frontiers | Research Topics

Adoption of artificial intelligence in human and clinical genomics, volume II

Edited by
Li Zhang, Guangming Zhang and Piyush Shukla

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-5055-7
DOI 10.3389/978-2-8325-5055-7

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Adoption of artificial intelligence in human and clinical genomics, volume II

Topic editors

Li Zhang — University of London, United Kingdom

Guangming Zhang — University of Texas Health Science Center at Houston, United States

Piyush Shukla — Rajeev Gandhi Technical University, India

Citation

Zhang, L., Zhang, G., Shukla, P., eds. (2024). *Adoption of artificial intelligence in human and clinical genomics, volume II*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-5055-7

Table of contents

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1090180

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Zhe Song,
Taylor's University, Malaysia
Jun Zheng,
Baotou Teachers' College, China
Defei Liu,
Southwest University, China

*CORRESPONDENCE
Shuhua Xu,
jssdtsxsh@163.com

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 05 November 2022
ACCEPTED 16 December 2022
PUBLISHED 04 January 2023

CITATION
Zhou G, Xu S, Liu X, Ge J, He Q, Cao W, Ding J and Kai X (2023), Relationship between the image characteristics of artificial intelligence and EGFR gene mutation in lung adenocarcinoma. *Front. Genet.* 13:1090180. doi: 10.3389/fgene.2022.1090180

Relationship between the image characteristics of artificial intelligence and EGFR gene mutation in lung adenocarcinoma

Guoping Zhou[1], Shuhua Xu[1]*, Xiaoli Liu[2], Jingjun Ge[3], Qiyu He[1], Weikang Cao[1], Junning Ding[1] and Xinghua Kai[1]

[1]Department of Cardiothoracic Surgery, Dongtai Hospital of Traditional Chinese Medicine, Dongtai, China, [2]Department of Pathology, Dongtai Hospital of Traditional Chinese Medicine, Dongtai, China, [3]Department of Radiology Imaging, Dongtai Hospital of Traditional Chinese Medicine, Dongtai, China

Lung Adenocarcinoma (LUAD) is a kind of Lung Cancer (LCA) with high incidence rate, which is very harmful to human body. It is hidden in the human body and is not easy to be discovered, so it brings great inconvenience to the treatment of LUAD. Artificial Intelligence (AI) technology provides technical support for the diagnosis and treatment of LUAD and has great application space in intelligent medicine. In this paper, 164 patients with primary LUAD who underwent surgery in Hospital A from January 2020 to December 2021 were selected as the study subjects, and the correlation between the imaging characteristics of LUAD and Epidermal Growth Factor Receptor (EGFR) gene mutation was analyzed. Finally, the conclusion was drawn. In terms of the study on the correlation between EGFR mutation of LUAD and the imaging characteristics of Computed Tomography (CT), it was concluded that there were significant differences between the patient's sex, smoking history, pulmonary nodule morphology and the EGFR gene, and there was no significant difference between the patient's tumor size and EGFR gene; in the study of the relationship between EGFR gene mutation and CT signs of LUAD lesions, it was found that there were significant differences between the symptoms of cavity sign, hair prick sign and chest depression sign and EGFR gene, but there was no significant difference between the symptoms of lobulation sign and EGFR gene; in the study of pathological subtype and EGFR gene mutation status of LUAD patients, it was concluded that the pathological subtype was mainly micropapillary. The mutation rate was 44.44%, which was the highest; in terms of CT manifestations of adjacent structures of lung cancer and the study of EGFR gene mutation status, it was found that there was a statistical difference between the tumor with vascular convergence sign and EGFR gene mutation, and pleural effusion, pericardial effusion, pleural thickening and other signs in tumor imaging were not significantly associated with EGFR gene mutation; in terms of the study of CT manifestations of adjacent structures of LCA and EGFR gene mutation status, it was concluded that pleural effusion, pericardial effusion, pleural thickening and other signs in tumor images were not significantly associated with EGFR gene mutation; in terms of analysis and cure of LUAD, it was concluded that the cure rate of patients was relatively high, and only a few people died of ineffective treatment. This paper provided a reference for the field of intelligent medicine and physical health.

KEYWORDS

lung adenocarcinoma, EGFR gene mutations, artificial intelligence, intelligent medicine, image characteristics

1 Introduction

The imaging features of LUAD play an important role in the treatment and research of LCA, and the research on them can provide reference for intelligent medical treatment and cure of LUAD. With the rapid development of artificial intelligence technology, intelligent medical treatment has brought great convenience to the treatment of LUAD. It has become a general trend to analyze the correlation between the image characteristics of LUAD and EGFR gene mutation. At the same time, due to the huge and complex data set of lung cancer gene expression profile, it is necessary to extract effective information through artificial intelligence methods.

The imaging features of LUAD have been deeply studied in the field of intelligent medicine. Yu Lingming used a mountain learning arrangement based on CT image characterization to predict the pathological staging of non-small cell LCA (Yu et al., 2019). Yoon Jiyoung predicted the effect of programmed acellular death ligand one presentation on advanced LUAD by CT imaging (Yoon et al., 2020). Koyasu Sho concluded that gradient tree boosting was useful for forecasting the pathologic subtypes of non-small cell LCA and the polyester/18F pentafluoromethacrylate interferon for EGFR mutation status was useful (Koyasu et al., 2020). Pascoe Heather M analyzed the multifaceted nature of LUAD (Pascoe et al., 2018). She Yunlang investigated the predictable worth of CT-based radiology in distinguishing inert LUAD from aggressive LUAD in patients with lung nodules (She et al., 2018). Gertych Arkadiusz found that convolutional neural networks could discriminate accurately between the four tissue growth patterns of LUAD in digital slides (Gertych et al., 2019). Abdul Jabbar Khalid believed that geographic immune variability illuminated the differential evolution of LUAD (AbdulJabbar et al., 2020). There are many researches on LUAD, but there is no research on artificial intelligence in this field.

EGFR gene mutation has a certain correlation with cancer treatment, and many scholars have also made achievements in this field. Li Yajun analyzed the effect of CT slide depth and convoluted kernels on the performance of radiological models for forecasting EGFR statuses in non-small cell LCA (Li et al., 2018). Jia Tian-Ying modeled by using the characteristics of radiology and random forest, and identified the EGFR mutation of LUAD through non-invasive imaging (Jia et al., 2019). Mei Dongdong analyzed CT images of lung adenocarcinoma and explored whether cell characteristics could become a substitute biomarker for EGFR mutation (Mei et al., 2018). Tulchinsky Eugene analysed the escape mechanism of EGFR-targeted treatment in LCA (Tulchinsky et al., 2019). Kazue Yoneda studied the treatment process and protocols for non-small cell LCA with EGFR deletions (Yoneda et al., 2019). Cicek Tugba analysed the adequacy of Endotracheal Ultrasonography-tobronchial Needle Aspiration (EBUS-TBNA) specimens for mutation analysis in LCA (Cicek et al., 2019). Vyse Simon elucidated the structure of the crystal structure of mutant kinases with EGFR exon 20 insertion and revealing a unique mechanism of kinase initiation and spatial configuration, which determined the lack of reaction of these EGFR synapses to commercially approved EGFR injectants (Vyse and Huang, 2019). Although there are many studies on EGFR and LCA, the research in this area is not deep enough.

This study retrospectively analyzed 164 patients with primary LUAD who underwent surgery in Hospital A from January 2020 to December 2021. The EGFR gene detection results of these patients are available, which could improve the research speed of LUAD and accelerate its progress. The purpose of this study is to explore the relationship between these patients and general clinical features in order to better guide clinical management.

2 Feature extraction of LUAD image based on artificial intelligence

2.1 Artificial intelligence and medical health

The wide application of medical imaging is mainly driven by the progress of computer vision technology. However, there is a serious shortage of imaging and radiotherapy doctors, and there is a shortage of doctors with rich and high-quality clinical experience. Due to the vision and experience evaluation of imaging doctors, there are many cases of misdiagnosis (Bianconi et al., 2019; Wang et al., 2019). The speed of medical imaging operators reading films and the speed of radiation therapists drawing target areas are time-consuming. In medical imaging, the use of artificial intelligence can help doctors read films and draw targets, which would save doctors a lot of time and improve the accuracy of diagnosis, radiotherapy and surgery. With the support of artificial intelligence technology, feature extraction of lung adenocarcinoma image has more possibilities.

2.2 Feature recognition algorithm of LUAD image based on artificial intelligence

The design process of feature recognition algorithm for LUAD image based on artificial intelligence is recorded in Figure 1.

The multi-scale fractal feature is a fractal parameter change measurement function (Wu et al., 2020). The multi-scale fractal feature can be understood as the change degree of D dimension area $(K(x,y,\varepsilon))$ within the scale range of $\varepsilon_{\max}$. The multi-scale fractal feature is used to highlight the difference in fractal features between man-made targets and natural background

$$MFFK(x,y)=\sum_{\varepsilon=2}^{\varepsilon_{\max}}\left[K(x,y,\varepsilon)-\frac{1}{\varepsilon_{\max}-1}\sum_{\varepsilon=2}^{\varepsilon_{\max}}K(x,y,\varepsilon)\right]^2 \quad (1)$$

The threshold value is set to 5, and the binary LUAD image is obtained by segmentation of the enhanced LUAD image as follows:

$$\theta=\frac{\tan^{-1}\left(\frac{2\left(\frac{M_{11}}{M_{00}}-x_c y_c\right)}{\left(\frac{M_{20}}{M_{00}}-x_c^2\right)-\left(\frac{M_{02}}{M_{00}}-y_c^2\right)}\right)}{2} \quad (2)$$

In the formula, M_{00} is the zero matrix, and (x_c, y_c) represents the centroid of LUAD image.

The n-order m-fold Zernike moment of a discrete image is defined as follows:

$$Z_{n,m}=\frac{(n+1)}{\pi(N-1)^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}I(x,y)R_{n,m}(\rho)e^{jim\theta} \quad (3)$$

In the formula, $n = 0,1,2,...; 0 \le |m| \le n; n-|m|$ is an even number. (ρ,θ) is the polar coordinate representation under the unit circle, and $R_{n,m}$ is the radial polynomial.

In order to optimize discrete problems, this paper proposes an artificial intelligence algorithm to update them:

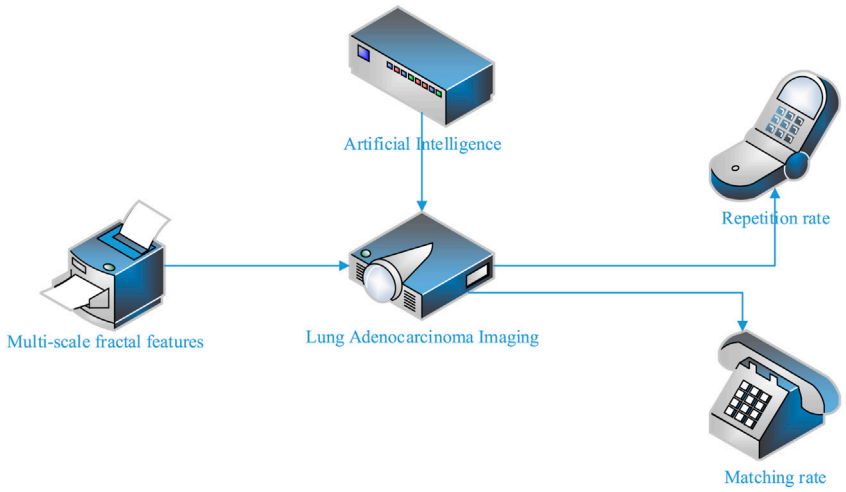


FIGURE 1
AI-based LUAD image feature recognition algorithm.

$$v_{ij} = \begin{cases} v_{ij}^{1}, \ x_{ij} = 0 \\ v_{ij}^{0}, \ x_{ij} = 1 \end{cases} \tag{4}$$

In the formula, v_{ij}^{1} and v_{ij}^{0} respectively represent the probability that the jth position of the ith particle becomes one or 0.

The calculation method of (Eq. 4) still has some defects, so this paper improves the particle position. The update formulas of particle position are as follows:

$$x_{ij}(t+1) = \begin{cases} \bar{x}_{ij}(t), \ r_{ij} \le v_{ij}^{'} \\ x_{ij}(t), \ r_{ij} > v_{ij}^{'} \end{cases} \tag{5}$$

$$v_{ij}^{'}(t) = sig\left(v_{ij}(t)\right) \tag{6}$$

In the formula, $\bar{x}_{ij}(t)$ represents the inversion of $x_{ij}(t)$ in binary, and r_{ij} is a random number in the $[0, 1]$ interval.

The repetition rate is the ratio of the number of repeated feature points extracted from two similar images to the total number of extracted points (Gong et al., 2020). The higher the repetition rate of feature points, the more stable the extracted feature points are and the higher the correct matching rate of feature points would be.

The formula for calculating the repetition rate is as follows:

$$R = \frac{N_3}{\min(N_1, N_3)} \times t \tag{7}$$

In the formula, R is the repetition rate; N1 and N2 are the number of feature points extracted from the left and right images respectively; N3 is the number of feature points from the left image projection to the right image; t is the overlap of similar images.

The ratio k between the number of matches obtained from fine matching of two images and the average value of feature points extracted from the two images is the image matching rate.

The matching rate is calculated as follows:

$$k = \frac{N}{N_1 + N_2} \times t \tag{8}$$

In the formula, N_1 and N_2 are the feature points extracted from the two images; t is the overlap of similar images.

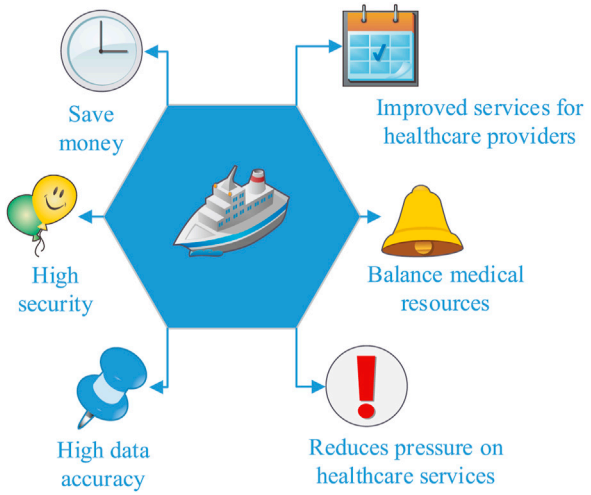


FIGURE 2
Advantages of smart healthcare.

3 Intelligent medical and biological information

With the development of technology and the aging of the population, AI would become a part of many industries in the future, including medical technology (Ge et al., 2019). Although computers would never replace humans, they are useful for certain tasks and can improve the patient's experience. However, the benefits are much greater. The combination of artificial intelligence and medical technology would simplify medical care and create new opportunities. In the case of declining profits and increasing government regulations, it saves money and promotes the survival and development of pharmaceutical enterprises. AI also provides the possibility to standardize many processes in the field of healthcare. Just

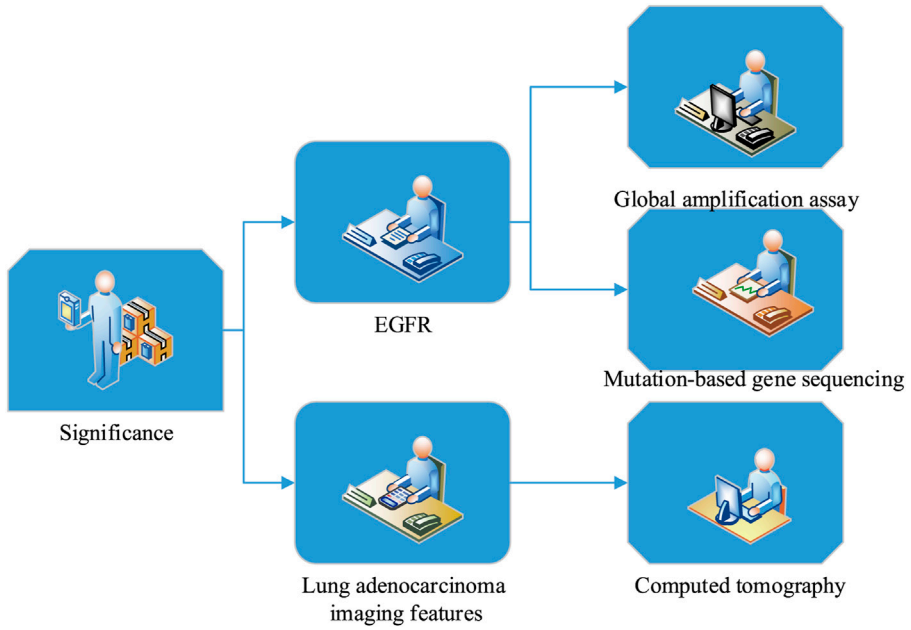


FIGURE 3
Significance of LUAD imaging features and EGFR gene mutation study.

as the specific procedures are different, many policies and practices are the same. Patients often do not understand their rights and responsibilities in health and safety and the accuracy of their data.

The development of AI medical platform would help medical institutions improve services and balance medical resources, thereby reducing the pressure on medical services, especially in areas with limited medical resources (Zhang et al., 2020). Medical institutions choose different construction methods to improve their medical services according to different information technology levels. The AI platform consists of data, computing power, open source systems and algorithms, and various technologies. The computing power ensures the speed of the AI platform. The advantages of AI in medical care are recorded in Figure 2.

In the field of data driven auxiliary medical diagnostic imaging, applications include brain tumor segmentation, pulmonary nodule detection, Alzheimer's disease detection, lymph node detection, pulmonary bronchiectasis detection, chest disease detection and liver ultrasound detection. Each patient has an average of 20–30 images, such as the medical imaging of pulmonary nodules. Computer vision models, such as residual neural networks, are usually used to automatically identify pulmonary nodules. It can train neural networks with dozens or even hundreds of layers, which requires high computing power. The huge amount of data increases the computing time. Therefore, the development of a supercomputing platform can not only reduce computing time, but also improve medical efficiency and reduce patient waiting time.

Knowledge based design creates high value medical diagnostic cards. The knowledge map is a typical product of the big data era. The combination of big semantic network, big data technology and deep learning technology is becoming the main driving force for the development of artificial intelligence. The medical knowledge map mainly uses four technologies: knowledge representation, knowledge extraction, knowledge fusion and knowledge reasoning.

Artificial intelligence has been applied to LUAD, thyroid cancer, breast nodules and other tumors, coronary artery plaque, skin cancer, liver pathology and many other fields. Early LCA is asymptomatic, but 70%–80% of patients with advanced LCA lost the opportunity of surgery and more and more of them developed into LCA. Although most pulmonary nodules are benign, the proportion of early LCA is high and benign.

4 Significance of imaging characteristics and EGFR gene mutation in LUAD

The research significance of imaging characteristics of LUAD and EGFR gene mutation is discussed from two aspects: LUAD imaging and EGFR gene mutation, as shown in Figure 3.

Lung cancer is one of the most common cancers in the world, with the highest mortality. Lung adenocarcinoma is the most common histological type of lung cancer. Small cell LCA accounts for 92% of LCA. EGFR mutation is the main subtype of LUAD. About 45% of non-smoking Non Small Cell LCA (NSCLC) patients have EGFR mutation. Although EGFR mutations are more common in women and non-smokers, there are no reliable clinical features that can accurately distinguish EGFR phenotype from LCA. In clinical practice, EGFR phenotype and LCA are distinguished by gene detection of tissue samples. Gene detection is not only costly and time-consuming, but also invasive due to tissue sampling. Computed tomography is the most popular and widely used imaging method in LCA, because it provides excellent spatial resolution information about the microstructure of LCA, which may reflect the genetic phenotype of the tumor.

EGFR has been proved to be abnormally or highly expressed in many solid tumors and can regulate various biological activities of tumor cells, which inhibit apoptosis through its downstream signal

TABLE 1 Basic information about the test subjects.

Basis of classification	Subgroup	Number
Gender	Male	86
	Female	78
Genotype	Mutant	64
	Wild type	100

transduction. The mutation of EGFR receptor gene may partly predict the sensitivity and efficacy of targeted therapy. At present, the mutation status of EGFR receptor is mostly determined by global amplification detection or mutation based gene sequencing, which requires expensive and complex surgery or puncture technology to obtain tissue samples. In recent years, molecular targeted therapy has become a new hot spot in the treatment of lung cancer, and EGFR gene mutation has attracted great public attention.

5 Correlation between imaging characteristics of LUAD and EGFR gene mutation

5.1 Selection of test objects

From January 2020 to December 2021, 164 patients with primary LUAD who underwent surgery in Hospital A were retrospectively analyzed and their basic information was recorded in Table 1.

From January 2020 to December 2021, the average age of 164 patients with primary LUAD who underwent surgery in Hospital A was (53.5 ± 2.2) years old. Of these patients, 86 were women and 78 were men. 64 patients were mutated and 100 patients were wild type. In general, the probability of men suffering from LUAD is higher than that of women, and the wild type patients are more than the mutant type patients.

Inclusion criteria: Patients with pulmonary nodules or masses found on standard chest CT; untreated lesions with clinical anti-cancer treatment before the study; surgical intervention was carried out after all chest CT scans, and tissue samples were collected for pathological classification and EGFR gene analysis.

Exclusion criteria: Chest CT scanning does not meet the inclusion criteria; chest CT scanning is performed in other institutions. There is no postoperative tissue available for EGFR gene detection, and the imaging data are incomplete.

5.2 CT measurement and film reading method

All patients were placed in supine position with arms raised, and spiral chest examination was performed with Siemens dual source CT scanner. Tumor features include tumor size, lung lobe sign, hair prick sign, pleural traction sign, air branch sign, bilateral lung metastasis and pleural effusion. The tumor size, lung lobe sign, hair prick sign, pleural traction sign, air bronchogram sign, the presence of bilateral lung metastasis and pleural effusion were recorded one by one on chest CT. During the examination, the size of the tumor is measured by the diameter of the longest lung window. If it is multiple lesions, it is measured by the diameter of the largest lesion.

The film would be read independently by three doctors above medical level. If the readings differ, the three doctors would discuss and resolve the issue. The film reading content includes: The lung lobe where the disease is located, the density of the disease, and the signs of lobation, cavitation, folding, pleural indentation and lymph node enlargement.

Lesion density: This is divided into pure ground glass density and density with solid components.

Lymph node enlargement: In a defined area, lymph nodes with a short diameter>1 cm are enlarged.

5.3 Correlation between imaging characteristics of LUAD and EGFR gene mutation

5.3.1 Correlation between EGFR mutation and chest CT imaging features in LUAD

In order to thoroughly analyze the correlation between the imaging characteristics of LUAD and EGFR gene mutation, this paper studied the correlation between the mutation type, sex, smoking history and tumor size based on the EGFR mutation and chest CT imaging characteristics of LUAD, and recorded the results in Table 2.

In 164 patients, 42 were male patients with mutation type, and the mutation rate was 48.8 (42/86). 56 female patients were mutated, the mutation rate was 71.79% (56/78). There was a significance difference between the both of them (p = .012); 35 patients had no smoking history, and the mutation rate was 35.35% (35/99). 29 patients had smoking history, the mutation rate was 44.62% (29/65). There was a significance difference between the both of them (p = .024); there were 49 patients with tumors ≤ 3cm, and the mutation rate was 44.95% (49/109). 15 patients had tumors>3cm, and the mutation rate was 27.27% (15/55). There was no meaningful difference between them (p > .05); the number of patients with EGFR mutation type pulmonary nodules with partial solid morphology was 28, and the mutation rate was 36.84% (28/76). There were 36 patients with EGFR gene mutation whose pulmonary nodules were completely solid, and the mutation rate was 40.91% (36/88). There was a significance difference between the both of them (p = .024).

5.3.2 The relationship between EGFR gene mutation and CT findings in LUAD

Mutation of EGFR gene in lung adenocarcinoma lesions represents the symptoms of lung adenocarcinoma, which can analyze the relationship between lung adenocarcinoma symptoms and CT signs and provide reference for correlation analysis between imaging features of lung adenocarcinoma and EGFR gene mutation. The test results of the relationship between EGFR gene mutation and CT signs of LUAD lesions are summarized in Table 3.

From January 2020 to December 2021, among 164 patients with primary LUAD who underwent surgery in Hospital A, 81 patients showed lobulation sign, with a mutation rate of 37.04% (30/81). There was no meaningful difference (p > .05); 51 patients showed cavitation sign, and the mutation rate was 35.29% (18/51). The difference was statistically meaningful (p = .023); 73 patients showed spicule sign, and

10.3389/fgene.2022.1090180

TABLE 2 Correlation of EGFR mutations in LUAD with chest CT imaging features.

Group	Gender		Smoking history		Tumor size		Pulmonary nodule pattern	
	Male	Female	No	Yes	Pulmonary nodules≤3 cm	Pulmonary masses>3 cm	Partially solid	Completely solid
Genome group	42	56	35	29	49	15	28	36
Wild group	44	22	64	36	60	40	48	52
Total	86	78	99	65	109	55	76	88
Mutation rate/%	48.84	71.79	35.35	44.62	44.95	27.27	36.84	40.91
p	012		024		325		024	

TABLE 3 Relationship between EGFR mutations and CT signs in LUAD lesions.

Group	Lobar sign		Cavity sign		Burr sign		Pleural depression sign	
	No	Yes	No	Yes	No	Yes	No	Yes
Genome group	34	30	46	18	33	31	18	46
Wild group	49	51	67	33	58	42	51	49
Total	83	81	113	51	91	73	69	95
Mutation rate/%	40.96	37.04	40.71	35.29	36.26	42.47	26.09	48.42
p	316		023		034		005	

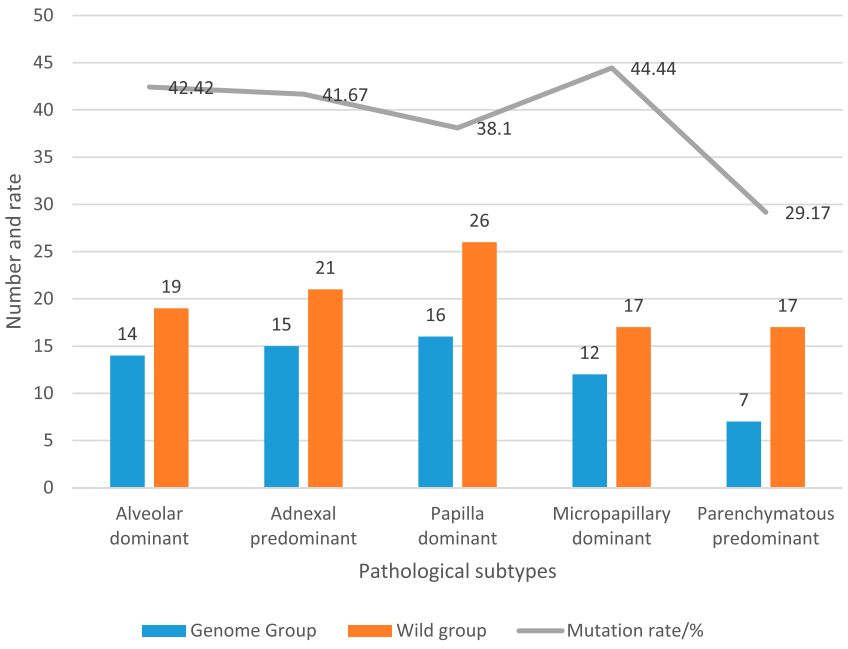


FIGURE 4
Pathological subtypes and EGFR mutation status in LUAD patients.

10.3389/fgene.2022.1090180

TABLE 4 CT presentation of adjacent structures and EGFR mutation status in LCA.

Parameters	Genome group	Wild group	p
Vascular cluster sign +	21	65	035
Vascular cluster sign-	43	35	
Pleural effusion+	27	21	35
Pleural effusion-	37	79	
Pericardial effusion+	33	48	65
Pericardial effusion-	31	52	
Thoracic thickening+	31	34	64
Thoracic thickening-	33	66	

The number of mutations in the tumor group with vascular convergence sign was 21, and the number of mutations in the group without vascular convergence sign was 43. There was a statistical difference between them ($p < 05$). In addition, pleural effusion, pericardial effusion, pleural thickening and other signs in tumor images were not significantly associated with EGFR, gene mutation ($p > 05$).

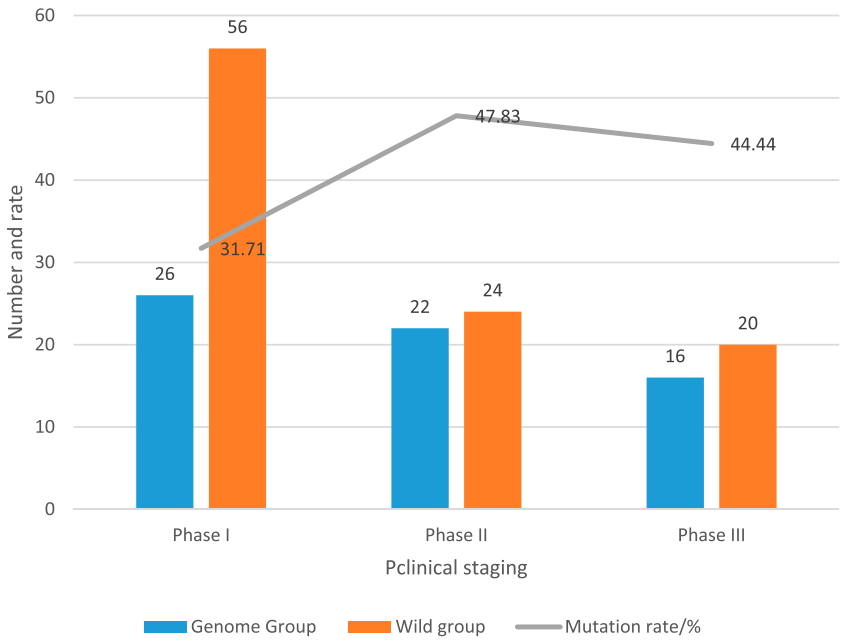


FIGURE 5
The relationship between EGFR gene mutations and clinical stages of LCA.

the mutation rate was 42.47% (31/73). The difference was statistically meaningful (p = .034); there were 95 patients with chest depression sign, and the mutation rate was 48.42% (46/95). The difference was statistically meaningful (p = .005).

5.3.3 Pathological subtypes and EGFR gene mutation in patients with LUAD

All cases of lung adenocarcinoma were classified into histological subtypes according to the international multidisciplinary classification of lung adenocarcinoma, including acinar dominated, adherent dominated, papillary dominated, micropapillary dominated and solid dominated. The determination results of pathological subtypes and EGFR gene mutation status of LUAD patients are recorded in Figure 4.

Among 164 patients with LUAD, 14 patients had acinar dominant pathological subtype, with a mutation rate of 42.42% (14/33); the pathological subtype of 15 patients was mainly adherent type, and the mutation rate was 41.67% (15/36); the pathological subtype of 16 patients was mainly nipple type, and the mutation rate was 38.1% (16/42); the pathological subtype of 12 patients was mainly micropapillary type, and the mutation rate was 41.38% (12/29); the pathological subtype of seven patients was mainly solid type, and the mutation rate was 29.17% (7/24).

5.3.4 CT manifestations of adjacent structures of LCA and mutation status of EGFR gene

LUAD belongs to non-small cell carcinoma. The growth of LCA cells would affect the changes in the structure of surrounding cells, so it

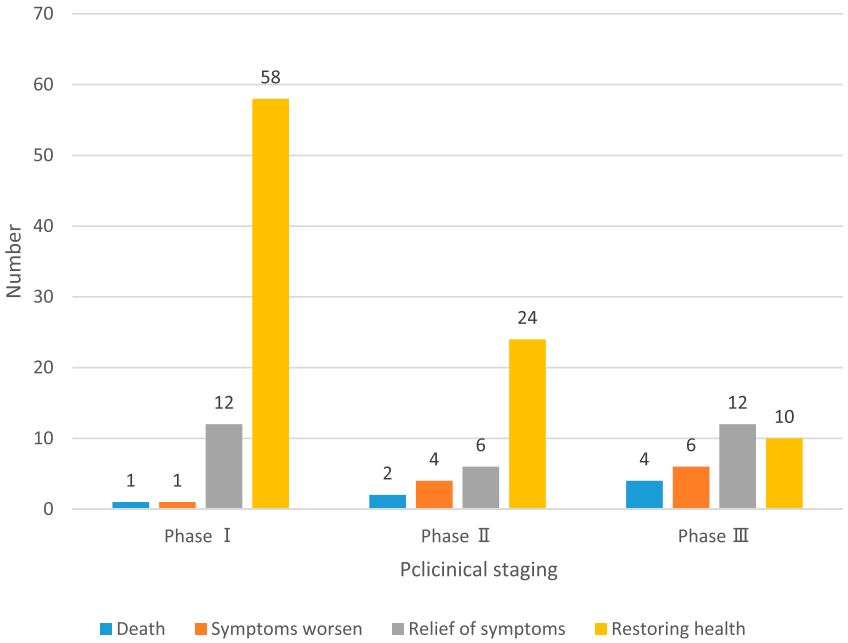


FIGURE 6
Analysis and cure of adenocarcinoma of the lung.

is of practical significance to investigate it. The CT findings of adjacent structures of LCA and the investigation results of EGFR gene mutation status are shown in Table 4.

5.3.5 The relationship between EGFR gene mutation and clinical stage of LCA in patients with LUAD

According to the symptoms and types of LUAD, the clinic can be divided into stage I, stage II and stage III. The relationship between EGFR gene mutation and clinical stage of LUAD patients is recorded in Figure 5.

26 patients with clinical stage Ⅰ had mutation, and the mutation rate was 31.71% (26/82); 22 patients in stage Ⅱ had mutations, and the mutation rate was 47.83% (22/46); there were 16 cases of mutation in stage III patients, and the mutation rate was 44.44% (16/36).

5.3.6 Analysis and cure of LUAD

In order to analyze the treatment of patients, 164 patients were interviewed and investigated to explore whether the image feature analysis of pulmonary adenocarcinoma based on artificial intelligence can cure the patients' pulmonary adenocarcinoma and promote their health. After excluding the patients who could not be contacted, the actual number of people in this survey was 140, and the physical conditions of these 140 people were recorded in Figure 6.

Among the patients in clinical stage I, 58 recovered, 12 were relieved, one was aggravated and one died; among the patients in clinical phase II, 24 recovered, six were relieved, four were aggravated and two died; among the patients in clinical phase III, 10 recovered, 12 were relieved, six were aggravated and four died. In general, the cure rate of patients is high, and only a few people died of ineffective treatment.

6 Conclusion

In order to analyze the correlation between image features of LUAD and EGFR gene mutation, and improve the cure rate of LUAD, this paper designed an artificial intelligence algorithm to analyze the correlation between image features of LUAD and EGFR gene mutation. It also designed a test to describe it, and finally reached a feasible conclusion. There were significant differences between the EGFR mutation rate and the sex, smoking history, pulmonary nodule morphology, patients with cavity sign, hair prick sign, thymus depression sign and tumor with vascular convergence sign groups of LUAD patients. There was no meaningful difference between tumor size, lobulated sign, pleural effusion, pericardial effusion, pleural thickening and EGFR mutation rate. According to the interview and investigation of patients, the cure rate of LUAD was greatly improved, which showed that AI could propose effective cure measures to improve the treatment plan of patients.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

GZ, SX, XL, and JG: Writing- original draft preparation. QH, WC, JD, and XK: Editing data curation, Supervision.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

AbdulJabbar, K., Raza, S. E. A., Rosenthal, R., Jamal-Hanjani, M., Veeriah, S., Akarca, A., et al. (2020). Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat. Med.* 267, 1054–1062. doi:10.1038/s41591-020-0900-x

Bianconi, F., Palumbo, I., Fravolini, M. L., Chiari, R., Minestrini, M., Brunese, L., et al. (2019). Texture analysis on [18F] FDG PET/CT in non-small-cell lung cancer: Correlations between PET features, CT features, and histological types. *Mol. imaging Biol.* 216, 1200–1209. doi:10.1007/s11307-019-01336-3

Cicek, T., Ozturk, A., Yilmaz, A., Aktas, Z., Demirag, F., and Akyurek, N. (2019). Adequacy of EBUS-TBNA specimen for mutation analysis of lung cancer. *Clin. Respir. J.* 132, 92–97. doi:10.1111/crj.12985

Ge, X., Li, G. Y., Jiang, L., Jia, L., Zhang, Z., Li, X., et al. (2019). Long noncoding RNA CAR10 promotes lung adenocarcinoma metastasis via miR-203/30/SNAI axis. *Oncogene* 3816, 3061–3076. doi:10.1038/s41388-018-0645-x

Gertych, A., Swiderska-Chadaj, Z., Ma, Z., Ing, N., Markiewicz, T., Cierniak, S., et al. (2019). Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci. Rep.* 91, 1483. doi:10.1038/s41598-018-37638-9

Gong, Jing., Liu, J., Hao, W., Nie, S., Zheng, B., Wang, S., et al. (2020). A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images. *Eur. Radiol.* 304, 1847–1855. doi:10.1007/s00330-019-06533-w

Jia, T-Y., Xiong, J. F., Li, X. Y., Yu, W., Xu, Z. Y., Cai, X. W., et al. (2019). Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. *Eur. Radiol.* 299, 4742–4750. doi:10.1007/s00330-019-06024-y

Koyasu, Sho., Nishio, M., Isoda, H., Nakamoto, Y., and Togashi, K. (2020). Usefulness of gradient tree boosting for predicting histological subtype and EGFR mutation status of non-small cell lung cancer on 18F FDG-PET/CT. *Ann. Nucl. Med.* 341, 49–57. doi:10.1007/s12149-019-01414-0

Li, Y., Lu, L., Xiao, M., Dercle, L., Huang, Y., Zhang, Z., et al. (2018). CT slice thickness and convolution kernel affect performance of a radiomic model for predicting EGFR status in non-small cell lung cancer: A preliminary study. *Sci. Rep.* 81, 17913. doi:10.1038/s41598-018-36421-0

Mei, D., Luo, Y., Wang, Y., and Gong, J. (2018). CT texture analysis of lung adenocarcinoma: Can radiomic features be surrogate biomarkers for EGFR mutation statuses. *Cancer Imaging* 181, 52–59. doi:10.1186/s40644-018-0184-2

Pascoe, H. M., Knipe, H. C., Pascoe, D., and Heinze, S. B. (2018). The many faces of lung adenocarcinoma: A pictorial essay. *J. Med. Imaging Radiat. Oncol.* 625, 654–661. doi:10.1111/1754-9485.12779

She, Y., Zhang, L., Zhu, H., Dai, C., Xie, D., Xie, H., et al. (2018). The predictive value of CT-based radiomics in differentiating indolent from invasive lung adenocarcinoma in patients with pulmonary nodules. *Eur. Radiol.* 2812, 5121–5128. doi:10.1007/s00330-018-5509-9

Tulchinsky, E., Demidov, O., Kriajevska, M., Barlev, N. A., and Imyanitov, E. (2019). Emt: A mechanism for escape from EGFR-targeted therapy in lung cancer. *Biochimica Biophysica Acta (BBA)-Reviews Cancer* 1, 29–39. doi:10.1016/j.bbcan.2018.10.003

Vyse, S., and Huang, P. H. (2019). Targeting EGFR exon 20 insertion mutations in non-small cell lung cancer. *Signal Transduct. Target. Ther.* 41, 5–10. doi:10.1038/s41392-019-0038-9

Wang, X., Zhao, X., Li, Q., Xia, W., Peng, Z., Zhang, R., et al. (2019). Can peritumoral radiomics increase the efficiency of the prediction for lymph node metastasis in clinical stage T1 lung adenocarcinoma on CT? *Eur. Radiol.* 2911, 6049–6058. doi:10.1007/s00330-019-06084-0

Wu, G., Woodruff, H. C., Shen, J., Refaee, T., Sanduleanu, S., Ibrahim, A., et al. (2020). Diagnosis of invasive lung adenocarcinoma based on chest CT radiomic features of part-solid pulmonary nodules: A multicenter study. *Radiology* 2972, E282–E458. doi:10.1148/radiol.2020209019

Yoneda, K., Imanishi, N., Ichiki, Y., and Tanaka, F. (2019). Treatment of non-small cell lung cancer with EGFR-mutations. *J. UOEH* 412, 153–163. doi:10.7888/juoeh.41.153

Yoon, J., Suh, Y. J., Han, K., Cho, H., Lee, H. J., Hur, J., et al. (2020). Utility of CT radiomics for prediction of PD-L1 expression in advanced lung adenocarcinomas. *Thorac. cancer* 114, 993–1004. doi:10.1111/1759-7714.13352

Yu, L., Tao, G., Zhu, L., Wang, G., Li, Z., Ye, J., et al. (2019). Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis. *BMC cancer* 191, 464. doi:10.1186/s12885-019-5646-9

Zhang, J., Zhao, X., Zhao, Y., Zhang, Z., Wang, J., and Han, J. (2020). Value of pre-therapy 18FFDG PET/CT radiomics in predicting EGFR mutation status in patients with non-small cell lung cancer. *Eur. J. Nucl. Med. Mol. imaging* 475, 1137–1146. doi:10.1007/s00259-019-04592-1

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1087273

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Cui Zhengyan,
Henan University of Animal Husbandry and Economy, China
Jun Zheng,
Baotou Teachers' College, China

*CORRESPONDENCE
Jian Zhao,
zhaojian@ccu.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 02 November 2022
ACCEPTED 01 December 2022
PUBLISHED 04 January 2023

CITATION
An T, Chen Y, Chen Y, Ma L, Wang J and Zhao J (2023), A machine learning-based approach to ERα bioactivity and drug ADMET prediction.
Front. Genet. 13:1087273.
doi: 10.3389/fgene.2022.1087273

A machine learning-based approach to ERα bioactivity and drug ADMET prediction

Tianbo An[1,2], Yueren Chen[1], Yefeng Chen[1], Leyu Ma[1], Jingrui Wang[1] and Jian Zhao[3]*

[1]College of Network Security, Changchun University, Changchun, Jilin, China, [2]Institute of Education, Xiamen University, Xiamen, Fujian, China, [3]College of Computer Science and Technology, Changchun University, Changchun, Jilin, China

By predicting ERα bioactivity and mining the potential relationship between Absorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) attributes in drug research and development, the development efficiency of specific drugs for breast cancer will be effectively improved and the misjudgment rate of R&D personnel will be reduced. The quantitative prediction model of ERα bioactivity and classification prediction model of Absorption, Distribution, Metabolism, Excretion, Toxicity properties were constructed. The prediction results of ERα bioactivity were compared by XGBoot, Light GBM, Random Forest and MLP neural network. Two models with high prediction accuracy were selected and fused to obtain ERα bioactivity prediction model from Mean absolute error (MAE), mean squared error (MSE) and R2. The data were further subjected to model-based feature selection and FDR/FPR-based feature selection, respectively, and the results were placed in a voting machine to obtain Absorption, Distribution, Metabolism, Excretion, Toxicity classification prediction model. In this study, 430 molecular descriptors were removed, and finally 20 molecular descriptors with the most significant effect on biological activity obtained by the dual feature screening combined optimization method were used to establish a compound molecular descriptor prediction model for ERα biological activity, and further classification and prediction of the Absorption, Distribution, Metabolism, Excretion, Toxicity properties of the drugs were made. Eighty variables were selected by the model ExtraTreesClassifier Classifie, and 40 variables were selected by the model GradientBoostingClassifier to complete the model-based feature selection. At the same time, the feature selection method based on FDR/FPR is also selected, and the three classification models obtained by the two methods are placed into the voting machine to obtain the final model. The experimental results showed that the model's evaluation indexes and roc diagram were excellent and could accurately predict ERα bioactivity and Absorption, Distribution, Metabolism, Excretion, Toxicity properties. The model constructed in this study has high accuracy, fast convergence and robustness, has a very high accuracy for Absorption, Distribution, Metabolism, Excretion, Toxicity and ERα classification prediction, has bright prospects in the biopharmaceutical field, and is an important method for energy conservation and yield increase in the future.

KEYWORDS
machine learning, ERα bioactivity, ADMET, breast cancer, drug development

1 Introduction

With the development of the times, more and more people began to pay attention to their own health problems. In today's society, breast cancer is one of the most common and lethal cancers. Estrogen receptors have a close correlation with the development of breast cancer, and according to related studies, estrogen receptor alpha (ERα) in estrogen has a strong promoting effect on the dominant characterization of breast cancer (Lempereur et al., 2016).

At present, the main treatment for breast cancer patients is to inhibit the expression of ERα gene using anti-hormone therapy, and then control the estrogen level in patients through the modulation of estrogen receptor activity, so as to achieve the inhibition of breast cancer spread and malignant trend, and gradually combine drugs and radiochemotherapy to achieve effective treatment of breast cancer (Ali and Coombes, 2002; Mohla et al., 2009; Huang et al., 2015). Drugs developed based on the corresponding compounds have been widely used in clinical treatment.

Drug R&D involves the discovery and development of new drugs, and the difference between these two stages is the determination of candidate drugs, candidate drugs represent active compounds involved in clinical research, so it is necessary to screen active compounds for drug research and development. Active compounds are compounds with certain biological or pharmacological activities obtained through various ways and methods, in the screening process, we should do research on biological activity, pharmacokinetics, toxicity analysis, etc. while synthesizing, so as to find the molecules needed for drug development. Then it is necessary to conduct tests related to biological activity and pharmacological data, that is, the quantitative structure activity relationship (QSAR) model (Bolboaca and Jäntschi, 1900; Singh et al., 2013; Ezugwu et al., 2021) of the compound. After the screening work is completed, in order to avoid the late risk of drug development (Samuel et al., 2021; Sun et al., 2022), also needs to verify whether it has ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) properties. In the compound database, the construction vector can select compounds with excellent ADMET properties and biological activities (Deng, 2013). Mining tacit knowledge in drug data and using machine learning prediction can reduce the R&D cost of pharmaceutical processes (Guo et al., 2022), and deep learning algorithms accelerate drug target recognition efficiency (Fenglei et al., 2021). Construction of prediction models for protein hotspot residues based on machine learning algorithms can assist drug development (Hu, 2019). Further mining the activity of drugs against tumor therapeutic targets and the sensitivity of tumor cell lines can efficiently develop novel tumor drugs (Li, 2021). The ADMET classification prediction model has good performance in predicting the properties of anti-breast cancer drugs (Yaqin et al., 2022). Scholars have found that some genes as well as core TFs can evaluate the efficacy of adjuvant therapy for breast cancer, of which E2F1 can regulate MAPK signaling pathways involved in pharmaceutical processes (Ye et al., 2022). Recognizing that AI is a very effective tool in disease assessment, patient data are collected through machine learning to develop mathematical models and predict outcomes (Suh and Lee, 2017; Fu et al., 2022; Zheng et al., 2022), a large number of researchers have investigated interpretable disease diagnostic models (Casteleiro-Roca et al., 2020; Tjoa and Guan, 2021; Xu et al., 2022).

Jiang and other scholars established an easy to understand OPLS-DA model based on only two descriptors, and the accuracy of the model was as high as 93% and 79% (Matsson et al., 2007; Jiang et al., 2020). Pan and other scholars trained 79, 99 and 780 compounds respectively, and achieved high accuracy (Jiang et al., 2020). Although good results have been achieved, there is still room for optimization in terms of descriptors and the number of compounds. Based on 1974 compounds (samples) and 729 molecular descriptors (variables) provided by candidate compound data set, this paper solves the problem of low reliability caused by previous model development based on relatively small data set. In addition, the application value of the existing ADMET model in specific drug screening work is still unclear, based on this, this study, first, a dual feature screening combination method is proposed to screen data variables; then, the combination of Bagging algorithm and random forest algorithm is used to realize ERα' Bioactivity prediction; finally, different target values are introduced to test the performance of the prediction model, so as to improve the development efficiency of breast cancer specific drugs, as well as reducing the error judgment rate of R&D personnel.

2 Materials and methods

2.1 Data acquisition

The main ways to obtain drug data include laboratories, online public publications, biochemical databases, and so on. In this paper, the optimized modeling (2021) dataset of anti-breast cancer drug candidates provided by the China Association for Science and Technology is used as the candidate compound dataset.

Taking anti-breast cancer drugs as an example to test the prediction framework for the following reasons: First, breast cancer is the largest cancer in the world, and breast cancer patients in China account for a relatively large proportion;

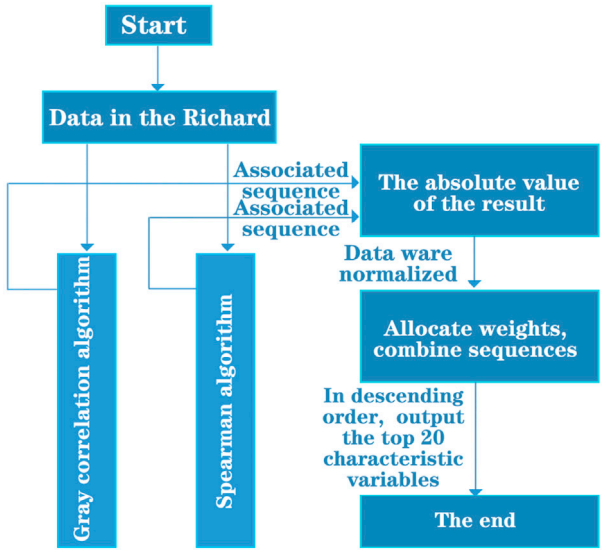


FIGURE 1
Flow chart of combined optimization method for dual feature screening.

second, researchers have accumulated a large number of R&D data of anti-breast cancer drugs over the years, laying the foundation for further machine-learning-based prediction research; third, predecessors have left sufficient literature data on breast cancer characteristic targets, which plays a guiding role in the establishment of the validation prediction framework.

2.2 Data processing

According to the data of 1974 compounds (samples) and 729 molecular descriptors (variables) provided by the candidate compound dataset, 20 main variables were screened, which made the selected main variables have a good effect on ERα biological activity. Data preprocessing is performed first, and column variables (molecular descriptors) are cleaned. Different screening algorithms were used to analyze the distribution characteristics of the cleaned data. In order to obtain more accurate screening results, gray relational degree method and Spearman rank correlation coefficient analysis are selected to reduce the dimension of the data and data screening, and the two are further combined analysis, and a dual feature screening combined optimization method is proposed. The specific analysis is divided into k steps, and the flow chart is shown in Figure 1.

Step 1. Relevance calculation. The data were processed using Algorithm I (grey correlation analysis) and Algorithm III (Spearman rank correlation analysis) to obtain the corresponding correlation coefficient sequences $X(i)$ and $Y(j)$.

Grey correlation analysis method, which takes the difference value between curves as the correlation degree. There are several comparison sequences $(x_1, x_2, \ldots, x_n)$ for a reference sequence x_0, correlation coefficient of each comparison sequence and reference sequence at each time (i.e., each point in the curve) $\beta(x_i)$ can be calculated by the following formula: ρ is the resolution coefficient, generally between 0 and 1, usually taken as 0.5. Δmin represents the second level minimum difference, Δmax represents the maximum difference between two levels. $\Delta_{oi}(k)$ represents the absolute difference between each point on the comparison series x_i curve and each point on the reference series x_0 curve. Therefore, the correlation coefficient $\beta(x_i)$ can also be simplified as follows:

$$\beta_{oi} = \frac{\Delta(\min) + \rho\Delta(\max)}{\Delta_{oi}(k) + \rho\Delta(\max)} \tag{1}$$

Because the correlation coefficient is the value of the correlation degree between the comparison sequence and the reference sequence at each time (i.e., each point in the curve), it has more than one number, and the information is too scattered to facilitate the overall comparison. Therefore, it is necessary to centralize the correlation coefficient of each time (i.e., each point in the curve) into one value, that is, to calculate its average value, as a quantitative expression of the correlation degree between the comparison sequence and the reference sequence, the formula of correlation degree c_i is as follows:

$$c_i = \frac{1}{N}\sum_{k=1}^{N}\beta_i(k) \tag{2}$$

Spearman rank correlation coefficient is generally considered as Pearson linear correlation coefficient between ranked variables, in actual calculation, there are simpler calculation methods. Assume that the original data x_i, y_i has been arranged from large to small, put x'_i, y'_i as the location of the original x_i, y_i data after arrangement, then $d_i = x'_i - y'_i$ represents the difference of rank between x_i, y_i.

