GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Medicine
and Life Sciences

Article

Exploration and Practice of Artificial Intelligence Empowering Diagnosis and Treatment of Hematological Diseases

Qing Yuan [1], Qiaoling Wang [1,*] , Yanwu Liu [1], Jiayu Liu [1], Mingxia Tang [1] and Jian Yang [1]

[1] Ya'an People's Hospital, Ya'an, Sichuan, China
* Correspondence: Qiaoling Wang, Ya'an People's Hospital, Ya'an, Sichuan, China

Abstract: Artificial intelligence technology has made great progress in the clinical treatment of blood diseases. This paper comprehensively discusses the core application of artificial intelligence in the diagnosis and treatment of blood diseases, summarizes the existing problems, such as non-standard data standards, small sample size, insufficient model interpretability and disconnection between the system and clinical process, and puts forward solutions, such as the establishment of high-quality data platform. Optimize convenient clinical decision assistance, strengthen the combination of model and procedure, and increase the training of compound talents to promote the intelligent and standardized direction of blood disease diagnosis and treatment.

Keywords: artificial intelligence; blood disease; machine learning; auxiliary diagnosis; clinical information processing

Received: 26 March 2025
Revised: 31 March 2025
Accepted: 22 April 2025
Published: 26 April 2025

1. Introduction

In the diagnosis and treatment of hematological diseases, it is a difficult problem that perplexes and restricts clinicians because of the complicated pathogenesis, various types and difficult treatment. However, with the rapid development of big data and AI technology, the traditional diagnosis and treatment model based on a small amount of data and relying on manual data operation has been reformed. AI can provide favorable support for the diagnosis and treatment of blood diseases in the direction of risk prediction, auxiliary diagnosis and individualized treatment through big data processing, pattern recognition and fusion, thus improving the diagnosis and treatment accuracy and efficiency. In this paper, the application mode, existing problems and practice path of AI in the field of blood diseases are discussed in detail, in order to provide reference for the intelligent development of blood disease diagnosis and treatment [1].

2. The Application Model of Artificial Intelligence Technology in the Diagnosis and Treatment of Blood Diseases

2.1. Blood Disease Risk Screening Based on Machine Learning

The machine learning model is applied to the risk screening of blood diseases. By collecting multidimensional information such as patients' basic data, blood examination indicators and family history, we build a model to achieve efficient and accurate prediction of risk. Common models include logistic regression, random forest and support vector machine. Taking logistic regression model as an example, if the eigenvector is $X = [x_1, x_2, \cdots, x_n]$, the corresponding weight is $W = [w_1, w_2, \cdots, w_n]$, and the bias term is b, then the prediction probability of blood disease risk $P(y = 1|X)$ can be expressed as:

$$P(y = 1|X) = \frac{1}{1+e^{-(W \cdot X+b)}} \quad (1)$$

The model is trained by minimizing the Loss function, the common loss function is Log Loss:

$$L = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \log\left(P^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - P^{(i)}\right)\right] \quad (2)$$

Where m is the number of samples, $y^{(i)}$ is the actual label, and $P^{(i)}$ is the prediction probability. This approach allows patients at significant risk for blood diseases to be pinpointed and interventions taken. In addition, the model can explain the major risk factors through the feature importance ranking function, which can be used as a reference for clinicians to make decisions [2].

2.2. Intelligent Image Recognition AIDS Diagnosis

The diagnosis of medical imaging such as bone marrow smear and blood routine image is subjective and time-consuming. Convolutional neural network deep learning algorithm is outstanding in image diagnosis, which can automatically and systematically extract and classify image features. The CNN structure generally includes convolution layer, pooling layer and fully connected layer. Assuming that the input image is I and the convolution kernel is K , the convolution operation can be expressed as:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) \times K(m,n) \tag{3}$$

The model minimizes the cross-entropy loss function through the backpropagation algorithm, specifically:

$$L = -\sum_{c=1}^{C} y_c \, log(\hat{y}_c) \tag{4}$$

Where, C is the number of categories, y_c is the true label, and $\hat{y}_c$ is the model prediction probability. This method has been widely used in auxiliary imaging diagnosis of different blood diseases such as acute leukemia and lymphoma, and its efficiency and accuracy have been greatly improved [3]. At the same time, in order to make the model more sensitive to details, attention mechanism and residual network are used to improve the robustness of the model in the face of a large number of complex image features.

