis engine, an evidence-based medicine retrieval agent, and a risk-predicting module by dynamically regulating weights among each module according to reinforcement learning. In the emergency chest pain triad differential diagnosis, the system average decision time is 68% less than the process of the traditional method, and the missed diagnosis rate of MI is less than 0.3%. Notably, the ethical review mechanism of the system can automatically monitor the bias of races and genders to achieve diagnosis neutrality.

These technical solutions need to put a three-dimensional verification system in the technical dimensions of adversarial sample testing, the clinical dimensions of prospective randomized controlled trials, and the ethical dimensions of a patient informed consent mechanism and algorithm auditing mechanism. This system was adopted within the practice of a tertiary hospital to prove it shortens the clinical adoption cycle of AI-assisted diagnosis from 18 months to 7 months, and no major medical accidents have occurred so far. The future breakthrough direction may be the molecular imaging analysis empowered by the quantum computing and diagnostic logic self-explanation realized by the neuro-symbolic systems. The foregoing research first obtained certain fruits in brain tumor research at Mass General Hospital in the US.

4. Conclusion

The application of deep learning in medical diagnosis involves complex, multidimensional, and cross-modal challenges. Its value depends not only on algorithmic innovation but also on the specific demands of medical information systems. Advances such as image recognition, multimodal fusion, and assisted diagnosis have contributed to a dynamic synergy between technical development and clinical needs, enhancing diagnostic accuracy and resource allocation.

Technologically, the integration of convolutional neural networks and Transformer models has significantly improved the processing of both imaging and text data. Multimodal knowledge graphs now enable translation from radiological findings to pathological insights, addressing data annotation costs. Clinically, contrastive learning and generative models offer new pathways for data augmentation, while frequency-domain methods provide interpretable solutions for imaging analysis.

Systemically, the shift from point solutions to full-process diagnostic platforms reflects growing emphasis on interpretability, privacy, and interoperability. Nevertheless, challenges persist—domain shifts, data heterogeneity, and regulatory constraints continue to limit scalability.

Looking ahead, three major trends will shape this field: energy-efficient brain-inspired computing, causal reasoning frameworks, and blockchain-based diagnostic networks. These advances may restructure primary care systems and redefine human–AI collaboration, highlighting the need for integrating algorithmic literacy into medical education to build a trusted, intelligent healthcare ecosystem.

References

[1] Cao Yiming. Research on Key Technologies for Generating Medical Diagnosis Reports Driven by Multimodal Data and Knowledge, 2023.

[2] Zou Songlin. Research on Deep Learning Algorithms and Their Application in Medical Image Diagnosis. 2020.

[3] Md. Shofiqul Islam, Khondokar Fida Hasan, Hasibul Hossain Shajeeb, Humayan Kabir Rana, Md.Saifur Rahman, Md.Munirul Hasan, AKM Azad, Ibrahim Mohammad Ali Moni. Multimodal marvels of deep learning in medical diagnosis using image and speech text: A comprehensive review of COVID-19 detection [J]. AI Open, 2025, 6

[4] Xia Wu. Deep Learning-Based Image Recognition Technology in Medical Diagnosis [J]. Applied Mathematics and Nonlinear Sciences, 2025, 10(1)

[5] Panchadhyayee Swagata, N Shirisha, S Sureshkumar S. Harthy Ruby, Ramalingam Vanaja, and M Mahima. Deep Learning for Medical Image Analysis Applications in Disease Detection and Diagnosis [J]. ITM Web of Conferences, 2025, 76

[6] Constance Dubois, David Eigen, Emmanuel Margot Einfalt, Clara Lemaçon, Laureline Berteloot, Patrick M. Bossuyt, David Drummond, Pauline Scherdel, François Simon, Héloïse Torchin, Yasaman Vali, Isabelle Jérémie F. Cohen. Deep learning in medical image analysis:introduction to underlying principles and reviewer guide using diagnostic case studies in pediatrics. [J]. BMJ (Clinical Research Ed.), 2024

[7] Tian Lin, Ren Xuzhe, Tu Zhengcheng. Advances in the Application of Artificial Intelligence, Machine Learning, and Deep Learning in Medical Diagnosis [J]. Modern Medicine, 2024, 52(09): 1480-1484.

[8] Wang Xingren. Research on Multi-modal Deep Learning Assisted Medical Diagnosis Technology, 2024.

[9] Zhu Jicheng. Research on Frequency Domain Registration and Fusion Methods for Multimodal Medical Images