If there is no same rank, Spearman rank correlation coefficient can be expressed as:

$$spc = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{3}$$

If the same rank exists, it is necessary to calculate Pearson's linear correlation coefficient between ranks:

$$spc = \frac{\sum_i (x_i - x')(y_i - y')}{\sqrt{\sum_i (x_i - x')^2 \sum_i (y_i - y')^2}} \tag{4}$$

Step 2. Take absolute value. Because the correlation degree derived by the two algorithms may be positive or negative, but in fact the correlation degree is similar to the concept of "distance,"

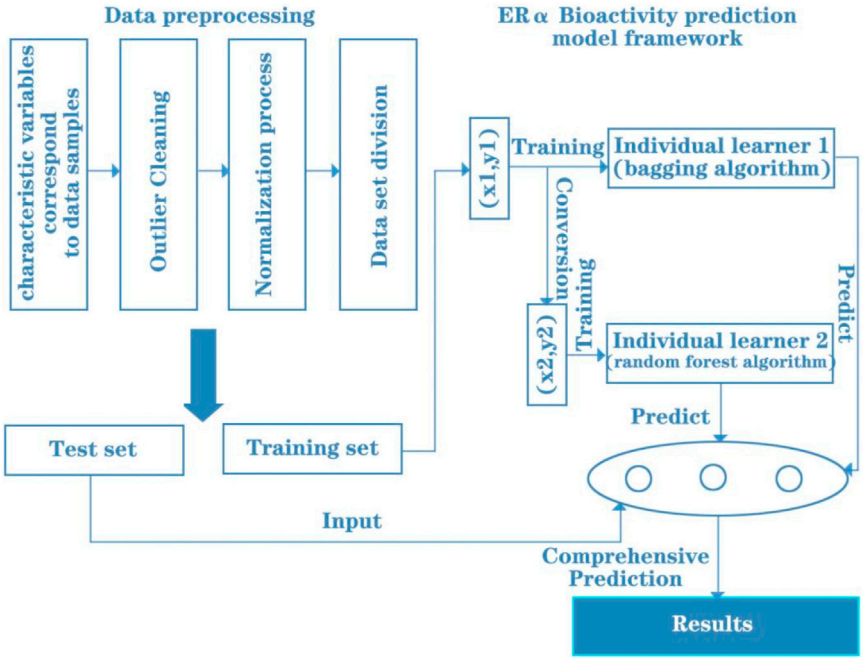


FIGURE 2
Framework of ERα bioactivity prediction model.

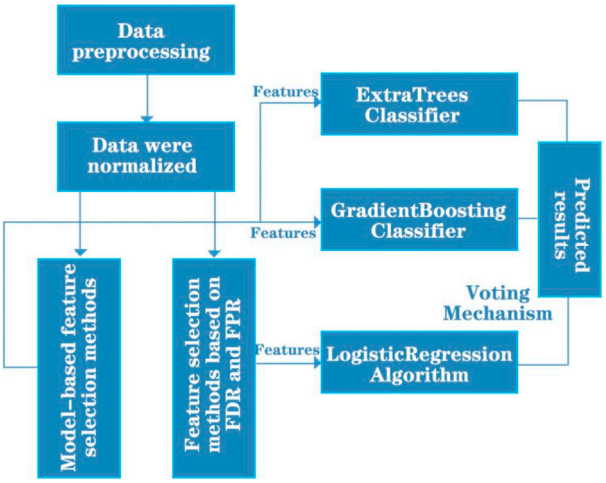


FIGURE 3
Schematic diagram of the classification prediction model framework.

neither positive correlation nor negative correlation affects the judgment of the correlation size. Therefore, the correlation degree is taken as the absolute value to more clearly represent the correlation between variables. The formula is shown below.

TABLE 1 Evaluation metrics for regression problems.

Full name	Abbreviations	Meaning
Mean squared error	MSE	Mean square error
Mear absolute error	MAE	Mean absolute error
Coefficient of determination	R^2	Decidability factor

$$X(i) = |X(i)| \\ Y(j) = |Y(j)| \tag{5}$$

Step 3. Data normalisation. As two different correlation analysis methods are used in this paper, the correlation values $X(i)$ and $Y(j)$ derived from the two algorithms need to be normalised in order to obtain the final screening results. The normalisation algorithm used in this paper is (0, 1) normalisation and the formula is shown below.

$$x_{normalization} = \frac{x - Min}{Max - Min} \tag{6}$$

Step 4. Assignment of weights. By comparing the magnitude and distribution of correlation under the two analysis methods, it is approximated that the screening process of the data set by Algorithm 1 (grey correlation analysis method) can obtain a

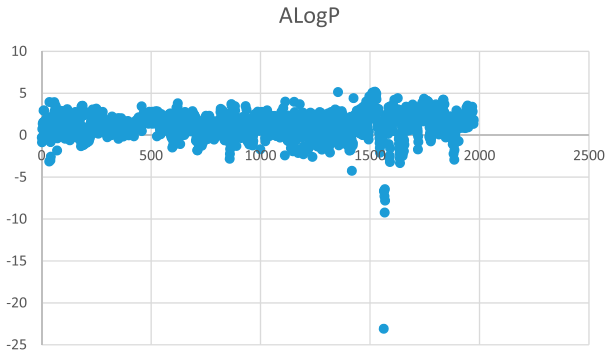


FIGURE 4
Scatter plot of ALogP.

higher correlation value. Therefore, a weight of 0.6 was assigned to sequence $X(i)$ and a weight of 0.4 to sequence $Y(j)$. By assigning weights, the final correlation degree sequence is obtained. The calculation formula is shown below.

$$Z(k) = 0.6X(k) + 0.4Y(k) \quad (7)$$

Step 5. Sorting. The normalised variable data is sorted according to the magnitude of the correlation values to produce the top 20 most significant sub-descriptors for biological activity.

2.3 Data modelling

The 20 main variables (i.e., molecular descriptors) obtained from data processing were used to model the prediction of ERα biological activity by molecular descriptors of compounds. Data standardisation was first performed to achieve uniformity of magnitude. The dataset was then partitioned by K-fold cross-validation to achieve adequate use of the dataset to fit the prediction model. Finally, integrated learning is used to combine the Bagging and Random Forest algorithms through the Stacking method to build the entire prediction model framework. In summary, the immediate ERα bioactivity prediction model framework is shown in Figure 2.

In order to make further classification predictions of the ADMET properties of the drug, the original 729 molecular descriptor variables were processed for data in this paper, and a total of 300 molecular descriptor variables were retained for data analysis. After normalizing the data, the data were subjected to model-based feature selection and FDR/FPR-based feature selection, respectively. After the two model selections, the classification models obtained from the two methods were placed into a voting machine to obtain the final model, and the entire prediction model framework was constructed. In summary, the classification prediction model framework for the five specific compound classes is shown in Figure 3.

TABLE 2 The 20 most significant sub-descriptors (i.e., variables) for biological activity.

Molecular descriptors	Grey correlation	Spearman's rank correlation coefficient	Normalised and weighted score values
MDEC-23	0.7997	0.5491	0.9999
MLogP	0.7943	0.5452	0.9872
LipoaffinityIndex	0.7954	0.5249	0.9707
CrippenLogP	0.7861	0.4738	0.9081
SwHBa	0.7895	0.4457	0.8883
nC	0.7641	0.4868	0.8821
nT6Ring	0.7809	0.4383	0.8668
n6Ring	0.7861	0.4281	0.8665
BCUTp-1h	0.7821	0.4329	0.8639
SP-5	0.7764	0.4427	0.8631
C2SP2	0.7686	0.4430	0.8499
SP-6	0.7775	0.4201	0.8444
ATSp4	0.7642	0.4444	0.8437
ATSp2	0.7601	0.4443	0.8365
ATSp5	0.7537	0.4514	0.8321
ATSp3	0.7648	0.4296	0.8312
maxsOH	0.7460	0.4619	0.8284
ATSp1	0.7550	0.4439	0.8274
nHaaCH	0.7632	0.4226	0.8221
naaCH	0.7632	0.4226	0.8221

2.4 Evaluation of the ERα bioactivity prediction model

For the ERα bioactivity prediction model, three evaluation metrics shown in Table 1 are used in this paper.

The mean squared error (MSE) reflects the degree of correlation between the independent and dependent variables; the MSE evaluates the degree of variation in the data.

Mean absolute error (MAE) is the average of the absolute errors and is often used to reflect the reality of the error in the predicted values.

TABLE 3 Descriptions of null-valued columns (e.g., nB).

Count	1974
Mean	0
Std	0
Min	0
25%	0
50%	0
75%	0
Max	0

TABLE 4 Evaluation data of the dual feature screening combination optimisation method.

Weighted score	MAE	MSE	R^2
MLPRegressor	1.3723	3.2465	−0.6000
GradientBoostingRegressor	0.6828	0.8280	0.5900
RandomForestRegressor	0.6104	0.6873	0.6607
AdaBoostRegressor	0.8500	1.1100	0.4500

This coefficient is often used to reflect the degree of reliability of changes in the dependent variable in a regression model. The higher the value, the better the predicted value fits the true value.

2.5 Evaluation of classification prediction models for ADMET properties of compounds

The experimental data was designed to reduce the rate of misclassification by drug developers, in order to ensure robustness. In this paper, five performance comparison metrics were selected: accuracy, precision, recall, F1 value, Cohen's Kappa coefficient and ExtraTreesClassifier accuracy. The formulae for each evaluation metric are shown below.

The accuracy rate is calculated as shown in Table 2.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where TP is a positive example judged to be positive, FP is a negative example judged to be positive, TN is a negative example judged to be negative and FN is a positive example judged to be negative.

The accuracy rate is calculated by the formula:

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{9}$$

The recall is calculated as:

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{10}$$

The F1 value is calculated using the formula:

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{11}$$

The Cohen's Kappa coefficient is calculated as:

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{12}$$

Where, p_0 represents the observed compliance rate and p_e represents the opportunity compliance rate.

2.6 ROC curve

Horizontal axis FPR: 1-TNR, 1-Specificity. The larger the FPR, the more actual negative classes in the predicted positive classes. Vertical axis TPR: Sensitivity. The larger the TPR, the more actual positive classes in the predicted positive classes. The closer the ROC curve is to the (0, 1) point, the more it deviates from the 45° diagonal, the better.

3 Results

3.1 Dataset description

The optimized modelling of anti-breast cancer drug candidates (2021) dataset provided by CCSA was used. The 729 molecular descriptors of 1974 compounds in it were used as dependent variables to find the molecular descriptors among them that could significantly affect ERα activity as feature variables for subsequent questions. Inevitably, there are some anomalies in the data used as dependent variables, which will interfere with the selection of the characteristic variables and have a collateral negative impact on the solution of the subsequent problem. For example, a large proportion of the numerical columns of the molecular descriptors have null (0) or almost null (0) values.

If there are too many null values, the data reliability of the variable will be low, and it is considered that the molecular descriptors corresponding to these data are unlikely to become feature variables and will waste arithmetic power in the subsequent screening of the feature vectors. Therefore, in this step, the numerator descriptors with 95% of the data items being null are eliminated. The data visualisation also revealed some outliers in the data, and a scatter plot of one of the columns is shown in Figure 4.

The raw data was normalised and after noise and dimensionality reduction to form the experimental dataset. Using the KS test, the study found that the data samples did not satisfy a normal distribution, which in turn led to the use of an outlier treatment based on box plot analysis.

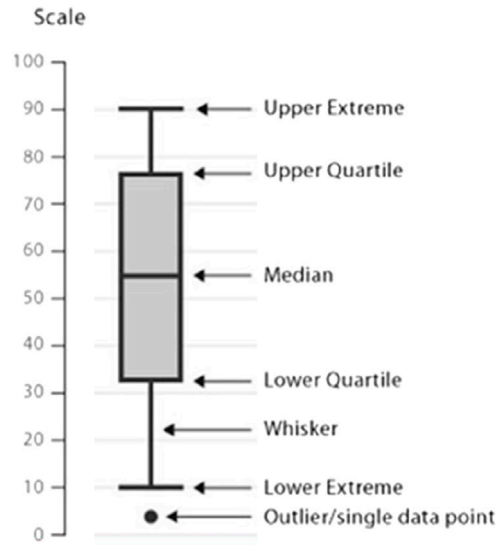


FIGURE 5
Schematic diagram of the box line diagram.

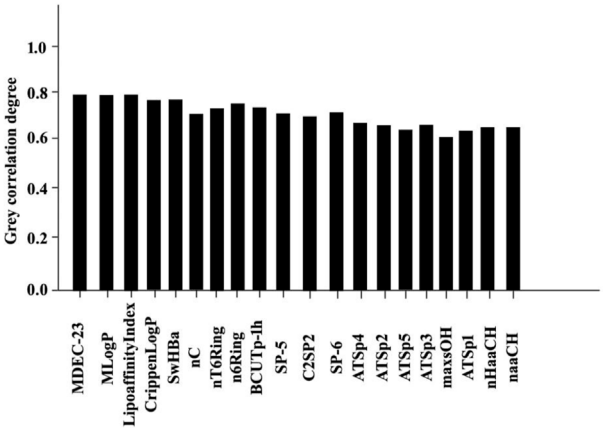


FIGURE 6
Statistical results of 20 bioactive molecules based on Grey correlation analysis.

TABLE 5 Evaluation metric values for each algorithm.

	MAE	MSE	R^2
GradientBoostingRegressor	0.6732	0.7927	0.6087
RandomForestRegressor	0.6180	0.6752	0.6600
Stacking	0.5940	0.6436	0.7002

After sorting each column of data from smallest to largest, the interval between the values of the first and third quartiles is taken as the acceptable range, where the upper bound is the third quartile—IQR (IQR = third quartile–first quartile) and the lower bound is the first quartile—IQR, and numbers outside this value area are considered outliers. These outliers are taken to correspond to the element values, with outliers over the upper bound being taken to the upper bound and those over the lower bound being taken to the lower bound, otherwise they remain as they are. Then go back to Step 1 and clear the column with the higher number of null values. As shown in Figure 5.

Through the above data processing, a total of 430 molecular descriptors were removed. The final 20 molecular descriptors (i.e., variables) with the most significant impact on bioactivity as a result of the dual feature screening combined optimisation method are shown in Table 3.

The results of the Spearman rank correlation coefficient processing for this study are shown in Figures 6, 7

3.2 Data set variable screening combination optimisation

The 20 features derived from the dual feature screening combination optimisation method, Spearman's algorithm and the grey correlation algorithm were substituted into the correlation regression model and the MAE, MSE and R^2 were used to evaluate the advantages and disadvantages of the three methods. The table below shows the results of the above experiments. From the data in the table, it is analysed that the dual feature screening combination optimisation method used in this paper has better results. As shown in Table 4.

3.3 Model performance

3.3.1 ERα bioactivity prediction model evaluation

By analysing the various algorithm evaluation metrics, we found that the MLP performed poorly regardless of the metrics, and the r^2 value of AdaBoostRegressor was consistently below 0.5, so it was also out of our selection range. The final model is the result of fusing the random forest and GradientBoostingRegressor after tuning the parameters through a grid search. As shown in Figure 8.

The results in Table 5 show that the model after stacking performs better than the original two models in all three scoring metrics, and also that this model is realistic in the field of machine learning when the R^2 value is greater than 0.7 and the fitted function is more realistic.

3.3.2 Evaluation of classification prediction models for the ADMET properties of compounds

For the classification prediction models of the ADMET properties of compounds, the results of the evaluation of the five target value related classification models are given below. It

TABLE 6 Evaluation of classification models for each algorithm with CYP3A4 as the target value.

Algorithms	Accuracy	Accuracy	Recall rate	F1 value	Cohen's Kappa coefficient
LogisticRegression	0.9367	0.9626	0.9529	0.9577	0.8321
ExtraTreesClassifier	0.9367	0.9564	0.9596	0.9580	0.8298
RandomForestClassifier	0.9468	0.9600	0.9697	0.9648	0.8560
Integrated learning models based on Stacking methods	0.9538	0.9668	0.9581	0.9634	0.8816

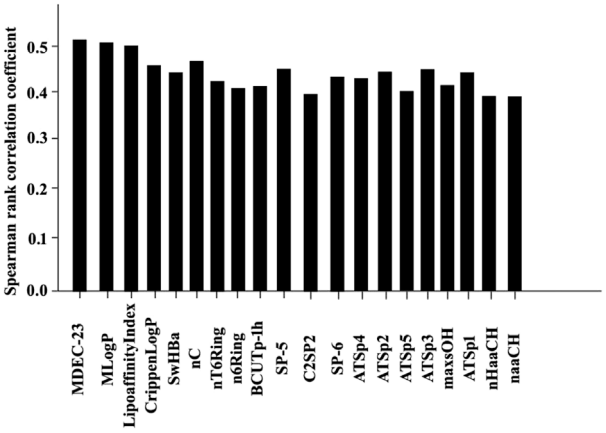


FIGURE 7
Statistical results of 20 bioactive molecules based on Spearman rank correlation coefficient analysis.

can be seen that in most cases, the model set up in this paper prevails. Of course, the effect is not significant due to the inherently low scores of the initial models. As shown in Table 6.

In the following results, the roc diagram about the two toxicity indicators is given. Figures 9, 10, it can be seen that the classification effect for these two indicators is good.

The resulting three classification models were placed into the voting machine, which is the final model. Five final models were obtained on five validation sets with accuracy: 0.92,0.95,0.91,0.76,0.96; F1 values: 0.91,0.96,0.92,0.80,0.97 respectively.

4 Discussion

In today's society, breast cancer is the most common and lethal cancer and is the leading cause of cancer death in women (Desantis et al., 2015), and related studies have confirmed that breast cancer is genetically risky (Lilyquist et al., 2018). The expression of estrogen receptor ERα is closely related to the development of breast cancer, and this gene is involved in the proliferation and differentiation of breast cancer cells (Geng, 2016). Currently ERα is considered an important target for the treatment of breast cancer (Xu, 2018). Researchers have found that the ADMET module ensures that the corresponding compound becomes a candidate with good pharmacokinetic properties and safety (Dejun et al., 2018). Thus, machine learning can effectively reduce the cost of drug research and development and improve the stability and accuracy of prediction models (Wang et al., 2006). Scholars have determined that the optimal cut-off values for E2 and FSH in serum can assess CIA in breast cancer populations in southern China (Yang et al., 2022). Estrogen receptors alpha (ERα) in estrogen has been found to be an important target for the treatment of breast cancer (Chang et al., 2013).

We collected data on anti-breast cancer drug candidates provided by the China Association for Science and Technology, which provided 974 compounds (samples) and 729 molecular descriptors (variables) information. The dataset is therefore highly representative and generalizable. In this study, random forest algorithm and Bagging regression algorithm were used to construct a quantitative prediction model of ERα bioactivity by compounds, and two model selection features, ExtraTreesClassifier and GradientBoostingClassifier, were selected. According to the actual situation, taking into account the properties of target value, fdr-based method is adopted to improve F1 value in feature selection. At the same time, the feature selection method based on model also has high accuracy. Both vote to ensure the accuracy of the final model.

It is shown that the weighted scores obtained by grey correlation and Spearman coefficient are more suitable for the data distribution characteristics of anti-breast cancer candidates. In this study, 20 molecular descriptors with strong antagonistic effects were selected from 729 molecular descriptors. These are MDEC-23, MLogP, LipoaffinityIndex, etc. Random forest algorithm and Bagging regression algorithm were selected to construct the model, and then Stacking model fusion method was used to establish the prediction model framework. The model showed good performance in MAE, MSE and three scoring indexes after Stacking, and the predicted result value of the model on the validation set was 0.7, which proved that the model could accurately predict. According to the classification prediction model of ADMET properties of compounds, two feature selection methods were used to screen the features, and three classification models obtained by the two methods were placed into the voting machine to obtain high accuracy and F1 values. Therefore, the model established in this study is of great significance in assisting prediction of ERα biological activity and improving the development efficiency of specific drugs for breast cancer.

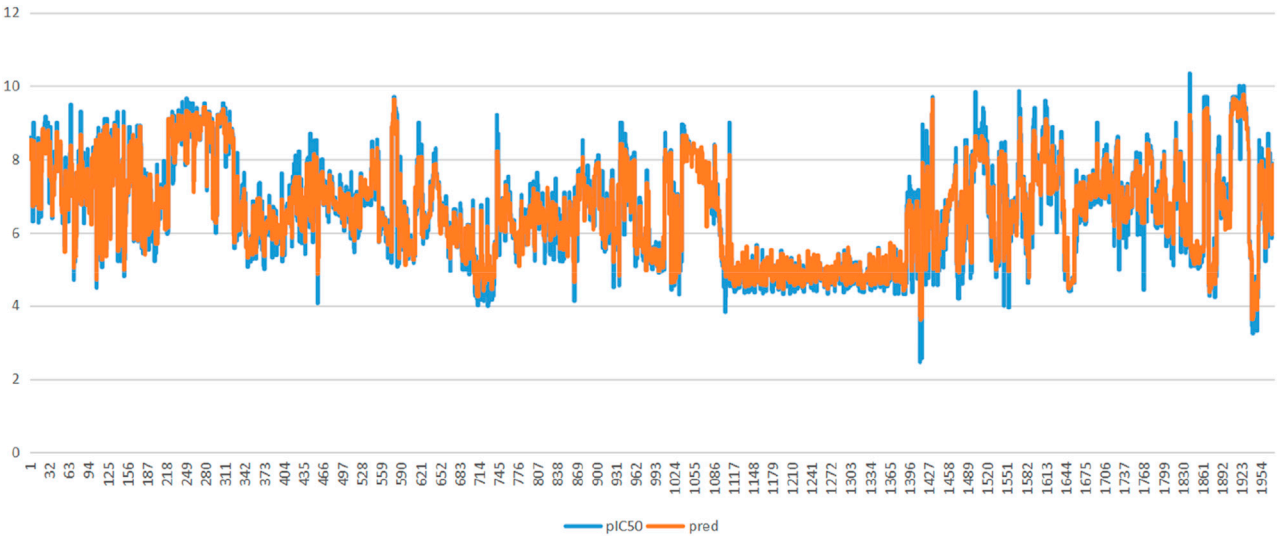


FIGURE 8
Feature variables and pIC50 fit results.

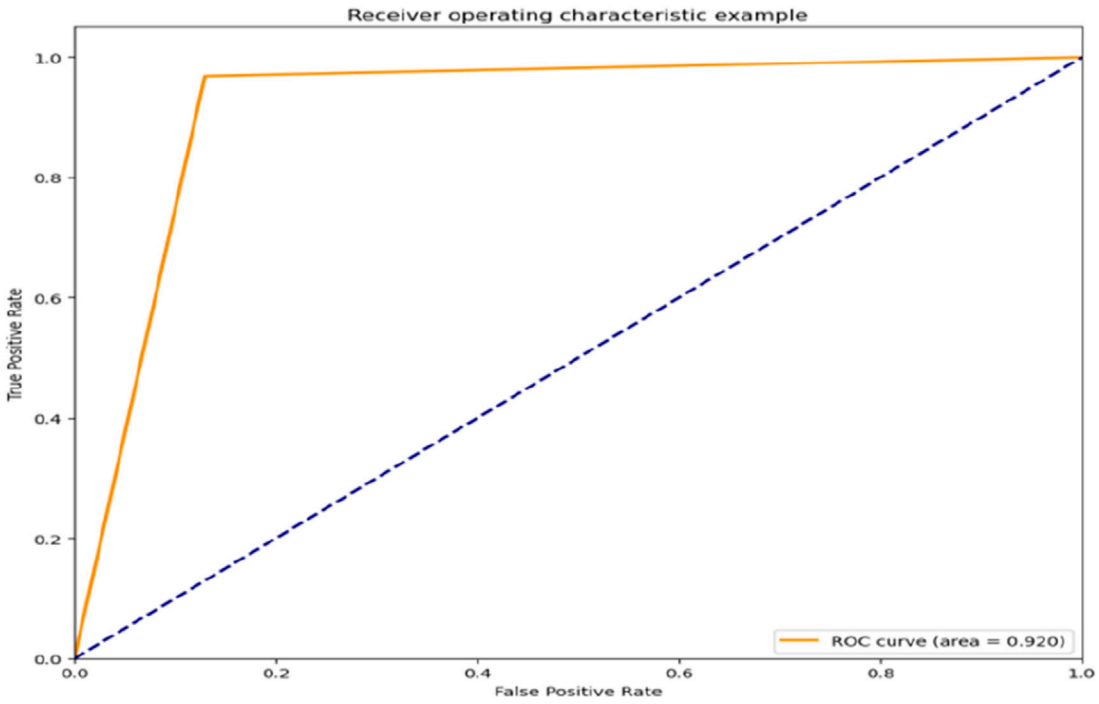


FIGURE 9
Classification model roc curve with hERG as the target value.

Despite the good results, there are limitations in this study. In future anti-breast cancer drug screening efforts, this study should attempt to collect more datasets for machine learning training, draw on expert opinion, and continuously optimise the feature generation tools to improve the accuracy and stability of the model. Through the interdisciplinary

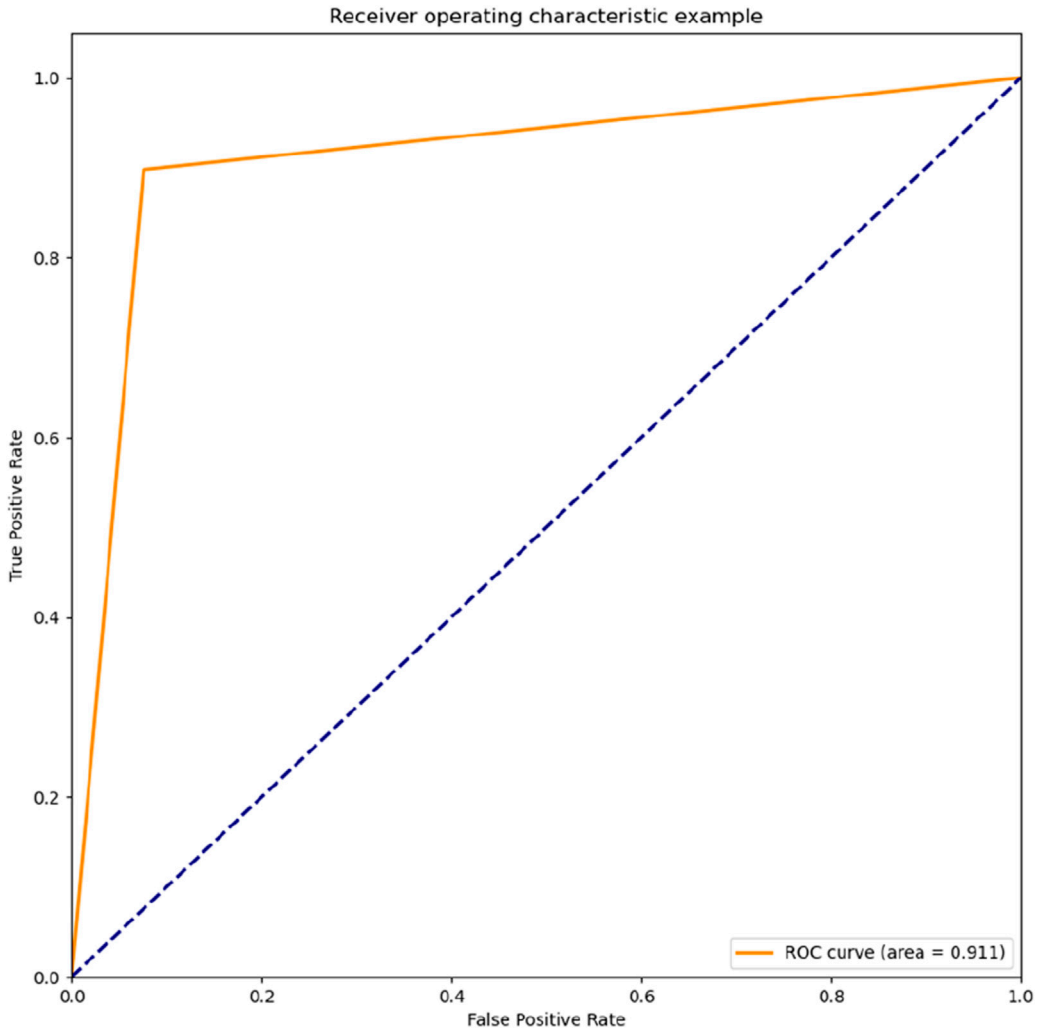


FIGURE 10
Classification model roc curve with MN as the target value.

collaboration between machine learning and biopharmaceuticals, we can reduce the cost of pharmaceuticals, reduce the error rate of drug developers, and provide a reference for similar drug development work.

5 Conclusion

A quantitative prediction model for ERα bioactivity and a classification prediction model for ADMET properties of compounds were developed, which can assist in the development of specific drugs for breast cancer. With the high accuracy of the models, the cost of drug development and the rate of misclassification by developers can be effectively reduced.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

TA contributed to the conception of the study; YuC performed the experiment; YeC contributed significantly to analysis and manuscript preparation; LM and JW performed the data analyses and wrote the manuscript; JZ helped perform the analysis with constructive discussions.

Funding

Key R&D Plan of Changchun Science and Technology Bureau (21ZGM29), Jilin Province Disabled Accessibility Technology and Intelligent Device Innovation Team (20200301054RQ), Jilin Provincial Key Laboratory for Human Health Status Identification and Function Enhancement (20200601004JC), and Science and Technology Development Plan Project of Jilin Provincial Science and Technology Department (20200404207YY).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1087273/full#supplementary-material

References

Ali, S., and Coombes, R. C. (2002). Endocrine-responsive breast cancer and strategies for combating resistance. *Nat. Rev. Cancer* 2 (2), 101–112. doi:10.1038/nrc721

Bolboaca, S. D., and Jäntschi, L. (1900). Comparison of quantitative structure-activity relationship model performances on carboquinone derivatives. *Sci. Worl. J.* 9, 1148–1166. doi:10.1100/tsw.2009.131

Casteleiro-Roca, J. L., Jove, E., Gonzalez-Cava, J. M., Mendez Perez, J. A., Calvo-Rolle, J. L., and Blanco Alvarez, F. (2020). Hybrid model for the ANI index prediction using Remifentanil drug and EMG signal. *Neural comput. Appl.* 32, 1249–1258. doi:10.1007/s00521-018-3605-z

Chang, Y. H., Chen, J. Y., Hor, C. Y., Chuang, Y. C., Yang, C. B., and Yang, C. N. (2013). Computational study of estrogen receptor-alpha antagonist with three-dimensional quantitative structure-activity relationship, support vector regression, and linear regression methods. *Int. J. Med. Chem.* 2013, 743139. doi:10.1155/2013/743139

Dejun, Y., Xiangcao, Y., Chongyuan, X., Shilin, Z., and Huang (2018). Molecular docking of uric acid-lowering activity and ADMET properties of small molecule compounds from red fennel. *Chin. J. Clin. Pharmacol.* 34 (23), 2750–2752+2777. doi:10.13699/j.cnki.1001-6821.2018.23.019

Deng, Z. L. (2013). *Prediction of compound activity based on biorelation spectra and its web service implementation*. Wuhan, China: Huazhong Agricultural University. Available at: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201401&filename=1013336348.nh&v=.

Desantis, C. E., Bray, F., Ferlay, J., Lortet-Tieulent, J., Anderson, B. O., and Jemal, A. (2015). International variation in female breast cancer incidence and mortality rates. *Cancer Epidemiol. Biomarkers Prev.* 24 (10), 1495–1506. doi:10.1158/1055-9965.EPI-15-0535

Ezugwu, A. E., Shukla, A. K., Agbaje, M. B., Oyelade, O. N., Jose-Garcia, A., and Agushaka, J. O. (2021). Automatic clustering algorithms: A systematic review and bibliometric analysis of relevant literature. *Neural comput. Appl.* 33, 6247–6306. doi:10.1007/s00521-020-05395-4

Fenglei, L., Qiaoyu, H., Ruofan, X., and Fang, B. (2021). A deep learning-based approach to drug design. *J. Nat.* 43 (5), 383–390.

Fu, D., Mo, K., Deng, W., Zhao, Y., Ding, Q., Hong, S., et al. (2022). Application value of machine learning method in measuring gray matter volume of AIDS patients. *Dis. Markers* 2022, 1210002. doi:10.1155/2022/1210002

Geng, C. (2016). *Mechanisms of curcumin promotion of tamoxifen sensitivity in ER alpha-negative breast cancer*. Jinan, China: Shandong University. Available at: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2016&filename=1016203756.nh&v=.

Guo, Y., Luo, M., and Xing, R. (2022). A knowledge discovery-oriented approach to drug ADMET intelligence prediction. *Intell. Sci.*, 1–14.

Hu, S. (2019). *Research on mining and prediction of drug interactions data*. Hefei, China: Anhui University. Available at: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2019&filename=1019130317.nh&v=.

Huang, B., Warner, M., and Gustafsson, J. Å. (2015). Estrogen receptors in breast carcinogenesis and endocrine therapy. *Mol. Cell. Endocrinol.* 418, 240–244. doi:10.1016/j.mce.2014.11.015

Jiang, D., Lei, T., Wang, Z., Shen, C., Cao, D., and Hou, T. (2020). ADMET evaluation in drug discovery. 20. Prediction of breast cancer resistance protein inhibition through machine learning. *J. Cheminform.* 12 (12), 16. doi:10.1186/s13321-020-00421-y

Lempereur, M., Majewska, C., Brunquers, A., Wongpramud, S., Valet, B., Janssens, P., et al. (2016). Tetrahydro-iso-alpha acids antagonize estrogen receptor alpha activity in MCF-7 breast cancer cells. *Int. J. Endocrinol.* 2016, 9747863. doi:10.1155/2016/9747863

Li, X. (2021). *Research on the mechanism of action of antitumor drugs based on deep learning*. University of Chinese Academy of Sciences, Shanghai Institute of Pharmaceutical Sciences, Chinese Academy of Sciences. Available at: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2021&filename=1021609355.nh&v=.

Lilyquist, J., Ruddy, K. J., Vachon, C. M., and Couch, F. J. (2018). Common genetic variation and breast cancer risk—past, present, and future. *Cancer Epidemiol. Biomarkers Prev.* 27 (4), 380–394. doi:10.1158/1055-9965.EPI-17-1144

Matsson, P., Englund, G., Ahlin, G., Bergström, C. A. S., Norinder, U., and Artursson, P. (2007). A global drug inhibition pattern for the human ATP-binding cassette transporter breast cancer resistance protein (ABCG2). *J. Pharmacol. Exp. Ther.* 323, 19–30. doi:10.1124/jpet.107.124768

Mohla, S., Stearns, V., Sathyamoorthy, N., Rosenfeld, M. G., and Nelson, P. (2009). The biology of hormone refractory breast and prostate cancer: An NCI workshop report. *Cancer Biol. Ther.* 8 (21), 1975–1985. doi:10.4161/cbt.8.21.9918

Samuel, Y., Garg, A., and Mulugeta, E. (2021). Synthesis, DFT analysis, and evaluation of antibacterial and antioxidant activities of sulfathiazole derivatives combined with in silico molecular docking and ADMET predictions. *Biochem. Res. Int.* 2021, 7534561. doi:10.1155/2021/7534561

Singh, S., Das, S., Pandey, A., Paliwal, S., and Singh, R. (2013). Quantitative structure activity relationship studies of topoisomerase I inhibitors as potent antibreast cancer agents. *J. Chem.* 2013, 1–9. doi:10.1155/2013/849793

Suh, K. H., and Lee, E. C. (2017). Contactless physiological signals extraction based on skin color magnification. *J. Electron. Imaging* 26 (6), 1. doi:10.1117/1.jei.26.6.063003

Sun, J., Liu, B., Wang, R., Yuan, Y., Wang, J., and Zhang, L. (2022). Computation-based discovery of potential targets for rheumatoid arthritis and related molecular screening and mechanism analysis of traditional Chinese medicine. *Dis. Markers* 2022, 1905077. doi:10.1155/2022/1905077

Tjoa, E., and Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Trans. Neural Netw. Learn. Syst.* 32 (11), 4793–4813. doi:10.1109/TNNLS.2020.3027314

Wang, Z., Rao, H., and Li, Z. (2006). Machine learning approach for the prediction model of selective cyclooxygenase-2 inhibitor activity. *Chem. Res. Appl.* (11), 1317–1321.

Xu, H., Zheng, Q., Zhu, J., Xie, Z., Cheng, H., Li, P., et al. (2022). A deep learning model incorporating knowledge representation vectors and its application in diabetes prediction. *Dis. Markers* 2022, 7593750. doi:10.1155/2022/7593750

Xu, Z. W. (2018). *Molecular mechanism of CHES1 affecting breast cancer proliferation by regulating ERα activity*. Dalian, China: Dalian University of Technology. Available at: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2019&filename=1019006898.nh&v=.

Yang, L., Li, R., Yu, M., Huang, F., Zeng, J., Lu, Y., et al. (2022). Determining the optimal cut-off values of serum E2 and FSH for evaluating the menopausal status of breast cancer patients in a southern Chinese population. *Dis. Markers* 2022, 8716160. doi:10.1155/2022/8716160

Yaqin, Q., Yulan, X., Mengyuan, L., Jinrui, W., and Jiming, X. (2022). Predictive modeling of ADMET properties of anti-breast cancer active compound. Natural Science Edition. *J. Yunnan Univ.*, 1–8.

Ye, X., Zhou, J., Tong, D., Wang, D., Wang, H., Guo, J., et al. (2022). E2F1 affects the therapeutic response to neoadjuvant therapy in breast cancer. *Dis. Markers* 2022, 8168517. doi:10.1155/2022/8168517

Zheng, G., Shi, L., Liu, J., Zhao, Y., Du, F., He, Y., et al. (2022). Registered trials of artificial intelligence conducted on chronic liver disease: A cross-sectional study on ClinicalTrials.gov. *Dis. Markers* 2022, 6847073. doi:10.1155/2022/6847073

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fgene.2022.1094332

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Peibo Bao,
Tongji University, China
Li Xiang,
Dalian Neusoft University of Information, China
Haitao Yang,
Beijing University of Technology, China

*CORRESPONDENCE
Yuanyuan Yang,
164102113@stu.cuz.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 10 November 2022
ACCEPTED 27 December 2022
PUBLISHED 11 January 2023

CITATION
Yang Y, Guo Q, Lu M, Huang Y, Yang Y and Gao C (2023), Expression of miR-320 and miR-204 in myocardial infarction and correlation with prognosis and degree of heart failure.
Front. Genet. 13:1094332.
doi: 10.3389/fgene.2022.1094332

Expression of miR-320 and miR-204 in myocardial infarction and correlation with prognosis and degree of heart failure

Yuanyuan Yang[1,2]*, Qiongya Guo[3,4], Min Lu[1,2], Yansheng Huang[1,2], Yu Yang[5] and Chuanyu Gao[1,2,6]

[1]Department of Cardiology, People's Hospital of Zhengzhou University, Zhengzhou, Henan, China, [2]Department of Cardiology, Henan Provincial People's Hospital, Zhengzhou, Henan, China, [3]Department of Gastroenterology, People's Hospital of Zhengzhou University, Zhengzhou, Henan, China, [4]Department of Gastroenterology, Henan Provincial People's Hospital, Zhengzhou, Henan, China, [5]Department of Geriatrics, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China, [6]Department of Cardiology, Fuwai Central China Cardiovascular Hospital, Zhengzhou, Henan, China

Myocardial infarction is a very dangerous cardiovascular disease with a high mortality rate under the modern developed medical technology. miRNA is a small molecule regulatory RNA discovered in recent years, which can play an important role in many cancers and other diseases. Medical data, machine learning and medical care strategies supporting the Internet of Things (IoMT) have certain applications in the treatment of myocardial infarction. However, the specific pathogenesis of myocardial infarction is still unclear. Therefore, this paper aimed to explore the expression of microRNA-320 and microRNA-204 in myocardial infarction and used the expression of microRNA-320 and microRNA-204 to predict the prognosis of patients with myocardial infarction. In order to discuss the expression of microRNA-320 and microRNA-204 in myocardial infarction in more detail. In this paper, 40 patients in the trial period were selected for clinical research, and 10 patients with normal cardiac function were selected in NHF group as control group. 10 patients with heart failure were selected as AMHF group. 10 patients with acute myocardial infarction were selected as AMNHF group. 10 patients with heart failure after old myocardial infarction were selected as OMHF group. AMHF group, AMNHF group and OMHF group were taken as the case group. This paper analyzed the difference of miR between different groups and determined that there were significant differences in the expression of miR-320 and miR-204 between different groups. Finally, the expression and prognosis of miR-320 and miR-204 in myocardial infarction were analyzed. The analysis results showed that the expression of microRNA-320 and microRNA-204 can inhibit the activity of myocardial cells. On the fifth day, the corresponding expression of microRNA-320 and microRNA-204 reduced the optical density of myocardial cells to 1.75 and 1.76, which was significantly lower than that on the first day. Moreover, excessive miR-320 expression and excessive miR-204 expression can increase the apoptosis rate of myocardial cells. The above results indicated that the high expression of microRNA-320 and microRNA-204 can be a bad prognostic factor in patients with myocardial infarction, showing that medical data, machine learning and medical care strategies supporting IoMT can play a role in the treatment of myocardial infarction. Therefore, it is urgent to understand the pathogenesis of heart failure after myocardial infarction and find new treatment schemes to improve the positive prognosis.

KEYWORDS

myocardial infarction, miR-320 and miR-204, healthcare strategy, medical data, internet of things (IoMT) medical data

1 Introduction

In developed countries, heart failure after myocardial infarction is an important cause of hospitalization and death. With the development of medical technology, the survival rate of patients has also been improved and the life cycle of patients with heart failure has also been extended. However, the aging population in China has led to an increasing incidence of heart failure year by year. Heart failure is a major public health problem facing China in the 21st century. In the past 10 years, although the treatment strategy of heart failure has changed from simply short-term hemodynamics to the goal of inhibiting and delaying the development of cardiac remodeling, its 5-year mortality is still close to that of malignant tumors. Therefore, to seek new therapeutic targets and reduce the mortality of patients are the most important for heart failure after myocardial infarction.

Myocardial infarction is still one of the diseases with the highest mortality rate, and there are many researches on the treatment of myocardial infarction. Fan Xiaoming studied the role of interleukin-10 in reducing renal injury after myocardial infarction in diabetes. The experimental results suggested that interleukin-10 is expected to become a detection index of myocardial infarction in diabetes (Fan et al., 2022). Reed Grant W studied the treatment and nursing of acute myocardial infarction. Through the analysis of the follow-up treatment results of 100 patients with acute myocardial infarction, he suggested that acute myocardial infarction should be nursed with traditional Chinese medicine to alleviate the pain of patients (Reed et al., 2017). Qiao Xiaoying explored the role of renal cortex matrix metalloproteinase-9 in the clinical treatment of myocardial infarction and believed that renal cortex matrix metalloproteinase-9 is an important factor to promote the recovery of myocardial injury after acute myocardial infarction (Qiao et al., 2022). Although there are many studies on the etiology and treatment of myocardial infarction, the specific pathogenesis is not particularly clear.

For the specific pathogenesis of myocardial infarction, many scholars have discussed the expression level of related factors. Dai Wangde showed the specific mechanism of myocardial ischemia in rats through the rat model of myocardial infarction, and he found that some factors had obvious relationship with the area of myocardial infarction and cardiac blood return in rats (Dai et al., 2022). Goncalves Meire S found that nerol can cause acute myocardial infarction in rats and can reduce the acute myocardial infarction induced by isoproterenol in rats by studying the rat model (Goncalves et al., 2022). Li Jianwei's research indicated that the lack of breast cancer activating factor can aggravate the inflammatory reaction in mice with acute myocardial infarction. He analyzed the expression of breast cancer activator before and after inflammation through the mouse myocardial infarction model (Li et al., 2022). It can be found that most scholars focus on common factors for the expression of myocardial infarction related factors, and there are not many studies on the relationship between MicroRNA (miRNA) and myocardial infarction.

This paper analyzed the expression of microRNA-320 and microRNA-204 in myocardial infarction, which was rarely studied by predecessors. In addition, microRNA-320 and microRNA-204 are frequently thought to have a stronger connection to cancer and have a role in the development of cancer. The expression of microRNA-320 and microRNA-204 in myocardial infarction, however, has not received much attention. Therefore, the research in this paper is of great significance for the treatment of future myocardial infarction and the subsequent research on miRNA.

2 miR-320, miR-204 and myocardial infarction

2.1 Medical data, machine learning and healthcare strategies supporting IoMT

In myocardial infarction, metabolic remodeling is regarded as a compensation of adaptive ability. However, as time goes on, it may become a harmful result, which can eventually destroy the energy metabolism of the heart and the systolic frequency of the heart. At the same time, the decrease of the fatty acid intake and oxidation level in patients with myocardial infarction and the increase of blood glucose concentration have an important relationship with metabolic remodeling.

TABLE 1 Main parameters required for the experiment.

Serial no	Abbreviation	Full name
1	LVEDV	Left ventricular end diastolic volume
2	LVESV	Left ventricular end systolic volume
3	IVS	Interventricular septal
4	PWTs	Posterior wall thickness at end-systole
5	LVIDd	Left ventricular internal diameter at end-diastole
6	LVIDs	Left ventricular internal diameter at end- systole
7	LVEF	Left ventricular ejection fraction
8	LVFS	Left ventricular fractional shortening
9	cESS	Circumferential end-systolic wall stress

LVEF, LVFS, cESS, and MEE, are finally calculated by the relevant indicators in Table 1. The calculation process is as follows.

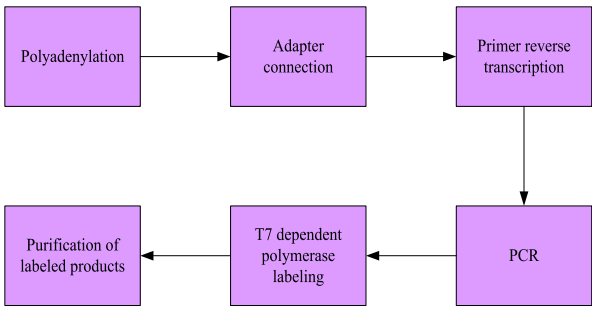


FIGURE 1
miRNA labeling steps.

The myocardial diameter and the quality of myocardial cells in the late diastole are also related to metabolic remodeling. In myocardial cells of patients with myocardial infarction, the protein content is decreased, and main fatty acid transportation β Oxidase and the expression of miRNA are also decreased. In the myocardial biopsy, the blood glucose intake and metabolism of patients with dilated cardiomyopathy have been significantly improved. In addition, some studies have shown that improving the utilization of fatty acids and sugars is an effective means to treat heart failure after myocardial infarction. In recent years, scholars from China and other countries have discussed the changes of small molecule metabolite content, aiming to explore the pathogenesis of heart failure and find new therapeutic targets (Redfors et al., 2021).

Amino acids play a vital role in human vitality and biosynthesis. Some amino acid metabolism abnormalities, such as miRNA and branched chain amino acids, can activate the signal transduction pathway of rapamycin target protein, thus promoting myocardial remodeling and causing heart failure (Blankenberg et al., 2018). In this paper, the difference of plasma amino acid metabolite miRNA between patients with heart failure and those without heart failure after myocardial infarction was analyzed by liquid chromatography tandem mass spectrometry. The most different metabolites and the pathogenesis of heart failure were investigated by cell test. This paper analyzed the changes of amino acid metabolism in plasma of patients with heart failure after myocardial infarction were and found the related metabolic indexes, and the pathogenesis was discussed.

2.2 miR-320 and miR-204

miRNA is tiny molecules that contain about 22 different kinds of coding RNAs (Ferhat et al., 2022). MiRNA is not only involved in the regulation of cell metabolism, differentiation, growth and apoptosis, but also in the occurrence and development of many diseases, such as tumors. In recent years, it has been new discoveries about miRNA in terms of heart failure, arrhythmia, cardiomyocyte apoptosis, vascular regeneration, *etc.* At present, the function of miRNA in human heart disease has been studied from patients with heart failure. Heart failure after myocardial infarction is mainly due to heart pump failure, which can not provide sufficient blood supply for the heart. Heart failure and sudden cardiac death are related to pathological myocardial hypertrophy. Therefore, it is very meaningful to explore its molecular mechanism and find new targets. Cardiac hypertrophy is mainly due to the activation of signal transduction and transcriptional regulation in cells, which leads to the enlargement of myocardial cells and the increase of interstitial synthesis. From 2006 to now, the correlation between miRNA and myocardial hypertrophy and heart failure has received extensive attention (Palsson et al., 2020). Cell culture experiments have shown that different miRNAs play an important role in the pathological changes of cardiac hypertrophy.

The expression level of miRNA can predict the development of the disease and may be a new therapeutic target for cardiovascular disease. However, it is very difficult to detect the myocardial tissue abnormalities in patients at an early stage. Therefore, it is very important to find new miRNA research methods. In this paper, different stages of heart failure caused by myocardial infarction were selected through miRNA array method to detect and compare the miRNA array of each group. Several miRNAs with the most significant differences between groups were screened to find specific antibodies to heart failure at different stages. New markers of heart failure and new miRNA regulatory pathways have been explored.

2.3 Calculation method of heart failure degree

The degree of heart failure can be evaluated by myocardial energy expenditure (MEE) (Swain et al., 2020). MEE detection in clinical practice is simple, safe, cheap, stable and reusable. At present, there are many indicators to evaluate MEE, such as maximum myocardial oxygen consumption, left ventricular systolic wall pressure, systolic wall stress time integral, pressure volume work, maximum left ventricular pressure, average velocity of myocardial annular shortening, average velocity of myocardial annular contraction, pressure work index, pressure volume area, *etc.*

The left atrial end systolic stress is a non-invasive measurement method. In addition, the multispectral imaging technology can analyze the blood flow through the aorta and accurately measure the stroke volume and ejection time. Therefore, circumferential end-systolic wall stress (cESS) can be used to replace the left ventricular systolic tension, so that patients with myocardial infarction can be evaluated non-invasive. In addition, studies have proved that the left ventricular external work is in direct proportion to the average cardiac pressure and cardiac output (Cerisano et al., 2019).

The patient should lie on the left side to be examined by PHILIP 7500 color Doppler ultrasound. At the same time, ECG signals should be recorded. The standard profile recommended by the American Ultrasonic Association is adopted, and the S3 ultrasonic probe is adopted, with the probe frequency of 2.5 MHz. The main parameters required for the experiment are shown in Table 1.

LVEF:

$$LVEF = \frac{LVEDV - LVESV}{LVEDV} \times 100\% \quad (1)$$

LVFS:

$$LVFS = \frac{LVM}{BSA} \quad (2)$$

$$LVM = 0.8 \times 1.4 \times \left[(LVIDd + IVS + PWTs)^3 - LVIDd^3 \right] + 0.6 \quad (3)$$

$$BSA = 0.0061 \times height + 0.0128 \times weight - 0.1529 \quad (4)$$

Among them, LVM represents the mass of left ventricle and BSA represents the body surface area.

TABLE 2 General clinical data between different groups.

	NFH(10)	AMHF(10)	OMHF(10)	AMNHF(10)	*p*-value
Age	56.26	58.54	59.33	57.36	1.171
Height m)	1.62	1.64	1.65	1.65	1.986
Weight (kg)	55.34	55.87	57.54	52.78	.765
Male (proportion)	5 (50%)	4 (40%)	6 (60%)	5 (50%)	.875
BMI(kg/m^2)	22.76	22.65	22.34	23.65	1.232

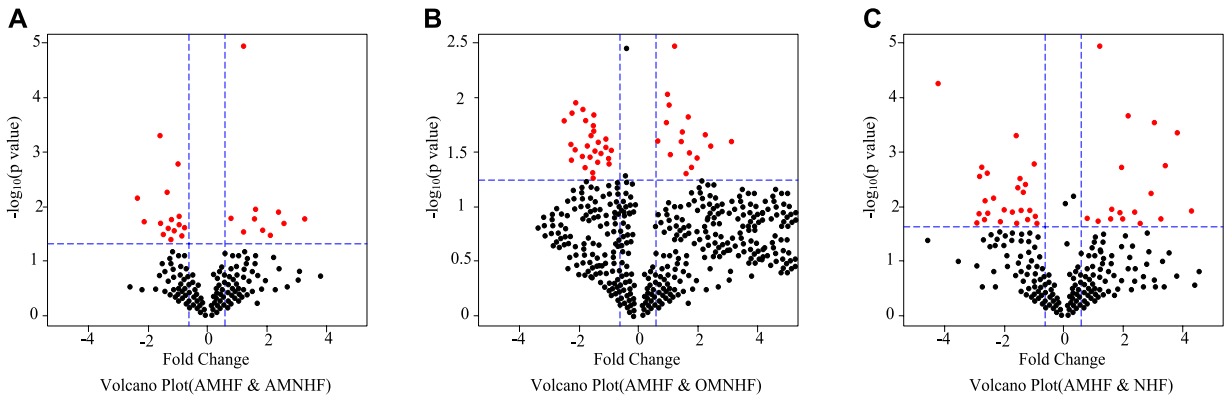


FIGURE 2
Volcanic map of differential expression miRNA. **(A)** Volcano Plot (AMHF & AMNHF) **(B)** Volcano Plot (AMHF & OMNHF) **(C)** Volcano Plot (AMHF & NHF).

cESS:

$$cESS = \frac{SBP \times (LVIDs/2)^2 \times \left\{1 + \frac{(LVIDs/2+PWTs)^2}{(LVIDs/2+PWTs/2)^2}\right\}}{(LVIDs/2 + PWTs)^2 - (LVIDs/2)^2} \quad (5)$$

SBP is the systolic pressure measured by Mercury cuff sphygmomanometer.