2.3. Clinical Information Extraction Based on Natural Language Processing

Clinical electronic medical records contain a large amount of structured and unstructured data, and traditional manual processing of these data is time-consuming and error-prone. NLP technology is used to analyze and extract information from medical records automatically. Especially in the field of hematology, named entity recognition and relationship extraction models are often used to assist processing. Specifically, given the medical record text sequence $T = [t_1, t_2, \dots, t_n]$, the entity label sequence is predicted by Bi-LSTM and CRF:

$$h_t = \text{Bi-LSTM}(t_t) \tag{6}$$

$$P(y|T) = \frac{e\sum_{t=1}^{n}\left(A_{y_{t-1},y_t} + P_t(yt)\right)}{\sum y' \in Y e^{\sum_{t=1}^{n}\left(A_{y'_{t-1},y'_t} + P_t(y'_t)\right)}} \tag{7}$$

Where, A is the state transition matrix and P_t is the predicted score at the current moment. The model can automatically extract keywords such as the type of blood disease (such as leukemia, lymphoma), test indicators (such as hemoglobin, bone marrow cell ratio) and drug names. At the same time, we also use the relationship extraction model to realize the study of the relationship between drugs and symptoms.

2.4. Individualized Treatment Plan Recommendation and Prognosis Assessment

The formulation of an individualized treatment plan emphasizes the consideration of characteristic differences between patients, and recommendation systems and predictive models can assist physicians in making optimal decisions, often using collaborative filtering or reinforcement learning algorithms. Set the feature vector of patients as F_p and the feature of treatment plan as F_t, and construct a score matrix R using matrix decomposition technology to predict the adaptability score of patients to treatment plan:

$$\hat{R}_{pt} = F_p^T \cdot F_t \tag{8}$$

The model is trained on the therapeutic effects of many historical cases, and strategies such as the appropriate drug combination, duration of treatment, and method of medication are provided to the current patient in real time. In addition, regularization terms can be added to prevent overfitting, thus ensuring the scientific nature of the recommendation. Survival analysis models, such as the Cox proportional risk model, can be used to further evaluate the prognosis of treatment:

$$h(t|X) = h_0(t) \times e^{\beta X} \tag{9}$$

Where, $h(t|X)$ is the individual risk function in time, $h_0(t)$ is the baseline risk, X is the patient characteristics, and β is the model parameter. This technology can monitor the treatment plan in real time and accurately predict the treatment outcome, which is a safe, rational and individualized way to manage patients with blood diseases.

3. The Challenge of AI-Assisted Diagnosis and Treatment of Blood Diseases

3.1. Multi-Source Data Standards are not Uniform

The diagnosis and treatment of hematological diseases requires a variety of clinical and auxiliary laboratory data, including hematological tests, genomic data, histopathology, radiography, and electronic medical records. However, because the various detection data come from different instruments and equipment, there are great differences in data acquisition methods, data structures and code systems, resulting in very difficult data fusion. For example, the blood test data obtained by various medical institutions may be named with different fields, different units of measurement, and different reference ranges of test reports [4]. Although some pathological image data comply with DICOM standard, there is still no consensus on the standardization of image labeling and classification system. In addition, while genetic sequencing data has a common format (such as a VCF or BAM file), the detailed description also varies from institution to institution, resulting in repeated cleaning and transformation of data input and processing, which affects the effectiveness of AI model training (Table 1).

Table 1. Overview of Multi-source Data Standards for diagnosis and treatment of blood diseases.

Data source	Common format	Degree of standardization	Inconsistent field ratio	Missing value ratio
Gene sequencing data	VCF/BAM	Low	27%	15%
Pathological image data	DICOM/JPEG	Medium	19%	10%
Blood test data	HL7/Excel	Low	34%	12%
Electronic medical record data	XML/JSON/text	Extremely low	42%	18%
Clinical follow-up data	Table/text mix	Extremely low	38%	22%

The data show that the proportion of field inconsistency and missing values in different data sources is high, the standardization level of electronic medical record data and follow-up data is the lowest, the field inconsistency rate is greater than 35%, and the proportion of missing values is generally high. The inconsistency of multi-source data is the key point that hinders the data unification of AI systems.

3.2. Insufficient Sample Size and Labeling

Artificial intelligence models have high requirements for sample size and accurate labeling of blood diseases. The low incidence of blood diseases, small number of patients, rare disease subtypes and the protection of their own privacy leads to the time-consuming collection of blood disease samples. In addition, complex operations such as pathological image analysis and gene sequencing require the involvement of hematologists or pathology specialists, which consumes a lot of time and labor. For example, for tasks such as cell classification and gene variation interpretation, it is difficult to find a large-scale and efficient labeling process due to high accuracy requirements, which also exacerbates the problem of insufficient labeling data (Table 2).

Table 2. Data summary of AI model training samples for blood diseases.

Data type	Total sample size (National multi-center)	Annual new sample	Expert labeling time/sample	Labeled rate
Pathological section image data	20,000 cases	1500 cases	20 minutes	45%
Genetic test data	15,000 cases	1200 cases	30 minutes	52%
Blood routine and biochemical data	50,000 cases	5000 cases	5 minutes	68%
Clinical follow-up text data	10,000 cases	800 cases	25 minutes	40%

As shown in the table, the number of labeled samples is small and the labeling rate is too low. The labeling rate of most important data is less than 50%, and the problems such as high labeling time and difficult data growth cause the shortage of training data, which restricts the progress of the training model.

3.3. Model Interpretability and Physician Trust Disorders

Most of the existing artificial intelligence models use complex deep learning structures in the diagnosis and treatment of blood diseases. Although they have high diagnostic accuracy, they lack the interpretability of reasoning steps, so doctors do not know the basis for making judgments. Doctors will often rely more on tools with clear reasoning processes and medical explanations to make critical decisions, and when the data provided by AI systems is not interpretable, their trust will be reduced. Secondly, the data feature dimension used in the model training process is too high to provide an intuitive display of the feature contribution weight, resulting in a low sense of identification between doctors and patients with the decisions made by AI (Table 3).

Table 3. Survey of hematology doctors' trust in AI models.

Model type	Diagnostic accuracy	Interpretability score (out of 10)	Doctor Trust (out of 10)	Actual adoption rate
Traditional machine learning models	82%	6.8	7.0	65%
Black box deep learning model	91%	4.2	5.5	48%
Convolutional neural network image recognition model	88%	4.9	6.0	52%

It can be seen from the table that although some AI models have high accuracy, their interpretability scores are low. In particular, the interpretability of the deep learning model is only 4.2 points, which leads to the low trust and acceptance of the model by doctors, and can not play a practical effect.

3.4. The System is Disconnected from Clinical Processes

The development and application of the AI-assisted blood disease treatment system has not been well aligned with real-world clinical workflows. Many AI tools are only used as auxiliary application tools, and are not integrated into information systems such as HIS, EMR or LIS, which makes it more difficult to invoke data. Moreover, after the system is actually applied, the operation process of doctors will become complicated and cumbersome, which will increase their work burden [5]. In addition, due to the unreasonable setting of nodes for the intervention of the AI system in the diagnosis and treatment process, the daily work rhythm of doctors is also affected, and the frequency of actual use of the system is low (Table 4).

Table 4. Current status of AI system and clinical process interface.

Hospital information system module	Whether to connect to the AI system	Data call delay time	Doctor usage frequency (times/week)	System call rate
HIS system	Partial docking	15 seconds	3 times	42%
LIS system	unjointed	Unable to call	0 times	0%
EMR system	Partial docking	20 seconds	4 times	50%
PACS system	unjointed	Unable to call	0 times	0%

As shown in the table, AI systems are currently only connected in a few HIS systems and EMR systems, and there is a lag in data invocation. LIS and PACS systems have not been connected, the actual utilization rate is not high, and the system call rate is less than 50%, which shows that the system integration is low.