MEE:

$$MEE = cESS \times LVET \times LVSV \times (heart\ rate) \times 4.3 \times 10^{-7} \quad (6)$$

LVSV is left ventricular stroke volume.

3 Experiment on expression of Mir-320 and Mir-204 in myocardial infarction

3.1 Test object

3.1.1 Case group

The patients with myocardial infarction hospitalized in a hospital from October 2020 to October 2021 were selected as the research objects, and 40 patients were randomly selected as the observation objects in this experiment. According to the myocardial infarction's progression, they were split into four groups. One group is the NHF group, which stands for the control group with normal myocardial function. Patients with heart failure, those with acute myocardial infarction, and those with heart failure following an older myocardial infarction make up the other three case groups. These groups are designated as AMHF group, AMNHF group, and OMHF group, respectively. All participants are aware of the experimental content and signed the informed consent form to agree and cooperate with the experiment.

3.1.2 Inclusion criteria

Patients with myocardial infarction who meet the diagnostic criteria of coronary heart disease: there are typical chest tightness, chest pain and elevation of S and T segments of ECG (electrocardiogram), and coronary angiography result shows acute coronary artery occlusion. Acute myocardial infarction is evaluated according to the killip grade. The killip score of patients in the AMHF group is more than II, and the killip score of patients with AMNHF is I (Abbasloo et al., 2020).

Patients who have had an old myocardial infarction and are experiencing heart failure must fulfill two criteria. First, they should have a history of coronary heart disease and coronary intervention. Second, they should have a cardiac function score of Grade 3 or above as determined by the New York Heart Association (Engel and Rockson, 2020).

3.2 Experimental data collection

3.2.1 Collection and numbering of plasma samples

The enrolled patients should keep stomach empty in the morning of the next day after entering the group. They were collected 3 mL of

TABLE 3 Differentially expressed miRNAs between NHF group and AMNHF group.

	Name	Fold change	*p*-value
Up regulation of expression	Hsa-mir -409-3p	6.653	.013
	Hsa-mir-26b-5p	5.765	.032
	Hsa-mir-369-3p	5.543	.035
	Hsa-mir-148b-3p	4.432	.041
	Hsa-mir-329	4.132	.047
Down regulation of expression	Hsa-mir-320	.765	.001
	Hsa-mir-204	.643	.005
	Hsa-mir- 20b-3p	.543	.014
	Hsa-mir-767-5p	.532	.018
	Hsa-mir -299-5p	.412	.023

TABLE 4 Differentially expressed miRNAs between NHF group and OMHF group.

	Name	Fold change	*p*-value
Up regulation of expression	Hsa-mir-3941	6.753	.015
	Hsa-mir-338-5p	5.755	.034
	Hsa-mir-433	4.546	.041
	Hsa-mir-19a-3p	4.432	.045
	Hsa-mir-329	4.165	.049
Down regulation of expression	Hsa-mir-320	.865	.003
	Hsa-mir-204	.748	.007
	Hsa-mir-2110	.663	.015
	Hsa-mir-l247-5p	.542	.019
	Hsa-mir-296-5p	.432	.025

10.3389/fgene.2022.1094332

TABLE 5 Differentially expressed miRNAs between NHF group and AMHF group.

	Name	Fold change	*p*-value
Up regulation of expression	Hsa-mir-3667-5p	8.753	.009
	Hsa-mir-4275	7.751	.013
	Hsa-mir-647	7.548	.019
	Hsa-mir-1284	6.433	.023
	Hsa-mir-1306-3p	6.165	.024
Down regulation of expression	Hsa-mir-320	.785	.007
	Hsa-mir-204	.758	.009
	Hsa-mir-205-5p	.675	.012
	Hsa-mir -423-5p	.641	.013
	Hsa-mir-21-3p	.572	.021

peripheral venous blood through an ethylene diamine tetraacetic acid anticoagulant catheter, and plasma was separated 2 h later. The collected plasma was transferred to a 1 mL standard non RNA enzyme centrifuge tube. In different centrifuge tubes, the plasma of five patients is 100 ul each, a total of 500 ul, which should be put in the refrigerator at - 80°C.

3.2.2 miRNA labeling

The flow chart of miRNA labeling is shown in Figure 1. Specifically, it includes six steps: 1) 3′polyadenylation; 2) 5′adapter connection; 3) the oligo-dT primer containing a 3′adapter at the 5′end is used for reverse transcription; 4) PCR (polymerase chain reaction) detection is performed with primers containing promoters and complementary sequences of 3′and 5′adapters to the connected miRNA samples; 5) Polymerase dependent labeling, 6) purification of labeled products.

3.2.3 Data analysis

The scanned image is extracted by GenePixPro6.0 software, and the average value is obtained by repeating 10 times. The miRNAs with a gene expression intensity of 50 or higher in all samples are medially normalized. The independent *t*-test is used to screen miRNAs. The Volcano Plot is obtained by taking Fold Change as the abscissa and the negative logarithm of P-log10 (*p*-value) as the ordinate. When the variation is more than 1.5 times, and $p < .05$, the gene expression difference between the two samples can be directly obtained.

All measurement data are expressed in terms of average standard deviation. When two measurement data conform to normal and variance, two independent t-tests are used respectively. If they do not conform, a modified *t*-test is used. S When multiple groups of measurement data conform to the positive and the square difference, the analysis of variance method is used; otherwise, the Welch test is used. PSS13.0 statistical software is used to analyze the data.

3.3 Differentiated miRNA

The general clinical data of different groups are shown in Table 2. It can be seen that the patients in the four groups have similar physical conditions. The BMI is about 22.5, and the *p*-value is 1.232. There is no difference. In the basic situation, such as age, height and sex, the *p*-value is greater than .05, indicating that there is no difference in the physical condition of each group of patients, and subsequent experiments can be carried out.

3.3.1 Volcanic map for screening differential miRNA expression

Volcanic images are an effective means to intuitively display the differences between various samples. The index includes expression variation multiple and *p*-value, which can clearly reflect the relationship between different magnification (variation measurement) and statistical significance (taking into account the variation degree and variation degree). According to different data, they can be isolated from a group of special genes. In order to reveal miRNAs with significant differences in gene expression, volcanic maps of different groups were selected. The results are as shown in

TABLE 6 Target gene segment.

	Name	Seq
1	hsa-mir-204-p1	AGCTGTACAAGTAAGCCTGATCATGTACCCATAGG
2	hsa-mir-204-p1	GGGAGAGGGGCTTAGCTTATGGGACAGTTA TGGGC
3	hsa-mir-320c	TAAACTCGAGGAGGGGGAAAAAAAGACCTGAT

TABLE 7 PCR reaction system.

Serial no	Components	Volume (ul)
1	Upstream primer	.25
2	Downstream primer	.25
3	cDNA	1.00

Figure 2. The blue dotted line is perpendicular to the coordinate axis, indicating a rise or decrease of 1.5 times, and the red dot indicates significant differential expression of miRNA. As shown in Figure 2A, the differentially expressed miRNAs in AMHF and AMNHF groups are more concentrated. As shown in Figures 2B,C, the differentially expressed miRNAs in AMHF and OMNHF groups are more dispersed.

3.3.2 VDifferentially expressed miRNAs

MiRNA microarray analysis is a method to detect the difference of miRNA expression between patients with heart failure and normal control group by using peripheral venous blood plasma detection technology, so as to screen out differential miRNAs. The successful detection of differential miRNAs can find new heart failure markers, thus laying the foundation for new therapeutic goals.

As shown in Table 3, the five miRNAs with the largest difference in up regulation of expression between NHF group and AMNHF group are miR-409-3p, miR-26b-5p, miR-369-3p, miR-148b-3p and miRNA-329, respectively. The five miRNAs with the largest difference in down regulation of expression are miR-320, miR-204, miR-20b-3p, miR-767-5p and miR-299-5p, respectively.

As shown in Table 4, the five miRNAs with the largest difference in up-regulated expression between NHF group and OMHF group are miR-3941, miR-338-5p, miR-433, miR-19a-3p and miR-329 respectively. The five miRNAs with the largest down-regulation differences were miR-320, miR-204, miR-2110, miR-l247-5p and miR-296-5p, respectively.

As shown in Table 5, the five miRNAs with the largest difference in expression up regulation between NHF group and AMHF group are miR-3667, miR-4275, miR-647, miR-1284 and miR-1306-3p, respectively. The five miRNAs with the largest difference in expression down regulation are miR-320, miR-204, miR-205-5p, miR-423-5p and miR-21-3p, respectively.

By comparing the difference miRNAs between the control group and different case groups, it can be found that there are significant differences in miR-320 and miR-204 between patients with normal myocardial function and patients with myocardial infarction, and this difference occurs in each course of disease. Therefore, it is meaningful to make targeted diagnosis and treatment of myocardial infarction through miR-320 and miR-204.

3.4 Expression of miRNA in myocardial infarction

3.4.1 Construction and packaging of miRNA lentivirus

The selected gene information was miR-204-3p and miR-320C. The carrier information is GV254. The sequence of elements is Ubi-EGFP-MCS-IRES-Puromycin, and the cloning site is Nhe I.

The acquisition of target gene fragments is shown in Table 6.

The target gene is extracted from the plasmid containing the target gene by PCR technology, and the target vector is digested. After electrophoretic extraction, it is exchanged, and the final result is bacterial competent cells. The isolated clone is detected by PCR technology, and then the positive clone detected by PCR method is compared with it. If there is no error, the target gene can be obtained.

3.4.2 Cell resuscitation and culture

The cells were separated from liquid nitrogen and quickly poured into 37°C water for rapid dissolution. It was then transferred to a centrifuge containing medium and centrifuged at room temperature for 5 min. The speed is 1,000 rpm. After centrifugation, the supernatant was removed and blown with fresh medium and then transferred to the culture bottle for culture in the incubator. 10% bovine serum protein and 1% penicillin streptomycin double antibody solution were used to culture in the medium at 37°C and saturated humidity. Cells with logarithmic growth cycles were used in the experiment.

In the case of high density, cell passage must be carried out. The suspension cells were discarded at a ratio of 1:2-1:4 and then cultured in a new complete medium. During passage, the old medium was discarded and washed with normal saline, and then .25% trypsin was added into the culture dish. When the cell volume increased, the complete culture medium was added. The cells were dried and then sent to a centrifuge, and then the supernatant was poured out, resuspended, and divided into a new medium at a ratio of 1:2-1:4.

3.4.3 Cell cryopreservation

The cells were centrifuged at normal temperature for 5 min and then resuspended and transferred to a labeled refrigerated container. The container should be placed in a freezer at minus 80°C to transfer to liquid nitrogen the next day for long-term freezing.

3.4.4 Cell proliferation test

Myocardial cells in logarithmic growth cycle were counted after suspension from body weight. Cells in the 24 well plate were seeded into the 24 well plate according to the density of 105 cells per well. Myocardial infarction differential miRNA and different concentrations of myocardial infarction drugs were added. At the same time, the blank culture medium and blank control group were set (only the same number of cells were added, without any other treatment) and placed in the culture dish for 48 h. The cells in the 24 well plate were transferred to the 96 well plate, 200 ul for each well, and four well plates in each group were incubated at 37°C for 3 h. The microplate reader was used to measure the optical density (OD) absorption value at the wavelength of 490 nm. Under stable transfection conditions, 200ul × 104 cells were placed on 96 well plates, and 20 ul nutrient solutions were added at different culture times. The OD value was determined after incubation at 37°C for 3 h.

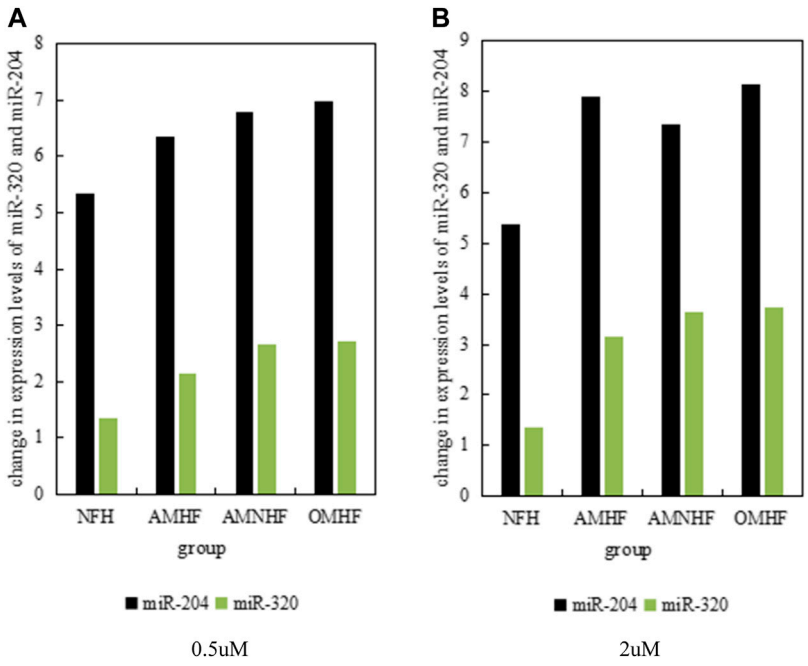


FIGURE 3
Expression of miRNA in each group at different concentrations. **(A)** 0.5uM **(B)** 2uM.

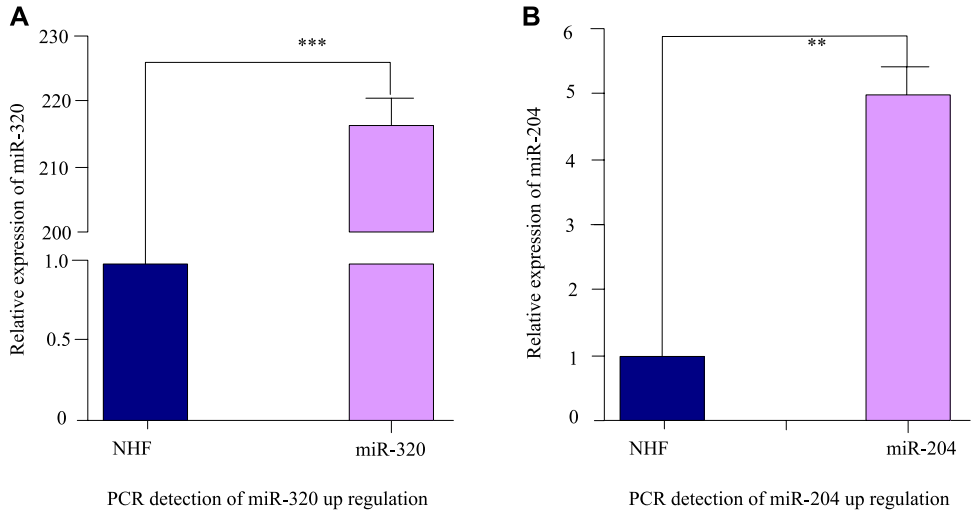


FIGURE 4
Fluorescence quantitative PCR detection results after miRNA transfection. **(A)** PCR detection of miR-320 up regulation. **(B)** PCR detection of miR-204 up regulation.

3.4.5 Detection of protein expression by western blot

3.4.5.1 Extraction of whole cell protein

The cells of the three same treatment groups in the 6-well plate were gathered together for centrifugation and washed through the pre cooled PBS (phosphate buffered saline) to absorb the supernatant as much as possible. Next, the inhibitor of 1% phosphatase and protease was added, and the cells were subjected to ultrasonic wave for 5 s on the ultrasonic cell grinder after repeated blowing.

3.4.5.2 Determination of sample protein concentration

18 ul PBS and two ul protein supernatant were added to the 96 well plate hole, and 200 ul premixed water-soluble compound working solution was added to fully mix and put in the 37°C incubator for

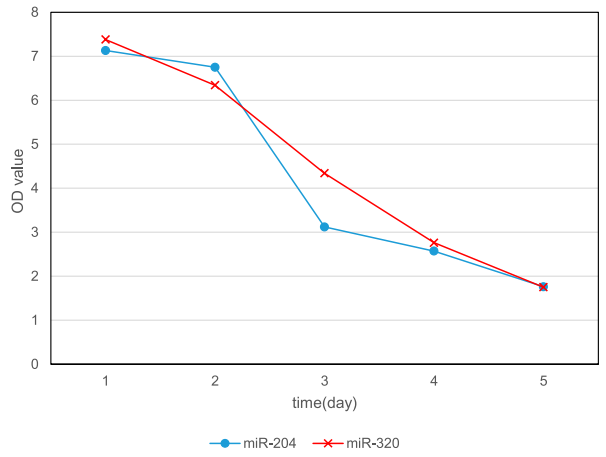


FIGURE 5
Change of OD value of cells in 5 days.

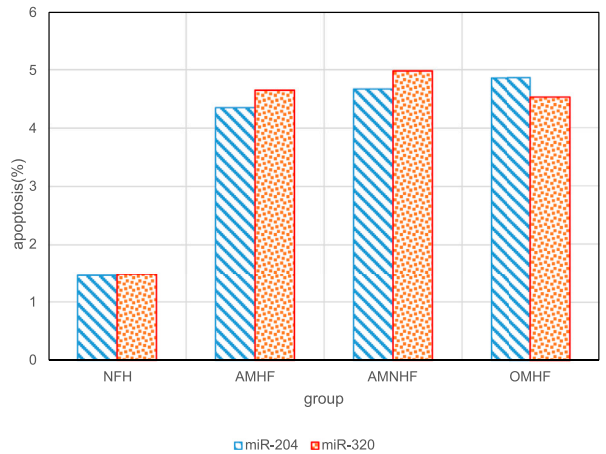


FIGURE 6
Relationship between overexpression of miR-320 and miR-204 and apoptosis in different groups.

30 min. The protein concentration of each sample is calculated according to the standard curve formula and then multiplied by the corresponding dilution multiple to obtain the protein concentration.

3.4.5.3 Protein electrophoresis

The bottom of 10% of the preform was torn to get a white adhesive strip, and then the rubber was added to the electrophoresis buffer in the electrophoresis tank. The liquid wass injected beyond the conductor. The comb was pulled out, and the comb hole was cleaned with electrophoresis buffer to remove foam and adhesive. According to the preset sampling sequence and feeding amount, the protein sample shall be placed in the comb hole marked with protein. At the beginning of electrophoresis, the constant voltage should be 80 V. When the protein is gradually separated, the voltage should be increased to 140 V until the electrophoresis is completed.

3.4.5.4 Protein transfer

After electrophoresis, the glue, polyvinylidene fluoride membrane, filter paper and sponge pad were cleaned in the membrane rotation buffer solution to ensure that no bubbles are generated between each component. Next, the electrophoresis plate was placed in the groove of the rotary film buffer, ensuring that the adhesive faces black (cathode), and the polyvinylidene fluoride film faces white (anode), with a constant current of 350 mA. The membrane transfer time is usually 90–180 min, depending on the molecular weight of the target protein.

3.4.6 Detection of target gene expression level

The myocardial cells were reverse transcribed into RNA and complex DNA. The PCR reaction system is shown in Table 7.

3.4.7 Expression of microRNA-320 and microRNA-204 in cardiomyocytes treated with demethylated drugs

Demethylated drugs can play a certain role in the treatment of patients with myocardial infarction. It can be seen from the above text that compared with patients with normal myocardium, the expression levels of microRNA-320 and microRNA-204 in patients with myocardial infarction are significantly reduced. Therefore, .5 umol and two umol of demethylated drugs were used to treat myocardial cells, and the expression levels of microRNA-320 and microRNA-204 genes were studied. The results are as shown in Figure 3, and Figure 3A shows the situation that the drug is .5 umol; Figure 3B shows the situation where the drug is 2umol. It can be clearly found from Figure 3 that different concentrations of demethylated drugs have certain protective effects on myocardial cells and have certain effects on the expression levels of microRNA-320 and microRNA-204.

3.4.8 Exogenous oligonucleotide transfection is highly utilized to express miRNA expression levels in myocardial cell lines

The results are shown in Figure 4. In subsequent experiments, oligonucleotides were used to transfect cardiomyocytes, and the expression of microRNA-320 and microRNA-204 in cardiomyocytes was detected by fluorescence quantitative PCR within 48 h. U6 was analyzed as an internal reference gene. As shown in Figure 4A, the level of miR-320 in myocardial cell lines transfected with oligonucleotides has been greatly improved. Compared with normal myocardial cells, the level of miR-320 has increased by about 215. As shown in Figure 4B, the miR-204 level of myocardial cell line transfected with oligonucleotides is 5 times higher than that of normal myocardial cells, which is significantly increased.

3.4.9 Overexpression of miRNA inhibits the cell activity of myocardial cell lines

In order to analyze the specific effects of miRNA on cardiomyocytes, overexpression of miRNA was used to culture cardiomyocytes. The experiment was divided into 5 days, and the expression levels of microRNA-320 and microRNA-204 were gradually increased to observe the cell activity. The change of OD value for five consecutive days is shown in Figure 5. It can be found that with the increase of miR-320 and miR-204 expression levels over time, the OD values continued to decline, and the OD values on the fifth day were 1.75 and 1.76, respectively. This shows that the cell activity continues to decline, which means that the overexpression of miR-320 and miR-204 can reduce the activity of myocardial cells.

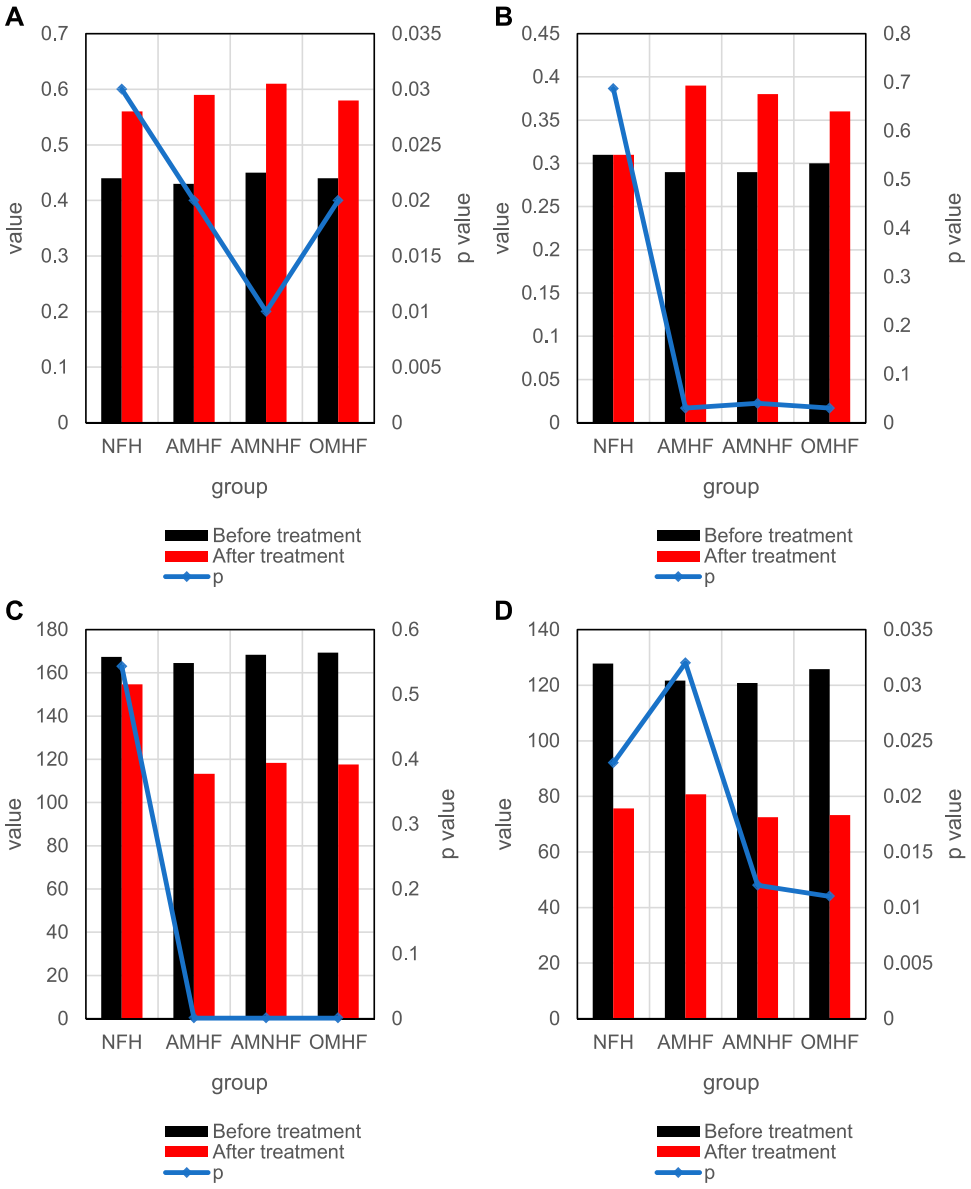


FIGURE 7
Comparison of differences between groups before and after treatment. **(A)** Changes of LVEF indicators in each group before and after treatment. **(B)** Changes of LVFS indicators in each group before and after treatment. **(C)** Changes of cESS indicators in each group before and after treatment. **(D)** Changes of MEE indicators in each group before and after treatment.

In order to analyze the impact of overexpression of miR-320 and miR-204 on patients in different groups, the apoptosis rate of myocardial cells in the case of overexpression of miR-320 and miR-204 was tested respectively. The results are shown in Figure 6. The overexpression of miR-320 and miR-204 has a significant impact on the apoptosis rate of myocardial infarction patients.

3.4.10 Comparison of myocardial function indexes between the two groups after treatment

At present, the function of miRNA in human heart disease has been studied from patients with heart failure. Heart failure is due to the pump failure of the heart, which can not provide sufficient blood supply for the heart and usually can lead to pathological cardiac hypertrophy and death.

Cardiac hypertrophy is an important manifestation of myocardial cells in various aspects such as blood pressure overload, endocrine disorder, myocardial injury, myocardial tissue and contractile protein gene mutation. Heart failure and sudden cardiac death are related to pathological myocardial hypertrophy (Kloner et al., 2020). Therefore, it is very meaningful to explore its molecular mechanism and find new targets. Cardiac hypertrophy is mainly due to the activation of signal transduction and transcriptional regulation in cells, leading to the enlargement of myocardial cells and the increase of interstitial synthesis.

Figure 7 shows the changes of heart failure degree indicators in each group before and after treatment. Among them, Figure 7A shows the changes of LVEF indicators between groups before and after treatment, and Figure 7B shows the changes of LVFS indicators

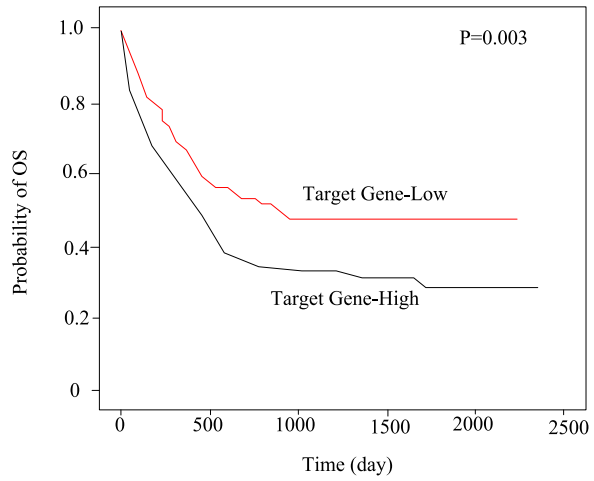


FIGURE 8
Survival graph.

between groups before and after treatment; Figure 7C shows the changes of cESS indicators between groups before and after treatment, and Figure 7D shows the changes of MEE indicators between groups before and after treatment. As shown in Figure 7, the indicators LVFS and cESS are significantly different between the control group and the case group before and after treatment, which can be used as indicators to measure myocardial infarction.

The relationship between the expression level of target genes and the prognosis of patients with myocardial infarction.

In this paper, Kaplan-Meier method was used to compare the expression level of target genes in 30 patients with myocardial infarction. The results showed that the total survival period (OS) of the group with high expression of target genes is lower, as shown in Figure 8.

By analyzing the treatment response of patients with myocardial infarction in the high expression group and the low expression group of target gene, it can be found that the total survival period of patients with myocardial infarction in the high expression group is shorter. This shows that the role of high-level target genes in myocardial infarction is an important factor to affect the prognosis of myocardial infarction.

4 Conclusion

The human genome has more than 1,000 miRNAs, which regulate 1/3 of the human genes. MiR-1 and miR-133 are less expressed in genes, but MiR-126-3p, miR-125a, miR-125a, MiR-143, miR-320 and miR-204 are mainly expressed in tissue cell types. In other words, miRNA expression is different in different diseases. Therefore, the differential expression of miRNAs can be used to predict and diagnose diseases. In recent years, the abnormal expression of gene mRNA is closely related to cardiovascular diseases such as tumor, diabetes, heart failure, *etc.* In this paper, experiments have proved that miR-320 and miR-204 have differential expression in patients with myocardial infarction. Compared with people with normal myocardial cells, the expression levels of microRNA-320 and microRNA-204 in patients with myocardial infarction are significantly down regulated. The level of target gene expression corresponding to miR-320 and miR-204 has an important impact on the prognosis of myocardial infarction, and the expression of target gene transcripts is a negative factor for the prognosis of myocardial infarction. In this paper, there are still some deficiencies in the experiment, so in the follow-up research, more about medical data, machine learning and healthcare strategies supporting IoMT should be learned, hoping to better understand the pathogenesis of myocardial infarction.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval for this study and written informed consent from the participants of the study were not required in accordance with local legislation and national guidelines.

Author contributions

All author contribute this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbasloo, E., Hamid, N., and Vakili, A. (2020). Chronic treatment with apelin, losartan and their combination reduces myocardial infarct size and improves cardiac mechanical function. *Clin. Exp. Pharmacol. Physiology* 47.3, 393–402. doi:10.1111/1440-1681.13195

Blankenberg, S., Neumann, J. T., and Westermann, D. (2018). Diagnosing myocardial infarction: A highly sensitive issue. *Lancet* 392.10151, 893–894. doi:10.1016/S0140-6736(18)31996-2

Cerisano, G., Buonamici, P., Gori, A. M., Migliorini, A., Santini, A., Marcucci, R., et al. (2019). Relation between post-myocardial infarct remodelling and gelatinase activity in patients enrolled in the TIPTOP trial. *Eur. J. heart Fail.* 21.1, 127–129. doi:10.1002/ejhf.1361

Dai, W., Amoedo, N. D., Perry, J., Le Grand, B., Boucard, A., Carreno, J., et al. (2022). Effects of OP2113 on myocardial infarct size and No reflow in a rat myocardial ischemia/reperfusion model. *Cardiovasc. Drugs Ther.* 36.2, 217–227. doi:10.1007/s10557-020-07113-7

Engel, G., and Rockson, S. G. (2020). Feasibility and reliability of rapid diagnosis of myocardial infarction. *Am. J. Med. Sci.* 359.2, 73–78. doi:10.1016/j.amjms.2019.12.012

Fan, X., Zhang, X., Liu, L. C., Kim, A. Y., Curley, S. P., Chen, X., et al. (2022). Interleukin-10 attenuates renal injury after myocardial infarction in diabetes. *J. investigative Med.* 705 (2022), 1233–1242. doi:10.1136/jim-2021-002008

Ferhat, E., Gultekin, K., Murat, G., Can, O., Altintas, M. S., Aydin, R. T., et al. (2022). Association of changes in the infarct and remote zone myocardial tissue with cardiac remodeling after myocardial infarction: A T1 and T2 mapping study. *Int. J. Cardiovasc. Imaging* 38.2, 363–373. doi:10.1007/s10554-021-02490-y

Goncalves, M. S. S., Silva, E. A. P., Santos, D. M., Santana, I. R., Souza, D. S., Araujo, A. M., et al. (2022). Nerolidol attenuates isoproterenol-induced acute myocardial infarction in rats. *Naunyn-Schmiedeberg's Archives Pharmacol.* 395.3 (2022), 353–363. doi:10.1007/s00210-022-02202-w

Kloner, R. A., Shi, J., Dai, W., Carreno, J., and Zhao, L. (2020). Remote ischemic conditioning in acute myocardial infarction and shock states. *J. Cardiovasc. Pharmacol. Ther.* 25.2, 103–109. doi:10.1177/1074248419892603

Li, J., Qu, H., and Wang, Y. (2022). Aib1 deficiency exacerbates inflammatory responses in acute myocardial infarction mice. *J. Mol. Med.* 100.8, 1181–1190. doi:10.1007/s00109-022-02231-1

Palsson, S. H., Engstrom, C., Enochsson, L., Osterlund, E., and Sandblom, G. (2020). Risk factors for postoperative myocardial infarct following cholecystectomy: A population-based study. *HPB* 22.1, 34–40. doi:10.1016/j.hpb.2019.06.018

Qiao, X., Bhave, S., Swain, L., Zweck, E., Reyelt, L., Crowley, P., et al. (2022). Myocardial injury promotes matrix metalloproteinase-9 activity in the renal cortex in preclinical models of acute myocardial infarction. *J. Cardiovasc. Transl. Res.* 15.2, 207–216. doi:10.1007/s12265-021-10114-y

Redfors, B., Simonato, M., Chen, S., Vincent, F., Zhang, Z., Thiele, H., et al. (2021). Ambient temperature and infarct size, microvascular obstruction, left ventricular function and clinical outcomes after ST-segment elevation myocardial infarction. *Coron. Artery Dis.* 33.2, 81–90. doi:10.1097/MCA.0000000000001099

Reed, G. W., Rossi, J. E., and Cannon, C. P. (2017). Acute myocardial infarction. *Lancet* 389, 10065197–10065210. doi:10.1016/S0140-6736(16)30677-8

Swain, L., Reyelt, L., Bhave, S., Qiao, X., Thomas, C. J., Zweck, E., et al. (2020). Transvalvular ventricular unloading before reperfusion in acute myocardial infarction. *J. Am. Coll. Cardiol.* 76.6, 684–699. doi:10.1016/j.jacc.2020.06.031

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fgene.2022.1086379

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Lei Shi,
Luliang University, China
Tiefeng Wu,
Qingdao University of Technology, China
Fenghui Dong,
Nanjing Forestry University, China

*CORRESPONDENCE
Xiaopei Cao,
kana0720@zjnu.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 01 November 2022
ACCEPTED 09 December 2022
PUBLISHED 13 January 2023

CITATION
Zhang C and Cao X (2023), Biological gene extraction path based on knowledge graph and natural language processing.
Front. Genet. 13:1086379.
doi: 10.3389/fgene.2022.1086379

Biological gene extraction path based on knowledge graph and natural language processing

Canlin Zhang[1] and Xiaopei Cao[2]*

[1]Sorenson Communications, Salt Lake City, UT, United States, [2]College of Creative Culture and Communication, Zhejiang Normal University, Jinhua, Zhejiang, China

The continuous progress of society and the vigorous development of science and technology have brought people the dawn of maintaining health and preventing and controlling diseases. At the same time, with the update and iteration of bioinformatics technology, the current biological gene research has also undergone revolutionary changes. However, a long-standing problem in genetic research has always plagued researchers, that is, how to find the most needed sample genes from a large number of sample genes, so as to reduce unnecessary research and reduce research costs. By studying the extraction path of biological genes, it can help researchers to extract the most valuable research genes and avoid wasting time and energy. In order to solve the above problems, this paper used the Bhattacharyya distance index and the Gini index to screen the sample genes when extracting the characteristic genes of breast cancer. In the selected 49 public genes, 6 principal components were extracted by principal component analysis (PCA), and finally the experimental results were tested. It was found that when the optimal number of characteristic genes was selected as 5, the recognition rate of genes reached the highest 90.31%, which met the experimental requirements. In addition, the experiment also proved that the characteristic gene extraction method designed in this paper had a removal rate of 99.75% of redundant genes, which can greatly reduce the time and money cost of research.

KEYWORDS

biological gene extraction, knowledge graph, natural language processing, path research, biological gene

1 Introduction

In a rapidly developing world, various diseases are increasingly threatening people's healthy life. Many of these diseases can be resisted through physical defense, but many diseases come from human genes. How to study human-related diseases and find solutions for corresponding diseases is a difficult problem that cannot be ignored in current life sciences. Each genetic disease has its own characteristics, and the ever-changing gene expression is the key to preventing medical staff from curing such diseases. Coincidentally, how to extract the most needed genes when studying disease genes is also the key for researchers to solve the problem of genetic diseases. From the above description, the importance of biological gene extraction pathway research can be

found. In order to solve this problem, this paper selects the theory-assisted research of knowledge graph and natural language processing.

This paper adopts the knowledge graph and natural language processing theory for the research on the extraction path of biological genes, the purpose of which is to apply the technology of knowledge extraction of knowledge graph to the extraction of biological genes. In this way, the extraction of genes in clinical medicine and biological research can be better achieved, thereby further promoting the development of life sciences. The Bhattacharyya distance index and Gini index used in this paper have a good screening effect on massive research samples, and the PCA method can also further realize the extraction of characteristic genes. The innovation of this paper includes the following aspects: 1) It is not limited to the activity research of specific genes, but seeks to reduce the research cost and optimize the research process of similar research from the methodological level. 2) The combination of knowledge graph and gene extraction path is realized, which provides a more mature theoretical basis for optimizing the gene extraction path.

2 Literature review

It is not a novel thing to study biological genes to help people better understand and use biological genes. Xu L B discovered in his research on biological genes that biological rhythms are an important mechanism for organisms to receive external signals and regulate their own behavior. To this end, he summarized some core biological rhythms including biological clocks to assist biological breeding and human disease prevention and control (Xu et al., 2020). Ebigwai J K found that most of the existing protein research tools are only useful for the study of protein interaction PPI, so he proposed a new comprehensive PPI information extraction tool for support vector machine classifier, which is very important for protein gene research (Ebigwai et al., 2020). In order to better identify and predict the encoding of the genome, Abbas B proposed a new tool to rank, compare and identify the recurring properties in the hidden Markov model (HMM) (Abbas et al., 2019). Hasan M was opposed to synthetic gene editing and transgenic technologies that are widely used in food research and development. He hoped that the relevant departments can formulate more perfect biosafety regulations to reduce the risks related to food safety (Hasan et al., 2017). Do H pointed out that the mining technology of association rules can promote the study of different gene expression. To this end, she designed CPMiner, a data mining method for processing biological data, and dedicated to a unified framework for extracting gene expression using association rules (Do et al., 2020). Although these studies have carried out research on biological genes to a greater or lesser extent, they are all limited to a specific type or

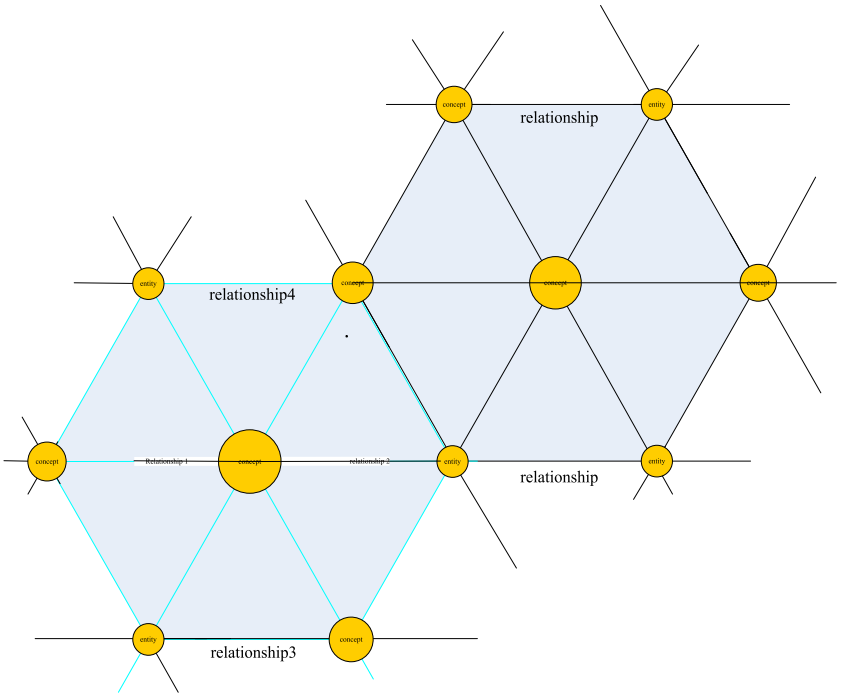


FIGURE 1
Knowledge graph network.

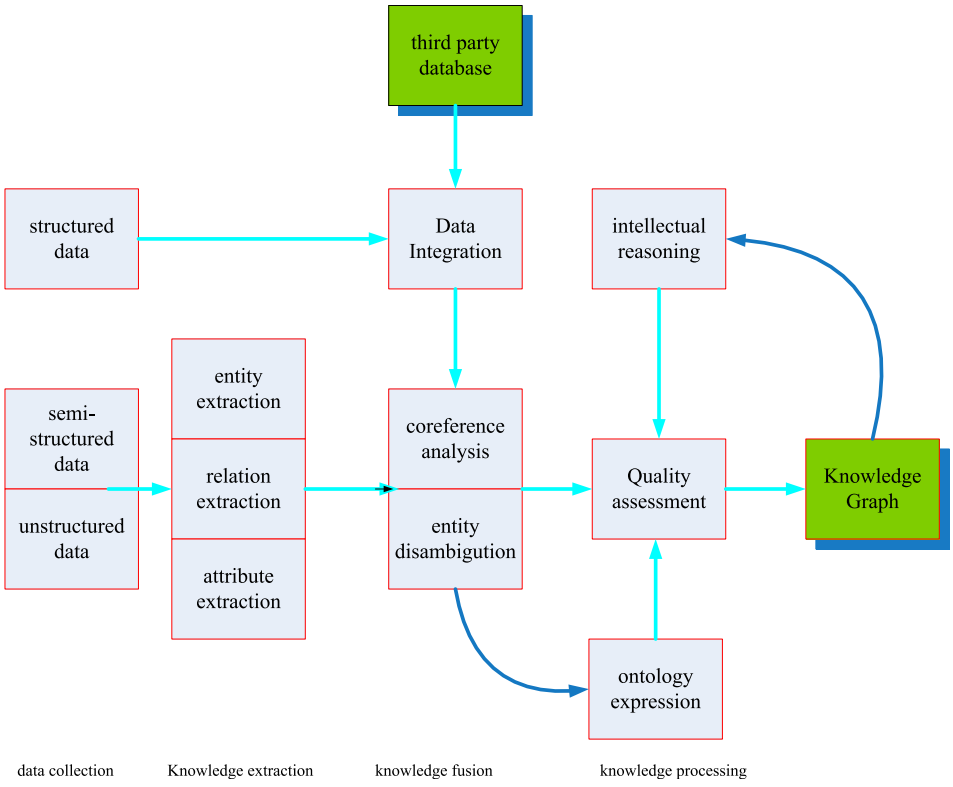


FIGURE 2
Construction process of the knowledge graph.

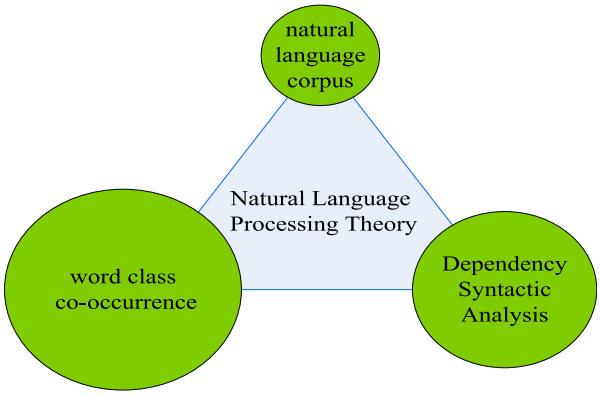


FIGURE 3
Natural language processing theory.

even a specific gene. However, this paper takes a unique approach to study the extraction path of biological genes, and provides a new research direction for the current gene research.

In recent years, with the spread of the concept of knowledge graph, more and more people have begun to conduct in-depth research on knowledge graph. Am A pointed out that the types of knowledge graphs are very broad, including public and private types. Knowledge graphs can also be obtained through the Internet, which requires statistical and linguistic methods (AlMarshad et al., 2021). Cai J L pointed out that a new multi-layer convolutional network model can be used in the model construction and connection prediction of large-scale knowledge graphs, which can build highly complex knowledge graphs based on its high parameter efficiency (Cai et al., 2019). Nuaima R H applied a new learning framework in the study of reasoning large-scale knowledge graphs (KG), which can construct knowledge graphs with continuous states in the KG vector space more accurately, diversely and efficiently (Nuaima et al., 2018). Shi J found that existing self-symptom detectors are far from meeting the needs of clinical decision support systems. Therefore, she studied an automated knowledge graph that can automatically learn diseases and conditions in electronic medical records to identify patients' diseases (Shi et al., 2017). Lin Y Z found that the time limit is an invisible challenge to the mining and exploration of large-scale knowledge graphs. To this end, she proposed an online mining algorithm, which can greatly improve the speed and accuracy of mining knowledge graphs within a certain period of time (Lin et al., 2017a). Such a comprehensive

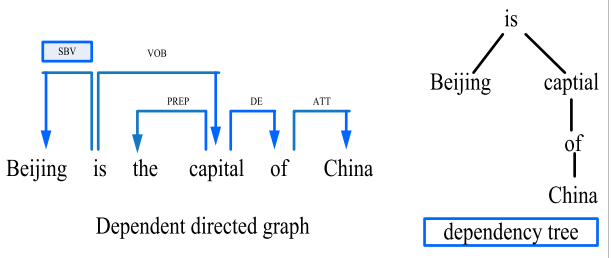


FIGURE 4
Process of dependency parsing.

knowledge map research provides a pioneering idea for the combination of knowledge map and biological gene extraction path in this paper. This paper looks forward to solving the problem of biological gene extraction path that has puzzled researchers for a long time by using the related theory of exponential graph and natural language processing.

3 Methods of biological gene extraction research based on knowledge graph and natural language processing

3.1 Knowledge graph and natural language processing

3.1.1 Knowledge graph

Knowledge graph is a semantic network proposed by Google in 2012 to describe the relationship between concepts in the real world in order to improve the search ability of its own engine (Paulheim and Cimiano, 2017). In fact, the essence of the knowledge graph is a conceptual network woven by related relationships between physical machines in the real world. As shown in Figure 1, on this network, nodes represent real objects or concepts, and the lines connecting each node are network edges and represent the correlation between them.

By combining the knowledge network with the Internet, it is possible to establish the relationship between things on the Internet with the help of the knowledge graph, and on this basis, to combine the all-encompassing information on the Internet to create new knowledge, which is a major breakthrough in artificial intelligence. The knowledge graph has been inseparable from artificial intelligence since it was proposed, and is widely used in search engines, intelligent question answering, and personalized recommendation (Lin et al., 2017b). The combination of the two can reduce the tediousness of manual search, and use Internet tools to visualize the searched knowledge.

The construction process of the knowledge graph is based on the original data and use technical means to extract objective knowledge facts from the original data, then extracting knowledge elements from the knowledge facts and storing knowledge elements in a database. The construction of the knowledge graph hides the update of the knowledge graph (Zhang et al., 2017). Because the real world is constantly moving and changing, and new concepts are produced every moment, along with the demise of concepts, the construction of knowledge on this basis also means the renewal of knowledge. As shown in Figure 2, the construction of knowledge graph includes three stages, namely knowledge extraction, knowledge fusion and knowledge processing. Each knowledge construction is an iterative process of updating.

Among them, information extraction is the starting link of knowledge graph construction (Natthawut and Ryutaro, 2018). This link is mainly the process of extracting entities and the interrelationships of entity attributes from real entities. Through this link, a large number of discrete entities and their interrelationships can be achieved. Then, through the knowledge fusion link, the data obtained in the information extraction link is cleaned and classified. The redundant and repeated data are eliminated, and errors are eliminated to improve the hierarchy and logic of knowledge. Actually this coincides with the research on the biological gene extraction pathway studied in this paper. The research topic of this paper is to extract the core genes from the complex genome. The third link is the knowledge processing link. In this link, knowledge needs to be processed to obtain systematic and structured knowledge. In addition, a dynamic knowledge network needs to be finally obtained by constructing ontology, inferring knowledge and evaluating quality.

3.1.2 Natural language processing

Natural language processing (NLP) is an important research tool for computers, especially in artificial intelligence. It mainly studies the means of communication between humans and computers, that is, natural language (Jia et al., 2017). In the information age, people can access information on the Internet more quickly because of natural language processing, and can also respond more quickly to changes in online life. The most typical example is the interception of e-mail spam. With the help of natural language processing, people can easily set up blocking words, so that the mailbox can block spam by locking key words to avoid harassment. The processing of natural language is inseparable from the natural language corpus. Natural language corpus is an indispensable foundation for processing natural language. Meanwhile, with the increasing development of natural language processing related work, the theory of natural language processing has gradually matured. As shown in Figure 3, the processing theory of natural language includes three parts: natural language corpus, dependency syntax analysis and part-of-speech co-occurrence (Zhu et al., 2017).

TABLE 1 Sample gene datasets.

Datasets	Category	Sample amount	Gene amount
Leukemia data set	ALLI.ALLB.AWL	46	6844
Polio data set	4	22	3629
Breast cancer data set	3	50	2000
Colon cancer data set	ALL.AML	82	96548

TABLE 2 Basic description of sample data.

Breast cancer sample data division	Lesion sample	Normal sample
Sample size	28	22
Training set	20	10
Test set	8	12

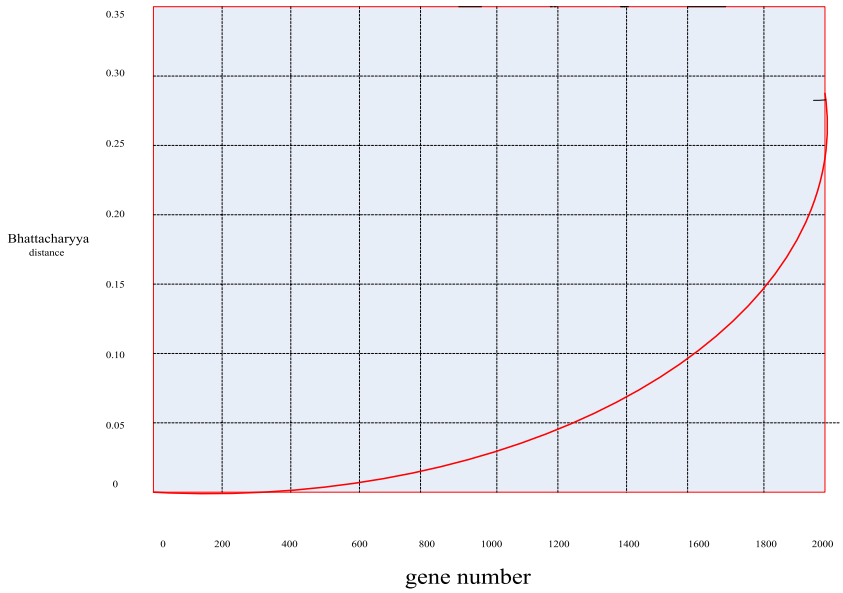


FIGURE 5
Bhattacharyya distance distribution curve of breast cancer genes.

The study of natural language corpus comes from the interdisciplinary subject of corpus linguistics. Corpus linguistics is a hybrid science that combines linguistics, computer science and applied linguistics (Wang et al., 2018). Through the study of a large number of real language materials, this discipline summarizes some abstract language laws that are detached from specific languages and words, and applies these language laws to natural language processing, which plays an important role in improving the language learning ability of the machine. Corpus linguistics has been around for a hundred years since its inception, and so far corpora are still an integral part of natural language processing.

Dependency parsing is an important part of natural language processing. Its working principle is mainly to reveal its intrinsic syntactic structure by analyzing the dependencies between various language components, and to assist the research by constructing a corresponding dependency tree. The workflow of dependency parsing is based on the existence of dependencies

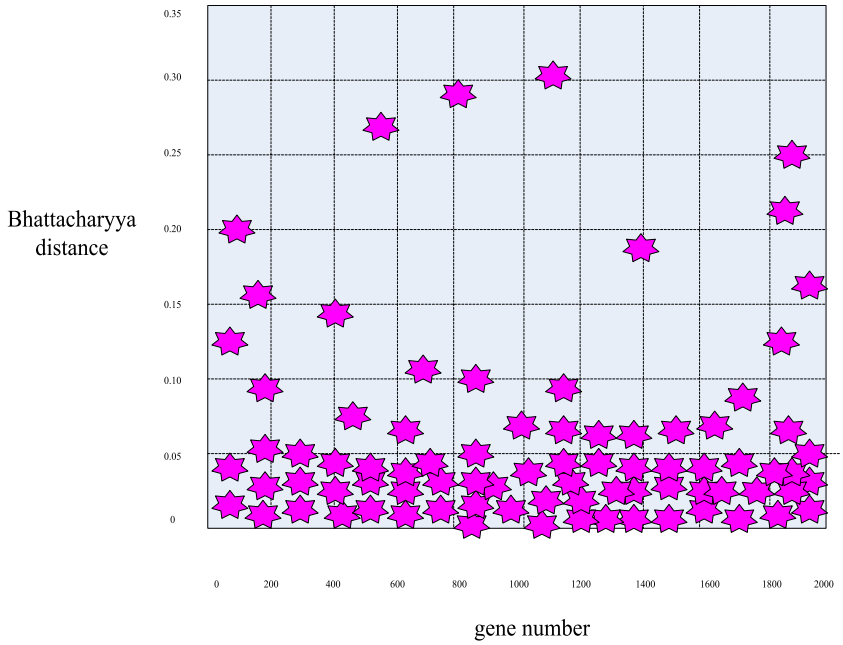


FIGURE 6
Scatter plot of the distribution of Bhattacharyya distance for breast cancer genes.

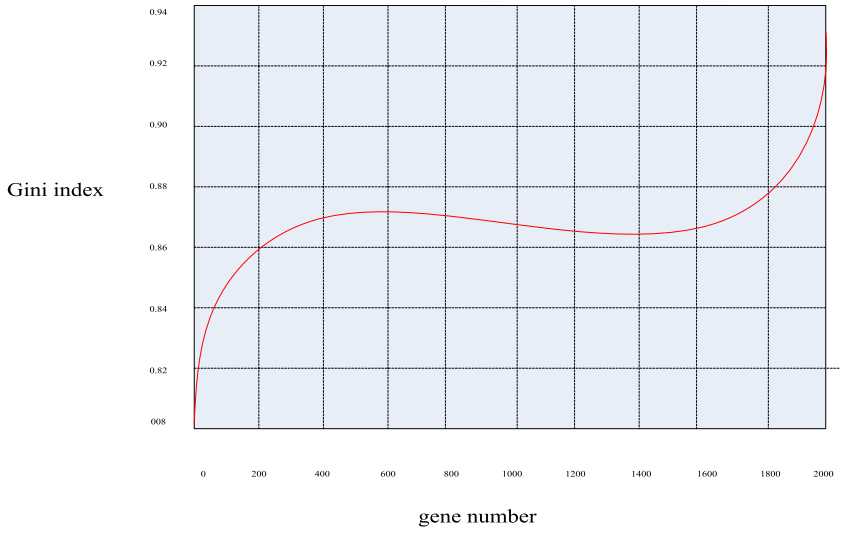


FIGURE 7
Gini index distribution curve of breast cancer genes.

between different natural language units. Generally speaking, a dependency relationship consists of two parts, the core word and the modifier. Researchers can analyze the structure of sentences through dependency syntax analysis technology to determine the main components of sentences, so as to help machines better understand natural language and achieve more efficient human-computer interaction. As shown in Figure 4, the analysis process of dependency parsing is as follows:

Co-occurrence theory, literally, is a theory that studies the relationship of things that occur together (Tom et al., 2018).

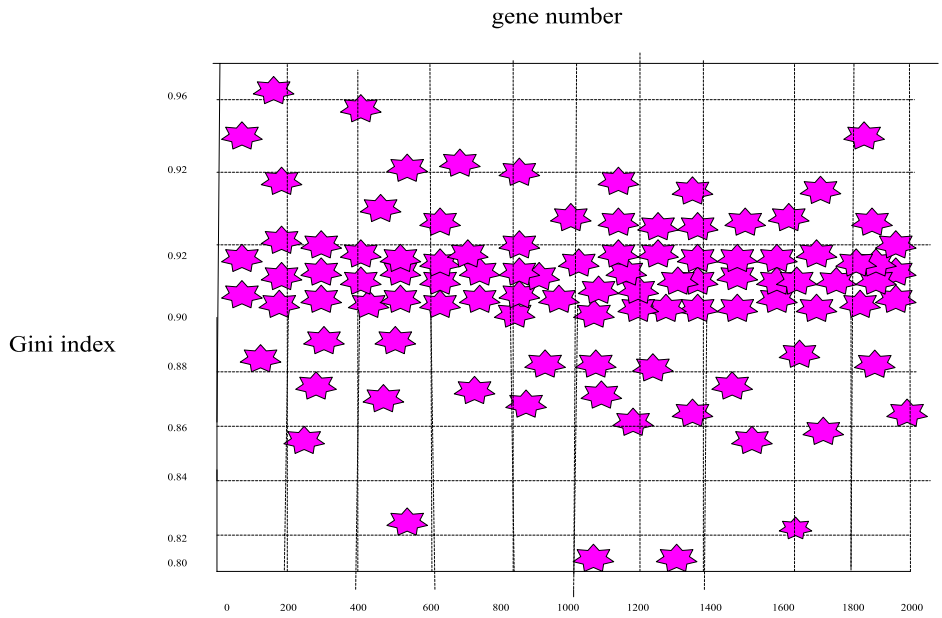


FIGURE 8
Scatter plot of the distribution of Gini index distribution of breast cancer genes.

Generally speaking, the co-occurrence relationship is the most basic relationship in the entity relationship. In fact, the co-occurrence relationship refers to inferring the affinity between different entities by counting the number of co-occurrences between them in the same text, and this relationship is also called co-occurrence relationship in the field of natural language. Co-occurrence relationship can be divided into broad and narrow sense. Different from the limited scope of the co-occurrence relationship in the narrow sense, the generalized contribution relationship expands the scope of this relationship to time and space, and when analyzing this generalized co-occurrence relationship, it can further increase the prediction of the real world. For example, Walmart found that sales of egg tarts and flashlights can increase during the hurricane season, and the increase in sales and the arrival of a hurricane are co-occurrences in a broad sense. Based on this broad co-occurrence, Walmart would put together egg tarts and anti-hurricane supplies before the hurricane to increase sales.

3.2 Feature gene extraction based on knowledge graph and natural language processing

3.2 1 Feature gene extraction

Feature gene extraction refers to finding out the genes that are not relevant for classification, extracting those representative genes. Then these genes are grouped into a subset for classification (Wong et al., 2018). There are two feature selection methods, namely feature selection and feature extraction. In a broad sense, feature extraction refers to a kind of mapping. The main methods are principal component method and partial least squares method. Feature selection only removes redundant genes and noise genes. Feature selection is difficult to achieve the optimal classification effect because the misclassification rate is too large, which is not considered in this paper.

3.2 2 Scoring guidelines for gene expression data

In view of the high-dimensional and high-noise characteristics of gene expression, this paper believes that dimensionality reduction and denoising of gene expression are very necessary. Scoring guidelines are integral to this process. This paper mainly uses feature score criterion (FSC) and Fisher discriminant method. The scoring formulas are shown in Formulas 1, 2:

$$FSC(g_i) = \left| \frac{\eta_i^+ - \eta_i^-}{\omega_i^+ - \omega_i^-} \right| \tag{1}$$

$$FDR(g_i) = \frac{(\eta_i^+ - \eta_i^-)^2}{(\omega_i^+)^2 - (\omega_i^-)^2} \tag{2}$$

In Formula 1 and Formula 2, η_i^+ and η_i^- represent the mean, while ω_i^+ and ω_i^- represent the standard deviation.

TABLE 3 List of principal component characteristic data of public genes.

Main ingredient	Eigenvalues	Contribution rate (%)	Feature contribution rate (%)
A	28.174	39.658	39.658
B	10.742	18.649	49.165
C	9.645	12.359	57.761
D	8.742	8.542	61.952
E	4.568	3.406	71.962
F	2.691	2.864	77.294
G	1.669	2.173	79.429
H	1.485	1.749	81.263
I	1.293	1.599	83.648

3.2.3 Feature gene extraction method

The purpose of eigengene extraction is to extract the most eigengenes from a group of genes, and the number of a group of eigengenes is extremely large in the experiment. Therefore, the selection of the extraction method of eigengenes is very important. The commonly used eigengene selection methods are as follows:

Filtering method: The filtering method refers to the selection according to the data of the information itself. This paper mainly uses the following two filtering methods, namely the signal-to-noise ratio and the t-statistic method. The signal-to-noise ratio is the simplest filtering method. The criterion for selecting feature vectors by using the signal-to-noise ratio method is the degree of difference in the vector attribute space in different categories of samples. Assuming that there is a gene g), the calculation formula of its signal-to-noise ratio $d(g)$ is as Formula 3:

$$d(g) = \left| \frac{\eta_g^+ - \eta_g^-}{\omega_g^+ - \omega_g^-} \right| \tag{3}$$

In Formula 3, η_g^+ and η_g^- represent the mean, while ω_g^+ and ω_g^- represent the standard deviation.

The t-statistic method refers to the use of the t-part theory to infer the probability of the occurrence of differences, and its calculation formula is shown in Formula 4:

$$t_i = \frac{\bar{x}_i - \overline{y_i}}{\sqrt{\frac{s_{1i}^2}{n_1} + \frac{s_{2i}^2}{n_2}}} \tag{4}$$

In Formula 4, $\bar{x}_i$ and $\overline{y_i}$ represent the mean, and s_{1i}^2 and s_{2i}^2 represent the variance. It can be seen that the larger the calculated t value, the greater the difference, that is, the greater the influence of gene i on the sample classification.

3.3 Path of feature gene extraction based on PCA

Gene expression data are characterized by small samples, high dimensionality, and high noise, and there are very few genes associated with diseases. The presence of a large number of redundant genes makes the expression difference between the diseased and normal samples small, and the contribution to the classification is also small. Therefore, it is necessary to use a certain gene microarray technology to propose genes from a large number of genes, and extract biological genes that truly meet the characteristics (Balsmeieri et al., 2018).

3.3.1 Construction of gene expression data

Gene data covers the basic characteristics of genes and is a display of the physiological state of biological cells. The gene expression profile can be obtained by using gene chip technology. The expression profile data can be expressed as Formula 5:

$$X = \left(x_{ij}\right) = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1j} \\ x_{21} & x_{22} & \ldots & x_{2j} \\ \vdots & \vdots & \ldots & \vdots \\ x_{i1} & x_{i2} & \ldots & x_{ij} \end{bmatrix} \tag{5}$$

In Formula 5, x_{ij} represents the j-th sample for the i-th gene.

3.3.2 Elimination of irrelevant genes

There are tens of thousands of genes in a biological sample, and at least a few are needed for the experiment. Other unrelated genes can cause gene redundancy and interfere with the experiment. Therefore, this paper chooses a hybrid gene extraction method that combines the Bhattacharyya distance and the Gini index. The formula for calculating the Bhattacharyya distance of a gene is shown in Formula 6:

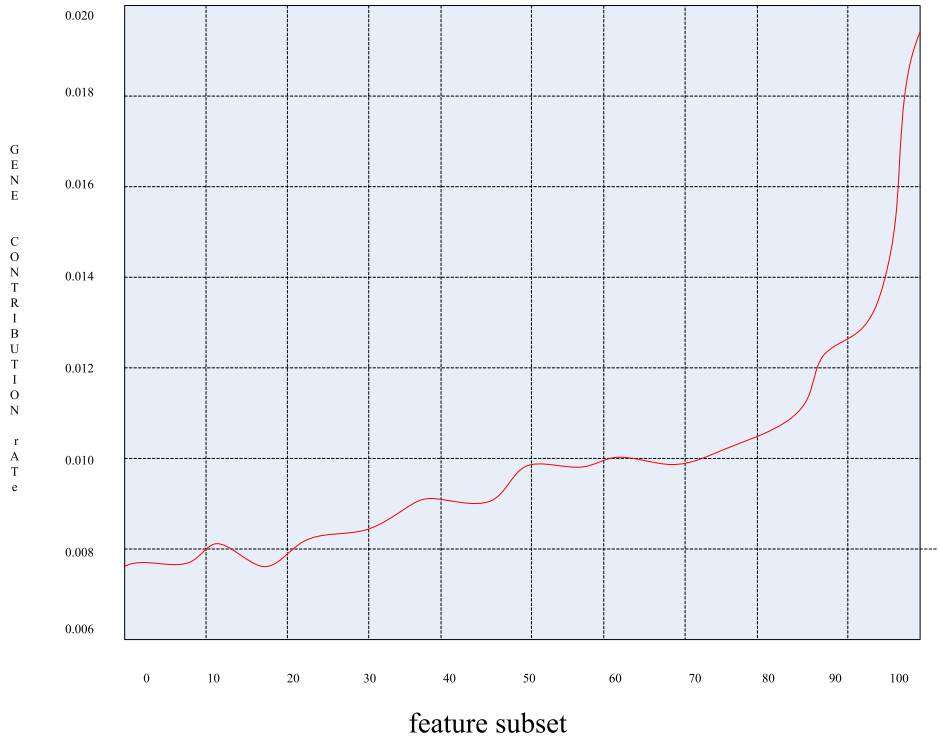


FIGURE 9
Graph of the contribution rate of each principal component.

$$B_d = \frac{1}{8}\frac{(\eta_1 - \eta_2)^2}{(\omega_1^2 + \omega_2^2)} + \frac{1}{2}\ln\left(\frac{\omega_1^2 + \omega_2^2}{2\omega_1\omega_2}\right) \tag{6}$$

Among them, η_1 and η_2 represent the mean of genes in the first and second samples respectively, while ω_1 and ω_2 represent the sample variance.

Gini index is an index commonly used in data mining to evaluate the goodness of classification nodes (Wi et al., 2017). Before calculating the Gini index, the original data should be discretized. The discrete formula is shown in Formula 7:

$$S_{ij} = Int\left[20 \times \frac{n_{ij} - \min(i)}{\max(i) - \min(i)} + 0.5\right] \tag{7}$$

Among them, $Int[]$ represents rounding, and $\max(i)$ and $\min(i)$ represent the maximum and minimum values of gene g_i.

The discretized data is substituted into Formula 8:

$$Gini(k) = 1 - \sum_{j=0}^{20}\left[p_{ij}\right]^2 \tag{8}$$

In Formula 8, $Gini(k)$ refers to the Gini index of k-type genes, and p_{ij} refers to the relative rate of level j in category k.

The formula for calculating the Gini index of gene g_i is as Formula 9:

$$Gini(g_i) = \sum_{k=1}^{2}\frac{n_k}{n}Gini(k), i \in [1, 2000] \tag{9}$$

In Formula 9, n is the total number of samples and n_k is the number of samples in the k-th class. Therefore, the smaller the Gini value of gene g_i, the more taxonomic information this gene contains, and *vice versa*.

The Bhattacharyya distance and the Gini index are combined together, and these two indicators are integrated as a standard to measure the gene classification information, which avoids the defect of "signal-to-noise ratio" and other methods that do not generate variance in the differential expression of genes (Diamantopoulos et al., 2017).

3.3.3 Selection of principal component feature

PCA refers to using the idea of dimensionality reduction to replace many original variables with new variables on the basis of keeping the principal components unchanged. These new variables can reflect most of the information of the original variables (Johnny et al., 2017; UzmaAl-Obeidat et al., 2022). The purpose of principal component analysis is to extract as few comprehensive variables with strong classification ability as possible as the representative of all variables, so as to retain the original variable information and reduce the number of

TABLE 4 Detection results of signature genes.

Number of genes	Classification accuracy (%)	Number of misjudgments
6	78.82	7
10	84.62	5
5	90.31	4
16	67.74	5
20	88.52	4

variables. Assuming that there are n samples, and each sample contains m variables, then there is Formula 10:

$$X_{m\times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix} \tag{10}$$

On this basis, the data is standardized to obtain Formula 11, Formula 12, Formula 13:

$$x_{ij} = \frac{x_{ij} - \overline{x_j}}{h_j} \tag{11}$$

$$\overline{x_j} = \frac{1}{n}\sum_{n=1}^{1} x_{ij} \tag{12}$$

$$h_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2} \tag{13}$$

$\overline{x_j}$ means the mean of the j-th attribute, and h_j means that the variance is normalized to have a mean of 0 and a standard deviation of 1 for each data.

Then the correlation coefficient matrix is obtained as Formula 14:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1t} \\ P_{21} & P_{22} & \ldots & P_{2t} \\ \vdots & \vdots & \vdots & \vdots \\ P_{m1} & P_{m2} & \ldots & P_{mt} \end{bmatrix} \tag{14}$$

The correlation coefficient p_{ij} represents the correlation between x_i and x_j, and the calculation formula is as Formula 15:

$$p_{ij} \frac{\sum_{k=1}^{n}\left(x_{ki} - \overline{x_i}\right)\left(x_{kj} - \overline{x_j}\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{ki} - \overline{x_i}\right)^2 \sum_{k=1}^{n}\left(x_{kj} - \overline{x_j}\right)^2}}, \quad i, j = 1, 2, ..., m \tag{15}$$

Solving the characteristic formula according to the correlation coefficient matrix P, there is Formula 16:

$$\left|\gamma I - P\right| = 0 \tag{16}$$

The characteristic formula is solved to get different eigenvalues. Sorting these eigenvalues yields Formula 17:

$$\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_m \geq 0 \tag{17}$$

Then the standard orthonormal vector is calculated by the existing eigenvalues to obtain Formula 18:

$$e_i = (e_{1i}, e_{2i}, .., e_{mi})^T \tag{18}$$

On this basis, the analysis of different principal component contribution rates obtains Formula 19:

$$\lambda_i = \frac{\gamma_i}{\sum_{k=1}^{m}\gamma_k}, i \tag{19}$$

After calculating the contribution rate of each principal component, the cumulative sum is obtained to obtain the cumulative contribution rate as shown in Formula 20:

$$\lambda_{(i)} = \frac{\sum_{k=1}^{i}\gamma_k}{\sum_{k=1}^{m}\gamma_k}, \; i = 1, 2, ..., m \tag{20}$$

Finally, the gene contribution rate Formula 21 can be found:

$$g_j = \frac{\sum_{i=1}^{t}\left|d_{ij}\right|}{\sum_{i=1}^{t}\sum_{k=1}^{m}\left|d_{ik}\right|}, \; j = 1, 2, .., m \tag{21}$$

4 Experiment and results of feature gene extraction based on PCA

4.1 Experimental data acquisition and processing

With the fast-paced work and life, the living pressure of modern women has greatly increased. Staying up late and insomnia are quite common. In such a stressful life rhythm, breast cancer, the killer disease of women's health, has quietly appeared in more and more women. In order to alleviate this situation, this paper took breast cancer as the research object, and analyzed the characteristic gene extraction pathway of breast

cancer, hoping to enrich the breast cancer gene extraction pathway and contribute to the control and prevention of breast cancer.

In order to verify the feature gene extraction path proposed in this paper, several common disease gene datasets were collected in this paper, as shown in Table 1:

As shown in Table 1, the total number of samples in the selected four diseases was 200 cases. It can be seen that the number of samples of breast cancer genes was large, but the total number of genes was relatively small. The former satisfied the quantitative requirements of the experiment, and the latter represented the less complex number of genes, which was convenient for screening. Even though all four of these diseases were common, the choice of breast cancer genes was more beneficial than other options.

Through the investigation, it can be seen that among the 50 samples in this breast cancer data set, 28 were diseased genes, and the remaining 22 were normal samples. In this regard, 30 samples were used as the training set, and the remaining 20 samples were used as the test set. The description of the samples is shown in Table 2:

4.2 Experimental process and results

4.2.1 Elimination of irrelevant genes

Based on the characteristics of sample genes with small number of samples and many redundant genes, although there were 2000 genes in 20 samples, there were still a large number of genes that were not related to the content of the experiment. Therefore, this paper combined the two methods of Bhattacharyya distance and Gini index to measure to promote the standardization of gene classification. This method avoided the defect that methods such as "signal-to-noise ratio" did not express differential gene expression due to variance. On this basis, this paper focused on the former and the latter as a supplement, so that the priority and secondary could better extract public genes.

First, according to the distance calculation formulas mentioned above, the Bhattacharyya distance value of these 2000 sample genes was calculated to obtain the distribution curve diagram of the Bhattacharyya distance of these genes in Figure 5 and the distribution scatter diagram in Figure 6.

First, it can be seen from Figures 5, 6 that in general, only a small number of genes have high information content, while most of the genes have a Bhattacharyya distance index between 0 and 0.05. This indicated that the genes within this range had no significant difference in mean and variance between normal samples and cancer samples, so it can be easily deleted as an irrelevant gene. Considering the validity and accuracy of the experiment, in this experiment, the Bhattacharyya distance index of these 2000 genes was sorted from small to large, and the corresponding genes ranked in the bottom 200 were selected as further experimental objects, eliminating the other 1800 irrelevant genes.

Similarly, according to the calculation formula of the Gini index mentioned above, the Gini index values of these 2000 genes were calculated, and the Gini index distribution curve diagram of the sample genes can be obtained in Figure 7 and the distribution scatter diagram in Figure 8.

It can be seen from Figures 7, 8 that the Gini index values of these 2000 sample genes were mostly between 0.88 and 0.92. In fact, the smaller the Gini index of a gene, the more equal the data expression, and the greater the amount of classification information of the gene. Since the experimental design of this paper was mainly based on the Bhattacharyya distance ranking, supplemented by the Gini index ranking, the same sorting method was used to eliminate irrelevant genes, and 200 spare genes were extracted.

After selecting 200 genes for each of the Bhattacharyya distance index and the Gini index as experimental subjects, among these 400 genes, the common genes were found whose duplicates were regarded as both. It could be determined that the amount of taxonomic information in these public genes far exceeded that of the excluded irrelevant genes. In the end, 49 common genes were obtained.

4.2.2 Extraction of characteristic genes

The dimensionality reduction idea of PCA was used to select the features of the 49 common genes finally selected, and the feature genes could be selected by the contribution rate of each gene in the finally selected principal components.

The eigenvalue, contribution rate and feature contribution rate of each principal component obtained by PCA of the 49 public genes are shown in Table 3:

It can be seen from Table 3 that the eigenvalues of the first six principal components were all greater than 2, and the eigenvalues corresponding to the seventh principal component were all less than 2. The cumulative contribution rate of the first six principal components has reached 85.478%. Therefore, only the first six principal components could be further analyzed.

When further analyzing the first six principal components, the contribution rate of each principal component has been calculated, and Figure 9 was obtained by arranging the contribution rate of each principal component in ascending order:

It can be seen from Figure 9 that the contribution rate of these 49 genes varied greatly, and the gene contribution rate varied greatly between 0.010 and 0.018. That is to say, those genes with a large gene contribution rate contain more classification information of the sample and contribute more to the classification. It can be seen that these genes contain a great amount of information, and the corresponding sample genes can be extracted as characteristic genes.

4.2.3 Classification inspection

On the basis of extracting the eigengenes above, this paper used the support vector machine (SVM) radial basis kernel function to detect the extracted eigengenes. The results are shown in Table 4:

It can be seen from Table 4 that when the number of selected eigengenes was 5, the recognition rate of genes reached a maximum of 90.31%. Therefore, it can be determined that only 5 genes in the 2000 sample genes finally contained rich information and meet the experimental requirements. The 5 most useful genes were extracted from 2000 genes, and the rate of removing redundant genes reached 99.75%, which just proved that the characteristic gene extraction path designed in this paper has a good classification effect on breast cancer genes.

5 Conclusion

This paper studied the extraction path of biological genes related to knowledge graph and natural language processing. The main significance of this research was to provide more abundant channels for gene extraction in modern medicine. In the past biological gene extraction research, most people were limited to the activity of the extracted genes, but ignored how to accurately extract the characteristic genes from the source to better meet the experimental needs, which is also the starting point of this study. In this regard, the sample genes of the research object were standardized by two screening criteria, the Bhattacharyya distance index and the Gini index. Then the PCA method was used to further extract the eigengenes, and finally the eigengenes were verified. It was proved that the redundant gene removal rate of the characteristic gene extraction method in this paper reached 99.75%, which provides a certain reference value for modern biological research.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

CZ: write paper XC: revise paper.

Conflict of interest

CZ was empolyed by Sorenson Communications.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abbas, B., Tjolli, I., Dailami, M., and Munarti (2019). Phylogenetic of sago palm (Metroxylon sagu) and others monocotyledon based on mitochondrial nad2 gene markers. *Biodiversitas J. Biol. Divers.* 20 (8), 2249–2256. doi:10.13057/biodiv/d200820

AlMarshad, L. K., AlJobair, A. M., Al-Anazi, M. R., Bohol, M. F. F., Wyne, A. H., and Al-Qahtani, A. A. (2021). Association of polymorphisms in genes involved in enamel formation, taste preference and immune response with early childhood caries in Saudi pre-school children. *Saudi J. Biol. Sci.* 28 (4), 2388–2395. doi:10.1016/j.sjbs.2021.01.036

Balsmeieri, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., et al. (2018). Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. *J. Econ. Manag. Strategy* 27 (3), 535–553. doi:10.1111/jems.12259

Cai, J. L., Yan, Y. F., Feng, G. W., and Jing, S. (2019). Dynamic change in the gene expression profile of rat benign prostate hyperplasia tissue after complete denervation. *Zhonghua nan ke xue = Natl. J. Androl.* 25 (11), 971–977.

Diamantopoulos, T., Roth, M., Symeonidis, A., and Klein, E. (2017). Software requirements as an application domain for natural language processing. *Lang. Resour. Eval.* 51 (2), 495–524. doi:10.1007/s10579-017-9381-z

Do, H., Thu, T., and Tran, N. (2020). Indigenous Lien Minh chicken of Vietnam: Phenotypic characteristics and single nucleotide polymorphisms of GH, IGFBP and PIT candidate genes related to growth traits. *Biodiversitas J. Biol. Divers.* 21 (11), 5344–5352.

Ebigwai, J. K., Ferdinand, A., and Ubi, G. M. (2020). Resolving taxonomic ambiguity between two morphological similar plant taxa using maturase K gene analysis. *J. Biol. Sci.* 20 (1), 13–21. doi:10.3923/jbs.2020.13.21

Hasan, M., Siddique, M. A., Hossain, M. A., and Rahman, M. S. (2017). 16S rRNA gene sequence based identification of Vibrio spp. in shrimp and tilapia hatcheries of Bangladesh. *Dhaka Univ. J. Biol. Sci.* 26 (1), 45–58. doi:10.3329/dujbs.v26i1.46349

Jia, Y., Wang, Y., Jin, X., Lin, H., and Cheng, X. (2017). Knowledge graph embedding: A locally and temporally adaptive translation-based approach. *ACM Trans. Web* 12 (2), 1–33. doi:10.1145/3132733

Johnny, D., Velupillai, S., George, G., Holden, R., Kikoler, M., Dean, H., et al. (2017). Detection of suicidality in adolescents with autism spectrum disorders: Developing a Natural Language Processing approach for use in electronic health records. *AMIA Symp.* 2017, 641–649.

Lin, Y. Z., Ou, D. L., Chang, H. Y., Lin, W. Y., Hsu, C., and Chang, P. L. (2017). Simultaneous visualization of the subfemtomolar expression of microRNA and microRNA target gene using HILO microscopy. *Chem. Sci.* 8 (9), 6670–6678. doi:10.1039/c7sc02701j

Lin, Z. Q., Xie, B., Zou, Y. Z., Zhao, J. F., Li, X. D., Wei, J., et al. (2017). Intelligent development environment and software knowledge graph. *J. Comput. Sci. Technol.* 32 (002), 242–249. doi:10.1007/s11390-017-1718-y

Natthawut, K., and Ryutaro, I. (2018). An automatic knowledge graph creation framework from Natural Language text. *Ieice Trans. Inf. Syst.* 101 (1), 90–98. doi:10.1587/transinf.2017swp0006

Nuaima, R. H., Roeb, J., Hallmann, J., Daub, M., Otte, S., and Heuer, H. (2018). Effector gene vap1 based DGGE fingerprinting to assess variation within and among *Heterodera schachtii* populations. *J. nematology* 50 (4), 517–528. doi:10.21307/jofnem-2018-055

Paulheim, H., and Cimiano, P. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8 (3), 489–508. doi:10.3233/sw-160218

Shi, J., Li, W., Gao, Y., Wang, B., Li, Y., and Song, Z. (2017). Enhanced rutin accumulation in tobacco leaves by overexpressing the NtFLS2 gene. *Bioence Biotechnol. Biochem.* 81 (9), 1721–1725. doi:10.1080/09168451.2017.1353401

Tom, Y., Devamanyu, H., Soujanya, P., and Cambria, E. (2018). Recent trends in deep learning based Natural Language Processing. *IEEE Comput. Intell. Mag.* 13 (3), 55–75. doi:10.1109/mci.2018.2840738

UzmaAl-Obeidat, F., Tubaishat, A., Shah, B., and Halim, Z. (2022). Gene encoder: A feature selection technique through unsupervised deep learning-based clustering for large gene expression data. *Neural Comput. Applic* 34, 8309–8331. doi:10.1007/s00521-020-05101-4

Wang, C., Ma, X., and Chen, J. (2018). Information extraction and knowledge graph construction from geoscience literature. *Comput. Geosciences* 112, 112–120. doi:10.1016/j.cageo.2017.12.007

Wi, C. L., Sohn, S., Rolfes, M. C., Seabright, A., Ryu, E., Voge, G., et al. (2017). Application of a Natural Language Processing algorithm to asthma ascertainment: An automated chart review. *Am. J. Respir. Crit. Care Med.* 196 (4), 430–437. doi:10.1164/rccm.201610-2006OC

Wong, A., Plasek, J. M., Montecalvo, S. P., and Zhou, L. (2018). Natural Language processing and its implications for the future of medication safety: A narrative review of recent advances and challenges. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* 38 (8), 822–841. doi:10.1002/phar.2151

Xu, L. B., Zhao, Z. G., Xu, S. F., Zhang, X. X., Liu, T., Jing, C. Y., et al. (2020). The landscape of gene mutations and clinical significance of tumor mutation burden in patients with soft tissue sarcoma who underwent surgical resection and received conventional adjuvant therapy. *Int. J. Biol. Markers* 35 (3), 14–22. doi:10.1177/1724600820925095

Zhang, C., Miao, Z., Xiao, H., Hu, Z., and Ji, Y. (2017). Knowledge graph embedding for hyper-relational data. *Tsinghua Sci. Technol.* 22 (2), 185–197. doi:10.23919/tst.2017.7889640

Zhu, Y., Zhou, W., Xu, Y., Liu, J., and Tan, Y. (2017). Intelligent learning for knowledge graph towards geological data. *Sci. Program.* 2017 (1), 1–13. doi:10.1155/2017/5072427

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 25 January 2023
DOI 10.3389/fgene.2023.1113095

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Saba Battelino,
University Medical Centre Ljubljana, Slovenia
Xiaofeng Li,
Heilongjiang International University, China
Defei Liu,
Southwest University, China
Nengyun Zhang,
Ningbo University, China

*CORRESPONDENCE
Min Huang,
huangm7709@163.com

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 01 December 2022
ACCEPTED 06 January 2023
PUBLISHED 25 January 2023

CITATION
Wu X, Huang M, Huang W, Zhao S, Xie J, Liu G and Chang S (2023), Preliminary investigation of the diagnosis and gene function of deep learning PTPN11 gene mutation syndrome deafness.
Front. Genet. 14:1113095.
doi: 10.3389/fgene.2023.1113095

Preliminary investigation of the diagnosis and gene function of deep learning PTPN11 gene mutation syndrome deafness

Xionghui Wu[1], Min Huang[1]*, Weiqing Huang[2], Sijun Zhao[1], Jiang Xie[1], Guangliang Liu[1] and Shuting Chang[2]

[1]Department of Otorhinolaryngology Head and Neck Surgery, Hunan Children's Hospital, Changsha, Hunan, China, [2]Department of Neonatology, Hunan Children's Hospital, Changsha, Hunan, China

Syndromic deafness caused by PTPN11 gene mutation has gradually come into the public's view. In the past, many people did not understand its application mechanism and role and only focused on non-syndromic deafness, so the research on syndromic deafness is not in-depth and there is a large degree of lack of research in this area. In order to let the public know more about the diagnosis and gene function of deafness caused by PTPN11 gene mutation syndrome, this paper used deep learning technology to study the diagnosis and gene function of deafness caused by syndrome with the concept of intelligent medical treatment, and finally drew a feasible conclusion. This paper provided a theoretical and practical basis for the diagnosis of deafness caused by PTPN11 gene mutation syndrome and the study of gene function. This paper made a retrospective analysis of the clinical data of 85 deaf children who visited Hunan Children's Hospital,P.R. China from January 2020 to December 2021. The conclusion were as follows: Children aged 1–6 years old had multiple syndrome deafness, while children under 1 year old and children aged 6–12 years old had relatively low probability of complex deafness; girls were not easy to have comprehensive deafness, but there was no specific basis to prove that the occurrence of comprehensive deafness was necessarily related to gender; the hearing loss of patients with Noonan Syndrome was mainly characterized by moderate and severe damage and abnormal inner ear and auditory nerve; most of the mutation genes in children were located in Exon1 and Exon3, with a total probability of 57.65%. In the course of the experiment, it was found that deep learning was effective in the diagnosis of deafness with PTPN11 gene mutation syndrome. This technology could be applied to medical diagnosis to facilitate the diagnosis and treatment of more patients with deafness with syndrome. Intelligent medical treatment was also becoming a hot topic nowadays. By using this concept to analyze and study the pathological characteristics of deafness caused by PTPN11 gene mutation syndrome, it not only promoted patients to find diseases in time, but also helped doctors to diagnose and treat such diseases, which was of great significance to patients and doctors. The study of PTPN11 gene mutation syndrome deafness was also of great significance in genetics. The analysis of its genes not only enriched the gene pool, but also provided reference for future research.

KEYWORDS

syndrome deafness, deep learning, intelligent medicine, PTPN11 gene mutation, diagnosis and gene function

1 Introduction

The diagnosis of deafness in PTPN11 gene mutation syndrome has not been studied too deeply at present. Some scholars only analyzed non-comprehensive deafness, but they did not analyze the situation of comprehensive deafness, which also led to that people were not familiar with syndrome deafness and did not often pay attention to it. In order to remedy this defect, it is of practical significance to study the diagnosis and gene function of deafness caused by PTPN11 gene mutation syndrome.

Deafness is a multiple disease, which has been studied by some scholars. Xu H Y analyzed the relationship between PTPN11 gene and deafness (Xu et al., 2019). Gao Y analyzed the clinical phenotype and gene of syndromic deafness with PTPN11 gene mutation (Gao et al., 2022). GaoXue believed that congenital sensorineural hearing loss was the initial manifestation of PTPN11 related NS, which was accompanied by multiple spots or NS (Gao et al., 2021). Huang S S analyzed the case report and diagnosis of multi spot NS with deafness as the main clinical feature (Huang et al., 2019). Wu Jie used targeted genome sequencing to make molecular diagnosis for a large number of hearing loss people (Wu et al., 2022). Mutai Hideki analyzed the whole exome analysis of Japanese hearing loss patients, which showed a high degree of heterogeneity between responsible and new candidate genes (Mutai et al., 2022). Cesca Federica analysised of the frequency of Usher gene mutations in non-syndromic hearing loss (Cesca et al., 2020). These studies on deafness have not yet involved PTPN11 gene mutation syndrome.

There are also many researches on deep learning in the field of deafness. Bing D adopted the prognosis classification of sudden deafness based on deep learning method (Bing et al., 2018). Huang Xin used machine learning and deep learning methods to predict the ototoxicity caused by drugs in silicon (Huang et al., 2021). Waldo Nogueira remixed the cochlear implant users with the deep learning model (Gajęcki and Nogueira, 2018). However, there are few researches on the diagnosis and gene function of PTPN11 gene mutation syndrome deafness in the field of deep learning in deafness disease.

In order to enhance the public's understanding of comprehensive deafness, this paper analyzed the data of 85 deaf children who visited Hunan Children's Hospital, P.R. China from January 2020 to December 2021, and analyzed the patient's age, gender, hearing loss, imaging examination, deafness gene and number of NS patients. Finally, the feasibility conclusion was drawn.

2 Investigation of comprehensive deafness

Deafness is a common disease that affects health and human normal life (KimOh et al., 2020), which affects 1%–3% of newborns. Deafness can be divided into two categories: syndromic deafness and non-syndromic deafness. Non-syndromic deafness is mainly caused by genetic deafness caused by deafness gene mutation. About 70% are non-syndromic deafness, and the remaining 30% are syndromic deafness. Syndromic deafness refers to the problem of other organs of the body besides hearing loss. Common comprehensive deafness includes the following types, as shown in Figure 1:

2.1 Bird face syndrome

Bird face syndrome is a syndrome of mandibular dysplasia and deafness caused by embryonic agenesis of the first and second

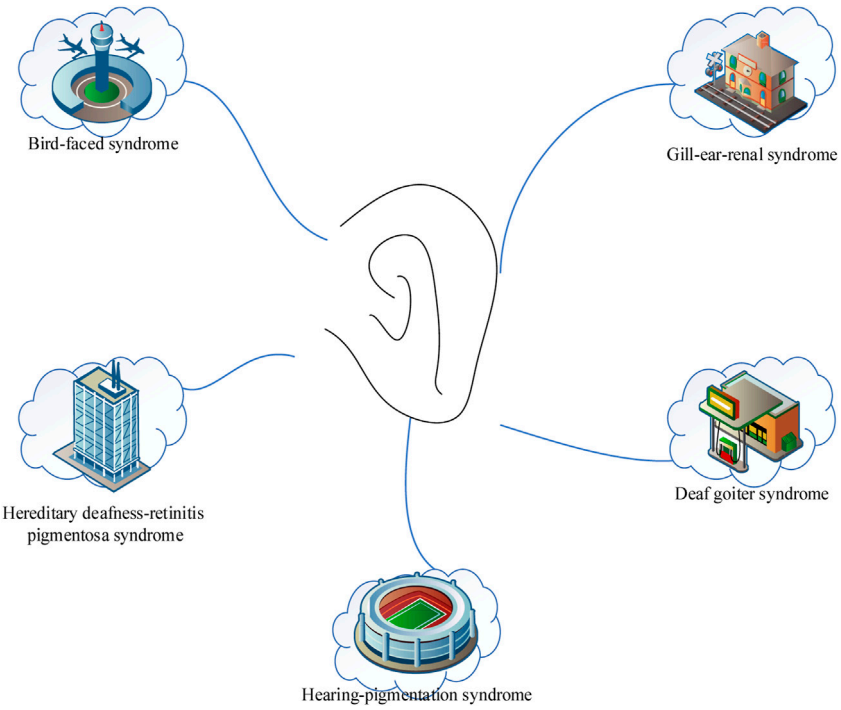


FIGURE 1
Common comprehensive deafness.

branchial arches (Barsottini et al., 2019). The syndrome is characterized by drooping eyelids, hypoplasia of facial bones (especially mandible and zygoma), malformations of the outer and middle ears, and conductive deafness. The vast majority of cases are autosomal dominant inheritance, and up to 70% of cases are sporadic cases. There is no clear family history, but married patients with children have a 53% chance of their children suffering from this disease.

2.2 Hereditary deafness-retinitis pigmentosa syndrome

Hereditary deafness-retinitis pigmentosa syndrome is a hereditary disease characterized by congenital sensorineural hearing loss and progressive retinitis pigmentosa (usually from late childhood to adolescence), resulting in visual loss (Deep et al., 2021). The typical clinical symptoms of hereditary deafness and retinitis syndrome are hearing loss and retinitis pigmentosa. Hereditary deafness and retinitis syndrome can be divided into three types according to the degree of hearing loss and the incidence and degree of vision loss. Among them, type I and type II are the most common (about 34%–44% respectively), while type III is relatively rare (12%–32%). The clinical features of hereditary deafness-retinitis pigmentosa syndrome type I include severe congenital deafness, language disorder, vestibular loss or abnormality. The onset age is less than 10 years old, and sometimes there are ataxia, intellectual disability and pseudopsychosis. Hereditary deafness retinitis pigmentosa syndrome type II is characterized by moderate congenital hearing loss and normal vestibular function. This often occurs between the ages of 20 and 30, without language disorders, ataxia, mental retardation or psychosis.

2.3 Hearing-pigment syndrome

Hearing-pigment syndrome is characterized by sensorineural hearing loss and abnormal pigmentation (abnormal pigmentation of eyes, skin and hair) (Taneja, 2020). The syndrome is also characterized by abnormal development of the middle nose and forehead, including ectopia of the inner canthus (wide eye distance), eyebrows, broad bridge of the nose and other abnormalities of the middle nose and forehead, as well as skeletal muscle defects in the limbs and congenital megacolon.

Clinical manifestations include inner ear, iris, skin pigment, hair pigment, epicanthus, limb skeletal muscle dysplasia and megacolon. Among them, the inner ear is deformed, and vestibular cistern and semicircular canal are the most common malformations. Some patients may also have cochlear malformations, which are characterized by flattening and slight bending of the cochlea. Skin pigmentation is defined as a large number of brown freckles on the face, neck and/or trunk and limbs. In some cases, there are local white patches on the skin. Hair pigment disorder is manifested as white hair on the forehead or less white hair, which is also manifested as premature lightening of eyebrows and body hair. Ectopia of inner canthus is characterized by wide eye distance, which may be accompanied by abnormalities in the middle of nose and forehead, such as eyebrow linking, broad bridge of nose, etc. Abnormalities of skeletal muscles of the extremities are manifested in muscular dysplasia of the extremities and contracture of the elbow joint (finger joint). Hirschsprung's disease is congenital Hirschsprung's disease or gastrointestinal atresia.

2.4 Deafness goiter syndrome

Most patients with deafness and thigh syndrome have severe to extreme hearing loss at birth (Kral and Sato., 2020). In a few cases, hearing loss occurs in the first few years after birth. After a cold or serious physical shock, hearing loss develops from serious to deep. Goiter is an important feature of deaf goiter. About 80% of patients with deaf goiter syndrome have goiter. It usually occurs in adolescence and is most severe between the ages of 21 and 31. The size of the goiter is not related to the degree of hearing loss.

2.5 Gill-ear-kidney syndrome

Gill-ear-kidneysyndrome, also known as hereditary progressive nephritis, is characterized by hematuria, proteinuria and progressive hypoalgesia. Some patients may have extrarenal manifestations, such as hearing loss, eye abnormalities, and tumors of esophageal smooth muscle.

3 PTPN11 gene and Noonan Syndrome

PTPN11 gene encodes Src homology two domain-containing protein tyrosine phosphatase (SHP2), which is located on chromosome 12 (12q24). It is expressed in most fetal and adult tissues and plays a key role in cell proliferation, differentiation, survival and death. Diseases associated with mutations in PTPN11 include NS and Noonan syndrome with multiple lenses (NS-ML). NS is a disease with autosomal dominant inheritance. Typical cases are between 1/2,500 and 1/1,000. Both men and women can occur, which often has a family history. In the neonatal period, the main manifestation is the appearance abnormality: The head circumference is large, and the face is small. The forehead is high, and the hair is sparse; the distance between the eyes is large, and the eyeballs are prominent. The inner eye is ridge, and the upper eyelid is drooping. The cleft palate is horizontal or downward inclined; the nose is short, and the bridge of the nose is low. The tip of the nose is full; the ear is low and the earlobe is oval; the midline is deep, and the upper lip is full. The upper jaw is high, and the lower jaw is small; the neck is short, and the neck is arched. The skin is loose, and the hair line on the back is low.

PTPN11 is the first known pathogenic gene and the most common NS gene, accounting for about 50% of all cases. SHP2 is a protein product of PTPN11 and a non-receptor protein tyrosine phosphatase, which is composed of the N and C terminal domains of SH2 and the catalytic domain of protein tyrosine phosphatase. Most mutations of PTPN11 leading to NS are located in residues related to the interaction between the N-terminal SH2 domain of SHP-2 and the domain of protein tyrosine phosphatase. Mutations in this region destroy the stability of the catalytic inactive form of SHP2 and the ability of the protein to switch from the active conformation to the inactive conformation, which leads to increased signal transmission through the RAS-MAPK pathway.

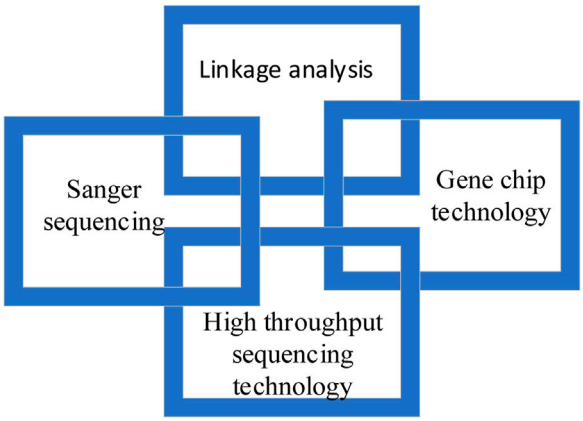


FIGURE 2
Common molecular diagnostic techniques for deafness.

SHP2 is involved in the regulation of RAS-MAPK signal pathway and the expression of various biological factors and cytokines, such as growth hormone, insulin-like growth factor and fibroblast growth factor. These factors cause various clinical symptoms in children with CF, including growth retardation and bone deformity. Patients with PTPN11 mutation are more likely to have short stature, thoracic malformation, cryptorchidism and mild bruising than those without PTPN11 mutation. N-methyl-D-aspartate receptors are critical for synaptic plasticity, learning, and memory. Tyrosine phosphorylation of N-methyl-D-aspartate receptor subunit regulates the function of N-methyl-D-aspartate receptor. SHP2 dephosphorylates SN related glutamate receptors, thereby regulating the function of N-methyl-D-aspartate receptors and leading to cognitive impairment of SN. Imaging showed that the volume of gray matter on both sides of the striatum decreased, and the area of the temporal region decreased. The thickness of the frontal cortex increased, while the thickness of the marginal cortex decreased, including the marginal structures that are important to the hippocampus. Hippocampus affects the attention and memory during the development of human brain, which is the expression of severe attention deficit, hyperactivity and memory disorder. The existence of PTPN11 mutation in NS patients is closely related to their hematoma susceptibility. The disorder of SHP2 peroxidase would increase the adverse reaction of platelets and the phenotypic coagulation dysfunction. The activated PTPN11 mutation plays a leukemic role in the stem cell microenvironment. Therefore, the over activation of interleukin 1β and other proinflammatory cytokines that can be produced by hematopoietic stem cell monocytes leads to the deterioration of myeloproliferative tumors and donor cell mediated myeloproliferative tumors in children after stem cell transplantation.

4 Common molecular diagnostic techniques for deafness

Common molecular diagnosis technologies for deafness include linkage analysis, gene chip technology, Sanger sequencing and high-throughput sequencing technology, as shown in Figure 2.

4.1 Linkage analysis

Linkage analysis is a classical method to study single gene disease. Genes are arranged linearly on chromosomes, while linkage analysis is based on the principle that different genes are linked in linkage groups. That is to say, the gene of interest is linked with another gene or genetic marker on the same chromosome, so as to find the disease causing gene. However, linkage analysis has some limitations. One is that it is not suitable for finding mutations in small families, because mutations in small families are difficult to separate. The other is that linkage analysis usually involves millions of chromosomes containing hundreds of genes and base pairs, which cannot be used to detect diseases. In order to remedy the defects of linkage analysis, gene chip technology is developed.

4.2 Gene chip technology

In recent years, gene chip technology has developed rapidly, which is not only limited to DNA detection, but also widely used in non-coding RNA, expression profiling, methylation and other fields. In the field of hereditary hearing loss, the chip hybridization of oligoribonucleic acid is usually used to detect hearing loss genes in the population in clinical practice, and the three-step method of multiple PCR amplification, chip hybridization and result digitization is used. Due to the high throughput, accuracy and low cost, gene chip technology for deafness is widely used in clinical practice.

4.3 Sanger sequencing

Sanger sequencing is also known as “first generation sequencing”. It uses DNA polymerase to lengthen the primers attached to the template sequence until the fluorescent labeled dideoxynucleotides enter the polymerization reaction, so as to stop the polymerization reaction and produce DNA fragments of different lengths. Sanger sequencing has the advantage of high precision, but it is gradually being replaced by high-throughput sequencing due to its short reading time, high cost and low throughput.

4.4 High throughput sequencing technology

High throughput sequencing technology is also known as “next-generation”sequencing technology. This is an important milestone of human genomics, which has reduced the cost and time of human genome design and is now widely used in clinical and scientific research.

5 Application of deep learning and intelligent medicine in medicine

5.1 Deep learning

At present, deep learning technology has been applied in many fields, and gradually developed from the initial simple applications such as processing still images and filling in text information to large-

TABLE 1 Hearing loss in children under 2 years of age level of hearing loss.

Sound intensity	Level of hearing impairment
≤ 30dBnHL	Normal
31–50dBnHL	Mild hearing loss
51–70dBnHL	Moderate hearing loss
71–90dBnHL	Severe hearing loss
>90dBnHL	Very severe hearing loss

TABLE 2 Hearing loss in children over 2 years of age level of hearing loss.

Sound intensity	Level of hearing impairment
≤ 25dB HL	Normal
25–40dB HL	Mild hearing loss
41–60dB HL	Moderate hearing loss
61–80dB HL	Severe hearing loss
> 80dBHL	Very severe hearing loss

scale applications such as processing video and audio data and from incomplete information to creative work. Based on big data, artificial intelligence can quickly deduce the general laws of underlying data processing and preprocessing methods by analyzing previous data and preprocessing methods, and can process and integrate these laws. It can also learn the processing principles, and create new work on this basis. The surface creative activities include painting, writing, language production, music creation, etc. High level activities may include intelligent management, situation assessment, strategic development, etc.

Deep learning is used to solve many problems with big data, such as computer vision, pattern recognition and natural language processing. Artificial intelligence technology based on deep learning is being integrated into clinical practice and medical care. At present, it is widely used in automatic medical image analysis and auxiliary diagnosis, such as classification of benign and malignant lung tumors and detection of benign and malignant breast lesions. It is also analyzing and testing bone, cartilage and other structures in the musculoskeletal field. The accuracy of diagnosis can be improved through the joint analysis of several test equipment; even electrocardiogram analysis can supplement the diagnosis of dysfunction, such as arrhythmia and cognitive impairment, by analyzing arrhythmia, cognitive impairment and other health problems.

5.2 Intelligent medical treatment

Smart medicine is to realize the interconnection between patients and medical personnel, medical facilities and medical equipment by establishing the medical records of the regional health information platform and using the latest deep learning technology, so as to gradually realize continuous informatization (Srinivasa et al., 2018). In the near future, more and more artificial intelligence, sensor technology and other advanced technologies would be introduced

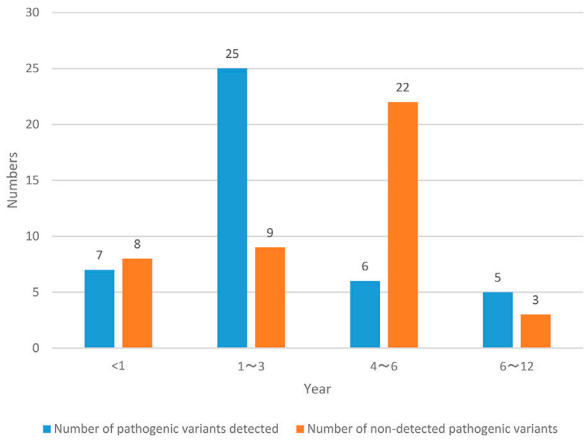


FIGURE 3
Age profile of patients.

into the medical and health industry, which would make medical and healthcare become truly intelligent and contribute to the successful development of medical and health undertakings.