4. The Practice of AI-Assisted Diagnosis and Treatment of Blood Diseases

4.1. To Build a High Quality Blood Disease Diagnosis and Treatment Data Platform

The establishment of high-quality blood disease diagnosis and treatment data platform is the core link of using AI for blood disease diagnosis. We integrate and manage the electronic medical record system (EMR), laboratory examination, imaging system (PACS), genetic testing platform and follow-up data to form a structured and unstructured big data set. At the same time, we need to establish unified data standard rules and interface rules to ensure unified data standards, good scalability and compatibility among various platforms. For the distortion of data, the accuracy and completeness of data can be improved by means of data cleaning, removing abnormal numbers and supplementing gaps, and experts can be involved in the annotation of important samples to increase their medical significance. At the same time, data security cannot be ignored, and it is necessary to use anonymization, encrypted storage and hierarchical authorization to ensure data privacy and security. On this basis, a shared open data exchange platform is established to promote data sharing among medical institutions, scientific research institutions, and enterprises, and provide reliable data protection for AI model training and clinical decision-making.

4.2. Building Clinically Friendly AI-Assisted Decision Systems

Creating clinically friendly AI-assisted decision systems is at the heart of truly applying AI to the diagnosis and treatment of blood diseases. In the system design, it is necessary to combine the actual needs and usage habits of medical staff to improve the applicability of functions, ease of process, interpretability of results and continuous optimization. The overall path design is shown in Figure 1.

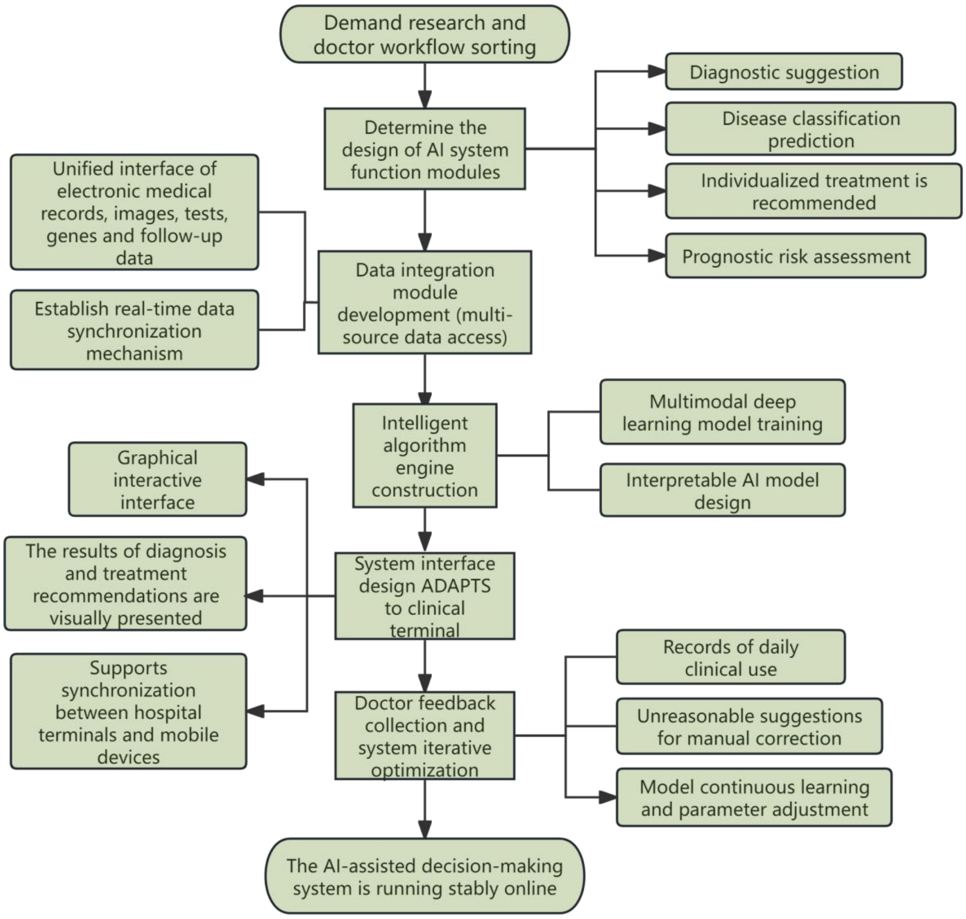

Figure 1. AI-assisted decision system development and application path diagram.