5.2.1 Medical equipment and drug information query

The barcode or two-dimensional code on the medical device or package is printed by the laser printer, so that the purchaser can easily check the authenticity of the medical device and drugs, and understand the use of the medical device and the recommended manufacturer, so as to obtain drug information, product, date, and series, and query drug sales information (Gopal et al., 2019; Sodhro et al., 2020). This information can effectively help buyers understand medical equipment.

5.2.2 Monitoring of medical waste disposal

Medical waste is different from ordinary waste, so the treatment of medical waste has a complete process: Deep learning technology can track the process of medical waste from transportation to treatment in real time, which effectively prevents violations in medical waste treatment.

5.2.3 Digitization of medical information

Deep learning technology can also promote the rapid development of mobile medicine. With intelligent medical equipment, people's important health indicators can be regularly uploaded to the medical information center. The information center can establish a health map for everyone, and monitor people's health indicators, so as to report people's physical conditions through mobile communication and provide relevant suggestions on people's lifestyle, nutrition, etc.

5.2.4 Telemedicine

Deep learning technology can upload patient information from different regions or even countries to the data center. For some severe patients or complex cases, authorized expert advisory teams can provide remote or virtual consultation, so that patients can receive timely, advanced and scientific treatment. In rural areas with low level of medical care and infrastructure, telemedicine based on deep learning technology can improve the level of rural medical care.

TABLE 3 Diagnostic criteria for Noonan syndrome.

Clinical features	Major	Secondary
Facial	Typical face	Suggestive face
Cardiac	Pulmonary stenosis and/or ECG abnormalities	Other abnormalities
Height	< P3	< P10
Chest wall	Pectus excavatum or funnel chest	Thoracic widening
First degree relatives	Confirmed diagnosis of Noonan's syndrome	Prompt Noonan syndrome
Other	Intellectual disability, cryptorchidism and lymphatic duct dysplasia present together	Intellectual disability, cryptorchidism and lymphatic duct dysplasia

6 Experimental evaluation of gene function of deafness caused by PTPN11 gene mutation syndrome

6.1 General information of the research object

At present, most of the deaf patients with PTPN11 gene mutation syndrome are young children, so this paper conducted a retrospective analysis of the clinical data of 85 deaf children who visited Hunan Children's Hospital,P.R. China from January 2020 to December 2021. All subjects received detailed medical history and general physical examination, as well as systematic hearing test, inner ear X-ray examination and coronal thin slice computed tomography, so as to exclude hearing loss caused by unilateral deafness and otitis media, meningitis, occupational injury and trauma. After these comprehensive assessments, the study participants were diagnosed with bilateral sensorineural hearing loss.

6.2 Hearing assessment of patients

Listening assessment is based on the principles of test matching and cross validation. Auditory physiology and behavioral audiometry were conducted, namely, brainstem evoked response, stable auditory evoked response, paradoxical otoacoustic emission, acoustic conductance and behavioral audiometry. The results were confirmed at the end of the auditory evaluation, and there was a good correlation between the test and retest. The hearing loss of children under 2 years old is determined according to the threshold of V wave response, and the hearing loss of children over 2 years old is determined according to the average of the four frequencies (500 Hz, 1,000 Hz, 2,000 Hz, and 4,000 Hz) felt in the behavioral audiometry test (Shoham et al., 2019). The assessment principles for hearing loss of children under 2 years old can be shown in Table 1.

The assessment principles for hearing loss of children over 2 years old can be shown in Table 2.

6.3 Methods of gene detection

Genomic DNA is extracted from peripheral venous blood of children and their parents, and high-throughput full exome sequencing is performed using Illumina NovaSeq6000 sequencing platform. Under the 150 PE sequencing mode, Agilent V6 chip is used to extract about 20,000 genes from the human genome and enrich them in the exon region and adjacent transverse regions. At least 10 Gb–12 Gb of data is generated from each sample.

The criteria for evaluating NS are shown in Table 3.

6.4 Experimental results of gene function in deafness with PTPN11 gene mutation syndrome

6.4.1 Patient's age

NS patients are divided into two groups according to the number of pathogenic variants detected and the number of pathogenic variants not detected, and the age status of NS patients is recorded in Figure 3.

The NS patients investigated in this paper were all 12 years old and below. Among them, the number of NS patients less than 1 year old detected pathogenic variants was 7, with a probability of 8.24% (7/85); the number of NS patients in the age range of 1–3 years old was 25, with a probability of 29.41% (25/85); the number of NS patients in the age range of 4–6 years old was 6, with a probability of 7.06% (6/85); the number of NS patients aged 6–12 years old with pathogenic variants was 5, with a probability of 5.88% (5/85); the number of NS patients less than 1 year old without pathogenic variation was 8, with a probability of 9.41% (8/85); the number of NS patients aged 1–3 years without pathogenic variation was 9, with a probability of 10.59% (9/85); the number of NS patients aged 4–6 years without pathogenic variation was 22, with a probability of 25.88% (22/85); the number of NS patients aged 6–12 years old with pathogenic variants was 3, with a probability of 3.53% (3/85). The number of NS patients aged from 1 to 3 years with pathogenic variation detected and the number of NS patients aged from 4 to 6 years without pathogenic variation detected are the largest, and the probability is also high.

6.4.2 Patient's gender

The results of gender status of NS patients are shown in Figure 4.

49 NS children were boys, with a probability of 57.65% (49/85). Among them, the number of pathogenic variants detected was 23, with a probability of 27.06% (23/85); the number of people without pathogenic variation was 26, with a probability of 30.59% (26/85). 36 NS children were girls, with a probability of 42.35% (36/85). Among them, the number of pathogenic variants detected was 20, with a probability of 23.53% (20/85); the number of pathogenic variants detected was 16, with a probability of 18.82% (16/85). In general, boys are prone to NS, and the probability of detecting pathogenic variants is lower than that of not detecting pathogenic variants.

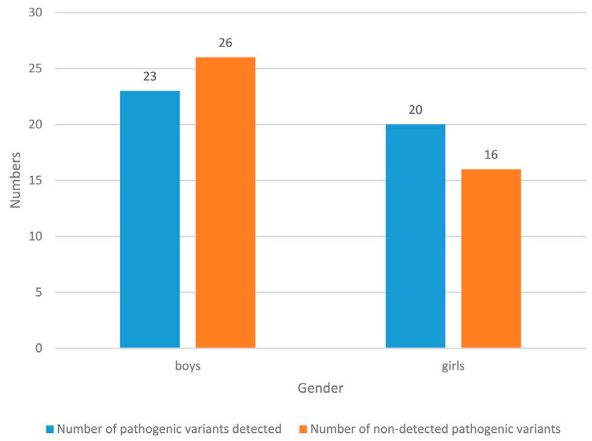


FIGURE 4
Gender status of patients.

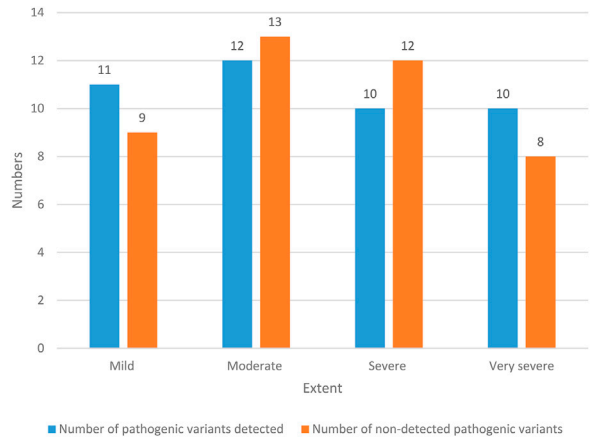


FIGURE 5
Hearing impairment of patients.

6.4.3 Hearing loss of patients

The hearing loss of NS patients is shown in Figure 5.

There were 20 NS patients with mild hearing loss, and the probability was 23.53% (20/85). Among them, 11 NS patients with mild hearing impairment detected pathogenic variation, with a probability of 12.94% (11/85); there were nine patients with mild hearing impairment NS who had no pathogenic variation, with a probability of 10.59% (9/85). There were 25 NS patients with moderate hearing loss, with a probability of 29.41% (25/85). Among them, 12 NS patients with moderate hearing loss detected pathogenic variation, with a probability of 14.12% (12/85). There were 13 NS patients with moderate hearing loss who had no pathogenic variation, with a probability of 15.29% (13/85). There were 22 NS patients with severe hearing loss, with a probability of 25.88% (22/85). Among them, 10 NS patients with severe hearing impairment had pathogenic variants, with a probability of 11.76% (10/85). There were 12 NS patients with severe hearing impairment who had no pathogenic variation, with a probability of 14.12% (12/85). There were 18 NS patients with extremely severe hearing loss, with a probability of 21.18% (18/85). Among them, 10 patients with extremely severe hearing impairment NS had pathogenic variants, with a probability of 11.76% (10/85). There were 8 NS patients with extremely severe hearing impairment who had no pathogenic variation, with a probability of 9.42% (8/85). In general, the hearing loss of NS patients is mostly moderate and severe.

6.4.4 Imaging examination

The imaging findings of 85 NS patients are recorded in Figure 6.

The imaging findings of 48 NS patients were abnormal in the inner ear and auditory nerve, with a probability of 56.47% (48/85). Among them, pathogenic variants were detected in 26 NS patients with abnormal inner ear and auditory nerve, with a probability of 30.59% (26/85). There were 22 NS patients with abnormal inner ear and auditory nerve, and no pathogenic variation was detected, with a probability of 25.88% (22/85). There were 37 NS patients with no abnormal imaging results, and the probability was 43.53% (37/85). In these NS patients, pathogenic variants were detected in 17 patients, with a probability of 20% (17/85). There were 20 patients without pathogenic variation, and the probability was 23.53% (20/85). It can be seen that NS patients often have abnormal inner ear and auditory nerve.

6.4.5 Deafness gene and number of cases in NS patients

The deafness genes and cases of 85 NS patients are recorded in Table 4.

Among 85 NS children with PTPN11 gene mutation, 15 children had a gene mutation of Exon13, and the nucleotide mutation was c.1510A>G. The amino acid variation was p.M504V, with a probability of 17.65% (15/85); there were 24 children with Exon1 gene mutation, and the nucleotide mutation was c.10C>G. The amino acid variation was p.R4G, with a probability of 28.24% (24/85); the gene mutation of 25 children was Exon3, and the nucleotide variation was c.236A>G. The variation of amino acid was p.Q79R, and the probability was 29.41% (25/85); the gene mutation position of 21 children was Exon11, and the nucleotide variation was c.1277A>G. The amino acid variation was p.H426R, with a probability of 24.7% (21/85). In general, most of the mutation genes in children are located in Exon1 and Exon3, with a combined probability of 57.65%.

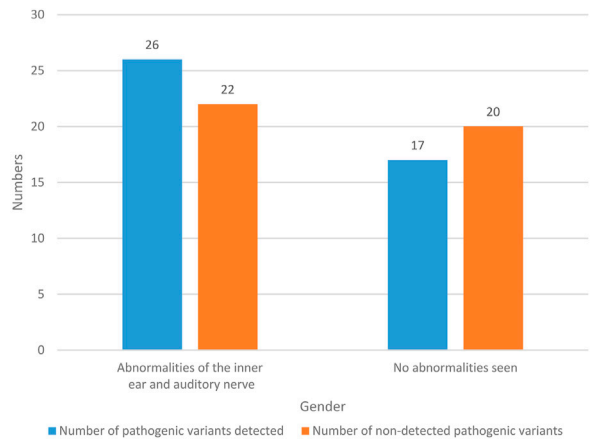


FIGURE 6
Imaging studies.

TABLE 4 Deafness causative genes and number of cases in NS patients.

Gene	Type of mutation	Mutation location	Nucleotide variant	Amino acid variant	Source	Number
PTPN11	Missense mutation	Exon13	c.1510A>G	p.M504V	Spontaneous	15
PTPN11	Missense mutation	Exon1	c.10C>G	p.R4G	Spontaneous	24
PTPN11	Missense mutation	Exon3	c.236A>G	p.Q79R	Spontaneous	25
PTPN11	Missense mutation	Exon11	c.1277A>G	p.H426R	Spontaneous	21

7 Conclusion

In order to study the diagnosis and gene function of deafness caused by PTPN11 gene mutation syndrome, this paper first introduced the syndrome deafness, and then analyzed the PTPN11 gene and Noonan syndrome. The common molecular diagnostic techniques for deafness were introduced, and the application of deep learning and intelligent medicine in medical treatment was analyzed. Finally, the gene function of deafness caused by PTPN11 gene mutation syndrome was analyzed, and the conclusion was drawn. The number of NS patients aged from 1 to 3 years with pathogenic variants detected and the number of NS patients aged from 4 to 6 years without pathogenic variants detected were the largest. Boys were prone to NS, and the probability of the number of people with pathogenic variants detected was lower than that of the number of people without pathogenic variants detected. The hearing loss of NS patients mostly showed moderate and severe damage, and NS patients often had abnormal inner ear and auditory nerve. Most of the mutation genes of the children were located in Exon1 and Exon3. Deep learning technology had an excellent application in analyzing the gene of NS patients.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval for this study and written informed consent from the participants of the study were not required in accordance with local legislation and national guidelines. This study does not violate and does not involve moral and ethical statement.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This study was supported by the department of Science and Technology of Hunan Province, Major Project, Hunan Science Project (2019) No. 49 (2019SK1015).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Barsottini, O. G., Pedroso, J. L., Martins, C. R., Franca, M. C., and Albernaz, P. M. (2019). Deafness and vestibulopathy in cerebellar diseases: A practical approach. *Cerebellum* 18 (6), 1011–1016. doi:10.1007/s12311-019-01042-4

Bing, D., Ying, J., Lan, L., Guan, J., Xie, L. Y., Zhao, L. D., et al. (2018). Classification of sudden deafness prognosis based on deep learning method. *Lin Chuang er bi yan hou tou Jing wai ke za zhi= J. Clin. Otorhinolaryngology, Head, Neck Surg.* 32 (15), 1125–1129. doi:10.13201/j.issn.1001-1781.2018.15.001

Cesca, F., Bettella, E., Polli, R., Leonardi, E., Aspromonte, M. C., Sicilian, B., et al. (2020). Frequency of usher gene mutations in non-syndromic hearing loss: Higher variability of the usher phenotype. *J. Hum. Genet.* 65 (10), 855–864. doi:10.1038/s10038-020-0783-1

Deep, N. L., Gordon, S. A., Shapiro, W. H., Waltzman, S. B., Roland, J. T., Jr, and Friedmann, D. R. (2021). Cochlear implantation in children with single-sided deafness. *Laryngoscope* 131 (1), E271–E277. doi:10.1002/lary.28561

Gajęcki, T., and Nogueira, W. (2018). Deep learning models to remix music for cochlear implant users. *J. Acoust. Soc. Am.* 143 (6), 3602–3615. doi:10.1121/1.5042056

Gao, X., Huang, S. S., Qiu, S. W., Su, Y., Wang, W. Q., Xu, H. Y., et al. (2021). Congenital sensorineural hearing loss as the initial presentation of PTPN11-associated noonan syndrome with multiple lentigines or noonan syndrome: Clinical features and underlying mechanisms. *J. Med. Genet.* 58 (7), 465–474. doi:10.1136/jmedgenet-2020-106892

Gao, Y., Li, Z. C., Ma, X. L., Xiao, Y., Dai, X., Mac, J., et al. (2022). The clinical phenotype and gene analysis of syndromic deafness with PTPN11 gene mutation. *Zhonghua er bi yan hou tou Jing wai ke za zhi= Chin. J. Otorhinolaryngology Head Neck Surg.* 57 (3), 317–323. doi:10.3760/cma.j.cn15330-20210525-00294

Gopal, G., Suter-Crazzolara, C., Toldo, L., and Eberhardt, W. (2019). Digital transformation in healthcare–architectures of present and future information technologies. *Clin. Chem. Laboratory Med. (CCLM)* 57 (3), 328–335. doi:10.1515/cclm-2018-0658

Huang, S. S., Huang, B. Q., Gao, X., Yuan, Y. Y., Su, Y., Wang, G. J., et al. (2019). Case report and diagnosis of Noonan syndrome with multiple lentigines with deafness as its main clinical feature. *Lin Chuang er bi yan hou tou Jing wai ke za zhi= J. Clin. Otorhinolaryngology, Head, Neck Surg.* 33 (9), 804–807. doi:10.13201/j.issn.1001-1781.2019.09.003

Huang, X., Tang, F., Hua, Y., and Li, X. (2021). *In silico* prediction of drug-induced ototoxicity using machine learning and deep learning methods. *Chem. Biol. Drug Des.* 98 (6), 248–257. doi:10.1111/cbdd.13894

KimOh, B. J. D-Y., Han, J. H., Oh, J., Kim, M. Y., Park, H. R., Seok, J., et al. (2020). Significant Mendelian genetic contribution to pediatric mild-to-moderate hearing loss and its comprehensive diagnostic approach. *Genet. Med.* 22 (6), 1119–1128. doi:10.1038/s41436-020-0774-9

Kral, A., and Sato., M. (2020). Nature and nurture in hearing: Critical periods for therapy of deafness. *Acoust. Sci. Technol.* 41 (1), 54–58. doi:10.1250/ast.41.54

Mutai, H., Momozawa, Y., Kamatani, Y., Nakano, A., Sakamoto, H., Takiguchi, T., et al. (2022). Whole exome analysis of patients in Japan with hearing loss reveals high heterogeneity among responsible and novel candidate genes. *Orphanet J. rare Dis.* 17 (1), 114–117. doi:10.1186/s13023-022-02262-4

Shoham, N., Lewis, G., Favarato, G., and Cooper, C. (2019). Prevalence of anxiety disorders and symptoms in people with hearing impairment: A systematic review. *Soc. psychiatry psychiatric Epidemiol.* 54 (6), 649–660. doi:10.1007/s00127-018-1638-3

Sodhro, A. H., Malokani, A. S., Sodhro, G. H., Muzammal, M., and Zongwei, L. (2020). An adaptive QoS computation for medical data processing in intelligent healthcare applications. *Neural Comput. Appl.* 32 (3), 723–734. doi:10.1007/s00521-018-3931-1

Srinivasa, K. G., Sowmya BjShikhar, A., Utkarsha, R., and Singh, A. (2018). Data analytics assisted internet of things towards building intelligent healthcare monitoring systems: Iot for healthcare. *J. Organ. End User Comput. (JOEUC)* 30 (4), 83–103. doi:10.4018/joeuc.2018100106

Taneja, M. K. (2020). Prevention and rehabilitation of old age deafness. *Indian J. Otolaryngology Head Neck Surg.* 72 (4), 524–531. doi:10.1007/s12070-020-01856-3

Wu, J., Cao, Z., Su, Y., Wang, Y., Cai, R., Chen, J., et al. (2022). Molecular diagnose of a large hearing loss population from China by targeted genome sequencing. *J. Hum. Genet.* 67 (11), 643–649. doi:10.1038/s10038-022-01066-5

Xu, H. Y., Yuan, Y. Y., and Dai, P. (2019). PTPN11 and the deafness. *Lin Chuang er bi yan hou tou Jing wai ke za zhi= J. Clin. Otorhinolaryngology, Head, Neck Surg.* 33 (9), 830–834. doi:10.13201/j.issn.1001-1781.2019.09.008

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 February 2023
DOI 10.3389/fgene.2023.1126212

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Yitong Niu,
Sofia University, United States
Xiaofeng Li,
Heilongjiang International University, China

*CORRESPONDENCE
Xionghui Wu,
w13787186439@163.com
Sijun Zhao,
zhaosj3991@163.com
Weiqing Huang,
15873167512@163.com
Min Huang,
3150612129@qq.com
Jiang Xie,
xiejiang2292@qq.com
Guangliang Liu,
1184847276@qq.com
Shuting Chang,
changst2008@163.com

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 17 December 2022
ACCEPTED 23 January 2023
PUBLISHED 08 February 2023

CITATION
Wu X, Zhao S, Huang W, Huang M, Xie J, Liu G and Chang S (2023), Relationship between single nucleotide polymorphism of NOS2 gene and inheritance of allergic rhinitis in children.
Front. Genet. 14:1126212.
doi: 10.3389/fgene.2023.1126212

Relationship between single nucleotide polymorphism of NOS2 gene and inheritance of allergic rhinitis in children

Xionghui Wu[1]*, Sijun Zhao[1]*, Weiqing Huang[2]*, Min Huang[1]*, Jiang Xie[1]*, Guangliang Liu[1]* and Shuting Chang[2]*

[1]Department of Otorhinolaryngology Head and Neck Surgery, Hunan Children's Hospital, Changsha, Hunan, China, [2]Department of Neonatology, Hunan Children's Hospital, Changsha, Hunan, China

Allergic rhinitis is a common chronic disease, and its high incidence has a great negative impact on the quality of life of many people, especially children. In this paper, through in-depth analysis of NOS2 gene polymorphism, the protective mechanism of NOS2 gene against AR was studied to provide theoretical and scientific basis for the diagnosis of children with AR. It was concluded that the concentration of Immunoglobulin E (IgE) in rs2297516 was 0.24 IU/mL compared with that in normal children. rs3794766 specific IgE concentration in the children group was increased by 0.36 IU/mL, which was higher than that in the healthy children group; the difference of rs7406657 specific IgE concentration between the children group and the healthy group was 0.03 IU/mL. The total serum IgE concentration in the healthy children group was lower than that in the infant group, and the change of Rs3794766 was the least, followed by rs2297516 and rs7406657. This means that rs7406657 is the highest, rs2297516 had general genetic correlation with AR patients, and rs3794766 had the least genetic correlation with AR patients. Among the three groups of SNP loci, the healthy children group was higher than the patient children group, indicating that AR reduces the gene frequency of the three loci, and the reduction of gene frequency will also increase the susceptibility of children to AR, because the frequency of gene occurrence will affect the gene sequence. In conclusion, smart medicine and gene SNPS can promote the detection and treatment of AR.

KEYWORDS

allergic rhinitis, gene single nucleotide polymorphism, genetic analysis, intelligent medical treatment, gene and inheritance of allergic rhinitis

1 Introduction

Rhinitis is widely defined as inflammation of nasal mucosa, which is a common disease affecting about 40% of the population. AR is a common chronic rhinitis, accounting for 10%–20% of the world population. Severe AR may have more adverse effects on quality of life, sleep and work, because severe AR leads to respiratory obstruction in children. Intelligent medical treatment can improve the detection of influencing factors of AR, and detect AR related genes through intelligent devices, thus improving the detection accuracy. AR not only makes life difficult for adults, but also affects the quality of life of many children. Therefore, the study of AR has important clinical and social significance.

AR can affect respiratory tract infection in children. Bousquet Jean put forward the first guide based on recommendation grading evaluation and development evaluation, integrating the

application of supporting research in AR management. The purpose was to improve the accuracy of the diagnosis strategy and treatment plan in allergic rhinitis (Bousquet et al., 2020a). Small Peter believed that allergen immunotherapy was an effective immunomodulation therapy. If the drug treatment of allergic rhinitis was ineffective or intolerable, or the patient chooses, it should be recommended (Small et al., 2018). Schuler IV Charles Frank studied the influencing factors and causes of children's AR, and proposed some coping strategies (Schuler et al., 2021). Okubo Kimihiro believed that it was necessary to understand AR through a guide and use this knowledge to develop treatment plans. He also discussed the concept of AR and cost-benefit analysis (Okubo et al., 2020). Scadding Glenis K believed that the combined treatment of intranasal corticosteroids and intranasal antihistamines was more effective than any one alone, which could provide second-line treatment for rhinitis patients with poor control after single drug treatment (Scadding et al., 2017). Meng Yifan covered recent studies on SNP, DNA methylation, regulatory B cells, immunotherapy and the role of biological agents in AR (Meng et al., 2019). Zhang Yuan needed to better understand the prevalence and characteristics of AR, sensitization mode and related risk factors, so as to improve treatment and develop effective allergic rhinitis prevention strategies (Zhang and Zhang, 2019). The above studies all described the hazards and influencing factors of AR, but did not combine with the SNP of the gene.

SNP is closely related to AR. Chen Min-Li studied the relationship between SNP and AR through meta-analysis of SNP and AR risk in interleukin-13 and differentiation 14 gene clusters (Chen et al., 2018). Falahi Sara used polymerase chain reaction restriction fragment length polymorphism to determine the relationship between SNP of interleukin-33 gene and AR (Falahi et al., 2022). Amarin Justin Z used binomial logistic regression to study the association based on genotype under the general, recessive and dominant models of disease penetrance, and believed that the marker selection in the future genetic association study of asthma and allergic rhinitis should include functional polymorphism (Amarin et al., 2017). Ke Xia aimed to evaluate the potential association between gene SNP and AR in Han population. His results showed that it was significantly related to AR risk in Han population (Ke et al., 2017). The above studies all described the relationship between SNP and AR, but there are still some deficiencies in genetic research.

The treatment plan for AR can effectively relieve the symptoms of patients, which is usually safe and positive. However, it still cannot meet people's expectations for the treatment of diseases, because the current treatment plan still cannot completely cure allergic rhinitis. Even after treatment, the quality of life of patients still suffers from some adverse effects. Intelligent medical treatment can improve the detection of influencing factors of AR, and detect AR related genes through intelligent devices, thus improving the detection accuracy. Therefore, in the aspect of AR treatment, it is necessary to discuss the treatment mechanism and pathogenesis in detail. By measuring the concentration of total IgE in patients' serum, the degree of association between allergic factors and NOS2 gene polymorphism is analyzed.

2 Factors of AR genetics

2.1 Etiology and genetic factors of AR

AR is a multifactorial disease caused by the interaction between gene and environment. Genetic research shows that AR is a complex genetic disease, in which environmental factors are mainly related to various allergens in the human environment (Numminen, 2017). Many symptoms of AR are highly genetically controlled and are polygenic. Allergic rhinitis is affected by a variety of allergens. AR genetic factors are a variety of disease related genes and their related transcription factors, involving the selection of immunoglobulin candidate genes, important transcription factors, cytokines, T cell surface antigens and other pathogenic genes, as shown in Figure 1. From the genetic point of view, the impact of environmental factors on allergic rhinitis can be determined. In addition, environmental changes allow the innate and acquired immunity of the respiratory mucosal immune system to be regulated through various genetic mechanisms, thereby increasing the vulnerability of patients to allergic genes.

2.2 Inducers of AR

Figure 2 shows the specific IgE antibody reaction caused by environmental allergens (mainly inhalation allergens and food allergens), among which inhalation allergens are the main cause of AR (Blaiss et al., 2018). Inhalation allergens include fungal spores, pollen, internal mites, animal feces, etc. Their concentrations have a great relationship with the severity of allergic respiratory symptoms. The higher the concentration of allergen, the greater the probability of infection. Allergens cause skin allergy, gastrointestinal and nasal symptoms, but rarely AR. For infants, food allergy is mainly related to milk and soybeans. Common food allergies in adults include peanuts, nuts, fish, eggs, milk, soybeans, apples, pears, etc. In addition, it should be noted that allergies to certain vegetables and fruits may cause cross reactions with plant pollen.

2.3 Treatment of AR

The treatment principle is to avoid contact with allergens, reasonably use antihistamines and glucocorticoids, and give special immunization to patients. In addition, IgE monoclonal antibody is very effective for severe AR, but it is expensive, because the preparation process of monoclonal antibody is complex. Active and effective AR treatment can prevent and reduce asthma attack. Although AR has not been completely cured, the standardization of combined treatment is helpful to optimize symptom control and significantly improve the quality of life of patients. The treatment of AR includes acute treatment, routine treatment and drug treatment, as shown in Figure 3. When the patient's rhinitis obstructs nostril breathing or shock occurs, a doctor should be immediately arranged for emergency treatment. General treatment includes avoiding contact with allergens, especially identified allergens, and avoiding contact as much as possible. People with pollination allergy cannot walk during pollination. In addition, tools can be used to fight allergies, reduce nasal inhalation or allergic contact with the shell, and reduce nasal and eye symptoms. Allergies to indoor microorganisms and mites should be ventilated regularly to keep the room clean and dry. People who are allergic to dandruff, feathers and feces should avoid contact with animals as much as possible. The characteristics of drug therapy are very different, lacking the best, fast and effective drug use methods. In addition to ordinary generic drugs, under the guidance of doctors, full consideration must be given to personal conditions to select the most appropriate drugs, mainly through the appropriate dosage for targeted treatment. In addition, drug therapy can greatly shorten the treatment time and achieve rapid results (Bousquet et al., 2020b).

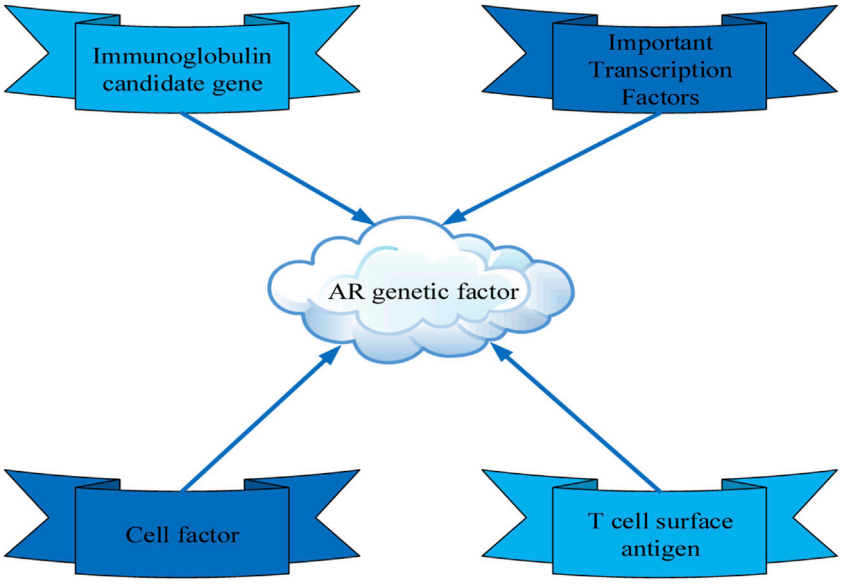


FIGURE 1
Genetic factors of AR.

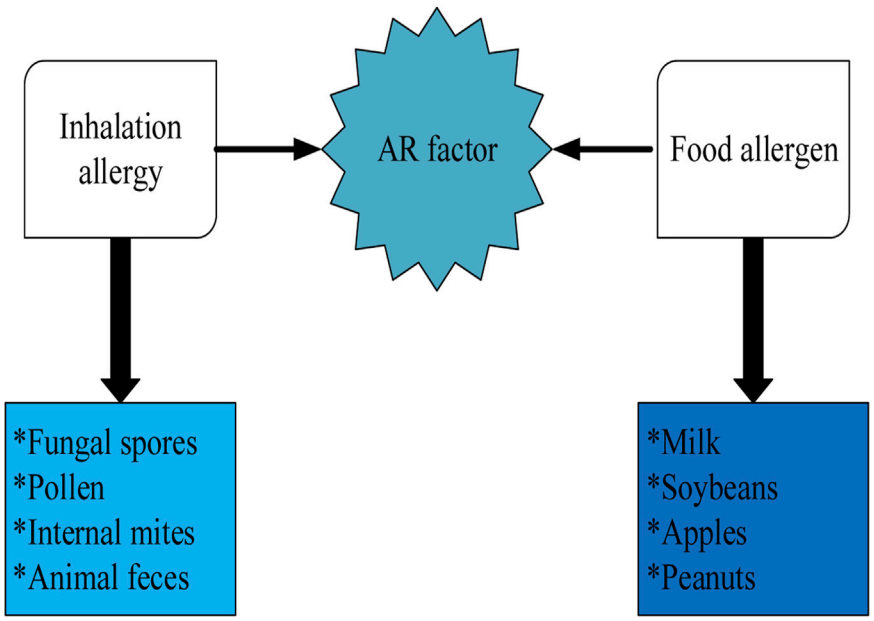


FIGURE 2
Inducement factors of AR.

2.4 Application of intelligent medical treatment in AR

The current intelligent medical treatment can use ultrasound and radio frequency for AR treatment. In particular, ultrasound therapy is mainly used for focused ultrasound technology, rather than invasive technology. It allows the ultrasound generated outside the body to focus on specific human targets, and the focus area has high energy. The biological effect of ultrasound on nasal submucous tissue can reduce the sensitivity of nerves, blood vessels and clay glands. It can immediately change the structure and function of target tissues to reduce the hypersensitivity reaction of nasal mucosa, thus increasing the effective nasal respiratory area. By selectively destroying parasympathetic nerve fibers, AR can significantly improve nasal itching, sneezing, nasal mucus and other diseases, and improve the symptoms of patients (Tartarisco et al., 2017). Moreover, the idea of

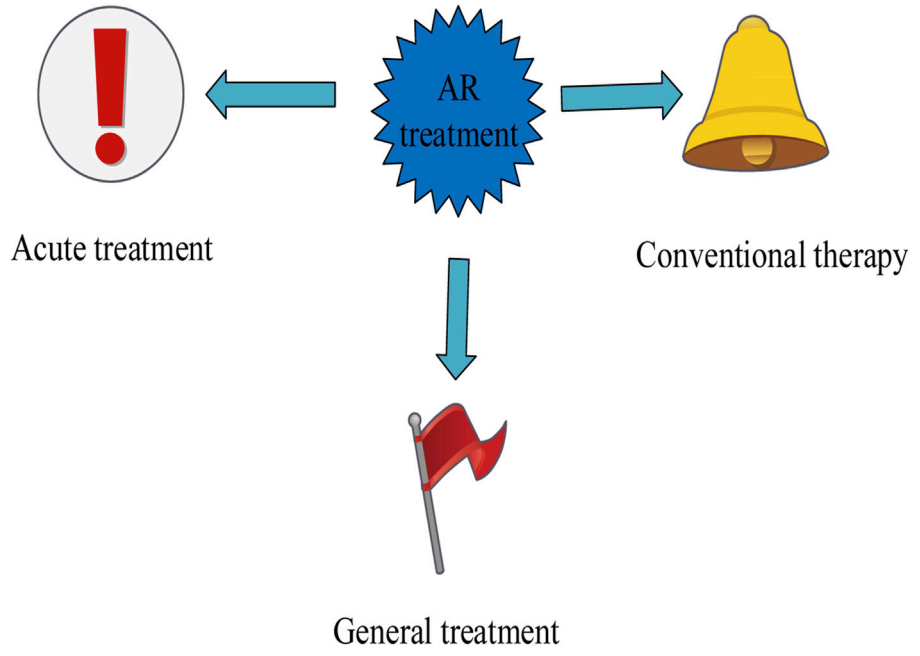


FIGURE 3
Treatment of AR.

non-invasive treatment has become that there is no damage to the invasive tissues. The ultrasonic rhinitis treatment instrument uses non-invasive ultrasonic therapy to treat allergic rhinitis, and provides doctors with new options for effective treatment of complications because non-invasive ultrasound therapy is more convenient. High frequency therapy is a kind of high-frequency electromagnetic band. Among them, the magnetic field can spread directly in the atmosphere or vacuum without being affected by the conductor. Low temperature plasma is the most widely used RF device in clinical application. The conductive medium forms a high concentration plasma area around the electrode, which is composed of highly ionized particles. These ionized particles have enough energy to tear the organic molecular chains in the tissue, separate molecules and reduce the tissue volume. The advantages of high-frequency treatment are relatively small damage to surrounding tissues, low treatment temperature, ablation of tissue targets, low thermal conductivity, less free radical release, less inflammatory reaction, and no carbon and environmental pollution.

3 Experimental data and methods of genetic association between NOS2 gene SNP and children AR

3.1 Clinical data

According to AR diagnosis and treatment methods, 285 children with AR are selected from the otorhinolaryngology department of Hunan Children's Hospital, P.R. China, from December 2020 to December 2021, including 167 boys and 118 girls, aged from 4 to 14 years. Then, 352 healthy children are searched again, including 157 boys and 195 girls, who are divided into two groups: the patient children group and the healthy children group. All children in the patient's children group have many symptoms, such as nasal congestion, poor breathing, sneezing and nose itching. Allergen skin prick test (ASPT) is AR positive, and each patient has a history of allergy inheritance. All children in the healthy children group have no genetic history of allergy, and ASPT is AR negative. In addition, all children in this experiment have no other allergic diseases and allergic genetic history, and the accuracy of the experiment is improved by excluding other criteria.

3.2 Experimental methods

1) The ASPT experiment and grading are carried out according to the normal classification standard, and the ASPT grading of the patient group is ≥1.
2) The patient's serum IgE is detected using intelligent detection equipment system.
3) All patients' genomic DNA is extracted by elbow vein blood test and refrigerated with a unified anti coagulation test tube.
4) There are three kinds of SNP site selection: rs2297516, rs3794766, rs7406657.
5) The SNP classification uses an intelligent experimental platform for classification and detection. The chip with the sample is tested internally in the mass spectrometer, and the locus genotype of the gene sequence is detected with the support of intelligent equipment.

3.3 Statistical methods

Statistical analysis is carried out with data analysis software. The measurement data are expressed in the way of mean value. The gender

TABLE 1 Analysis of gender differences between the two groups of children.

	Patient children group	Healthy children group
Boy	167	157
Girl	118	195
x^2	0.251	0.025
A	0.458	0.001

difference between the two groups is compared with x^2 Inspection. The degree of fit is used to test whether the gene distribution conforms to Hardy Weinberg equilibrium. x^2 is used to test the correlation between allele frequency, haplotype, multilinear regression analysis SNP and total serum IgE, and modify according to age and sex to determine the main genetic model of A<0.05.

4 Experimental results of genetic association between NOS2 gene SNP and children's AR

First, the gender difference between the two groups of children in the genetic association analysis of NOS2 gene SNP and children's AR is analyzed to determine whether the two groups of children met the test criteria. The specific differences are shown in Table 1.

According to the data described in Table 1, there are 167 boys and 118 girls in the patient children group, with a difference of 49 men and women. There are 157 boys and 195 girls in the healthy children group, with a difference of 38 men and women. There are 10 differences between boys in the patient children group and boys in the healthy children group, and 77 differences between girls and girls in the healthy children group. The $x^2$of the patient children group is 0.251, and the $x^2$of the healthy children group is 0.025. The difference between the two groups is 0.226. The A of the patient children group is 0.458, and the A of the healthy children group is 0.001. The difference between the two groups is 0.457. Through the analysis and comparison of $x^2$and A, it can be known that the gender differences between the two groups meet the statistical standards. Then, the allele frequency difference of NOS2 between the two groups of children is analyzed, mainly using single locus analysis. The specific analysis is shown in Figure 4.

According to the allele frequency difference of NOS2 between the two groups of children described in Figure 4, the allele frequency of NOS2 in the healthy children group is higher than that in the patient children group. It can be seen from the figure that the data of rs2297516 in the patient children group under single point detection is 315, while the data of rs2297516 in the healthy children group under single point detection is 384. The data of healthy children group is 69% higher than that of patient children group. The data of rs3794766 in the patient children group under single point detection is 118, while the data of rs3794766 in the healthy children group under single point detection is 121. The data of the healthy children group is 3 times higher than that of the patient children group. The data of rs7406657 in the patient children group under single point detection is 250, while the data of rs7406657 in the healthy children group under single point detection is 283. The data of the healthy children group is 33% higher than that of the patient children group. The SNP loci of the three groups are higher in the healthy children group than in the patient children group, which indicates that AR reduces the gene frequency of the three loci, and the decrease in gene frequency also leads to an increase in the susceptibility of children to AR. Then, the difference of the distribution of NOS2 gene SNP genotype between the patient children group and the healthy children group is analyzed. The specific comparison is shown in Table 2.

According to the data described in Table 2, rs2,297,516 genotypes are AA, CC and AC respectively. The AA, CC, and AC of the patients' children group are 78, 110 and 175 respectively, while the AA, CC, and AC of the healthy children group are 83, 125 and 214 respectively. Compared with the two groups, the AA of the healthy children group is 5. CC of healthy children is 15% higher than that of patients. The AC of healthy children is 39. The distribution of these three genes is higher in the healthy children group than in the patient children group, and A is 0.551, indicating that the differences between the two groups meet the statistical standard. The distribution of rs3794766 genotype is TT, CC and CT respectively. TT is 5, CC is 269, and CT is 108 in the patient children group, while TT is 9, CC is 308, and CT is 101 in the healthy children group. Compared with the two groups, TT in the healthy children group is 4. The CC of healthy children is 39. The CT of the healthy children group is 7. The distribution of TT and CC genes in healthy children is higher than that in patients. However, CT shows that the healthy children group is lower than the patient children group, and A is 0.974, indicating that the difference between the two groups conforms to the statistical standard. The distribution of rs3794766 genotype is CG, CC, and GG. The CG, CC, and GG of the patient children group are 135, 48, and 192 respectively, while the CG, CC, and GG of the healthy children group are 179, 44, and 204 respectively. Compared with the two groups, CG of the healthy children group is 44. The CC of healthy children is 4. The GG of healthy children group is 12. The distribution of CG and GG genes in healthy children is higher than that in patients. CC is lower in the healthy children group than in the patient children group, and A is 0.084, indicating that the difference between the two groups meets the statistical standard. Then, the correlation analysis between NOS2 gene SNP and serum total IgE concentration in patients' children group and healthy children group is analyzed. The specific changes are shown in Figure 5.

According to the comparative data in Figure 5, the total serum IgE concentration in rs2297516 of the patient's children group is 2.284 IU/mL, while that in rs2297516 of the healthy children group is 1.595 IU/mL, with a difference of 0.689 IU/mL. The total serum IgE concentration in rs3794766 is 2.291 IU/mL in children group, while that in rs3794766 in healthy children group is 1.654 IU/mL, with a difference of 0.637 IU/mL. The total serum IgE concentration in rs7406657 of the children group is 2.274 IU/mL, while the total serum IgE concentration in rs7406657 of the healthy children group is 1.551 IU/mL, with a difference of 0.723 IU/mL. The total serum IgE concentration of these three gene sequences is lower in the healthy children group than in the patient children group, and the change of rs3794766 is the smallest, followed by rs2297516, and finally rs7406657. This shows that rs7406657 has the highest genetic correlation with AR patients, while rs2297516 has a general genetic correlation with AR patients. The genetic correlation between rs3794766 and AR patients is the smallest. Finally, the association between NOS2 gene SNP and specific IgE is analyzed. The specific association analysis is shown in Figure 6.

According to the comparative data in Figure 6, the specific IgE concentration in rs2297516 of the patient's children group is 1.21 IU/mL, while the specific IgE concentration in rs2297516 of the healthy children group is 1.45 IU/mL, with a difference of 0.24 IU/mL. The specific IgE concentration in rs3794766 is 1.15 IU/mL in the children group, while that in the healthy children group is 1.51 IU/mL, with a difference of

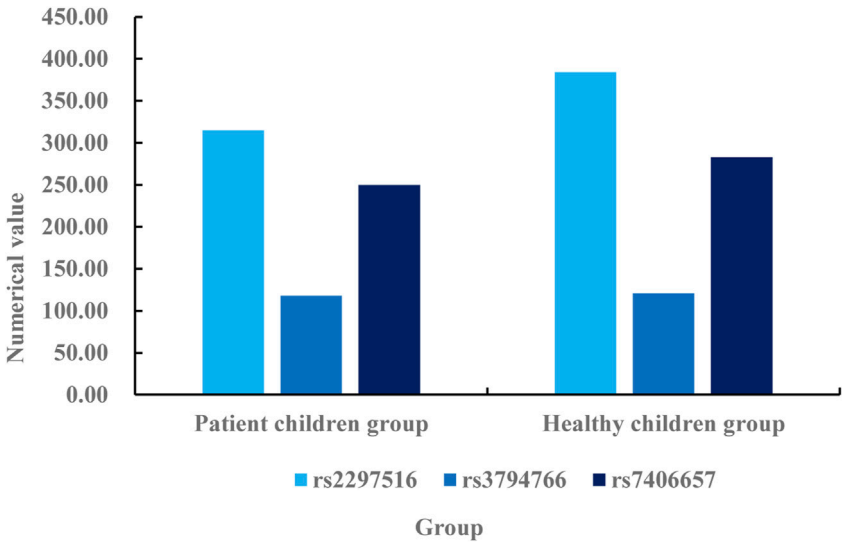


FIGURE 4
Comparison of allele frequency difference of NOS2 between two groups of children.

TABLE 2 Difference of distribution of NOS2 gene SNP genotype between patient children group and healthy children group.

SNP	Genotype	Patient children group	Healthy children group	A
rs2297516	AA	78	83	0.551
	CC	110	125	
	AC	175	214	
rs3794766	TT	5	9	0.974
	CC	269	308	
	CT	108	101	
rs7406657	CG	135	179	0.084
	CC	48	44	
	GG	192	204	

0.36 IU/mL. The specific IgE concentration in rs7406657 of the children group is 1.35 IU/mL, while the specific IgE concentration in rs7406657 of the healthy children group is 1.38 IU/mL, with a difference of 0.03 IU/mL. On the whole, the specific IgE concentration in the healthy children group is higher than that in the patient children group. This shows that healthy children are more specific to allergens and can be protected from the infection of these allergens. The children of the patients have low resistance to these allergens and are vulnerable to the infection of these substances. The total IgE concentration in the serum of the patient's children is relatively low, which is unable to effectively respond to the invasion of allergens.

5 Discussion

AR is a disease determined by genetic and environmental factors, which plays an important role in the occurrence of AR. Special immunotherapy is a method to treat allergic diseases and the only way to change the natural process of AR, which is to improve patients' allergen resistance by changing the role of target cells. At present, AR is widely considered as a genetic and environmental disease caused by genetic and environmental factors. SNP is a common genetic factor that affects people's susceptibility to disease. Studies have shown that the sensitivity of antiretroviral drugs and the effectiveness of drug therapy are related to multiple gene SNPs. The DNA differences of different races, groups and individuals found in the research of human genome observation system and their importance fundamentally change the diagnosis, treatment and prevention of diseases. NOS2 gene is located on chromosome 17, 37 KB long, with 26 exons. The second external element has the starting point of the converter, and the 26th external element has the final code. It runs a subspace area with many built-in factors required to activate the initializer. NOS2 genotype has multiple overlapping open reading frames, and the activation of subspace may affect the expression of INOS. The study found that the rs2297516 data

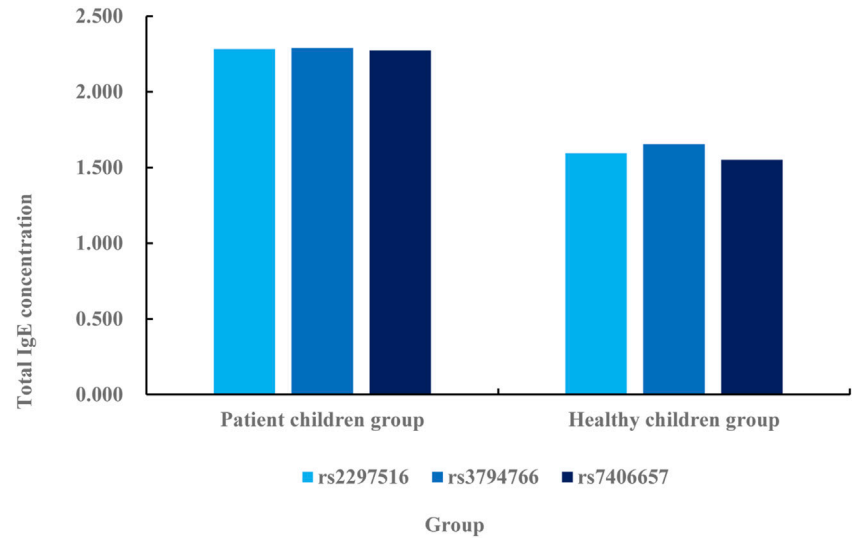


FIGURE 5
Correlation analysis between NOS2 gene SNP and total serum IgE concentration in patients and healthy children.

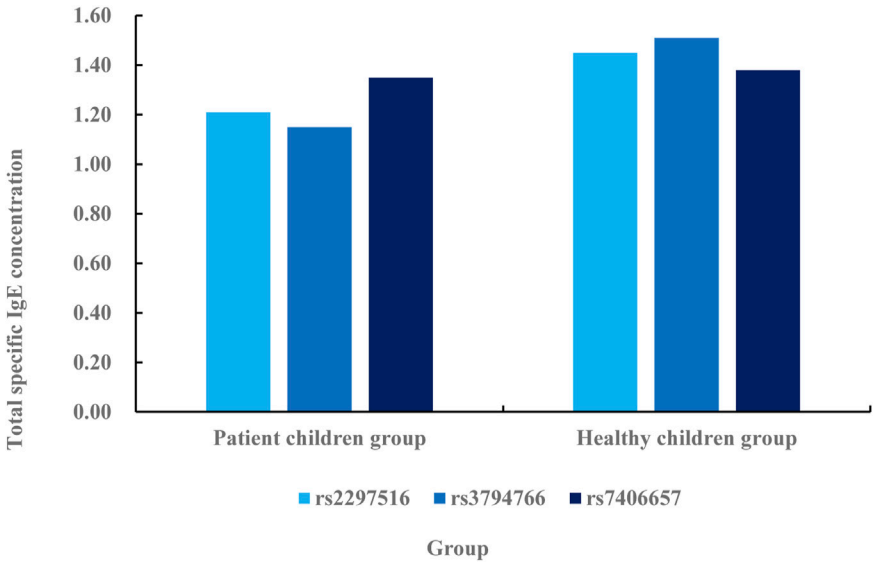


FIGURE 6
Association between NOS2 gene SNP and specific IgE.

of the healthy children group is 69% higher than that of the patient children group under the single point detection. The rs3794766 data of the healthy children group is 3% higher than that of the patient children group. The rs7406657 data of the healthy children group is 33% higher than that of the patient children group. The total serum IgE concentration of these three gene sequences is lower in the healthy children group than in the patient children group, and the change of rs3794766 is the smallest, followed by rs2297516, and finally rs7406657. This shows that rs7,406,657 has the highest genetic correlation with AR patients. The rs2,297,516 has the general genetic correlation with AR patients, and rs3,794,766 has the lowest genetic correlation with AR patients.

6 Conclusion

In the research of human genome SNP observation system, the DNA differences have found in different races, groups and individuals and their important properties fundamentally have changed the diagnosis, treatment and prevention of AR diseases. The SNP of NOS2 gene may be related to children's AR. The SNP genotype distribution of NOS2 gene analyzed in the experiment is very similar to the genetic sequence of allergic rhinitis. Intelligent medical treatment has improved the detection of influencing factors of AR, and AR related genes was detected through intelligent devices, thus improving the detection accuracy. In

addition, NOS2 gene SNP has a certain correlation with the total serum IgE concentration. The potential role of NOS2 gene SNP in the formation and development of AR has required a large number of samples, multi regions and multi countries to conduct in-depth research, which has laid a theoretical foundation for developing new AR control direction. However, given the limited number of samples collected in this study, it is necessary to increase the number of samples to confirm the reliability of the results. The severity of these problems requires more sampling and comprehensive analysis in future research, and comprehensive analysis of gene diversity of different chromosomes should also be considered.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Ethical Statement

Ethical approval for this study and written informed consent from the participants of the study were not required in accordance with local legislation and national guidelines.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This study was supported by the department of Science and Technology of Hunan Province, Major Project, Hunan Science Project (2019) No. 49 (2019SK1015).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amarin, J. Z., Naffa, R. G., Suradi, H. H., Alsaket, Y. M., Obeidat, N. M., Mahafza, T. M., et al. (2017). An intronic single-nucleotide polymorphism (rs13217795) in FOXO3 is associated with asthma and allergic rhinitis: A case–case–control study. *BMC Med. Genet.* 18, 132–136. doi:10.1186/s12881-017-0494-4

Blaiss, M. S., Hammerby, E., Robinson, S., Kennedy-Martin, T., and Buchs, S. (2018). The burden of allergic rhinitis and allergic rhinoconjunctivitis on adolescents: A literature review. *Ann. Allergy, Asthma & Immunol.* 121, 43–52. doi:10.1016/j.anai.2018.03.028

Bousquet, J., Anto, J. M., Bachert, C., Baiardini, I., Bosnic-Anticevich, S., Walter Canonica, G., et al. (2020). Allergic rhinitis. *Nat. Rev. Dis. Prim.* 6, 95–17. doi:10.1038/s41572-020-00227-0

Bousquet, J., Schunemann, H. J., Togias, A., Bachert, C., Erhola, M., Hellings, P. W., et al. (2020). Next-generation allergic rhinitis and its impact on asthma (ARIA) guidelines for allergic rhinitis based on grading of recommendations assessment, development and evaluation (GRADE) and real-world evidence. *J. Allergy Clin. Immunol.* 145, 70–80.e3. doi:10.1016/j.jaci.2019.06.049

Chen, M-L., Zhao, H., Huang, Q. P., and Xie, Z. F. (2018). Single nucleotide polymorphisms of IL-13 and CD14 genes in allergic rhinitis: A meta-analysis. *Eur. Archives Oto-Rhino-Laryngology* 275, 1491–1500. doi:10.1007/s00405-018-4975-7

Falahi, S., Mortazavi, S. H. R., Salari, F., Koohyanizadeh, F., Rezaeimanesh, A., and Gorgin Karaji, A. (2022). Association between IL-33 gene polymorphism (Rs7044343) and risk of allergic rhinitis. *Immunol. Investig.* 51, 29–39. doi:10.1080/08820139.2020.1804399

Ke, X., Song, S., Wang, X., Shen, Y., Kang, H., and Hong, S. (2017). Associations of single nucleotide polymorphisms of PTPN22 and Ctla4 genes with the risk of allergic rhinitis in a Chinese Han population. *Hum. Immunol.* 78, 227–231. doi:10.1016/j.humimm.2016.11.008

Meng, Y., Wang, C., and Zhang, L. (2019). Recent developments and highlights in allergic rhinitis. *Allergy* 74, 2320–2328. doi:10.1111/all.14067

Numminen, J. (2017). Allergic rhinitis." Duodecim; laaketieteellinen aikakauskirja. *Allerg. rhinitis* 133, 473–478.

Okubo, K., Kurono, Y., Ichimura, K., Enomoto, T., Okamoto, Y., Kawauchi, H., et al. (2020). Japanese guidelines for allergic rhinitis 2020. *Allergol. Int.* 69, 331–345. doi:10.1016/j.alit.2020.04.001

Scadding, G. K., Kariyawasam, H. H., MiRakian, R., Buckley, R. J., Dixon, T., et al. (2017). BSACI guideline for the diagnosis and management of allergic and non-allergic rhinitis (Revised Edition 2017; First edition 2007). *Clin. Exp. Allergy* 47, 856–889. doi:10.1111/cea.12953

Schuler, I. V., Frank, C., and Montejo, J. M. (2021). Allergic rhinitis in children and adolescents. *Immunol. Allergy Clin.* 41, 613–625. doi:10.1016/j.iac.2021.07.010

Small, P., Keith, P. K., and Kim, H. (2018). Allergic rhinitis. *Allergy asthma & Clin. Immunol.* 14, 1–11.

Tartarisco, G., Tonacci, A., Minciullo, P. L., Billeci, L., Pioggia, G., Incorvaia, C., et al. (2017). The soft computing-based approach to investigate allergic diseases: A systematic review. *Clin. Mol. Allergy* 15, 10–14. doi:10.1186/s12948-017-0066-3

Zhang, Y., and Zhang, L. (2019). Increasing prevalence of allergic rhinitis in China. *Allergy, asthma & Immunol. Res.* 11, 156–169. doi:10.4168/aair.2019.11.2.156

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fgene.2023.1132242

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Jiasheng Yang,
Changsha Medical University, China
Qi Bin,
Zhejiang Business Technology Institute, China
Dazhong Shu,
Sanya University, China
Lei Shi,
Luliang University, China

*CORRESPONDENCE
Jianwen Zhao,
15949947748@163.com

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 27 December 2022
ACCEPTED 24 January 2023
PUBLISHED 09 February 2023

CITATION
Ren S, Cao W, Ma J, Li H, Xia Y and Zhao J (2023), Correlation evaluation between cancer microenvironment related genes and prognosis based on intelligent medical internet of things.
Front. Genet. 14:1132242.
doi: 10.3389/fgene.2023.1132242

Correlation evaluation between cancer microenvironment related genes and prognosis based on intelligent medical internet of things

Shoulei Ren[1†], Wenli Cao[1†], Jianzeng Ma[1], Hongchun Li[2], Yutao Xia[1] and Jianwen Zhao[1*]

[1]Oncology Department, Yangguangronghe Hospital, Weifang, Shandong, China, [2]Nerosurgery Department, Yangguangronghe Hospital, Weifang, Shandong, China

The study of tumor microenvironment plays an important role in the treatment of cancer patients. In this paper, intelligent medical Internet of Things technology was used to analyze cancer tumor microenvironment-related genes. Through experiments designed and analyzed cancer-related genes, this study concluded that in cervical cancer, patients with high expression of P16 gene had a shorter life cycle and a survival rate of 35%. In addition, through investigation and interview, it was found that patients with positive expression of P16 and Twist genes had a higher recurrence rate than patients with negative expression of both genes; high expression of FDFT1, AKR1C1, and ALOX12 in colon cancer is associated with short survival; high expressions of HMGCR and CARS1 is associated with longer survival; overexpression of NDUFA12, FD6, VEZT, GDF3, PDE5A, GALNTL6, OPMR1, and AOAH in thyroid cancer is associated with shortened survival; high expressions of NR2C1, FN1, IPCEF1, and ELMO1 is associated with prolonged survival. Among the genes associated with the prognosis of liver cancer, the genes associated with shorter survival period are AGO2, DCPS, IFIT5, LARP1, NCBP2, NUDT10, and NUDT16; the genes associated with longevity are EIF4E3, EIF4G3, METTL1, NCBP1, NSUN2, NUDT11, NUDT4, and WDR4. Depending on the prognostic role of genes in different cancers, they can influence patients to achieve the effect of reducing patients' symptoms. In the process of disease analysis of cancer patients, this paper uses bioinformation technology and Internet of things technology to promote the development of medical intelligence.

KEYWORDS

tumor microenvironment, internet of things technology, smart healthcare, biological, correlation evaluation

1 Introduction

The correlation between cancer TME related genes and prognosis is an unclear field at present, and there are big problems at present. The data of cancer analysis not only comes from patients' information, but also needs to be analyzed in combination with patients' conditions to understand the relationship between genes behind them, so as to achieve a good cure effect and bring convenience to public life. It is necessary to analyze the correlation between cancer TME related genes and prognosis in this field.

TME is a topic of wide scholarly interest and has been studied by a number of scholars in this area. Hu et al. (2020) included 85 patients to investigate the effect of TME and related genes on patients with osteosarcoma. Survival analysis showed that patients with higher immune scores had good overall survival and disease-free survival. In addition, 769 genes were identified as TME-associated genes. It was concluded that TME was associated with the prognosis of patients with osteosarcoma. A prognostic model based on TME-related genes can effectively predict overall survival and disease-free survival in patients with osteosarcoma. Yuan et al. (2020) performed a systematic investigation of TME and identified TME-related genes with prognostic value in patients with lung adenocarcinoma, identifying a total of 281 prognostic TME-related genes. Subsequently, functional analysis and protein-protein interaction network analysis showed that these genes were mainly associated with immune response, inflammatory response and chemotaxis. Luo et al. (2021) performed survival analysis to identify colon cancer pivotal genes and used these pivotal genes with TNM staging to construct prognostic models. In addition, the correlation between pivotal gene expression and immune cell infiltration was assessed. Wang et al. (2021) investigated the potential relationship between co-expression modules and TME by weighted gene co-expression network analysis. Firstly, several TME-related genes were integrated and a risk prediction model was established. The model can accurately predict the progression and prognosis of BLCA and provide clinical implications for risk stratification, immunotherapy drug screening and treatment decisions. Liu et al. (2021) explored the relationship between TME and prognosis of colorectal cancer and identified prognostic genes associated with the microenvironment of colorectal cancer. Gene expression data were collected from the Cancer Genome Atlas and stromal/immune cell scores for colorectal cancer were calculated and their relationship with clinical outcomes. Li et al. (2021a) assessed the relationship between TME and prognosis and explored prognostic genes in rectal cancer. An expression data algorithm was used to estimate stromal and immune cells in malignant tumors to calculate the immune/stromal score. The correlation between immune/stromal score and survival time as well as clinical characteristics was assessed. Zhang et al. (2020) used the expression data algorithm to estimate cells and immune cells in tumor tissue to calculate the immune/stromal score. Patients with clear cell renal cell carcinoma in the Cancer Genome Atlas database were classified into low and high groups based on the score, and genes that were differentially expressed and significantly correlated with prognosis were identified. The results of functional enrichment analysis and protein-protein interaction networks indicated that these genes are mainly involved in the regulation of extracellular matrix and cellular activity. Finally, a series of microenvironment-related genes were obtained, which predict the prognosis of patients with clear cell renal cell carcinoma. In many studies, the analysis steps were similar, all of them investigated the prognosis of cancer, indicating the practical value of the investigation of cancer genes and prognosis.

The IoT technology has many applications in the medical field. Zhu et al. (2019) believed that the development of the IoT advanced medical care, and discussed different aspects of intelligent medical care, as well as health data and patient centered health management. Islam and Rahaman (2020) developed the intelligent medical monitoring system in the IoT environment. Javaid and Khan (2021) analyzed that the medical care enabled by the IoT could help meet the challenge of the COVID-19 pandemic. Kondka et al. (2022) used an intensive healthcare monitoring paradigm based on the IoT machine learning strategy. Zhang et al. (2018) proposed an intelligent IoT connection architecture for smart hospitals based on narrowband IoT, and introduced edge computing to deal with the delay requirements in the medical process. He developed an infusion monitoring system to monitor the real-time decline rate and residual drug volume during intravenous infusion, and discussed the challenges and future direction of building a smart hospital by connecting intelligent things. Li et al. (2021b) made a comprehensive investigation on the intelligent medical system of the IoT by using big data analysis of machine learning. Mohammed (2020) studied the detection and diagnosis system of novel coronavirus using the smart helmet based on the IoT. The analysis of IoT technology has not yet analyzed the correlation between cancer microenvironment related genes and prognosis.

In order to make the concept of intelligent medical care deeply rooted in people's hearts, this paper analyzed the IoT for intelligent medical care, tumor microenvironment and their applications. The correlation between cancer TME related genes and prognosis based on the intelligent medical IoT was evaluated, and four common cancers were took as examples to analyze their genes and prognosis effects. Finally, the feasibility conclusion was drawn. This paper provided a reference path for the correlation analysis between cancer microenvironment related genes and prognosis.

2 Intelligent medical IoT

2.1 IoT and intelligent healthcare

The structure of the IoT is very complex and consists of three main parts: The sensor layer is responsible for collecting information (it uses smart cards, Radio Frequency Identification (RFID) tags, two digit barcodes, sensors, etc.); the network layer is responsible for transmitting information (it uses wireless network, cellular network, wired network, RFID network, etc.); the application layer supports information analysis and processing and management decisions (Qadri et al., 2020).

Smart medical care comes into being under specific circumstances: In the new round of medical reform, the old medical system cannot meet the needs of today's constantly developing society, while the main reason for the new medical reform is that it is difficult to see a doctor and the cost is high. With the change of life style, acute and chronic diseases, aging population, and many hostile environmental threats, the demand for hospital information technology is growing. Intelligent medical management is to optimize the combination of QR code, RFID, wireless network and other technologies with the traditional hospital system integration.

2.2 Application of IoT in medical treatment

The IoT is a large-scale network established by connecting various information sensors to the Internet, which collects any object or process that needs real-time tracking, connection and interaction. It also collects sound, light, heat, electricity, mechanical, chemical and bioenergy, location and other important information. The goal is to connect objects to objects, objects to people, and all objects to the

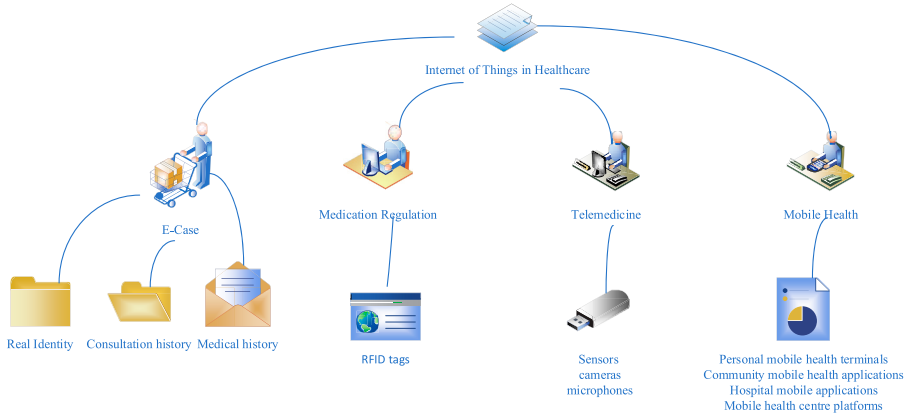


FIGURE 1
Internet of Things in healthcare.

network so that they can be more easily identified, managed, and controlled. The IoT is the extension and expansion of computer networks, which is a huge network formed by connecting various information sensors to the Internet. It enables all people and things to connect to the network and manage important events in a timely manner. The IoT has a wide range of applications, including intelligent security, intelligent logistics, intelligent medical and other fields (Habibzadeh et al., 2019). The main applications of IoT technology in the medical field are electronic documents, drug monitoring, telemedicine, etc., as shown in Figure 1.

2.2.1 Electronic case

The electronic case mainly includes checking the patient's real identity, consulting records, medical history, etc., In emergency situations, doctors can use electronic medical records to ensure efficient and timely care. Electronic records include information collection, storage, processing, and intelligent services. Personal electronic medical records collect and store clinical information of patients in outpatient and ward of different hospitals, as well as electrocardiogram, ultrasound and other examination information. In the consultation process, the patient only needs to bring a card, and the doctor can understand the patient's full medical history and quickly solve the problem.

2.2.2 Drug supervision

The IoT is increasingly being used in the field of drug monitoring, especially for obtaining information about drug production and circulation, and preventing counterfeiting and abuse (Singh et al., 2020). In recent years, medical malpractice has gradually become a serious problem in society. The workload of medical personnel is increasing, because they must constantly check the patients' drugs to avoid the wrong patient and drug combination. Traditional barcode recognition has some disadvantages, such as poor readability and reduced readability. Only RFID tags can be used after drug production, which can record all production information and drug flow information. They have tamper proof function. The advantage of RFID tag is that it can store more data and capture obscure objects without touching, so as to make them invalid. They can be used to help staff check drugs and issue alerts to prevent dispensing errors or drug abuse by patients.

2.2.3 Telemedicine

Telemedicine mainly involves remote monitoring and home care. Among them, wireless sensor network technology is used to monitor the physiological state and movement of the monitored object, so as to assist patients in the event of an accident. A telemedicine system usually consists of a ward, a home ward and a dedicated remote ward that can be connected *via* the Internet. In case of an accident, intelligent terminals such as mobile phones and computers can be used to communicate with remote experts in a timely manner. Remote experts can use various physiological information sensors, cameras, microphones, and other equipment to directly monitor and understand the patient's condition and provide timely treatment. This can save time and prevent patients from going to the hospital blindly.

2.2.4 Mobile medicine

Mobile medical needs a wireless network as the basis of a complete and efficient mobile medical system, including personal mobile medical terminals, community mobile medical applications, hospital mobile applications and mobile medical center platforms (Tian et al., 2019). In the era of limited medical resources, efficient medical resources can be integrated and maximized to better meet the challenges. Nowadays, with the rapid development of communication technology and smart phones, China's communication network has been greatly expanded. Medical software on smart phones is also being used more and more, and mobile medicine is making remarkable achievements. For example, the deployment, inventory and management of medical equipment or the use of mobile ward rounds, mobile medicine, electronic health records and drug management are becoming more and more popular, and these have also been applied for the first time in rural and remote mountain areas.

The intelligent sign in terminal is very useful for feeding patients, hanging water for nurses and registering doctors. This technology can be used to record patient information, prescription drugs and various physiological parameters. RFID technology can quickly retrieve relevant information through scanning codes and establish personal care standards. It effectively reduces the workload of medical staff and improves the implementation level of health education. Nurses are very important to the information and knowledge in the work process. Now, various types of healthcare management software are widely

used in mobile personal healthcare. Although it has not been widely used in China, many problems remain unsolved. However, it is believed that its huge advantages would promote the progress and development of medical and healthcare.

3 Tumor microenvironment and its application

3.1 Tumor microenvironment

TME refers to the internal environment where tumor cells originate and live. This includes not only the tumor cells themselves, but also their environment, fibroblasts, immune and inflammatory cells, glial cells, and other cells, as well as the stroma, microvessels, and biomolecules that infiltrate near the tumor. The most obvious signs of TME are hypoxia, low pH and high pressure. With the help of TME, tumors weaken anti-tumor immune response, and maintain proliferation, which also prevent apoptosis and maintain inflammatory environment and angiogenesis. The change of immune surveillance function from tumor clearance to tumor induction is a complex process, which involves several signal pathways mediated by cytokines expressed by tumor cells, immune, cells and other non-tumor cells, such as tumor related epithelial cells or fibroblasts in surrounding tissues.

The basis of tumor immune monitoring is fibroblasts, tumor related macrophages, tumor related neutrophils, bone marrow derived suppressor cells, tumor related fibroblasts, regulatory T cells and other cells that change the balance of immune cells in TME. This leads to increased inflammation and angiogenesis; phenotypic changes of neutrophils from N1 to N2, macrophages from MI to M2, and T cells from Thl to Th2; the number and activity of cytotoxic T cells and antigen presenting cells decreased. The sharp decrease of mature dendritic cells leads to the increase of monocyte progenitor cells, which helps to increase tumor related macrophages and bone marrow derived suppressor cells. The cytokine network formed between these immune cells is mutually reinforcing, which helps maintain the number of TME immune cells that contribute to tumor formation. In addition, transforming growth factor β, vascular endothelial growth factor, chemokines and inflammatory cytokines (especially Th2 mediated cytokines) have been proved to be involved in angiogenesis, inflammation, and tumor immunosuppression. Fibroblasts and regulatory B cells seem to be involved.

3.2 Proto oncogene and tumor suppressor gene

3.2.1 Cancer gene

Cancer genes, also known as proto oncogenes, mainly refer to DNA nucleotide sequences that have the ability to cause malignant transformation of cells, that is, the nucleic acid part that causes malignant transformation of cells. This concept was first discovered in the study of retroviral RNA viruses, which were initially called viral tumor genes. Later, it was found that the reverse transcriptional RNA of this nucleotide sequence also existed in the cell gene, and even found the carcinogenic base sequence that did not exist in the virus, so it was renamed as the cell tumor gene. Tumor gene is a highly conserved gene, which is expressed only during embryonic development. In adulthood, it would no longer express itself and would be silent. Tumor genes play an important role in regulating normal cell growth, proliferation, development, and differentiation during embryonic development. Therefore, oncogenes in cells are not only the cause of cancer, but also the genes necessary for life. Oncogenes can be produced only when the mutation of oncogene causes changes in its normal structure and function.

Cancer genes are usually different from normal genes in the following aspects: The expression level of tumor genes is often higher than that of proto oncogenes (non-mutated normal genes corresponding to tumor genes), and sometimes unnatural transcription occurs in differentiated cells. This difference in structure and function often leads to the loss of fusion protein activity, which leads to the imbalance of oncogene proteins and promotes unlimited cell proliferation and abnormal transformation. The multi site mutation of many carcinogenic genes results in the replacement of a single amino acid of the corresponding protein by other amino acids.

3.2.2 Tumor suppressor gene

Tumor suppressor genes are genes that inhibit excessive cell division, thus preventing the development and deterioration of cancer. This is because the normal function of its expression products would be to negatively stimulate cell growth and inhibit proto oncogenes. When carcinogenic genes are destroyed by internal or external factors during cell division and proliferation, such destruction would lead to partial or even complete deletion of genes, thus leading to loss of gene expression or inactivation of gene expression products. As a result, the product of gene expression lost its function, leading to the occurrence of cancer. Tumor suppressor genes generally have the following characteristics: The inhibition of tumor suppressor genes is often accompanied by non-functional mutations involving loss of heterozygosity; tumor suppressor genes often mutate in cancer prone syndromes; somatic mutation occurs in spontaneous tumor suppression; tumor suppressor genes can also inhibit cell growth *in vitro*. Tumor suppressor genes have many important functions. For example, they induce cell differentiation and maintain genetic stability, which also induce cell aging and normal programmed cell death and regulate cell growth. They also inhibit protease activity and change methylase activity, which also regulate histocompatibility antigens and blood vessels and promote the formation of cells. The inactivation of these functions or the deletion or mutation of these genes can lead to the transformation of malignant cells, thus leading to the occurrence of cancer.

3.3 Cytokines

Many cytokines form a complex network to regulate the immune response of the human body in the local tumor microenvironment, and promote or inhibit the growth, invasion or metastasis of tumors. Cytokines mainly include interleukin-8, leukocyte inhibitory factor and interleukin-6, which are summarized in Figure 2.

3.3.1 Interleukin-8

Interleukin-8 belongs to the CXC chemokine group, which is involved in neutrophil activation and blood taxis. Interleukin-8

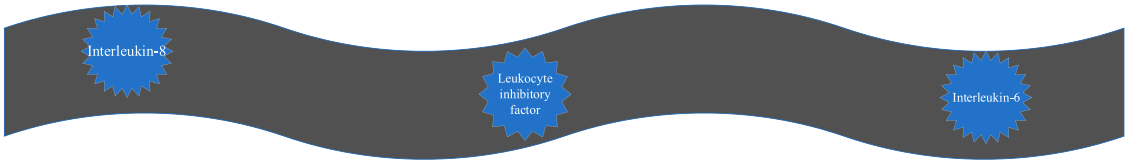


FIGURE 2
Cytokines.

TABLE 1 Cancers investigated in the paper.

Cancer type	Concept
cervical cancer	Early stage cervical cancer often has no obvious symptoms and signs, and the cervix may be smooth or difficult to distinguish from cervical columnar epithelial ectopic
Colon cancer	A common malignant tumour of the gastrointestinal tract occurring in the colon
Thyroid adenocarcinoma	Malignant tumours arising from the follicular or parafollicular epithelium of the thyroid
Liver Cancer	Malignant tumours of the liver

plays an autocrine role in pancreatic cancer tissue and stimulates vascular growth, which provides nutrition for pancreatic cancer cells and promotes the proliferation and invasion of pancreatic cancer cells.

3.3.2 Leukocyte inhibitory factor (LIF)

LIF is a growth factor that causes pancreatic cancer canceration. It promotes the proliferation of human pancreatic cancer cells by expressing c-fos gene, Jun-B gene and cyclin E. Clinically, the content of LIF in quasi pancreatic cancer cell line Hs-700T is low, which can induce STAT3 phosphorylation. However, LIF mediated pathway can be inhibited by protein kinase C, protein tyrosine kinase and calmodulin inhibitor.

3.3.3 Interleukin-6 (IL-6)

IL-6 is highly expressed in the matrix of many malignant tumors, so it is an important factor in the relationship between inflammation and tumor. In addition, IL-6 protects tumor cells from DNA damage, oxidative stress and apoptosis caused by treatment by promoting repair and inducing anti apoptosis pathway, which has anti-cancer effect. Therefore, the method of blocking IL-6 alone or in combination with traditional cancer

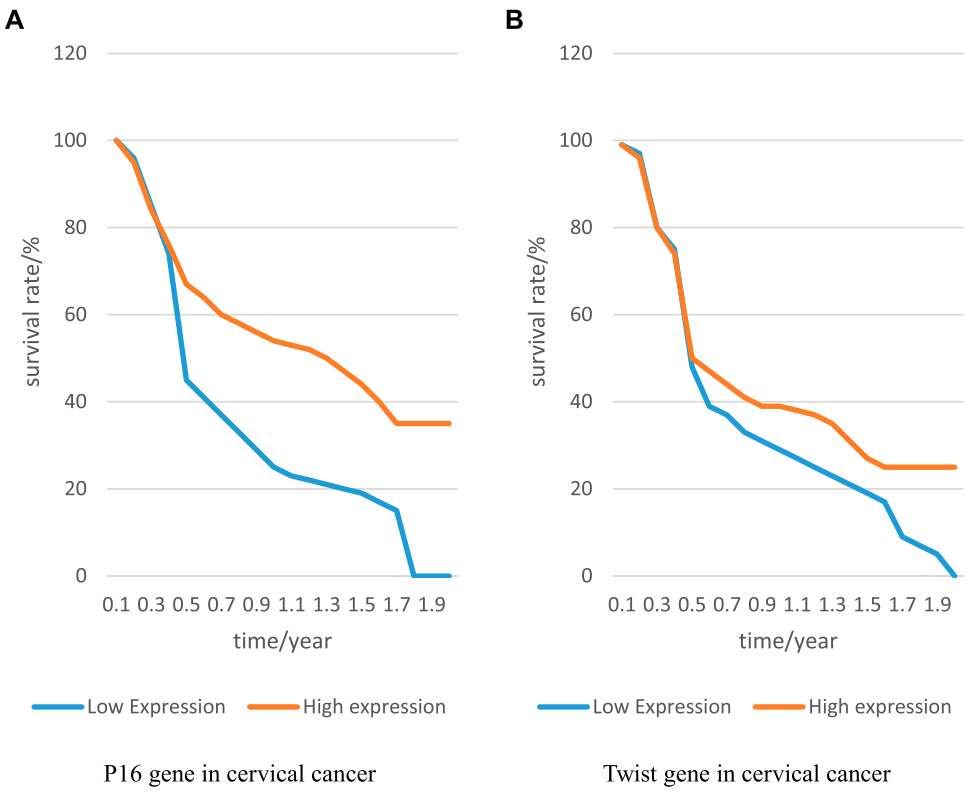


FIGURE 3
p6 and Twist genes and prognosis of cervical cancer. **(A)** P16 gene in cervical cancer, and **(B)** Twist gene in cervical cancer.

TABLE 2 Prognostic value of colon cancer genes.

Gene type	High expression is associated with shorter survival	High expression associated with prolonged survival
FDFT1	Yes	No
HMGCR	No	Yes
CARS1	No	Yes
AKR1C1	Yes	No
ALOX12	Yes	No

TABLE 3 Prognostic value of thyroid cancer genes.

Gene type	High expression is associated with shorter survival	High expression associated with prolonged survival
NDUFA12	Yes	No
NR2C1	No	Yes
FD6	Yes	No
VEZT	Yes	No
GDF3	Yes	No
FN1	No	Yes
PDE5A	Yes	No
GALNTL6	Yes	No
OPMR1	Yes	No
IPCEF1	No	Yes
AOAH	Yes	No
ELMO1	No	Yes

Among the 12 genes related to the prognosis of thyroid cancer investigated, the overexpression of NDUFA12, FD6, VEZT, GDF3, PDE5A, GALNTL6, OPMR1, AOAH was related to the shorter survival period; the high expression of NR2C1, FN1, IPCEF1, and ELMO1 was related to the longer survival period.

therapy may be a potential therapeutic strategy for the cancer where IL-6 signaling is dominant. IL-6 is highly expressed in some cancers, including breast cancer, lung cancer, stomach cancer, liver cancer and ovarian cancer. The high expression of IL-6 is associated with poor prognosis and can be used as a diagnostic marker of inflammation and malignant tumor.

4 Cancer microenvironment related genes based on intelligent medical IoT

The data in this paper were from the database of tumor genome map. The physical condition and physical health of cancer patients received extensive attention. In order to analyze cancer TME related genes, this paper used intelligent medical IoT technology to analyze the patient's physical data, and selected several common cancers for correlation analysis between cancer TME related factors and prognosis. The cancers investigated in this paper are summarized in Table 1.

4.1 Cervical cancer

The gene expression profile and clinical follow-up data of 20 cervical cancer patients were downloaded from the tumor genome map database, and the messenger RNA (mRNA) data of 188 cervical cancer patients with complete clinical follow-up information were downloaded from the two microarray datasets of the gene expression comprehensive database.

TISIDB database is an integrated storage portal for interaction between tumor and immune system. This study first determined the common prognostic related DEGs of TCGA dataset and GEO dataset, and then used TISIDB database to evaluate the interaction of common prognostic related DEGs in tumor and immune system.

The expression of P16 and Twist genes in cervical cancer was closely related to prognosis. This paper analyzed this cancer to explore the relationship between the expression of P16 and Twist genes and prognosis, and recorded the results to Figure 3:

In Figure 3A represents the relationship between P16 gene expression and patient survival, and B represents the relationship between Twist gene and patient survival. Compared with low expression of P16 gene, the life cycle of patients with high expression of P16 gene was prolonged, and it remained at 35% level; compared with low expression of Twist gene, the life cycle of patients with high expression of Twist gene was also prolonged, indicating that both genes had a great relationship with the prognosis of patients. Through the investigation and interview of patients, it was found that the recurrence rate of patients with positive expression of both genes was higher than that of patients

TABLE 4 Prognostic value of liver cancer genes.

Gene type	High expression is associated with shorter survival	High expression associated with prolonged survival
AGO2	Yes	No
DCPS	Yes	No
EIF4E3	No	Yes
EIF4G3	No	Yes
IFIT5	Yes	No
LARP1	Yes	No
METTL1	No	Yes
NCBP1	No	Yes
NCBP2	Yes	No
NSUN2	No	Yes
NUDT10	Yes	No
NUDT11	No	Yes
NUDT16	Yes	No
NUDT4	No	Yes
WDR4	No	Yes

Among the genes related to the prognosis of liver cancer, the genes related to the shorter survival period were AGO2, DCPS, IFIT5, LARP1, NCBP2, NUDT10, and NUDT16. The genes related to long survival period included EIF4E3, EIF4G3, METTL1, NCBP1, NSUN2, NUDT11, NUDT4, and WDR4.

with negative expression of both genes. This showed that if P16 gene was not controlled, cervical cancer would recur. Therefore, it is a target for the prognosis of cervical cancer. The presence or absence of Twist gene expression is related to the prognosis of patients and can be used as a preliminary prognostic marker.

4.2 Colon cancer

In this paper, the gene expression profile and clinical follow-up data of 20 colon cancer patients were downloaded from the tumor genome map database, and the survival results of five genes with similar prognostic value were recorded in Table 2:

Among the five genes related to the prognosis of colon cancer investigated, the high expression of FDFT1, AKR1C1, ALOX12 was related to the short survival period; high expression of HMGCR and CARS1 was associated with longer survival. The expression and function of genes had a very important relationship with the growth period of patients. Only by mastering the characteristics of different genes could the patient's symptom response be analyzed, thus bringing advantages to reduce the patient's symptoms. It could promote the low expression of FDFT1 gene, AKR1C1 gene, and ALOX12 gene, so as to improve the survival of patients. This kept HMGCR gene and CARS1 gene highly expressed to improve the survival of patients.

4.3 Thyroid carcinoma

In this paper, the gene expression profile and clinical follow-up data of 20 patients with thyroid cancer were downloaded from the tumor genome map database, and the survival results of 12 genes with similar prognostic value were recorded in Table 3.

4.4 Liver cancer

In this paper, the gene expression profile and clinical follow-up data of 20 patients with liver cancer were downloaded from the tumor genome map database, and the survival results of 15 genes with similar prognostic value were recorded in Table 4.

5 Conclusion

In order to analyze the correlation between cancer TME related genes and prognosis, this paper first introduced the intelligent medical IoT, and analyzed the application of the IoT in medicine. Subsequently, TME and its application were introduced. Later, the experiment was designed to analyze the prognosis related genes of patients with cervical cancer, colon cancer, thyroid cancer, and liver cancer. The conclusion was drawn. The relationship between different TME genes and prognosis was very different. The analysis of TME can improve the effect of reflecting cancer analysis, so when analyzing genes, it is necessary to grasp the characteristics of gene expression for cancer diagnosis. This analysis method can improve the prognosis and has potential value for cancer diagnosis and treatment. The use of IoT technology in biomedicine could greatly enhance the efficiency and quality of disease research, thus improving the analysis rate of cancer. In the future, the concept of intelligent medical treatment would gradually become popular. The use of novel research technologies and convenient research methods in this field could greatly improve the

effect of cancer analysis, so as to promote the public to keep healthy and meet their medical needs.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SR, WC, and JM: Writing-original draft preparation. HL, YX, and JZ: Editing data curation, Supervision.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Habibzadeh, H., Dinesh, K., Shishvan, O. R., Boggio-Dandry, A., Sharma, G., and Soyata, T. (2019). A survey of healthcare internet of things (HIoT): A clinical perspective. *IEEE Internet Things J.* 71, 53–71. doi:10.1109/jiot.2019.2946359

Hu, C., Liu, C., Tian, S., Wang, Y., Shen, R., Rao, H., et al. (2020). Comprehensive analysis of prognostic tumor microenvironment-related genes in osteosarcoma patients. *BMC cancer* 201, 814. doi:10.1186/s12885-020-07216-2

Islam, M., and Rahaman, A. (2020). Development of smart healthcare monitoring system in IoT environment. *SN Comput. Sci.* 13, 185. doi:10.1007/s42979-020-00195-y

Javaid, M., and Khan, I. H. (2021). Internet of Things (IoT) enabled healthcare helps to take the challenges of COVID-19 Pandemic. *J. Oral Biol. Craniofacial Res.* 112, 209–214. doi:10.1016/j.jobcr.2021.01.015

Kondka, L. S., Thenmozhi, M., Vijayakumar, K., and Kohli, R. (2022). An intensive healthcare monitoring paradigm by using IoT based machine learning strategies. *Multimedia Tools Appl.* 8126, 36891–36905. doi:10.1007/s11042-021-11111-8

Li, C., Liu, T., Liu, Y., Zhang, J., and Zuo, D. (2021). Prognostic value of tumour microenvironment-related genes by TCGA database in rectal cancer. *J. Cell. Mol. Med.* 2512, 5811–5822. doi:10.1111/jcmm.16547

Li, Wei., Chai, Y., Khan, F., Jan, S. R. U., Verma, S., Menon, V. G., et al. (2021). A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mob. Netw. Appl.* 261, 234–252. doi:10.1007/s11036-020-01700-6

Liu, Y., Cheng, L., Li, C., Zhang, C., Wang, L., and Zhang, J. (2021). Identification of tumor microenvironment-related prognostic genes in colorectal cancer based on bioinformatic methods. *Sci. Rep.* 111, 15040. doi:10.1038/s41598-021-94541-6

Luo, R., Guo, W., and Wang, H. (2021). A comprehensive analysis of tumor microenvironment-related genes in colon cancer. *Clin. Transl. Oncol.* 239, 1769–1781. doi:10.1007/s12094-021-02578-w

Mohammed, M. N. (2020). Novel COVID-19 detection and diagnosis system using IOT based smart helmet. *Int. J. Psychosoc. Rehabilitation* 247, 2296–2303.

Qadri, Y. A., Nauman, A., Zikria, Y. B., Vasilakos, A. V., and Kim, S. W. (2020). The future of healthcare internet of things: A survey of emerging technologies. *IEEE Commun. Surv. Tutorials* 22 (2), 1121–1167. doi:10.1109/comst.2020.2973314

Singh, R. P., Javaid, M., Haleem, A., and Suman, R. (2020). Internet of things (IoT) applications to fight against COVID-19 pandemic. *Diabetes & Metabolic Syndrome Clin. Res. Rev.* 144, 521–524. doi:10.1016/j.dsx.2020.04.041

Tian, S., Yang, W., Grange, J. M. L., Wang, P., Huang, W., and Ye, Z. (2019). Smart healthcare: Making medical care more intelligent. *Glob. Health J.* 33, 62–65. doi:10.1016/j.glohj.2019.07.001

Wang, Z., Tu, L., Chen, M., and Tong, S. (2021). Identification of a tumor microenvironment-related seven-gene signature for predicting prognosis in bladder cancer. *BMC cancer* 211, 692. doi:10.1186/s12885-021-08447-7

Yuan, J., Yuan, B., Zeng, L., Liu, B., Chen, Y., Meng, X., et al. (2020). Identification and validation of tumor microenvironment-related genes of prognostic value in lung adenocarcinoma. *Oncol. Lett.* 202, 1772–1780. doi:10.3892/ol.2020.11735

Zhang, Haibin., Li, J., Wen, B., Xun, Y., and Liu, J. (2018). Connecting intelligent things in smart hospitals using NB-IoT. *IEEE Internet Things J.* 53, 1550–1560. doi:10.1109/jiot.2018.2792423

Zhang, Z., Li, Z., Liu, Z., Zhang, X., Yu, N., and Xu, Z. (2020). Identification of microenvironment-related genes with prognostic value in clear cell renal cell carcinoma. *J. Cell. Biochem.* 1217, 3606–3615. doi:10.1002/jcb.29654

Zhu, H., Wu, C. K., Koo, C. H., Tsang, Y. T., Liu, Y., Chi, H. R., et al. (2019). Smart healthcare in the era of internet-of-things. *IEEE Consum. Electron. Mag.* 85, 26–30. doi:10.1109/mce.2019.2923929

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fgene.2023.1119990

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Youbao Ma,
Zhejiang University of Technology, China
Yanmin Wu,
Zhengzhou University of Light Industry, China
Jun Zheng,
Baotou Teachers' College, China

*CORRESPONDENCE
Rui Jiang,
192931103@st.usst.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 09 December 2022
ACCEPTED 02 February 2023
PUBLISHED 16 February 2023

CITATION
Wang J, Chen B, Zhu J, Zhang J and Jiang R (2023), Intelligent diagnosis value of preoperative T staging of colorectal cancer based on MR medical imaging. *Front. Genet.* 14:1119990. doi: 10.3389/fgene.2023.1119990

Intelligent diagnosis value of preoperative T staging of colorectal cancer based on MR medical imaging

Junqing Wang, Bingqian Chen, Jing Zhu, Junfeng Zhang and Rui Jiang*

Department of Radiodiagnosis, General Hospital of Western Warfare Zone, Chengdu, Sichuan, China

Colorectal cancer is a common malignant tumor in clinic. With the change of people's diet, living environment and living habits, the incidence of colorectal cancer has risen sharply in recent years, which poses a great threat to people's health and quality of life. This paper aims to investigate the pathogenesis of colorectal cancer and improve the efficiency of clinical diagnosis and treatment. This paper firstly introduces MR Medical imaging technology and related theories of colorectal cancer through literature survey, and then applies MR technology to preoperative T staging of colorectal cancer. 150 patients with colorectal cancer admitted to our hospital every month from January 2019 to January 2020 were used as research objects to carry out the application experiment of MR Medical imaging in the intelligent diagnosis of preoperative T staging of colorectal cancer, and to explore the diagnostic sensitivity, specificity and histopathological T staging diagnosis coincidence rate of MR Staging. The final study results showed that there was no statistical significance in the general data of stage T1-2, T3 and T4 patients ($p > 0.05$); for patients with preoperative T stage of colorectal cancer, the overall diagnosis coincidence rate of MR Was 89.73%, indicating that it was highly consistent with pathological T stage; compared with MR Staging, the overall diagnosis coincidence rate of CT for preoperative T staging of colorectal cancer patients was 86.73%, which was basically consistent with the diagnosis of pathological T staging. At the same time, three different dictionary learning depth techniques are proposed in this study to solve the shortcomings of long MR Scanning time and slow imaging speed. Through performance testing and comparison, it is found that the structural similarity of MR Image reconstructed by depth dictionary method based on convolutional neural network is up to 99.67%, higher than that of analytic dictionary and synthetic dictionary, which proves that it has the best optimization effect on MR Technology. The study indicated the importance of MR Medical imaging in preoperative T staging diagnosis of colorectal cancer and the necessity of its popularization.

KEYWORDS

medical imaging technology, magnetic resonance imaging, colorectal cancer, preoperative T staging, intelligent diagnosis and treatment

1 Introduction

1.1 Background and meaning

The recent changes in people's diet, living environment, and living habits have also caused a series of malignant diseases such as colorectal cancer. Colorectal cancer is a malignant tumor disease common in middle-aged and elderly people. It is a type of disease in the digestive system and has a very high incidence in our country and the world. Because its early symptoms are hidden, it is easy to miss early diagnosis and treatment. Most patients have reached the middle and late stages when they are diagnosed, so they have a higher mortality and disability rate. Nowadays, MR medical imaging technology is widely used in clinical practice and has achieved good results (Xiao and Ding, 2019). This article's goal is to examine the value of MR medical imaging in the accurate diagnosis of colorectal cancer's preoperative T staging and to discover a successful method for early detection and treatment of colorectal cancer. Improving the diagnosis rate of patients will also improve their quality of life.

1.2 Related work

The incidence of early colorectal cancer is insidious, the clinical symptoms are not prominent, and there are many uncertainties with the increase of cancer. Because of the great harm of colorectal cancer to humans (Zhu et al., 2020), colorectal cancer has been studied very early, and many methods of diagnosis and treatment of colorectal cancer have been explored. So J S, Cheong C, and Oh S Y once stated that for patients with colorectal cancer, preoperative staging and application of various imaging techniques are of great significance to formulating treatment plans and predicting prognosis. For this reason, they discussed the use of CT technology to diagnose colorectal cancer (So et al., 2017). Xu Jiayi, Wang Jinkai and Zhou Lu discussed the value of serum C-reactive protein (CRP), sugar chain antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA) in the preoperative diagnosis of colorectal cancer (Xu et al., 2017). Ma K investigated the use of CT in the treatment of common malignant tumors including lung and colorectal cancer. They emphasized that while CT technology does not significantly contribute to the management of cholangiocarcinoma, it is helpful in the detection and management of colorectal cancer (Ma et al., 2018). Jaramillo FA and Daniel Upegui Jiménez proposed that CT colorectal cancer is the fourth leading cause of death in the world and the fifth leading cause of cancer death in Colombia. They believe that MRI is an ideal method to evaluate colorectal cancer, especially for screening, because it can be staged by determining the degree of invasion of the muscle layer and adjacent organs, which is useful for determining candidates for chemotherapy or preoperative radiotherapy and planning surgery procedure is crucial (Jaramillo and Upegui Jiménez, 2016). In addition, Park SH et al. stated that the preoperative colorectal tumor location is essential for proper resection and treatment planning. In response to the low positioning accuracy of traditional colonoscopy, they proposed to develop several new positioning techniques. They reviewed the tumor localization error rates of several preoperative endoscopic techniques, combined information about localization errors and risk factors for surgery-related adverse events, and concluded an effective method for accurately localizing colorectal tumors (Park et al., 2017). It can be seen from the above research results that although there are various methods for the diagnosis and treatment of colorectal cancer, there are still relatively few studies using MR medical imaging technology to intelligently diagnose the preoperative staging of colorectal cancer. Therefore, this article attempts to use this technology to explore the clinical intelligent diagnosis and treatment of colorectal cancer, with a view to adding a new treatment method to clinical treatment.

1.3 Innovations in this article

The innovations of this article are mainly reflected in the following aspects: 1) A malignant gastrointestinal tumor, colorectal cancer is exceedingly dangerous to people's health and has a negative impact on patients' quality of life. The clinical aspects of colorectal cancer are covered in this article. Methods of diagnosis and treatment greatly aid in the improvement of patient quality of life, the rate of early diagnosis, and the social practical importance and value; 2) The use of MR medical imaging technologies in the preoperative T staging of colorectal cancer is explored in this study. It is also suggested to apply deep learning algorithms to address the shortcomings of sluggish imaging speed and lengthy MR scanning times, which is crucial for enhancing and optimizing the performance of MR technology and enhancing its use in clinical diagnosis.

2 MR medical imaging technology and its intelligent diagnostic value for preoperative T staging of colorectal cancer

2.1 MR medical imaging

(1) MR

MR, or magnetic resonance, is a physical phenomenon related to the gravitational theory of magnetic fields (Joany and Logashanmugam, 2018). The principle is to combine the externally applied radio frequency energy field with the proton energy field of the human body. In a strong magnetic field environment, the proton can be thought of as a spin-nucleus system that can reflect certain aspects of the system by absorbing the externally delivered matching radio frequency energy field (Song et al., 2017; Guo et al., 2019).

(2) MRI

MRI, is magnetic resonance imaging, that is, magnetic resonance medical image. It is a type of magnetic resonance imaging-based imaging technology used for clinical disease diagnosis and treatment. When protons in the human body or the spin-nucleus system are magnetized to form a macroscopic magnetization vector, they will be driven by the radio frequency field. After being forced to return to its equilibrium position, the magnetization vector will keep

circling the spin-nucleus system, eventually forming a closed coil. The coil will generate a magnetic resonance signal as a result of the nuclear system's steady rotation. Since this kind of magnetic resonance signal cannot distinguish between the various positions of the spin nucleus, gradient fields must be applied in different directions of the spin nucleus. Through independent marking of its spatial position information during rotation, the spin nucleus is able to do so (Ali et al., 2019; Hoshino et al., 2019).

(3) Three-dimensional modeling of MR medical images

MR medical imaging uses the gravitational force of a magnetic field to treat the protons of the human body as a spin nucleus system, and absorbs the externally applied matching RF power field to generate magnetic resonance signals, creating a two-dimensional image in the process. The spatial position of the MR image will continue to consistently correspond to distinct pixel coordinates because of the nuclear system's continual rotation (Yao et al., 2016; Pradhan et al., 2020). The patient's lesion location is eventually localized in three dimensions using a three-dimensional picture. Transverse, coronal, sagittal, and any other cross-sectional images of the human body are included in general MR three-dimensional medical imaging (Acuna et al., 2017).

(4) MR image processing

Before they can be used effectively, medical photographs typically need to be cleaned, denoised, improved, and subject to other processes. This is because, during the picture capture process, the image will be susceptible to varying degrees of external interference, and the final image will be fuzzy or noisy (Lu and Wang, 2018). The magnetic field of the human body interacts with the radio waves of the MRI scanner to produce magnetic resonance images. Magnetic resonance signals are produced when the protons in the human body interact with radio wave energy. The coordinates of the various places of the tissues being investigated are represented by these signals. A tissue image of the human body is created by the fusion of many locations (Shakeel et al., 2020; Xu et al., 2022). We frequently employ image data processing technology for picture processing, and the procedure is as follows:

1) Image denoising and filtering. There are two widely used algorithms: the top-hat transform algorithm and the dual-tree complex wavelet transform algorithm.

The dual-tree complex wavelet transform approach comes first. On the basis of the complex wavelet, the dual-tree complex wavelet is produced. The formula for one-dimensional data transformation is:

$$\Phi(t) = \Phi_h(t) + i\Phi_g(t) \tag{1}$$

t is the time, $\Phi(t)$ is the total function.

When the complex wavelet becomes a dual-tree complex wavelet, its two-dimensional data transformation formula is:

$$\Phi(a,b) = \Phi(a)\Phi(b) \tag{2}$$

where $\Phi(a,b)$ is a two-tree complex wavelet. Second: the top-hat transformation method. Open and close the original image as follows:

$$\begin{aligned} f \circ A &= (f \Theta A) \oplus A \\ f \bullet A &= (f \oplus A) \Theta A \end{aligned} \tag{3}$$

Then the top hat transformation is as follows:

$$\begin{aligned} WTH(a,b) &= f(a,b) - f \circ A(a,b) \\ BTH(a,b) &= f \bullet A(a,b) - f(a,b) \end{aligned} \tag{4}$$

WTH (.) is the form of image transformation, and BTH (.) is the result of transformation.

2) Image augmentation. Typically, frequency domain and air domain methods are used.

First: To achieve the goal of boosting the contrast, the spatial technique involves processing the picture pixels in the space where the image is placed. The airspace method is expressed as follows:

$$g(a,b) = U[f(a,b)] \tag{5}$$

g (a, b) is the image pixel in space, and U is the transformation function.

Second: The image is first transformed by the frequency domain method, the parameters are then improved, and lastly the enhanced image is changed back to the original region (Guo et al., 2018).

(6) Advantages and disadvantages of magnetic resonance imaging

At present, in addition to MR, the commonly used medical diagnostic imaging techniques in clinical use include CT and X-ray. X-ray is the earliest and most common medical imaging technology. It has strong penetrability and is suitable for imaging examinations of high-density tissues. It is generally used for bone examination, but the imaging is not clear and will produce harmful radiation to the human body. CT images are clearer than X-ray images and can be used to examine internal organs and brain tissue, but the radiation it produces is more harmful than X-rays. In contrast, MR not only has a high-definition resolution, but also does not generate harmful radiation to the human body. It can also perform arbitrary slices in different directions according to the patient's different body positions. However, MR images also have a shortcoming that cannot be ignored. The equipment takes a long time to scan, and the inspection of one part often takes a long time, which not only aggravates the patient's pain, but also increases its economic burden (Hu et al., 2020; Monroe et al., 2021). Therefore, in order to improve this problem and improve the speed and efficiency of MR image scanning, this article attempts to optimize and improve MRI. This content will be specifically introduced in the third section with colorectal cancer as an example.

2.2 Colorectal cancer

(1) The symptoms of colorectal cancer

TABLE 1 General information of patients.

Group	Man	Woman	Average age	Average age of disease
T1-2	27	14	51 ± 2.14	9 ± 2.17
T3	34	24	54 ± 3.22	7 ± 3.16
T4	27	20	53 ± 2.57	11 ± 2.34

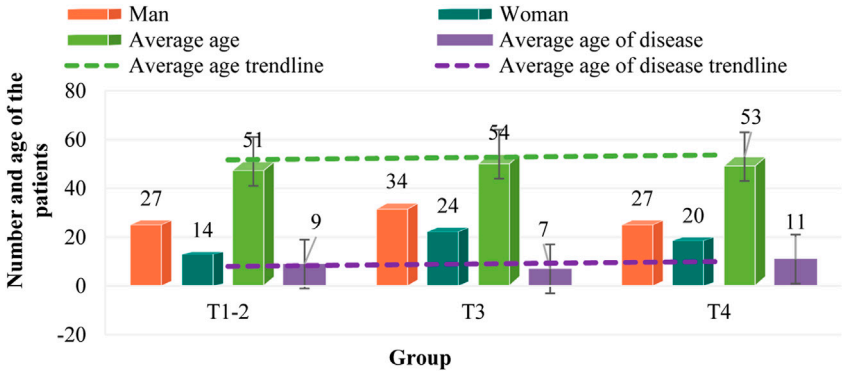


FIGURE 1
Comparison of general patient information.

Colorectal cancer, also known as colorectal cancer, is a common malignant tumor of the digestive tract, including colon cancer and rectal cancer. Its lesions often occur in the colorectal epithelial tissue. The high incidence of colorectal cancer is mainly middle-aged and elderly people between 54 and 81 years old, mostly in developed countries. At present, the pathological mechanism of colorectal cancer is not yet clear, but it is closely related to factors such as people's diet, living environment, family hereditary polyps and chronic inflammation. Studies have shown that people who eat high-fat, high-calorie, low-fiber foods are more likely to develop colon cancer (Yang et al., 2019). China has long been a country with a low incidence of colorectal cancer, but with the improvement of living conditions, people's diet and living conditions have undergone significant changes, and the incidence of colorectal cancer has gradually increased.

1) The early stage of the disease is difficult to detect, and the symptoms are extremely hidden, which often leads to missed and misdiagnosed phenomena. Many patients are often in the late stage of the disease when they are diagnosed. Therefore, the disease has a high mortality and disability rate;
2) The elderly are a frequent group of the disease, which is extremely harmful to the quality of life and health of the elderly;
3) The tumor cells in the lesion are prone to metastasis, which affects the normal physiological functions of the surrounding organs and other parts of the body;
4) Postoperative complications are obvious and difficult to cure, and the quality of life of patients has significantly decreased (Sun et al., 2017).