4.3. Promote the Deep Integration of Artificial Intelligence Models with Clinical Diagnosis and Treatment Processes.

Promoting the deep integration of AI models and clinical diagnosis and treatment processes is an important part of allowing AI technology to be effectively applied to practical operations. We should consider the construction of clinical diagnosis and treatment process as a whole, and systematically integrate the AI-assisted part into the whole process management of patients. Specifically, the artificial intelligence model is used to conduct preliminary analysis screening and assess risks at the first diagnosis of patients, so as to achieve rapid diagnosis and early warning. During diagnosis, multi-dimensional data such as patient image data, genomic data, and hematology are combined to optimize and improve the accuracy and efficiency of diagnosis through artificial intelligence analysis tools. In the formulation of treatment plan, through the deep learning of individual characteristics and a large amount of past data, to provide doctors with individual recommendations for treatment plan. In the stage of efficacy detection and regular review management, artificial intelligence is used to monitor the changes of patients' indicators in real time, give timely adjustment suggestions, and strengthen the scientific management of the disease. In addition, it is necessary to realize the link between AI and HIS, LIS, PACS and other systems, so as to seamlessly connect the data flow, avoid repeated entry, and improve the overall diagnosis and treatment efficiency.

4.4. Cultivate Compound Talents and Multi-Party Collaborative Innovation

The deep application of AI in the treatment of blood diseases requires cross-field manpower training and the mode of collaborative innovation of all parties to be guaranteed. At present, clinical talents have low awareness of AI technology, and AI researchers cannot fully understand clinical needs, which greatly affects the effectiveness of technology transformation. Therefore, we should focus on cultivating interdisciplinary medical science and technology talents, further develop universities, hospitals and scientific research institutions to set up AI foundation, data science, clinical application and other content teaching activities, in order to guide medical talents to learn AI technology, but also encourage AI researchers to explore the clinical environment, grasp the real use needs, and strengthen cross-field cooperation skills and team spirit. At the same time, we should establish a joint innovation system involving government agencies, scientific research institutions, hospitals and enterprises, and combine policy advantages, financial support, research and development technology and clinical pilot to complete the overall cooperation from technology research and development to clinical trials. It is suggested that the government make corresponding policies, ensure the support of talent training, project assistance, and achievement transformation, create a good innovation ecology, and promote the effective application of AI technology in the diagnosis and treatment of blood diseases.

5. Conclusion

With the development of AI, its application in the diagnosis and treatment of blood diseases is also increasing, including early risk screening, AI image recognition, clinical data extraction, personalized treatment recommendations, etc., so as to improve the accuracy and efficiency of diagnosis. However, there are still some problems to be solved, including the inconsistency of multiple data standards, too few labeling samples, low interpretability of models, and disconnection of system processes. To this end, this study puts forward some countermeasures: to build a high-quality data platform; Shaping clinically-friendly AI decision systems; Promote the model into the whole process of diagnosis and treatment; Cultivate multi-faceted talents. This helps to promote the AI process of blood disease diagnosis and treatment, and provides practical experience and innovative thinking for its further development.

References

1. V. Srivastava, R. Kumar, Y. M. Wani, et al., "Role of artificial intelligence in early diagnosis and treatment of infectious diseases," *Infect. Dis. (Lond.)*, vol. 57, no. 1, pp. 21–26, 2024, doi: 10.1080/23744235.2024.2425712.
2. B. Innocenti, Y. Radyul, and E. Bori, "The use of artificial intelligence in orthopedics: Applications and limitations of machine learning in diagnosis and prediction," *Appl. Sci.*, vol. 12, no. 21, p. 10775, 2022, doi: 10.3390/app122110775.
3. Y. Pan and Y. F. Jiao, "Application of artificial intelligence in the diagnosis and treatment of Kawasaki disease," *World J. Clin. Cases*, vol. 12, no. 23, pp. 5304–5307, 2024, doi: 10.12998/wjcc.v12.i23.5304.
4. C. Sharma et al., "Multisystem inflammatory syndrome in children and Kawasaki disease: a critical comparison," *Nat. Rev. Rheumatol.*, vol. 17, no. 12, pp. 731–748, 2021, doi: 10.1038/s41584-021-00709-9.
5. N. I. Yu and D. S. N., "Application of artificial intelligence in ophthalmology for the diagnosis and treatment of eye diseases," *Pattern Recognit. Image Anal.*, vol. 32, no. 3, pp. 477–482, 2022, doi: 10.1134/S1054661822030166.