(2) Clinical manifestations of colorectal cancer

1) Early clinical symptoms. In the early stages of the disease, the symptoms of colorectal cancer are not yet obvious. However, when the tumor in the intestine grows larger, the patient's bowel habits will gradually change, showing symptoms such as bleeding stool, diarrhea, alternating diarrhea and constipation, and local abdominal pain. The frequency of excretion increases, accompanied by a small amount of mucus and blood in the stool.
2) Middle and late clinical symptoms. In the middle and late stages, the tissues and organs around the colorectal have different degrees of necrosis and changes, and their functions are obviously impaired. For example, the surrounding tissues such as the bladder and prostate will have symptoms such as frequent urination, urgency and difficulty urinating. In severe cases, the tumor cells of colorectal cancer will also move and metastasize to distant tissues, such as liver and lungs. Patients with advanced colorectal cancer may also experience weight loss, loss of appetite, and anemia (Que, 2018).

(3) Preoperative T staging of colorectal cancer

T stage: The tumors located in the local intestinal mucosa were divided into T1, T2, T3 and T4 according to the depth of invasion. T4 suggested that the surrounding structures and tissues were invaded. The larger the number, the later the stage. Preoperative T staging is a preoperative preparation for the treatment of gastrointestinal diseases. Colorectal cancer is a malignant gastrointestinal tumor, and surgery is quite risky, coupled with the special physiological structure and function of the human rectum. Therefore, in order to increase the success rate of surgery and reduce the risk of surgery, it is very important to determine and stage the

TABLE 2 Comparison of MR staging and pathological T staging.

Pathological staging MR staging				T1-2	T3	T4	Total
T1-2				19	25	11	55
T3				17	18	21	56
T4				5	15	15	35
Total	41	58	47	146			

condition of colorectal cancer patients before surgery. The more accurate the preoperative staging is, the better it is for patients to undergo adjuvant chemotherapy before surgery, which can reduce the stage of malignant tumors, and also facilitate the selection and arrangement of surgical plans, thereby achieving the purpose of improving the success rate of surgery and the survival rate of patients. It is also very effective for the prognosis of patients, reducing the recurrence rate of colorectal cancer patients after surgery and greatly improving their quality of life (Frank et al., 2018).

2.3 Value of MR medical imaging on the intelligent diagnosis of preoperative T staging of colorectal cancer

With the development of medical technology, X-ray, CT and MR technologies have gradually been widely used in the medical field. Although MR has advantages over X-ray and CT in terms of imaging quality and impact on the human body, long scanning time and slow imaging speed are also its biggest disadvantages. High-quality images provide more accurate positioning for the preoperative staging of colorectal cancer. This article attempts to use colorectal cancer as a research sample to reconstruct its MR images by improving and optimizing MR imaging technology (Fan et al., 2017). Here, we will use the concept of deep learning to improve MR technology through deep learning algorithms, and use three different dictionary learning methods to reconstruct MR images. Deep dictionary method based on convolutional neural network can better process the structural details in the image to produce clearer results when training the convolutional neural network model.

(1) Reconstruction of magnetic resonance images using a parsing dictionary

$$TV(v) = \iint \nabla v(x,y)\|_1 dxdy = \iint \left|\frac{\partial v}{\partial x}\right| + \left|\frac{\partial v}{\partial y}\right| dxdy \tag{6}$$

$$TV(v) = \iint \left\|\nabla v(x,y)\right\|_2^2 + \Phi^2 dxdy$$

$$= \iint \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \Phi^2 dxdy \tag{7}$$

The MRI image is optimized and solved using the norm of total variation, as in formula (8)

$$\min_x \left\|F_u x - y\right\|_2^2 + \lambda TV(x)_1 \tag{8}$$

(2) Reconstruction of magnetic resonance images using a synthetic dictionary

In contrast to the analytical dictionary's quick and easy computation approach, the synthetic dictionary can describe more complicated images, has some adaptability, and can lessen the noise and filtering issues that the analytical dictionary has because of down sampling (Sibertinblanc et al., 2016). Here, we focus primarily on introducing the sparse representation, dictionary building, and synthetic dictionary reconstruction processes.

A transformation matrix created using the corresponding points of the magnetic resonance signal is the so-called dictionary. The magnetic resonance image's sparse representation is

$$y = D\alpha \tag{9}$$

The sparse coefficient is represented by a, while D is the dictionary matrix.

Create a synthetic dictionary after that, using the formula for the mathematical model (10)

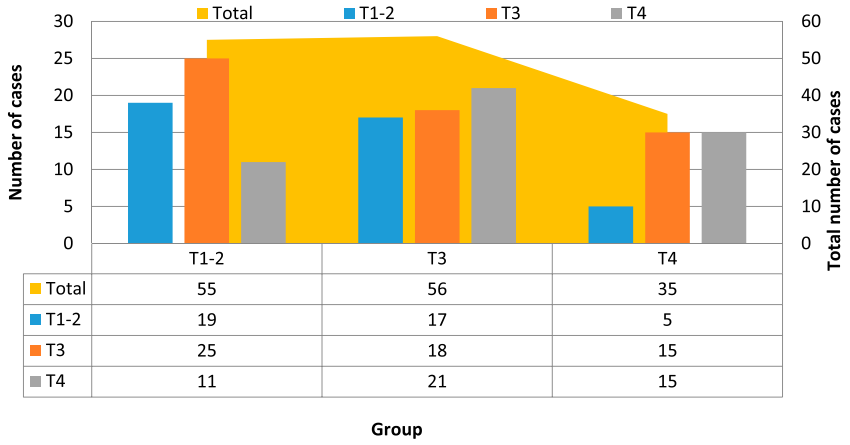


FIGURE 2
Comparison of MR staging and pathological T staging.

TABLE 3 Diagnostic coincidence rate, sensitivity, and specificity of MR staging.

	Sensitivity	Specificity	Diagnosis coincidence rate	*p*	X2
T1-2	46.34% (19/41)	90.67% (68/75)	89.73% (131/146)	0.021	0.85
T3	31.03% (18/58)	92.1% (70/76)	95.89% (140/146)	0.025	0.83
T4	31.91% (15/47)	96.67% (87/90)	99.32% (145/146)	0.042	0.81

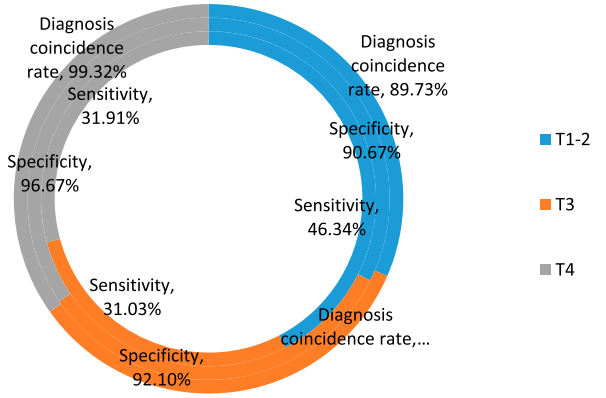


FIGURE 3
Sensitivity, specificity and diagnostic coincidence rate of MR staging.

TABLE 4 Comparison of CT staging and pathological T staging results.

Pathological staging CT staging	T1-2	T3	T4	Total
T1-2	14	18	9	41
T3	12	11	11	34
T4	10	8	5	23
Total	36	37	25	98

$$\min_{D,X} \|Y - DX\|_2^2, \mathrm{s,t}, \forall \mathrm{i}, \|\mathrm{x_i}\|_0 \le T \tag{10}$$

Split the model into a sparse representation problem of a single sample, and get:

$$\min_{x_i} \|y_i - Dx_i\|_2^2, \mathrm{s,t}, \forall \mathrm{i}, \|\mathrm{x_i}\|_0 \le T \tag{11}$$

(3) Deep dictionary-based magnetic resonance image reconstruction

Convolutional neural network algorithms are combined in this deep dictionary, which reconstructs MRI images by utilizing the powerful flexibility and self-learning capabilities of ANNs. We previously presented two strategies for learning dictionaries. While the synthetic dictionary has some adaptability, its ability to denoise and filter the image is superior. The analytical dictionary is a fixed transformation that can only handle a few simple image changes, and its sparse expressiveness is weak. Strong, however the reconstructed image effect is unsatisfactory when there is less image data (Park and Lee, 2017). So, we once more suggest a convolutional neural network-based deep dictionary approach.

$$H(R,b;Px,y) = \frac{1}{2}\|h_{r,b}(x) - y\|^2 \tag{12}$$

which represents the cost function, represents the connection parameters between the layers. Assuming a data set containing n samples, the overall cost function is:

$$H(W,b) = \left(\frac{1}{n}\sum_{i=1}^{n} H(W,b;x^{(i)},y^{(i)})\right) + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{j=1}^{s_l+1}\left(W_{ji}^l\right)^2$$
$$= \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\|h_{w,b}(x^{(i)} - y^{(i)})\|^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(W_{ji}^l\right) \tag{13}$$

The cost function is iterated by the dimensionality reduction method, and then

$$W_{ij}^l = W_{ij}^l - \alpha\frac{\partial}{\partial w_{ij}^l}H(W,b) \tag{14}$$

$$b_i^l = b_i^l - \alpha\frac{\partial}{\partial b_i^l}H(W,b) \tag{15}$$

Among them, α represents the rate of learning.

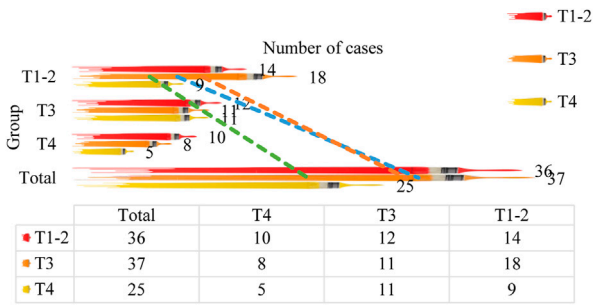
3 Application experiment of MR medical imaging in the intelligent diagnosis of preoperative T staging of colorectal cancer

3.1 General information collection

In order to specifically explore the application value of MR medical imaging in the preoperative T-analysis intelligent diagnosis of colorectal cancer, this article collected 150 colorectal cancer patients admitted to our hospital from January 2019 to January 2020. All patients were admitted to the hospital. Routine examinations were performed without any preoperative treatment. MR examinations were performed before the operation. Patients who could not tolerate MR technical examinations and surgical contraindications were excluded. Finally, 146 effective cases were selected, including 86 males and 60 females; age 38–73 years old, the average age is 56 ± 2.45 years, and the average age of illness is 14 ± 3.56 years. The general information of the patient is shown in Table 1.

3.2 MR inspection method

After grouping every patient in accordance with the preoperative T staging, they were subjected to preoperative MR technical



	Total	T4	T3	T1-2
T1-2	36	10	12	14
T3	37	8	11	18
T4	25	5	11	9

FIGURE 4
Comparison of CT staging and pathological T staging results.

examination to collect colorectal tumor cell images. The night before the examination, all patients took bowel cleansing drugs and fasted food and water. Use our own MRI scanner for operation, with 16-channel body phased array coils. The parameters of the scanner are: pulse sequence repetition time 2600–3400 ms, echo time 150 ms, field of view 400 mm, the thickness is 5 mm, the layer spacing is 1 mm, the matrix is 260 × 298, the number of excitations is 3, and the flip angle is 90°. First, the patients in each group were injected with 0.1 mmol/kg gabapentin meglumine as a contrast agent by intravenous injection, and kept advancing at a speed of 2 ml/s. MR scanning was performed immediately after the injection was completed. At the same time, the patient was instructed to keep breath-hold during the scan, first scan at a uniform speed, and then strengthen the local scan until a complete colorectal image is obtained.

3.3 Staging standards and observation indicators

Stage T1-2: there are obvious gaps in the fat at the lesion of the colorectal intestinal wall. Even the enhanced scan shows that the outer edge of the intestinal wall is smooth and there is no sign of nodules. T3: fat around the lesion on the intestinal wall. There are sparse markings, and the muscle layer has been invaded to a certain extent. When the scan is increased, the outer edge of the intestinal wall is uneven, and nodules are slightly prominent; stage T4: there are no gaps in the fat around the lesion of the intestinal wall and the adjacent tissues and organs. The boundary is blurred during enhanced scanning. The tumor has seriously invaded the colorectal fascia and surrounding organs.

According to the results of MR examination, the degree of invasion of each layer of the colorectal wall is evaluated, and the sensitivity, specificity, and rate of diagnostic coincidence for each colorectal cancer stage are assessed in comparison to the histological T staging.

3.4 Data processing

The data processing and analysis of this experiment used SPSS 22.0 statistical analysis software. The MR staging results of colorectal cancer and the pathological T staging results were tested for variance and chi-square, and the sensitivity, specificity and diagnostic coincidence rate of MR staging were compared. Among them, $p > 0.05$ indicates no significant statistical difference, $p < 0.05$ indicates significant statistical difference, $p < 0.01$ indicates extremely significant statistical difference; Chi-square 1 indicates high degree of agreement, 2 indicates general agreement, 3 indicates a low degree of compliance.

4 Discussion of experimental results

4.1 Comparison of differences in general patient information

According to the general data of colorectal cancer patients collected in the third part, the general conditions of each group of patients are counted, and the *p*-value and chi-square value of each group of patients on gender, age and disease age are calculated, and plotted as shown in Figure 1. The histogram to compare the differences in the general conditions of patients in each group.

As can be seen from Figure 1, in the samples collected in T1-2, there are 27 men and 14 women, with an average age of 51 years, and the average age of patients is 42 years. In the samples collected in T3, there are 34 men and 24 women, with an average age of 54 years, and the average age of patients is 47 years. In T4, there are 27 men and 20 women, with an average age of 53 years, and the average age of patients is 42 years. There was no significant difference in general information among T1-2, T3 and T4 patients ($p > 0.05$). The age of patients in each group is 50–60 years old, and the age of onset is about 10 years. The majority of men and women are men. Among the three groups, the number of patients in T3 group was the largest, with 58 cases, followed by T4 group with 47 cases, and T1-2 group with 41 cases.

4.2 Comparison of results between MR staging and pathological T staging

To ascertain the sensitivity, specificity, and diagnostic coincidence rate of MR staging, the patients' preoperative T staging was performed post-experimentally in accordance with the findings of the MR scan. The results were then compared with the pathological T staging performed prior to the experiment.

(1) Comparison of results between MR staging and pathological T staging

The results of MR staging and pathological T staging are compared as shown in Table 2 and Figure 2.

Table 2 and Figure 2 show that the pathological T staging detected and diagnosed by MR technology differs in certain ways from the T staging ($p < 0.05$). According to the pathological T staging, 146 patients were divided into T1-2 stages. 41 cases, 58 cases in T3 stage, 47 cases in T4 stage, and after MR examination, the T stage of the patients was 55 cases in T1-2 stage, 56 cases in T3 stage, and 35 cases in T4 stage. This shows that the symptoms of 14 patients were overestimated, of which 12 were misestimated as T4 stage, and 2 were misestimated as T3 stage. These results show that the steps and contents of the study need to be more specific, so this is only an individual case.

TABLE 5 CT staging's sensitivity, specificity, and rate of diagnostic concordance.

	Sensitivity	Specificity	Diagnosis coincidence rate	*p*	X2
T1-2	33.33% (12/36)	87.8% (36/41)	86.73% (85/98)	0.036	0.75
T3	21.62% (8/37)	76.79% (43/56)	87.76% (86/98)	0.026	0.79
T4	28% (7/25)	97.22% (70/72)	91.84%90/98)	0.038	0.81

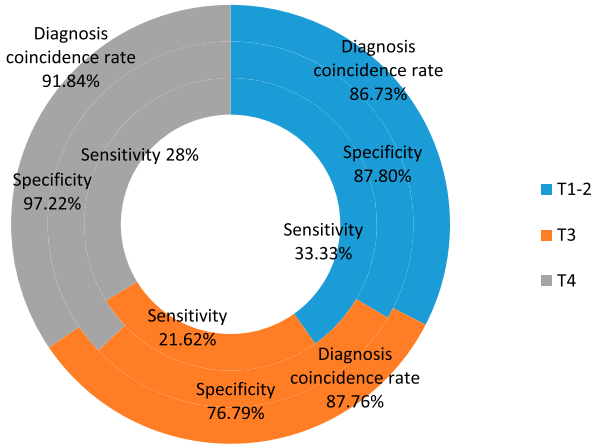


FIGURE 5
CT staging's sensitivity, specificity, and diagnostic coincidence rate.

(2) Evaluation of the sensitivity, specificity, and diagnostic coincidence rate of MR staging

The findings, which are shown in Table 3 and Figure 3, include evaluations of the sensitivity, specificity, and rate of diagnostic coincidence for MR staging as well as comparisons between pathological T staging results and MR staging results.

In Table 3, in T1-2 phase, the sensitivity, specificity and coincidence rate of MRI were 46.34%, 90.67% and 89.73% respectively. In T3 phase, their sensitivity, specificity and coincidence rate were 31.03%, 92.1% and 95.89% respectively. In T4 phase, their sensitivity, specificity and coincidence rate were 31.91%, 96.67% and 99.32% respectively.

From Table 3 and Figure 3, it can be seen that the total diagnostic coincidence rate of MR for preoperative T staging of colorectal cancer patients is 89.73%, the sensitivity of each stage is 46.34%, 31.03% and 31.91%, and the specificity is 90.67%, 92.1% and 96.67%, *p*-values were 0.021, 0.025, and 0.042, respectively, and the chi-square values were all greater than 0.8, which indicates that the preoperative T staging of MR and the pathological T staging have a higher degree of agreement, and the agreement and consistency strong.

4.3 Comparison of results between CT staging and pathological T staging

In order to further explore the application value of MR technology in the intelligent diagnosis of colorectal cancer preoperative T staging, and at the same time apply CT technology to the preoperative T staging diagnosis of colorectal cancer, compare the difference between CT staging and pathological T staging, and it is compared with MR staging to analyze the advantages of MR staging and CT staging. In addition, 98 patients with colorectal cancer admitted to our hospital were selected. The T stages were 36 cases in T1-2 stage, 37 cases in T3 stage, and 25 cases in T4 stage. According to the operation steps of MR examination, these 98 patients were examined by CT, and finally the CT staging was obtained, and the results were compared with the pathological T staging, as shown in Table 4 and Figure 4.

Table 4 and Figure 4 demonstrate that there are several differences between the T staging and the pathological T staging detected and diagnosed by CT technology ($p < 0.05$). According to the pathological T staging, 98 patients were divided into T1-2 stages. There were 36 cases in T3 stage, 37 cases in T4 stage, 25 cases in T4 stage. After CT examination, the T stage of patients was 41 cases in T1-2 stage, 34 cases in T3 stage, and 23 cases in T4 stage. This shows that the symptoms of five patients were overestimated, of which 3 were overestimated as T3 stage, and 2 were overestimated as T4 stage.

The *p*-value and chi-square value of the two were determined in accordance with the distinction between CT staging and pathological staging, and the sensitivity, specificity, and diagnostic coincidence rate of CT staging were examined. Table 5 and Figure 5 present the findings.

Table 5 and Figure 5 show that the overall CT diagnostic coincidence rate for colorectal cancer patients' preoperative T staging is 86.73%, the sensitivity of each stage is 33.33%, 21.62% and 28%, and the specificity is 87.8%, 76.79% and 97.22%, the *p*-values were 0.036, 0.026 and 0.038, and the chi-square values were all less than 0.8. This shows that although the diagnosis of CT staging and pathological T staging has a significant statistical difference, the agreement is average. Compared with MR staging, its agreement is slightly lower, which proves that MR technology is useful for preoperative T staging of colorectal cancer.

4.4 Effect of optimization of MR technology using various deep learning algorithms

According to the foregoing, among the existing clinical auxiliary diagnostic imaging technologies, although many studies have shown that MR technology is the most effective in auxiliary diagnostics, it also has serious problems such as long scanning time and slow acquisition of data and images. This article attempts to improve and optimize its core steps based on the working principle of MR technology. In order to improve the reconstruction of colorectal cancer pictures obtained by MR, three distinct dictionary learning

TABLE 6 Effect of optimization of MR technology using various deep learning algorithms.

Learning algorithm		30%	40%	50%	60%
Analytic dictionary	Peak signal to noise ratio(A)	32.25	36.29	38.47	41.56
	Structural similarity(A)	0.8521	0.8735	0.9174	0.9387
Synthetic dictionary	Peak signal to noise ratio(S)	33.54	37.55	41.45	44.67
	Structural similarity(S)	0.8741	0.9145	0.9279	0.9536

TABLE 7 Optimization results based on deep dictionary.

	50	100	150	200	250	300
Peak signal to noise ratio	34.47	34.47	36.55	41.89	42.38	45.54
Structural similarity	0.9118	0.9357	0.9472	0.9551	0.9678	0.9967
p-value	0.256	0.478	0.351	0.571	0.610	0.617
X2	0.897	0.880	0.857	0.872	0.810	0.835

methods are proposed (The complete calculation procedure and deduction steps refer to Section 2.3 of this article). In this section, we will compare the differences in peak signal-to-noise ratio and structural similarity of the reconstructed MR images produced by three different dictionary learning algorithms, and identify the algorithm that produces the best image reconstruction quality and the quickest imaging speed.

(1) Results of analytic and synthetic dictionary optimization

To sample MR pictures, use analytical dictionary and synthetic dictionary learning methods. Set the sampling rates to 20%, 30%, 40%, and 50% depending on the desired sampling rate. Under the same sample rate background, compare the two. Different learning is used to determine the reconstructed image's peak signal-to-noise ratio and structural similarity. Table 6 display the statistical findings. Analytical dictionary is 32.25% when sampling rates are 30%.

Table 6 shows that the peak signal-to-noise ratio and structural similarity of the synthetic vocabulary are higher than the values of the analytical dictionary under the same sampling rate, demonstrating that the synthetic dictionary has a superior influence on picture reconstruction.

(2) The optimization result of the deep dictionary

The deep dictionary algorithm we suggested is built on a convolutional neural network, based on the aforementioned. Sample training and testing are necessary depending on the neural network's properties. Select the first 200 MR images as the training group, the last 100 as the test group, and 300 MR images as the sample data. With 50, 100, 150, 200, 250, and 300 samples, respectively, iterative training and testing are conducted. Each sample interval is recorded. The results are displayed in Table 7 compare the peak-to-noise ratio and structural similarity of the image. They also compare the difference between the *p*-value and the chi-square value of each sample interval.

Table 7 show that when the sample size grows, the quantitative evaluation value of the deep dictionary-reconstructed image rises, showing a greater effect. The structural similarity also shows that the deep dictionary-reconstructed image is superior to the prior image. The two algorithms produce images that are more faithful to real MR scans. The structural similarity value reaches 0.9967 when the sample size is 300, which is significantly higher than the values obtained by the first two techniques. The sample training results are compatible with the test results (X2 > 0.8) and there is no statistically significant difference between the sample intervals, as can be observed from the *p*-value and Chi-square value ($p > 0.05$). As can be observed, the MRI picture generated by the deep dictionary approach based on the convolutional neural network has the best quality and the fastest imaging speed of the three dictionary learning methods.

5 Conclusion

Colorectal cancer is a malignant tumor disease that threatens human health and quality of life, especially among middle-aged and elderly people. Because its early symptoms are hidden and difficult to detect, its death and disability rates are high. Preoperative T staging of colorectal cancer is important for creating a precise surgical treatment plan, which is important for enhancing patient quality of life and increasing long-term survival rates.

MR is a cutting-edge imaging technique used in clinical and auxiliary medicine. In individuals with colorectal cancer, it is crucial to examine the tumor's location, its depth, and its connections to other tissues. Compared to CT and conventional X-rays, it offers additional benefits. Rectal cancer preoperative T staging offers significant diagnostic significance.

This study confirmed through experiments that MR has higher sensitivity, specificity and consistency in the diagnosis of each stage than CT. At the same time, it makes an algorithm based on convolutional neural network and compares the performance tests of analytic, synthetic, and deep dictionaries. A deeper lexicon is better for optimizing MR and is better for enhancing MR scanning and imaging speed. However, this article is not very proficient in the clinical diagnosis and treatment of colorectal cancer and needs further improvement. This is very helpful to increase the early diagnosis rate and improve the quality of life of patients. The inadequacy of the article is that it does not explain the situation of patients in each T stage, and how to care for patients in each T stage. The results of each stage are still not enough. I hope that we can strengthen, guide and formulate treatment plans and evaluate the prognosis in the future.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the General Hospital of the Western Theater of War, managed by the Health Service Department of the General Hospital of the Western Theater of War. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This study was supported by the Research and Development Program of The General Hospital of Western Theater Command (Grant No. 2021-XZYG-C04 and 2021-XZYG-C05).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Acuna, S. A., Elmi, M., Shah, P. S., Coburn, N. G., and Quereshy, F. A. (2017). Preoperative localization of colorectal cancer: A systematic review and meta-analysis. *Surg. Endosc.* 31 (6), 2366–2379. doi:10.1007/s00464-016-5236-8

Ali, M., Sarwar, A., Sharma, V., and Suri, J. (2019). Artificial neural network based screening of cervical cancer using a hierarchical modular neural network architecture (HMNNA) and novel benchmark uterine cervix cancer database. *Neural Comput. Applic* 31, 2979–2993. doi:10.1007/s00521-017-3246-7

Fan, Z., Wu, Z., and Li, R. (2017). The clinical application of endoscopic ultrasonography on preoperative staging of esophageal carcinoma. *J. Xinjiang Med. Univ.* 040 (005), 596–598.

Frank, J., Nina, G., Gerhard, S., Markl, B., Messmann, H., Anthuber, M., et al. (2018). Impact of primary tumor localization on the efficacy of bevacizumab in metastatic colorectal cancer. *Anticancer Res.* 38 (9), 5539–5546. doi:10.21873/anticanres.12889

Guo, Q., Xu, Z., and Cui, Y. (2019). The influence of different CEA level at colorectal cancer preoperative on prognosis and metastasis. *Pract. J. Cancer* 034 (004), 561–563.

Guo, Yu, Li, M., and Liu, X., (2018). The diagnostic value of CT-based radiomics in liver metastasis of colorectal cancer. *Chin. J. Clin. Med. Imaging* 029 (011), 798–802.

Hoshino, N., Sakamoto, T., Hida, K., and Sakai, Y. (2019). Diagnostic accuracy of computed tomography colonography for tumor depth in colorectal cancer: A systematic review and meta-analysis. *Surg. Oncol.* 30, 126–130. doi:10.1016/j.suronc.2019.08.003

Hu, K., Chen, K., He, X., Zhang, Y., Chen, Z., Li, X., et al. (2020). Automatic segmentation of intracerebral hemorrhage in CT images using encoder-decoder convolutional neural network. *Inf. Process. Manag.* 57 (6), 102352. doi:10.1016/j.ipm.2020.102352

Jaramillo, F. A., and Upegui Jiménez, Daniel (2016). MRI staging of colorectal cancer. *Rev. Colomb. De. Gastroenterol.* 31 (3), 273–282.

Joany, R. M., and Logashanmugam, E. (2018). Design and analysis of single pixel imaging on magnetic resonance imaging. *J. Med. IMAGING HEALTH Inf.* 8 (3), 596–601. doi:10.1166/jmihi.2018.2338

Lu, T., and Wang, Z. (2018). Prognostic value of preoperative fibrinogen and D-dimer levels in patients with colorectal cancer. *J. China Med. Univ.* 047 (006), 513–518.

Ma, K., Cheung, T., She, W., Chok, K. S. H., Chan, A. C. Y., Dai, W. C., et al. (2018). Diagnostic and prognostic role of 18-FDG PET/CT in the management of resectable biliary tract cancer. *World J. Surg.* 42 (3), 823–834. doi:10.1007/s00268-017-4192-3

Monroe, W. S., Skidmore, F. M., Odaibo, D. G., and Tanik, M. M. (2021). HihO: Accelerating artificial intelligence interpretability for medical imaging in IoT applications using hierarchical occlusion. *Neural Comput. Applic* 33, 6027–6038. doi:10.1007/s00521-020-05379-4

Park, S. H., Kim, H., Kim, E. K., Choi, D. K., Chung, Y. E., Kim, M. J., et al. (2017). Aberrant expression of OATP1B3 in colorectal cancer liver metastases and its clinical implication on gadoxetic acid-enhanced MRI. *Oncotarget* 8 (41), 71012–71023. doi:10.18632/oncotarget.20295

Park, S., and Lee, I. (2017). Impact of kras mutation & ercc1 over-expression on oxaliplatin-based chemotherapy in metastatic colorectal cancer patients. *Dis. Colon Rectum* 60 (6), E342.

Pradhan, N., Dhaka, V. S., Rani, G., and Chaudhary, H. (2020). Transforming view of medical images using deep learning. *Neural Comput. Applic* 32, 15043–15054. doi:10.1007/s00521-020-04857-z

Que, R. (2018). Value of CT and MRI in clinical diagnosis of preoperative local staging for rectal cancer. *Chin. Foreign Med.* 037 (017), 180–181.

Shakeel, P. M., Tolba, A., Al-Makhadmeh, Z., and Jaber, M. M. (2020). Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks. *Neural Comput. Applic* 32, 777–790. doi:10.1007/s00521-018-03972-2

Sibertinblanc, C., Fabre, A., Aparicio, T., Le Malicot, K., Bennouna, J., Ghiringhelli, F., et al. (2016). Impact of genetic polymorphisms of VEGF pathway on the response to bevacizumab in metastatic colorectal cancer (mCRC): Ancillary study of PRODIGE 9 trial. *J. Clin. Oncol.* 34 (15), 3534. doi:10.1200/jco.2016.34.15_suppl.3534

So, J. S., Cheong, C., Oh, S. Y., Lee, J. H., Kim, Y. B., and Suh, K. W. (2017). Accuracy of preoperative local staging of primary colorectal cancer by using computed tomography: Reappraisal based on data collected at a highly organized cancer center. *Ann. Coloproctology* 33 (5), 192–196. doi:10.3393/ac.2017.33.5.192

Song, X., Zhu, H., Pei, Q., Tan, F., Li, C., Zhou, Z., et al. (2017). Significance of inflammation-based indices in the prognosis of patients with non-metastatic colorectal cancer. *Oncotarget* 8 (28), 45178–45189. doi:10.18632/oncotarget.16774

Sun, Z., Liu, J., Jing, H., Dong, S. X., and Wu, J. (2017). The diagnostic and prognostic value of CHFR hypermethylation in colorectal cancer, a meta-analysis and literature review. *Oncotarget* 8 (51), 89142–89148. doi:10.18632/oncotarget.19408

Xiao, F., and Ding, W. (2019). Divergence measure of pythagorean fuzzy sets and its application in medical diagnosis. *Appl. Soft Comput.* 79, 254–267.

Xu, J., Wang, J., and Lu, Z. (2017). The diagnostic value of proactive of serum CRP, CA19-9 and CEA in patients with colorectal cancer. *J. Ningxia Med. Univ.* 039 (006), 641–644.

Xu, Z., Kamruzzaman, M. M., and Shi, J. (2022). Method of generating face image based on text description of generating adversarial network. *J. Electron. Imaging* 31 (5), 051411. doi:10.1117/1.jei.31.5.051411

Yang, K., Zhou, B., Yi, F., Chen, Y., and Cheng, Y. (2019). Correction to: Colorectal cancer diagnostic algorithm based on sub-patch weight color histogram in combination of improved least squares support vector machine for pathological image. *J. Med. Syst.* 43 (12), 333. doi:10.1007/s10916-019-1449-4

Yao, L., Zhao, L., and Yan, G. (2016). Value of magnetic resonance imaging(MRI) in local staging of colorectal cancer before surgery. *China Mod. Dr.* 054 (020), 92–94.

Zhu, Y., Hu, B., Xu, L., Yang, L., Wang, C., Zhang, S., et al. (2020). Tumor-specific nascent nine-peptide-epitopes prediction and bioinformatics characterization in human colorectal cancer. *J. Med. IMAGING HEALTH Inf.* 10 (6), 1338–1345.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 17 February 2023
DOI 10.3389/fgene.2023.1119181

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Jun Zheng,
Baotou Teachers' College, China
Xiaoxi Liu,
Xi'an Jiaotong University, China

*CORRESPONDENCE
Zhengfang Chen,
czf4419@suda.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 08 December 2022
ACCEPTED 31 January 2023
PUBLISHED 17 February 2023

CITATION
Song G, Zhang Y, Jiang Y, Zhang H, Gu W, Xu X, Yao J and Chen Z (2023), Signaling pathway of targeting the pancreas in the treatment of diabetes under the precision medicine big data evaluation system. *Front. Genet.* 14:1119181. doi: 10.3389/fgene.2023.1119181

Signaling pathway of targeting the pancreas in the treatment of diabetes under the precision medicine big data evaluation system

Ge Song[1,2], Yiqian Zhang[1], Yihua Jiang[1], Huan Zhang[1], Wen Gu[1], Xiu Xu[1], Jing Yao[1] and Zhengfang Chen[1]*

[1]Department of Endocrinology, Soochow University Affiliated Changshu First People's Hospital, Changshu, Jiangsu, China, [2]Department of Endocrinology, Changshu Hospital Affiliated to Soochow University, Changshu No.1 People's Hospital, Medical College of Soochow University, Changshu, Jiangsu, China

Diabetes is a chronic noncommunicable disease, which is related to lifestyle, environmental and other factors. The main disease of diabetes is the pancreas. Inflammation, oxidative stress and other factors can interfere with the conduction of various cell signaling pathways, thus inducing pancreatic tissue lesions and diabetes. Precision medicine covers epidemiology, preventive medicine, rehabilitation medicine and clinical medicine. On the basis of precision medicine big data analysis, this paper takes pancreas as the target to analyze the signal pathway of diabetes treatment. This paper analyzes from the five aspects of the age structure of diabetes, the blood sugar control standard of type 2 elderly diabetes mellitus, the changes in the number of diabetic patients, the ratio of patients using pancreatic species and the changes in blood sugar using the pancreas. The results of the study showed that targeted pancreatic therapy for diabetes reduced the diabetic blood glucose rate by approximately 6.94%.

KEYWORDS

diabetes disease, pancreatic target, precision medicine big data, signal path, targeting the pancreas

1 Introduction

The organic combination of precision medicine and big data technology is conducive to the analysis and sharing of medical data, so as to better meet the needs of medical development and lay a solid foundation for the development of related medical research. Diabetes is an endocrine inflammatory disease. At present, the treatment presents the characteristics of multiple pathways and multiple targets. The pancreas is an important target for diabetes treatment. The application of precision medicine big data to the research on the treatment of diabetes with the pancreas as the target is conducive to promoting the progress of diabetes medicine.

The incidence of diabetes is getting higher and higher, and many scholars have studied it. Bensellam M presented the identified molecular mechanisms involved in the dedifferentiation of cells in the diabetic pancreas that produce and release the hormone insulin. The roles and inhibitors of differentiation proteins were discussed and the emerging role of non-coding RNAs (Ribonucleic Acid) was highlighted (Bensellam et al., 2018).

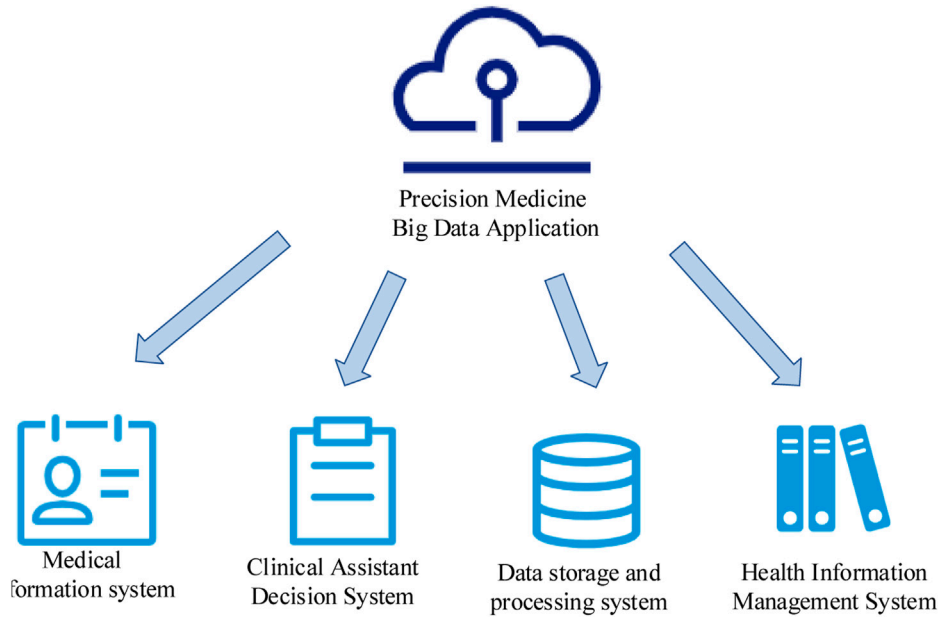


FIGURE 1
Big data applications in precision medicine.

Tuttolomondo A studied several factors that affect overall cerebrovascular risk to varying degrees in people with diabetes. Diabetes may lead to more insidious brain damage represented by lacunar infarcts, which would increase the risk of dementia and lead to a dramatic decline in cognitive function (Tuttolomondo, 2018). Hu C believed that the decreased ability of insulin to promote the processing and storage of glucose in muscle is due to impaired activation of glycogen synthase. Decreased glucose storage may occur due to decreased glucose uptake, and safety has been cognitively impaired secondary (Hu and Jia, 2018). In the Cardiovascular Health Study, Cho N H assessed the relationship between patients with subclinical cardiovascular disease, patients with diabetes and impaired glucose tolerance and normal subjects and the risk of clinical vascular disease (Cho et al., 2018). Liu believed the prevalence of diabetic lesions is high, and these complications are often associated with poor medication adherence and uncontrolled diabetes. The aim of the study was to determine medication adherence in patients with uncontrolled diabetes and to compare characteristics and identified barriers between patients with good and poor adherence to medication (Liu et al., 2018). Lane W mentioned that people with diabetes are prone to foot ulcers. If these ulcers do not heal, the patient may also undergo foot amputation, with diabetes preventing postoperative wound healing (Lane et al., 2017). Feig D S investigated whether microalbuminuria predicts later development of increased proteinuria and early mortality in patients with type 2 diabetes. Morning urine samples from diabetes clinic patients aged 50–75 years were examined by radioimmunoassay (Feig et al., 2018). Although there are many studies on the theory of diabetes, further research is needed on the treatment of diabetes.

The pancreas-targeted treatment of diabetes is widely used in medicine. Research by Marathe P H demonstrated that permanent neonatal diabetes may be caused by a complete lack of glucokinase activity. It reported three new cases of glucokinase-related permanent neonatal diabetes. Autosomal recessive inheritance and enzyme deficiencies are typical features of inborn errors of metabolism that occur in the glucose-insulin signaling pathway in these subjects (Marathe et al., 2017). Ogurtsova K explored the etiology of diabetes-related cognitive decline. The etiology involves insulin receptor downregulation, neuronal apoptosis, and glutamatergic neurotransmission (Ogurtsova et al., 2017). Bragg F studied the role of signal transduction and transcriptional activator protein signaling pathway in autoimmune diabetes (Bragg et al., 2017). Rowan J A provided a new strategy for diabetes prevention and treatment by studying the vascular endothelial cell nuclear factor signaling pathway in diabetes (Rowan et al., 2018). Wang Q mentioned that diabetes impairs the mobilization of hematopoietic stem cells from the bone marrow, thereby worsening the outcome of hematopoietic stem and progenitor cell transplantation and diabetic complications (Wang et al., 2017). Willeit explored the potential effects of curcumin on cardiomyocyte hypertrophy, possible mechanisms of nuclear transcription factor signaling in diabetes, hyperglycemia- and insulin-induced cardiomyocyte hypertrophy, and antihypertrophic effects of curcumin in primary culture (Willeit et al., 2017). Chamberlain JJ mentioned that insulin acutely controls metabolism in adipocytes, but also nuclear transcription *via* proline-directed serine/threonine kinase-mediated “mitotic” signaling (Chamberlain et al., 2017). Although pancreas-targeted therapy for diabetes is widely used in medicine, there are still problems in its application.

By combining computer technology with medicine, new knowledge can be discovered and new modalities of diagnosis and treatment can be created. There have been many research reports on diabetes signaling pathways. This article took pancreatic islets as the target, and systematically discussed the related signaling pathways. At the same time, through the use of

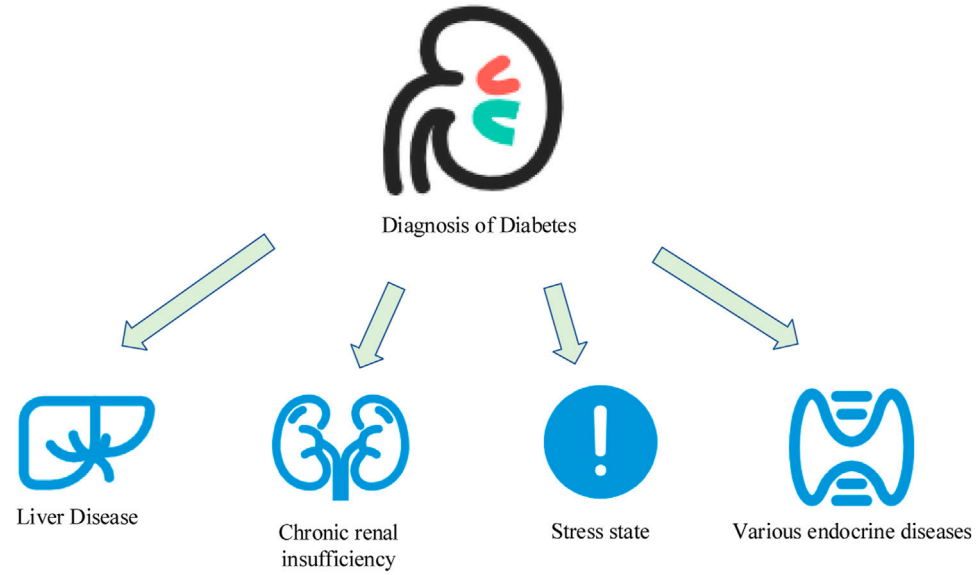


FIGURE 2
Diagnosis of diabetes.

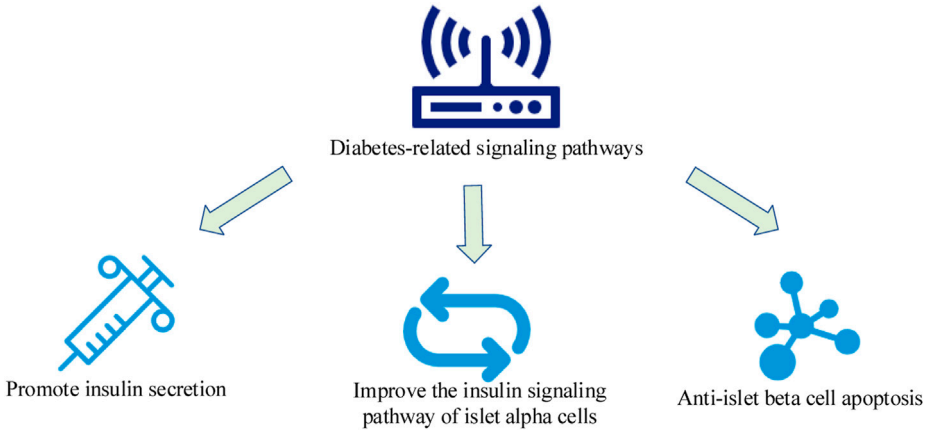


FIGURE 3
Pancreatic tissue as a target for the treatment of diabetes-related signaling pathways.

precision medicine big data analysis, some complex information was processed, so as to realize the diagnosis and prediction of diabetes.

2 Concepts related to precision medicine big data, diabetes and pancreas

2.1 Precision medicine big data

Precision medicine is an upgraded version of personalized medicine. According to the biological characteristics of patients, especially the data of genomics, and through modern technology, it provides accurate prevention, diagnosis and treatment for the clinic. It can not only improve the treatment effect, but also prevent inappropriate treatment, excessive treatment, waste of resources and other phenomena, and save medical expenses (Riddell et al., 2017). The implementation of precision medicine is based on the development of gene sequence detection technology, and on the basis of network and big data technology. This technology is able to treat patients with minimal medical damage and minimal medical costs, and it not only restores their physical functions, but also ensures their mental health. The big data application of precision medicine is shown in Figure 1.

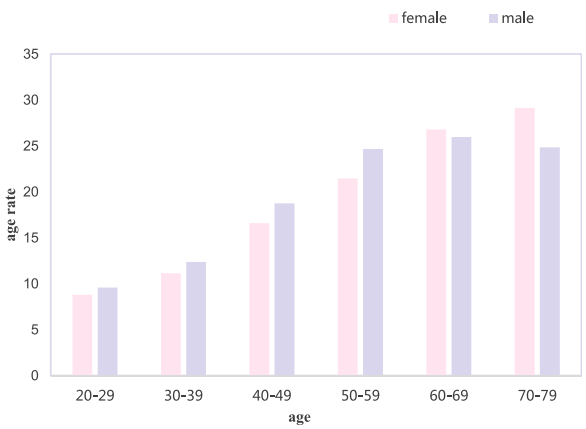


FIGURE 4
Age structure of diabetes.

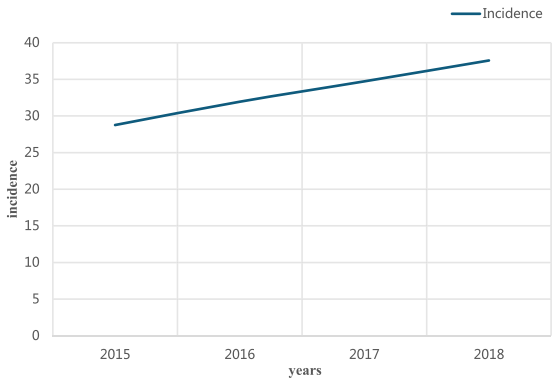


FIGURE 5
Number of people with diabetes.

2.1.1 Medical information system

In fact, medical information systems did not appear in recent years, but because of the complexity of medical services and the small coverage of medical information systems, they have not been widely used. There is a large amount of different medical data in different institutions, systems or databases. The challenge is how to integrate these data in an orderly manner without violating existing laws, regulations, and ethics; and how to conduct automatic computer analysis and mining without violating existing laws, regulations, and ethics. The construction and application of a high-quality medical information system is still an important prerequisite for the production of big data, and it is also an area where companies in the industry invest the most.

2.1.2 Clinical auxiliary decision-making system

Big data is not to replace experienced doctors, but to allow more doctors to make high-quality judgments before they have enough experience, which is also an important role of precision medicine. This technology can help doctors diagnose and treat patients' illnesses by analyzing various medical data such as patients' medical records, test results, medical images, etc.

2.1.3 Data storage and processing system

The development of the big data industry is inseparable from the support of infrastructure, and many companies have developed their own storage and processing platforms. However, the mainstream applications of software services are still on basic communication software such as cloud mailboxes and communication platforms. More software services in the medical field are also needed to improve the research and develop efficiency of pharmaceutical companies and strengthen the management of patients and clinical data by physicians.

2.1.4 Health information management system

At present, many health management systems have emerged in mobile health management, such as diabetes management, sleep quality management, and intestinal health management. In fact, they all rely on a large number of data processing and high-quality sensors to realize real-time monitoring of data, but many systems have no way to intervene in time when dealing with exceptions. However, big data is like a chicken-and-egg process. In the face of massive data, the correlation would gradually become prominent, and corresponding auxiliary decision-making would emerge. Big data technology can provide a health information management platform with information collection, analysis and processing functions by collecting various medical data files of patients in medical institutions and sharing various medical data through network technology.

2.2 Diabetes

Diabetes is a metabolic disease characterized by hyperglycemia. Hyperglycemia is caused by insufficient secretion of insulin and its biological function is impaired (Johal et al., 2017). Diagnosing diabetes is generally not difficult. The diagnosis can be made if the fasting blood glucose is greater than or equal to 7.0 mmol/L, or greater than or equal to 11.1 mmol/L within 2 h after a meal. The diagnosis is divided into type 1 and type 2 diabetes. The differential diagnosis of diabetes is shown in Figure 2.

2.2.1 Liver disease

Patients with liver cirrhosis usually have abnormal glucose metabolism, which is usually fasting or hypoglycemia, and the blood sugar would rise rapidly after meals. Patients with prolonged illness, fasting blood sugar would also increase.

2.2.2 Chronic renal insufficiency

There would be a slight abnormal glucose metabolism. Patients need to use some medications under the guidance of medical professionals to protect the kidney and repair the glucose metabolism function.

2.2.3 Stressed state

Under stressful conditions, such as heart, cerebrovascular accident, acute infection, trauma, etc., blood sugar would be excessively increased, and it would return to normal within 1–2 weeks after the stress factor disappears (Sorli et al., 2017).

2.2.4 Various endocrine diseases

Glucagonoma is a secondary cause of diabetes mellitus and has other characteristics besides hyperglycemia.

2.3 Concepts related to pancreas

In the upper abdomen of the human body, there is a small organ that is difficult to find, and that is the pancreas. Although the pancreas is small, it is very powerful and is one of the most important parts of the human body. The pancreas is a gland with endocrine and endocrine functions, and its physiological function and pathological changes are closely related to human health.

The pancreas is located retroperitoneally. In terms of secretion, although the pancreas is small, it contains many endocrine cells. These cells are also adjusting the physiological functions of the body during the process of digestion and absorption. If these cells change and secrete too much or too little, disease can result (Mario et al., 2017). The pancreas produces various digestive enzymes and insulin in the body to help the body break down proteins and other substances and lower blood sugar. Pancreatic juice is an exocrine substance that mainly includes alkaline bicarbonate and various digestive enzymes.

3 Pancreatic tissue as a target to treat diabetes signaling pathway

Diabetes is an endocrine inflammatory disease, and its clinical manifestations are multi-channel and multi-target regulation. The pancreas is the main lesion of diabetes, and it is also an important target for the treatment of diabetes. Its main functions are to promote the synthesis of hepatic glycogen and myoglycogen, the absorption and activation of glucose, and the inhibition of sugar xenobiogenesis. Figure 3 shows the relevant signaling pathways targeting pancreatic tissue.

3.1 Promotion of insulin secretion

3.1.1 Calcium atom channels and ATP-sensitive potassium channels

The cells in the pancreas that produce and release the hormone insulin are endocrine cells with electrical excitation. Its secretion is mainly due to an increase in intracellular calcium concentration, which allows it to produce insulin. L-type calcium channels are the main channel for glucose-induced insulin. Its variation may cause type 2 diabetes, and the state of energy metabolism in the body also affects the electrical activity of sensitive potassium channels. Under physiological conditions, glucose enters the pancreas to produce energy from cellular metabolism that produces and releases the hormone insulin. Adenosine triphosphate closes sensitive potassium channels in the cell membrane, opening calcium channels in the cell, thereby activating cells in the pancreas that produce and release the hormone insulin to secrete insulin (Kiran et al., 2017). Therefore, calcium channels and sensitive potassium channels may be effective drug targets for type 2 diabetes.

3.1.2 β-cell GLP-1 receptor signaling pathway

Glucagon-like peptide-1 (GLP-1) is a glucose concentration-based polypeptide hormone. It can regulate the gene expression, synthesis and secretion of cells in the pancreas that produce and release the hormone insulin, and can promote insulin secretion and inhibit insulin cell apoptosis. GLP-1 receptors are mainly distributed in the pancreas. After GLP-1 binds to its receptor, it can activate adenylate cyclase, thereby increasing intracellular cyclic adenylate and activating downstream protein kinases. GLP-1 activates the calcium atom signaling pathway through the acidity coefficient and cyclic adenylate binding protein pathway to promote the release of insulin.

3.2 Improvement of insulin signaling pathway in pancreatic islet cells by reducing glucagon secretion

Glucagon is a polypeptide hormone produced by pancreatic islet cells, and its abnormal secretion and metabolism may be related to the pathogenesis of type 2 diabetes. Under physiological conditions, cells in the pancreas that produce and release the hormone insulin block alpha cell glucagon production by paracrine (Arnold et al., 2018). The insulin receptor substrate, phosphatidylinositol kinase, is also expressed on cells, and insulin inhibits the gene expression of glucagon in islet cells through signal channels. In conclusion, the improvement of insulin resistance and the reduction of glucagon production is a new therapeutic approach.

3.3 Anti-islet beta cell apoptosis

The MAPK (mitogen-activated protein kinase signaling) pathway has a role in regulating cell growth, differentiation and apoptosis in mammalian cells. Different MAPK signaling pathways can be utilized by different extracellular stimuli, and their regulation can modulate various cellular responses. Stress-activated protein kinases and cell signaling cascades have important roles in cellular stress and cellular inflammatory factors. The results show that streptosporin can inhibit the occurrence of type 1 diabetes and

TABLE 1 Glycemic control criteria for type 2 diabetes in the elderly.

Health status	Reasonable saccharification (%)	Fasting or preprandial blood glucose (ol/L)	Blood sugar before bed (l/L)
healthy	<7.5	5.0–7.2	5.0–8.3
moderate health	<8.0	5.0–8.3	5.6–10.0
poor health	<8.5	5.6–10.0	6.1–11.1

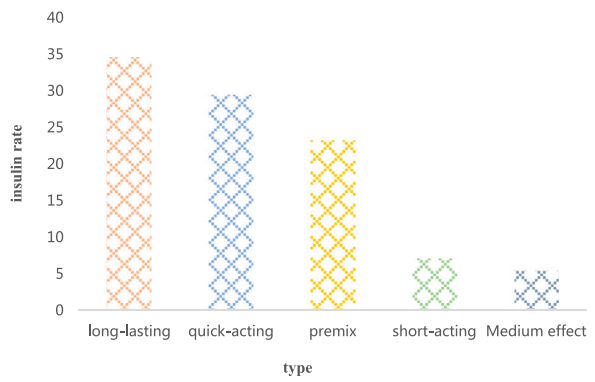


FIGURE 6
The proportion of patients using insulin type.

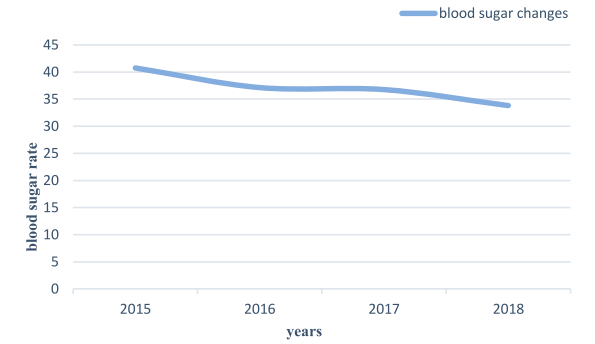


FIGURE 7
Blood glucose changes using the pancreas.

reduce the phosphorylation level of its kinase, thereby reducing its upstream kinase activity.

4 Algorithm of diabetes under the precision medicine big data evaluation system

4.1 Diabetes risk input expression

The diabetes risk input expression is a model that combines the user's risk index data (including past and present data). Since a patient has multiple medical records, multiple medical records appear in the same user's medical records. Therefore, it is necessary to comprehensively consider the weight of each medical record, that is, the importance of the sub-medical record.

Definition 1: collection of risk indicators

It is assumed that user G_i has lb sub-records (with time stamps), then all sub-records of Q_i can form a set of risk indicators, and the formula is:

$$Q_i = \{\langle k_{ab}, f_{ab}\rangle\}_{b\in\{1,2,\ldots,lb\}} \tag{1}$$

Among them, k_{ab} is the time, and f_{ab} is the feature vector of each index component; the entire data set can be represented as $Q = \{Q_i\}_{a\in\{1,2,\ldots,n\}}$, and n is the number of patients.

For the b-th record k_{ab}, f_{ab} in Q_i, its decay weight formula is:

$$Ph(b) = e^{-\infty b} \Big/ \sum_{k=0}^{l-1} e^{-\infty k} \tag{2}$$

Among them, ∞ is the adjustment weight range, and l is the time span.

Definition 2: risk input expression

The weighted average of each inspection record of 1 G_i based on the decay weight can be extracted as an input expression, which is the risk input expression:

$$x_a = \sum_{b=1}^{l_a} Ph(l_a - b) f_{ab} \tag{3}$$

It can be obtained that the risk index set $\{\langle k_{ab}, f_{ab}\rangle\}_{b\in\{1,2,\ldots,la\}}$ of user G_i can be normalized in the form of a matrix into the form of vector x_a, which can be used as the standard input of the calculation model.

The model is based on the support vector machine algorithm. Parameter r, g represent function $f(x)$, and its calculation method:

$$f(x) = rt + g \tag{4}$$

In the hyperplane, all the diabetes sample spaces are divided into two groups: one with diabetes and one without diabetes. Through mathematical transformation, it is the solution target:

$$\min_{r,g,\zeta} = \frac{1}{2} r^T r + C \sum_{k=1}^{m} \zeta^k \tag{5}$$

$$y_a \left(r^T \phi(x_a) + g\right) \geq 1, a = 1, \ldots, m \tag{6}$$

m is the number of training samples, and y_a is the class label.

Definition 3: diabetes index

The Diabetes Risk Index (DI) represents the risk of developing diabetes and has a value in the range [0,1]. The corresponding DI value can be obtained by probabilistic correction of the value of the judgment function. Probabilistic calibration is performed:

$$R(y = 1 \mid x) = 1/(1 + \exp(Sd(x) + F)) \tag{7}$$

Among them, S, F are conversion parameters.

DI is calculated as:

$$DI(x) = R(y = 1 \mid x) \tag{8}$$

4.2 Diabetes risk model based on genetic factors

The inheritance of diabetes is determined by the relatedness of family members, so a suitable indicator should be used to measure the genetic association of two people.

4.2.1 Genetic coefficient

In kinship, the blood of the child is inherited from the parents, so the genetic factor between the parent and the child is 1/2. Since the genetic factors represented by each side are all 1/2, the calculation formula of the genetic factors of the direct genetic relationship can be obtained:

$$k(X,Y) = (1/2)^M \tag{9}$$

Through the calculation of the direct genetic relationship, the paragenetic genetic factors can be obtained. First, it is necessary to find out the most recent common ancestors N_1 and N_2 of members X and Y, and calculate the genetic coefficients of X, Y and each ancestor separately. After that, the sum is done to get the heritability coefficients of X and Y:

$$k(X,Y) = \sum_{i=1}^{2} k(X,N)k(Y,N) \tag{10}$$

In Formula (10), n is the total number of diabetes-positive family members of user R.

4.2.2 Dynamic blood glucose expression

Ambulatory blood glucose sequence $g = (g_1, g_2, \ldots g_n)$ represents a time series of blood glucose values. The data $y(h)$ at time h can be predicted from the blood glucose value at time $h-1, h-2, \ldots h-n$, and there is a complex non-linear relationship $f(\cdot)$ between them, which is expressed by Formula (12):

$$y(h) = f\left(y(h-1), y(h-2)\right), \ldots, y(h-n) \tag{12}$$

It is assumed that the number of visible units and hidden units are m and n, respectively, then $z = (z_1, z_2, \ldots z_m)$, and $x = (x_1, x_2, \ldots x_n)$. b represents the bias of the visible layer, and c represents the bias of the hidden layer. The energy equation of the model is:

$$W(z, x \mid \theta) = -bz^T - cx^T - vwx^T \tag{13}$$

It is expressed in components as:

$$W(z, x \mid \theta) = -\sum_{i=1}^{m} b_i z_i - \sum_{i=1}^{n} c_j x_j - \sum_{j=1}^{n}\sum_{i=1}^{m} z_i W_{ij} x_j \tag{14}$$

The formula for calculating the joint probability distribution of the visible layer and the hidden layer is:

$$T(z, x \mid \theta) = \frac{1}{B(\theta)} e^{-w(z,x \mid \theta)} \tag{15}$$

$$B(\theta) = \sum_{z,x} e^{-w(z,x \mid \theta)} \tag{16}$$

The marginal probability distribution of ambulatory blood glucose sequence z to x is:

$$T(z \mid \theta) = \frac{1}{B(\theta)} \sum x e^{-w(z,x \mid \theta)} \tag{17}$$

When the visible layer unit state is given, the activation probability of the x_i-th hidden layer unit z_i is solved as:

$$T(x_i, \mid z_i, \theta) = sigm\left(b_j + \sum_{i=1}^{m} x_i W_{ij}\right) \tag{18}$$

Since cells are bidirectional, visible cells can be activated through hidden cells. The calculation method of the activation probability of the i-th visible unit z_i is:

$$T(z_i, \mid x, \theta) = sigm\left(c_j + \sum_{i=1}^{m} W_{ij} x_i\right) \tag{19}$$

By experimenting with predictive models and evaluating them predictively, the formula for calculating the root mean square error is:

$$RMSE = \sqrt{\frac{\sum_{a=1}^{n} (x_a - \hat{x}_a)^2}{n}} \tag{20}$$

In Formula (20), x_a is the predicted value of the a-th model; $\hat{x}_a$ is the actual value of the a-th model, and n is the capacity of the sample set.

5 Experimental exploration of the pancreas as the target for the treatment of diabetes under the precision medicine big data evaluation system

5.1 Experimental method

Under the precision medicine big data analysis system, this paper studied the signaling pathway of treating diabetes by targeting the pancreas. City A was selected as the research object, and the research was carried out from five aspects: the age structure of diabetes mellitus in 2015–2018, the blood sugar control standard of type 2 diabetes in the elderly, the change in the number of diabetic patients, the proportion of patients using pancreatic types, and the proportion of patients using pancreatic types.

5.2 Experimental analysis

5.2.1 Age structure of diabetes

The probability of disease in each age stage is different, and the age structure of diabetic patients in city A was investigated. The results are shown in Figure 4.

In the age structure of diabetes shown in Figure 4, the higher the female age, the greater the probability of developing diabetes. Men were more likely to develop diabetes at the age of 60–69, and there were more patients with diabetes over the age of 50. It can be seen that age is closely related to the occurrence of diabetes.

5.2.2 The blood sugar control standard of type 2 elderly diabetes mellitus

Older people often experience symptoms of hypoglycemia. In addition, elderly diabetic patients are prone to arteriosclerosis and cardiovascular disease. Hypoglycemia can lead to complications such as cerebral infarction and myocardial infarction. The blood sugar control standards for type 2 elderly diabetes mellitus are shown in Table 1.

For elderly diabetic patients, the efficacy and risk should be comprehensively considered, and the main purpose is to improve

their quality of life. In addition, personalized blood glucose control indicators should be developed to improve the health of the elderly.

5.2.3 Changes in the number of diabetic patients

Changes in the number of diabetic patients as the research subject, the number of diabetic patients in city A was investigated from 2015 to 2018. The specific results are shown in Figure 5.

Among the changes in the number of diabetic patients shown in Figure 5, the ratio of the number of diabetic patients showed an upward trend from 2015 to 2018. Among them, the proportion of diabetic patients in 2015 was about 28.76%; in 2018, the proportion of diabetic patients was about 37.57%, with an increase of 8.81%. It can be seen that in life, people should maintain a good diet and routine, and exercise regularly. With regular checkups, people can reduce the chances of developing diabetes.

5.2.4 Ratio of patients using pancreatic species

The main body of the investigation was the proportion of patients using pancreatic types in area A, and studies were carried out from long-acting, quick-acting, premixed, short-acting, and intermediate-acting. The results are shown in Figure 6.

In the proportion of patients using pancreatic types shown in Figure 6, more patients used pancreatic long-acting, fast-acting and premixed, and about 34.58% used long-acting patients; there was about 29.42% of the patients used quick-acting, and 23.25% of the patients used premix. The number of patients using pancreatic short-acting and intermediate-acting was relatively small; the number of patients using short-acting was about 7.21%, and the number of patients using intermediate-acting was about 5.54%.

5.2.5 Changes in blood sugar using the pancreas

The pancreas affects glucose metabolism in the human body and plays an important role in diabetes. Figure 7 shows the results of the research content of the pancreas-targeted treatment of diabetic blood sugar changes.

In the pancreas-targeted treatment of diabetes shown in Figure 7, blood glucose rates continued to decline from 2015 to 2018. In 2015, the blood sugar rate of pancreas-targeted treatment of diabetes was about 40.75%; in 2018, the blood sugar rate of pancreatic-targeted treatment of diabetes was about 33.81%, with a decrease of 6.94%. It can be seen that the pancreas would secrete insulin, which can regulate the level of glucose in the blood and brings blood sugar back to normal.

6 Conclusion

Through the analysis and sharing of precision medicine big data, it would help the development of current medical technology and improve the level of medical treatment. This paper systematically discussed the diabetes signaling pathway targeting the pancreas, which has certain reference value for guiding the development of new drugs and the use of a variety of clinical drugs. The pancreas is mainly used to treat diabetes, and its main function is to stimulate pancreatic secretion and inhibit the release of glucagon. Apoptosis of the pancreas is an important factor affecting the amount of insulin secreted, and apoptosis of the pancreas is related to many signaling pathways. Therefore, the signal pathway related to diabetes explored from the perspective of the pancreas has a certain guiding effect on the development and clinical treatment of new drugs for diabetes.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Arnold, S. E., Arvanitakis, Z., Macauley-Rambach, S. L., Koenig, A. M., Wang, H. Y., Ahima, R. S., et al. (2018). Brain insulin resistance in type 2 diabetes and Alzheimer disease: Concepts and conundrums. *Nat. Rev. Neurol.* 14 (3), 168–181. doi:10.1038/nrneurol.2017.185

Bensellam, M., Jonas, J. C., and Laybutt, D. R. (2018). Mechanisms of β-cell dedifferentiation in diabetes: Recent findings and future research directions. *J. Endocrinol.* 236 (2), R109–R517. doi:10.1530/JOE-17-0516

Bragg, F., Holmes, M. V., Iona, A., Guo, Y., Du, H., Chen, Y., et al. (2017). Association between diabetes and cause-specific mortality in rural and urban areas of China. *JAMA J. Am. Med. Assoc.* 317 (3), 280–289. doi:10.1001/jama.2016.19720

Chamberlain, J. J., Kalyani, R. R., Leal, S., Rhinehart, A. S., Shubrook, J. H., Skolnik, N., et al. (2017). Treatment of type 1 diabetes: Synopsis of the 2017 American diabetes association standards of medical care in diabetes. *Ann. Intern. Med.* 167 (7), 493–498. doi:10.7326/M17-1259

Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., et al. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res. Clin. Pract.* 138 (4), 271–281. doi:10.1016/j.diabres.2018.02.023

Feig, D. S., Donovan, L. E., Corcoy, R., Murphy, K. E., Amiel, S. A., Hunt, K. F., et al. (2018). Continuous glucose monitoring in pregnant women with type 1 diabetes (conceptt): A multicenter international randomised controlled trial. *Obstetrical Gynecol. Surv.* 73 (4), 199–201. doi:10.1097/01.ogx.0000532199.80944.24

Hu, C., and Jia, W. (2018). Diabetes in China: Epidemiology and genetic risk factors and their clinical utility in personalized medication. *Diabetes* 67 (1), 3–11. doi:10.2337/dbi17-0013

Johal, S. K., Jamsen, M. J., Bell, S., Mc Namara, K. P., Magliano, D. J., Liew, D., et al. (2017). Do statin users adhere to a healthy diet and lifestyle? The Australian diabetes, obesity and lifestyle study. *Eur. J. Prev. Cardiol.* 24 (6), 621–627. doi:10.1177/2047487316684054

Kiran, B., Mohanalakshmi, T., and Srikumar, R. (2017). C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in diabetes. *Int. J. Res. Pharm. Sci.* 8 (3), 476–479.

Lane, W. T., Bailey, S., Gerety, G., Gumprecht, J., Philis-Tsimikas, A., Hansen, C. T., et al. (2017). Effect of insulin degludec vs insulin glargine U100 on hypoglycemia in patients with type 1 diabetes: The SWITCH 1 randomized clinical trial. *Jama* 318 (1), 33–44. doi:10.1001/jama.2017.7115

Liu, Jo, Dumville, J. C., Hinchliffe, R. J., Cullum, N., Game, F., Stubbs, N., et al. (2018). Negative pressure wound therapy for treating foot wounds in people with diabetes mellitus. *Cochrane Database Syst. Rev.* 73 (18), 26–33. doi:10.1002/14651858.cd010318.pub3

Marathe, P. H., Gao, H. X., and Close, K. L. (2017). American diabetes association standards of medical care in diabetes 2017. *J. Diabetes* 9 (4), 320–324. doi:10.1111/1753-0407.12524

Mario, E. J., Richards, L. KS, Mcleod, H., Stanley, D., Yap, Y. A., Knight, J., et al. (2017). Gut microbial metabolites limit the frequency of autoimmune T cells and protect against type 1 diabetes. *Nat. Immunol.* 18 (5), 552–562. doi:10.1038/ni.3713

Ogurtsova, K. J., Fernandes, R., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N. H., et al. (2017). IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res. Clin. Pract.* 128 (6), 40–50. doi:10.1016/j.diabres.2017.03.024

Riddell, M. C., Gallen, I. W., Smart, C. E., Taplin, C. E., Adolfsson, P., Lumb, A. N., et al. (2017). Exercise management in type 1 diabetes: A consensus statement. *Lancet Diabetes & Endocrinol.* 5 (5), 377–390. doi:10.1016/S2213-8587(17)30014-1

Rowan, J. A., Rush, E. C., Plank, L. D., Lu, J., Obolonkin, V., Coat, S., et al. (2018). Metformin in gestational diabetes: The offspring follow-up (MiG TOFU) body composition and metabolic outcomes at 7–9 Years of age. *Obstetrical Gynecol. Surv.* 73 (10), 565–567. doi:10.1097/01.ogx.0000547170.99000.49

Sorli, C. S., Harashima, I. G., Tsoukas, M., Unger, J., Karsbol, J. D., Hansen, T., et al. (2017). Efficacy and safety of once-weekly semaglutide monotherapy versus placebo in patients with type 2 diabetes (SUSTAIN 1): A double-blind, randomised, placebo-controlled, parallel-group, multinational, multicentre phase 3a trial. *Lancet Diabetes & Endocrinol.* 5 (4), 251–260. doi:10.1016/S2213-8587(17)30013-X

Tuttolomondo, A. (2018). Relationship between diabetes and ischemic stroke: Analysis of diabetes-related risk factors for stroke and of specific patterns of stroke associated with diabetes mellitus. *J. Diabetes & Metabolism* 06 (5), 73–84. doi:10.4172/2155-6156.1000544

Wang, Q., Miao, Z., Torres, G., Wu, S., Ouyang, C., Xie, Z., et al. (2017). Metformin suppresses diabetes-accelerated atherosclerosis via the inhibition of drp1-mediated mitochondrial fission. *Diabetes* 66 (1), 193–205. doi:10.2337/db16-0915

Willeit, P, Skroblin, A., Moschen, A. R., Yin, X., Kaudewitz, D., Zampetaki, A., et al. (2017). Circulating MicroRNA-122 is associated with the risk of new-onset metabolic syndrome and type 2 diabetes. *Diabetes* 66 (2), 347–357. doi:10.2337/db16-0731

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 21 February 2023
DOI 10.3389/fgene.2023.1142795

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Dazhong Shu,
Sanya University, China
Cheng Tong,
Saigon University, Vietnam

*CORRESPONDENCE
Youshuang Cheng,
chengyoushuang@163.com
Ke Zhang,
zhangke96025@163.com

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 12 January 2023
ACCEPTED 06 February 2023
PUBLISHED 21 February 2023

CITATION
Yin Y, Lu J, Tong J, Cheng Y and Zhang K (2023), Relationship between early lung adenocarcinoma and multiple driving genes based on artificial intelligence medical images of pulmonary nodules. *Front. Genet.* 14:1142795. doi: 10.3389/fgene.2023.1142795

Relationship between early lung adenocarcinoma and multiple driving genes based on artificial intelligence medical images of pulmonary nodules

Yajun Yin[1], Jiawei Lu[1], Jichun Tong[1], Youshuang Cheng[2]* and Ke Zhang[1]*

[1]Department of Cardiothoracic Surgery, Changzhou Second People's Hospital, The Affiliated Hospital of Nanjing Medical University, Changzhou, Jiangsu Province, China, [2]Department of Cardiothoracic Surgery, Affiliated Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

Lung adenocarcinoma is one of the most common cancers in the world, and accurate diagnosis of lung nodules is an important factor in reducing its mortality. In the diagnosis of pulmonary nodules, artificial intelligence (AI) assisted diagnosis technology has been rapidly developed, so testing its effectiveness is conducive to promoting its important role in clinical practice. This paper introduces the background of early lung adenocarcinoma and lung nodule AI medical imaging, and then makes academic research on early lung adenocarcinoma and AI medical imaging, and finally summarizes the biological information. In the experimental part, the relationship analysis of 4 driver genes in group X and group Y showed that there were more abnormal invasive lung adenocarcinoma genes, and the maximum uptake value and uptake function of metabolic value were also higher. However, there was no significant correlation between mutations in the four driver genes and metabolic values, and the average accuracy of AI-based medical images was 3.88% higher than that of traditional images.

KEYWORDS

early lung adenocarcinoma, lung nodules, artificial intelligence, biological information discipline, medical, driver genes

1 Introduction

With the development of the staging of early lung adenocarcinoma, the detection rate of gene abnormalities increases, which plays an important role in the early occurrence of nodular lung adenocarcinoma. In recent years, with people's understanding of health and the continuous progress of medical imaging technology, the detection rate of pulmonary nodules is increasing every year. The previous research on the treatment, efficacy and prognosis of lung cancer is of great significance. At the same time, using a variety of biological information technology, it is found that the related genes related to the pathogenesis and prognosis of lung adenocarcinoma, which laid a foundation for the study of the molecular mechanism of lung adenocarcinoma.

Many scholars have studied the relationship between lung adenocarcinoma and multiple driving genes. Aisner Dara L believed that polygene analysis was a conventional treatment for patients with advanced lung adenocarcinoma. Lung cancer

variant tissue is a multi-sectoral study aimed at finding and treating cancer predisposing factors of lung cancer patients (Aisner et al., 2018). Chuang Chen-Hua, through the combination of human lung adenocarcinoma with mouse tumor barcode and unbiased gene analysis technology, found a transcription process that could enable patients to obtain metastasis and predict patient survival (Chuang et al., 2017). Chen Jianbin obtained that the genome of East Asian LUAD (lung adenocarcinoma) was more stable than that of European LUAD through research. Its characteristic was that the variation was small and the number of copies changes little (Chen et al., 2020). Gillette Michael A believed that proteome data set provided a unique shared resource for researchers and clinicians to find better understanding and treatment of lung adenocarcinoma (Gillette et al., 2020). Jamal-Hanjani Mariam believed that the tumor internal heterogeneity caused by chromosome instability was related to the increased risk of recurrence and death, which indicated that chromosome instability might be a prognostic factor (Jamal-Hanjani et al., 2017). Faruki Hawazin analyzed the molecular expression of lung AD (Alveolar Duct) and SCC (Squamous Cell Carcinoma), showing important and repeatable differences. Evaluation of tumor expression subtypes is a possible biomarker of immunotherapy (Faruki et al., 2017). Lindeman Neal I discussed the molecular analysis of lung cancer as a standard for treatment decisions of targeted inhibitors. New research led to the evaluation of alternative analysis techniques, target genes, patient populations and tumor types (Lindeman et al., 2018). The above research has achieved good results. However, with the continuous updating of technology, there are still some problems.

AI medical images are widely used in the treatment process. Pesapane Filippo used the radio diagnostic model to summarize AI science in the forensic field (Pesapane et al., 2018). Kaissis Georgios introduced a privacy based medical image analysis technology, which was an open-source software architecture that could perform joint learning and cryptographic reasoning under different private and secure aggregations (Kaissis et al., 2021). Alexander Alan believed that AI greatly changed the medical imaging market, thereby changing the working mode of imaging physicians, and helped imaging physicians speed up scanning to make more accurate diagnosis so as to reduce the workload (Alexander et al., 2020). Lee Cecilia S believed that although AI had made great progress, there were still some problems in its use in medical imaging due to the work pressure and limitations of experts' labels (Lee et al., 2019). Allen Jr Bibb analyzed that in the field of medical imaging, the development of machine learning technology had been rapidly developed in academic and industrial laboratories. The main AI tools in diagnosis imaging include disease detection and classification, image optimization, radiation reduction and workflow improvement (Allen et al., 2019). Arabi Hossein briefly introduced the application of AI technology in molecular imaging, radiotherapy and other fields in recent years, especially in depth learning. The application of AI was discussed from five aspects: PET (polyethylene terephthalate) instrument design, PET image quantization and segmentation, image denoising, radiation dosimetry, computer aided diagnosis and prediction (Arabi and Zaidi, 2020). Nakata Norio analyzed the latest computer vision research report and selected potential clinical applications from the perspective of radiologists, such as generating confrontation networks, knowledge extraction and general image data to guide learning (Nakata, 2019). The above research shows that the application of AI medical imaging has a positive effect, but there are still some problems.

In this paper, the relationship between AI technology and various driving genes in lung tumors on lung nodules was discussed. First, the role of AI technology in early lung adenocarcinoma of lung nodules was analyzed and simulated. This article also introduced the application of bioinformatics technology to study the factors affecting the occurrence, development, and prognosis of lung adenocarcinoma, and preliminarily discussed its role.

2 The study design of the relationship between early lung adenocarcinoma and multiple driver genes based on medical images of lung nodules

Due to the application of high-resolution computed tomography and the early detection of small pulmonary nodules, lung cancer has become the most common pathological type. Since EGFR gene mutation, targeted therapy has greatly improved the survival rate of lung cancer, especially breast cancer, and accurate molecular phenotype is a prerequisite for tumor treatment. CT examination is the main examination method of lung adenocarcinoma. The relationship between CT images and the expression level of control genes has certain reference value for the diagnosis and treatment of lung adenocarcinoma.

Study design: the factors related to artificial intelligence medical imaging are introduced to provide a theoretical basis for the following experiments; Then, we analyzed the application of artificial intelligence medical imaging in the early stage of lung adenocarcinoma of lung nodules, further discussed how to use medical imaging to assist the treatment of lung adenocarcinoma, and finally, combined with the materials described above, we made an empirical study on the relationship between early stage of lung adenocarcinoma of lung nodules and multiple driving genes.

3 Overview of experimental materials and methods for clinical application of early lung adenocarcinoma

(1) General materials

A total of 50 patients with lung adenocarcinoma were selected as the study object, from Changzhou Second People's Hospital. Among the 50 patients with lung nodules, 31 were male and 19 were female, aged 47–73 years with an average age of 60 years (60 ± 7.51).

A total of 50 patients with lung adenocarcinoma were selected as the study subjects, including 31 men of the 50 patients with pulmonary nodules, 19 women, aged 47–73 years and with a mean age of 60 years (60 ± 7.51), as shown in Table 1.

(2) Treatment

TABLE 1 Classification of patients with lung adenocarcinoma.

Patients with pulmonary nodules	Male	Female sex	Average age
Number of people	31	19	60 ± 7.51 years old
Proportion	62%	38%	100

1) The preoperative computed tomography (CT) scanning results of the patient were collected. The images were retrieved from the image library, and the CT scanning results were analyzed by several senior radiologists.
2) The original HE (hematoxylin eosin) in 50 pathological specimens were sectioned for pathological diagnosis and differentiation. Two pathologists with more than the qualification of attending should perform blind examination, and the inconsistent samples were discussed again and agreed.

3.1 The ethics committee statement

The data samples studied in this experiment are analyzed from relevant studies in this field. Animals or humans were not used as experimental subjects to promote progress in the medical field and to obtain written informed consent from the patients.

(3) Observations

It is better to have no artifact in the image, and it is better to have a small amount of artifact in the image of the tip, lung base and other parts. However, it has not any impact on the diagnostic results. The obvious artifact in the image is poor. At the same time, it also affects the internal structure and shape of the patient's pulmonary nodules, thus making the diagnosis wrong. The results of CT examination and pathological examination of the two groups were compared and the diagnostic coincidence rate of the two groups was compared.

(4) Statistical methods

In this study, SPSS18.0 statistical software was used to process the data and average x test was used to make statistics. $p < 0.05$ indicates that there is a significant difference between the two groups.

4 Results of early lung adenocarcinoma with multiple driver genes

4.1 Factors related to artificial intelligence medical imaging

(1) The significance of AI application in medical imaging

This paper summarizes the significance of three-point AI in medical imaging, as shown in Figure 1.

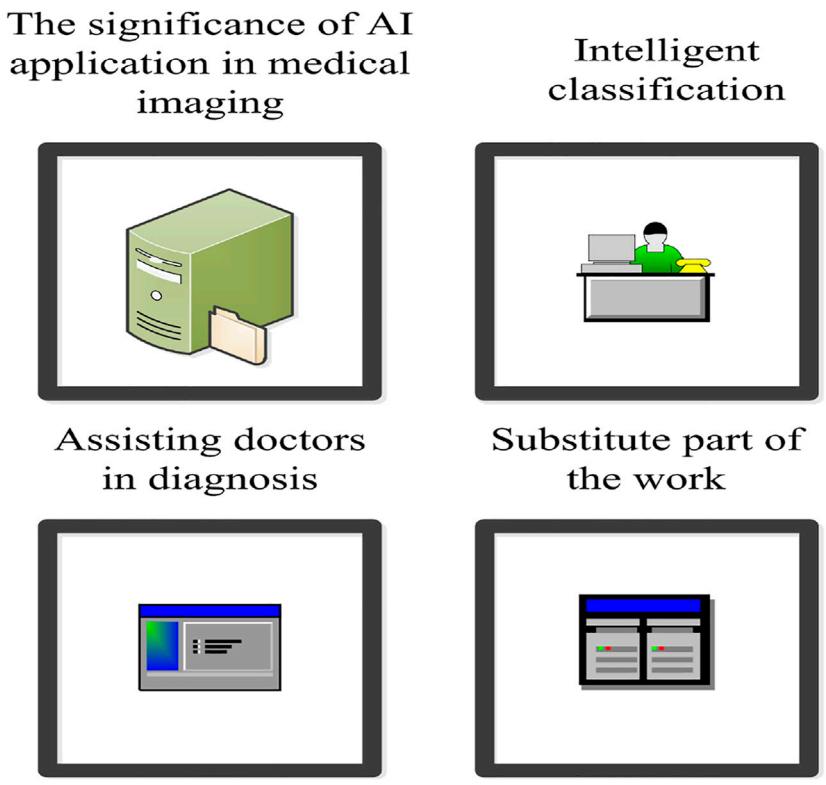


FIGURE 1
Significance of AI application to medical imaging.

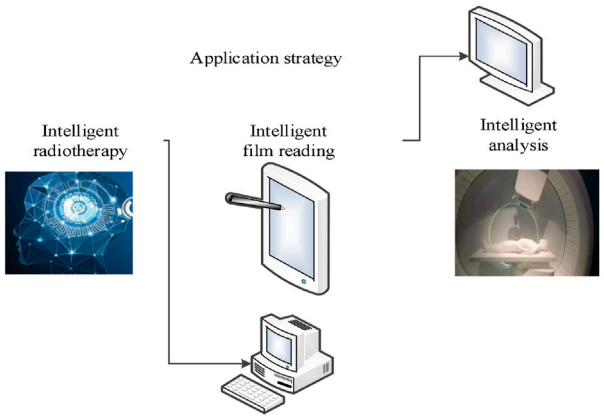


FIGURE 2
Application strategy of AI in medical imaging analysis.

1) Intelligent classification

AI is applicable to the type of intelligent detection. First, AI is more suitable for intelligent detection when the positive rate is low. Second, AI technology is more suitable for intelligent detection when the proportion of positive area data is low. Third, if the case image does not require too much professional knowledge, it is more suitable for AI intelligent recognition such as pulmonary nodules in physical examination. Therefore, AI technology used for intelligent classification detection can save a lot of negative case data, which can effectively reduce the work efficiency of medical personnel and reduce the waste of medical resources (Kim and Chung, 2020; Chiappelli and Penhaskashi, 2022).

2) Substitute part of the work

AI technology can replace part of the work of staff. This work can be divided into: First, if conditions are simple, it is more suitable to use AI technology to replace doctors' work, such as robot guidance. Second, if the structure of medical knowledge required by patients is relatively simple, then AI technology can be used to replace the work of doctors, such as AI doctors. Intelligent doctors can read patients' medical records and automatically output high-quality diagnostic results based on patients' chief complaints, symptoms, medical history, medical images and other data. Third, when the data structure of image analysis is relatively complex, AI can be used to replace the work of some doctors, such as intelligent diagnosis of medical images.

3) Assisting doctors in diagnosis

AI technology can improve doctors' abilities, such as auxiliary medical imaging diagnosis system. AI technology can also enhance doctors' ears, such as intelligent medical assistants. AI technology can still enhance doctors' brains, such as AI driven doctor assistance system (Sankaranarayanan et al., 2022).

(2) Application strategy of AI in medical image analysis

This paper summarizes the application strategy of three-point AI in medical image analysis, as shown in Figure 2.

1) Intelligent film reading

AI pattern recognition can be used to predict pathological changes in medical images and help doctors identify pathological changes and improve work efficiency (Sydorova et al., 2022). Professional imaging doctors need to spend about 10 min to review the CT images of pulmonary nodules. However, AI technology is more efficient, which can quickly identify and mark small nodules. It greatly reduces the doctor's inspection time and helps doctors make a diagnosis. In addition, the medical imaging system based on AI technology can also use the powerful learning ability of AI technology to carry out intelligent film reading and learning, so as to continuously improve the accuracy of lesion recognition. This helps doctors allocate time reasonably to better assist doctors in reading films and diagnosis.

2) Intelligent radiotherapy

At present, radiotherapy is an important method to treat cancer. Under traditional medical conditions, radiotherapy doctors are the main body of radiotherapy (Wang et al., 2022). However, due to the scarcity of radiologists, patients have to go to large hospitals. Radiotherapists need to mark the lesion manually during radiotherapy, which usually takes several hours in a conventional medical environment. With the help of AI, the lesion can be marked, which only needs to be confirmed by the radiation therapist, thus greatly improving the work efficiency.

3) Intelligent analysis

The number of pathologists in China is small. Pathologists not only spend a lot of time to make cell tissue sections, but also find small cancer cells in medical records containing billions of pixels. Therefore, even experienced case physicians makes mistakes. With the help of AI, the accuracy and efficiency of diagnosis can be greatly improved through intelligent analysis. Of course, it cannot be judged solely by AI, but also through pathological examination to avoid misdiagnosis.

4.2 Application of artificial Intelligence medical imaging in early lung adenocarcinoma of pulmonary nodules

(1) Application of imaging histology in lung cancer

Some scholars have studied that the appearance of surface image group can reflect the heterogeneity of tumors. Tumor type, T stage, gene expression pattern and prognosis are significantly related. Therefore, histologic imaging technology has become a hot spot in the field of lung cancer diagnosis. It has been widely used to differentiate benign and malignant pulmonary nodules and evaluate them comprehensively. This involves the analysis of histological type, cellular molecule and gene status, therapeutic effect and prognosis.

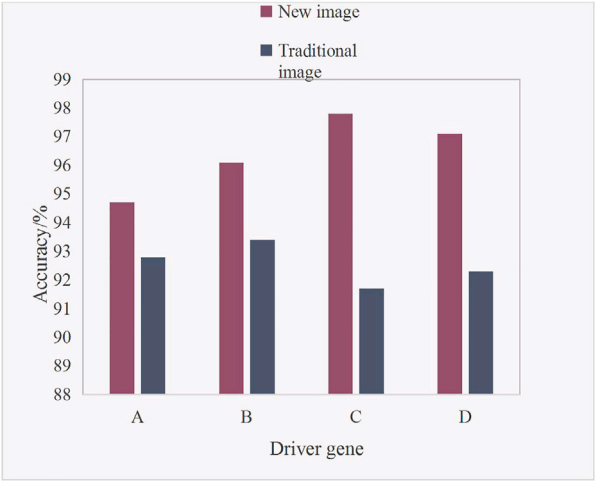


FIGURE 3
Accuracy analysis of early lung adenocarcinoma in lung nodules.

TABLE 2 Analysis of the relationship between groups X and Y and the four driver genes.

/	A	B	C	D	X value (%)
	(6)	(6)	(7)	(4)	
Group X	2 (33%)	2 (33%)	3 (43%)	1 (25%)	33.5
Group Y	5 (83%)	5 (83%)	5 (71%)	3 (75%)	78

1) Screening and differential diagnosis of pulmonary nodules

The guidelines for the diagnosis and management of pulmonary nodules recommend the use of low-dose CT for screening, but the false positive rate of low-dose CT is very high. Therefore, the qualitative diagnosis of pulmonary nodules is a challenging clinical problem. Histological imaging research provides more valuable information for the screening of nodules and the differentiation between benign and malignant nodules.

2) Auxiliary pathological diagnosis

Imaging group technology can combine imaging parameters with clinical manifestations. If tumor tissue or tissue slice cannot be obtained, pathological examination by non-invasive method is helpful to determine the next treatment plan. Some studies have shown that the performance of some image groups can be used as an independent influencing factor to predict the degree of tumor invasion, so as to achieve the goal of treatment stratification for early lung cancer patients.

3) Gene expression

With the continuous development of precision medicine, the diagnosis of lung cancer patients is more and more in-depth. However, it is difficult to find the mutant gene at very low concentration. Moreover, tumor heterogeneity can be tested by biopsy. Imagomics analysis can obtain useful information about gene variation. When surgery or biopsy cannot be performed, the combination of clinical manifestations can increase diagnostic value and assist in formulating treatment plans.

4) Efficacy evaluation

It is an important basis for formulating and adjusting treatment plans to accurately estimate and evaluate the efficacy in clinical practice. Some studies have shown that the changes in the imaging group during surgery can be used as the prognosis of neoadjuvant chemotherapy to distinguish between pathological complete remission and gross residual lesions.

5) Clinical prognosis

Lung cancer is the highest mortality malignant tumor in the world today. Even if it is treated in time, a series of problems such as local recurrence, distant metastasis, radiation induced lung injury and so on appear. Therefore, it is necessary to grade high-risk patients. Imageomics has a great predictive effect on the clinical prognosis of patients.

4.3 Analysis of the relationship between early lung adenocarcinoma of pulmonary nodules and multiple driving genes

(1) The relationship between groups X and Y and four driving genes

The experiment showed that 50 patients with lung adenocarcinoma admitted to the hospital were collected and divided into groups. The even number is the control group (group X), and the odd number is the study group (group Y). Amplification block mutation system method was used to detect the abnormalities of A, B, C, and D (A is HER1 is also called erb B1, EGFR, B is also called erbB2, neu, C is HER3 is also called erb B3, and D is HER4 is also called er-B 4) in the samples. The relationship between driver gene mutation and pathological classification was analyzed, as shown in Table 2.

The highest x value of the four driving genes in group Y was 83% and the average was 78%, which was significantly higher than that in group X. The highest X value of patients in the four driving genes was 43% and the average was 33.5%. The data were statistically significant.

(2) The relationship between abnormal driving gene and tissue type in early lung adenocarcinoma

Experiment description: 25 cases were invasive and 25 cases were non-invasive, and the analysis of A, B, C, and D driving genes was conducted by SUVmax and SUVindex analysis (SUV value is fully called standard uptake value (standard uptake value, SUV), which is a common semi-quantitative index of pet in tumor diagnosis, and refers to the radioactive activity of local tissue imaging agent and the average systemic injection activity. Invasive growth refers to the way tumor cells grow and destroy surrounding tissues, as shown in Table 3.

TABLE 3 Relationship between driver gene abnormalities and tissue type.

	Total	SUVmax	SUVindex	A	B	C	D
Infiltrative growth	25	7.62	4.79	0.241	0.342	0.354	0.412
Non-invasive growth	25	1.84	0.91				
Gene mutation	—						
positive	19	7.17	3.41	0.145	0.147	0.175	0.162
negative	25	6.72	4.12				

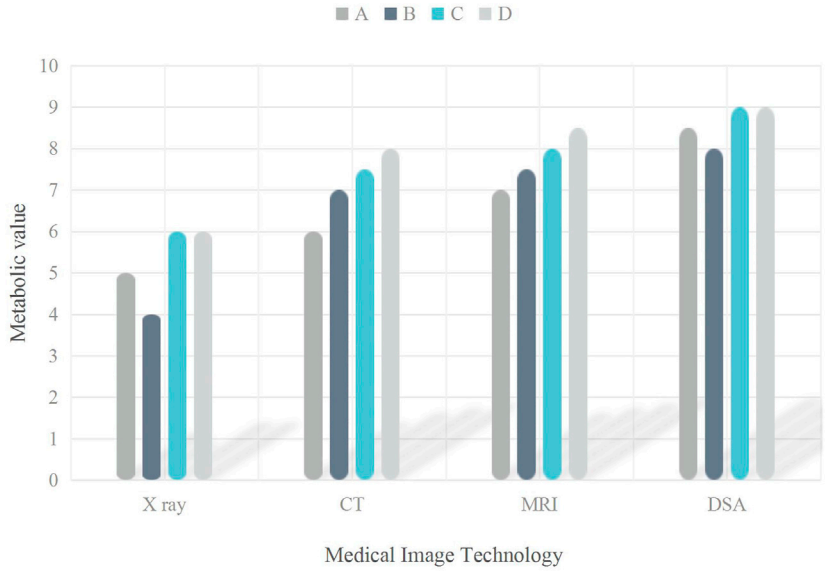


FIGURE 4
Medical image of the gene mutation in the lung adenocarcinoma in the driver gene.

It can be seen from Table 2 that the mutation values of the four driving genes were 0.241, 0.342, 0.354, and 0.412 respectively with the change of pulmonary nodule infiltration. The highest mutation rate of D was 41.2%, and the others were C, B, A in turn. The expression of D was consistent among the four genes, and the difference was statistically significant.

(3) Image accuracy

Figure 3 shows the accuracy analysis of four driving genes under AI medical imaging (new influence) and traditional imaging for early lung adenocarcinoma of lung nodules.

It can be seen from Figure 3 that the accuracy of the new image was higher than that of the traditional image in the four driving genes. The specific data analysis showed that the accuracy of the new image in driver gene A was 94.7%, and the accuracy of the traditional image in driver gene A was 92.8%. The accuracy of the new image in driver gene B was 96.1%, and the accuracy of the traditional image in driver gene B was 93.4%. The accuracy of the new image in driver gene C was 97.8%, and the accuracy of the traditional image in driver gene C was 91.7%. The accuracy of new image in driver gene D was 97.1%, and the accuracy of traditional image in driver gene D was 92.3%. It is concluded that the highest accuracy of C was 97.8% for the new image in the driver gene, while the highest accuracy of B was 93.4% for the traditional image in the driver gene. The average accuracy of the new image in the four driving genes was 96.43%, and the average accuracy of the traditional image in the four driving genes was 92.55%. The average accuracy of the new image was 3.88% higher than that of the traditional image.

(4) Effects of different imaging techniques on lung adenocarcinoma in four driving genes

Four kinds of medical imaging technologies were counted here, including X-ray, CT, MRI (Magnetic Resolution Imaging) and DSA (Digital subtraction angiography). The mutation of metabolic value in four driving genes to lung adenocarcinoma was analyzed, as shown in Figure 4.

It can be seen from Figure 4 that the metabolic values of the four medical imaging technologies in the four driving genes were different. In drive gene A, the metabolic value of X-ray was 5, that of CT was 6, that of MRI was 7, and that of DSA was 8.5. In drive gene B, the metabolic value of X-ray was 4, that of CT was 7, that of MRI was 7.5, and that of DSA was 8. In drive gene C, the metabolic

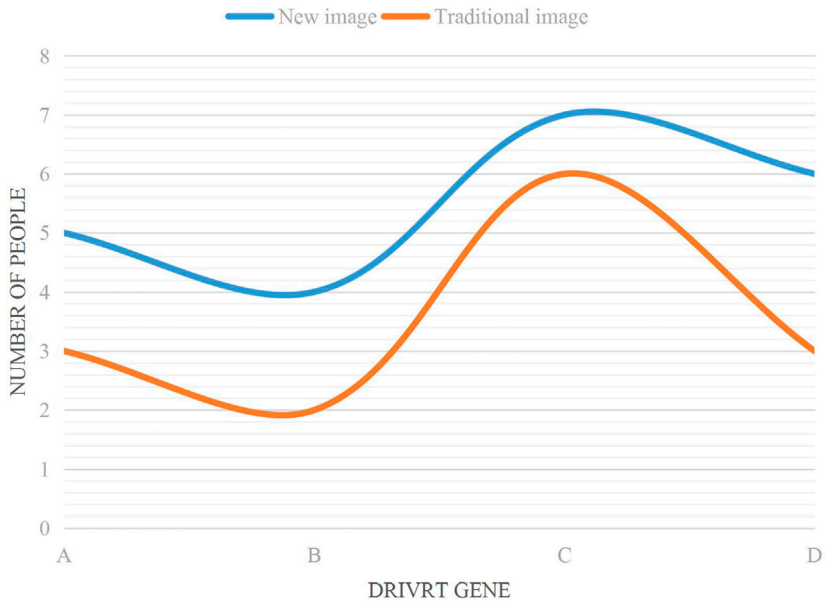


FIGURE 5
Number of gene mutations that can be detected.

value of X-ray was 6, that of CT was 7.5, that of MRI was 8, and that of DSA was 9. In driver gene D, the metabolic value of X-ray was 6, that of CT was 8, that of MRI was 8.5, and that of DSA was 9.

(5) Number of gene mutations

Gene mutation is a tumor specific somatic genetic change. Based on the above analysis of early lung adenocarcinoma and multiple driving genes of pulmonary nodules in AI medical images, 50 patients were divided into 25 patients each. The number of people whose gene mutations could be detected was analyzed by combining new medical images with traditional medical images, as shown in Figure 5.

It can be seen from Figure 5 that the trend of the two curves of the broken line chart is from A to B, then to C, and finally to D.

According to the data analysis in the trend combination chart, 20% of the total number of 5 people were detected by the new image combined with driver gene A. Combined with driver gene B, 16% of the total number of 4 people were detected. Combined with driver gene C, 28% of the total number of 7 people were detected. Combined with driver gene D, 24% of the total number of 6 people were detected. The new image detected 88% of the total number of 22 people under four driving genes. Traditional imaging combined with driver gene A detected 12% of the total number of 3 people. Combined with driver gene B, 8% of the total number of 2 people were detected. Combined with driver gene C, 24% of the total number of 6 people were detected. Combined with driver gene D, 12% of the total number of 3 people were detected. The new image detected 56% of the total number of 14 people under four driving genes. Through calculation, the new image detected 8 more people than the traditional image.

5 Discussion

Science and technology have made significant progress in recent years, and the application of AI technology has provided opportunities for development in various fields, including clinical medicine. Especially in the field of medical imaging, AI technology can be used to facilitate the visualization of imaging research results. Compared with traditional manual reading, AI reading has absolute advantages, which can not only reduce the reading time but also provide a highly sensitive and accurate reference for clinical diagnosis. In this paper, 50 patients with lung adenocarcinoma were admitted and grouped, and the patients in group Y had the highest x value of 83% among the four driver genes and the average value of 78%, which was significantly higher than that in group X. Patients had the highest X value of 43% and 33.5% mean value of 34.4 driver genes.

In order to further improve the accuracy of medical image diagnosis by using AI technology, this paper studied the application of AI in medical image analysis. First, the concept and advantages of AI were introduced. Then, the importance of AI application in medical imaging was analyzed. Finally, the strategy of applying AI in medical image analysis was proposed. This paper has presented a strategy for applying AI to medical image analysis. After a period of research and practice, it is found that this strategy has universal value, because it improves the reading efficiency of medical image analysis system, which effectively improves the reading depth of medical image analysis system and helps doctors diagnose accurately.

With the deepening of scientific understanding of tumor, people gradually realize that the occurrence and development of tumor are related to multiple driver genes, so in order to better evaluate the incidence and development of lung cancer. The EGFR (epidermal growth factor receptor) collision integration method can evaluate the effect of EGFR through a single hotspot mutation such as EGFRL858R, and can predict the different expression of genes

such as EGFRL858R and predict the effect of EGFR. Some scholars have conducted a single-region genome sequence analysis on many cases of lung cancer patients, and the results show that the mutation of this gene has a great relationship with the incidence of lung cancer. The prediction of NSCLC by 7-lncRNA (Long non-coding RNA) is particularly important for the early diagnosis of KRAS (Kirsten Rat Sarcoma Viral Oncogene Homolog) and EGFR.

Key driver genes and suppressor genes play important roles in the early detection and treatment of lung cancer. For early lung adenocarcinoma, a specific sequence analysis according to its clinical characteristics is combined with specific biomolecular markers to develop personalized treatment options and conduct extensive evaluation. However, as large-scale clinical data are still lacking, especially for patients with early lung adenocarcinoma, there are few data on outcome and most studies are retrospective and the follow-up time is very short. How to effectively overcome secondary drug resistance is an ongoing research topic. Furthermore, according to the Joint Committee on Cancer guidelines, many of the early-stage lung cancer studies reviewed here are rarely defined as AIS, minimally invasive adenocarcinomas, and partially appendicular growth. More attention is therefore expected to be paid to the prognosis of lung adenocarcinoma in the future.

6 Conclusion

In conclusion, AI based medical images are superior to traditional images in all aspects. At the same time, bioinformatics can be used to explore the pathogenesis of lung adenocarcinoma, so as to find appropriate biomarkers for early diagnosis and prognosis of lung adenocarcinoma.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

YY, JL, and JT wrote the manuscript. YC and KZ revise the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aisner, D. L., Sholl, L. M., Berry, L. D., Rossi, M. R., Chen, H., Fujimoto, J., et al. (2018). The impact of smoking and TP53 mutations in lung adenocarcinoma patients with targetable mutations—the lung cancer mutation consortium (LCMC2). *Clin. Cancer Res.* 245, 1038–1047. doi:10.1158/1078-0432.ccr-17-2289

Alexander, A., Jiang, A., Ferreira, C., and Zurkiya, D. (2020). An intelligent future for medical imaging: A market outlook on artificial intelligence for medical imaging. *J. Am. Coll. Radiology* 171, 165–170. doi:10.1016/j.jacr.2019.07.019

Allen, B., Jr, Seltzer, S. E., Langlotz, C. P., Dreyer, K. P., Summers, R. M., Petrick, N., et al. (2019). A road map for translational research on artificial intelligence in medical imaging: From the 2018 national institutes of health/RSNA/ACR/the academy workshop. *J. Am. Coll. Radiology* 169, 1179–1189. doi:10.1016/j.jacr.2019.04.014

Arabi, H., and Zaidi, H. (2020). Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy. *Eur. J. Hybrid Imaging* 41, 17–23. doi:10.1186/s41824-020-00086-8

Chen, J., Yang, H., Teo, A. S. M., Amer, L. B., Sherbaf, F. G., Tan, C. Q., et al. (2020). Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* 522, 177–186. doi:10.1038/s41588-019-0569-6

Chiappelli, F., and Penhaskashi, J. (2022). Toward a fractalomic idiotype/anti-idiotypic paradigm. *Bioinformation* 189, 730–733. doi:10.6026/97320630018730

Chuang, C-H., Greenside, P. G., Rogers, Z. N., Brady, J. J., Yang, D., Ma, R. K., et al. (2017). Molecular definition of a metastatic lung cancer state reveals a targetable CD109–Janus kinase–Stat axis. *Nat. Med.* 233, 291–300. doi:10.1038/nm.4285

Faruki, H., Mayhew, G. M., Serody, J. S., Hayes, D. N., Perou, C. M., and Lai-Goldman, M. (2017). Lung adenocarcinoma and squamous cell carcinoma gene expression subtypes demonstrate significant differences in tumor immune landscape. *J. Thorac. Oncol.* 126, 943–953. doi:10.1016/j.jtho.2017.03.010

Gillette, M. A., Satpathy, S., Cao, S., Dhanasekaran, S. M., Vasaikar, S. V., Krug, K., et al. (2020). Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* 1821, 200–225.e35. doi:10.1016/j.cell.2020.06.013

Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B. K., Veeriah, S., et al. (2017). Tracking the evolution of non–small-cell lung cancer. *N. Engl. J. Med.* 37622, 2109–2121. doi:10.1056/NEJMoa1616288

Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., et al. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nat. Mach. Intell.* 36, 473–484. doi:10.1038/s42256-021-00337-8

Kim, J-C., and Chung, K. (2020). Neural-network based adaptive context prediction model for ambient intelligence. *J. Ambient Intell. Humaniz. Comput.* 114, 1451–1458. doi:10.1007/s12652-018-0972-3

Lee, C. S., Tyring, A. J., Wu, Y., Xiao, S., Rokem, A. S., DeRuyter, N. P., et al. (2019). Generating retinal flow maps from structural optical coherence tomography with artificial intelligence. *Sci. Rep.* 91, 5694–5711. doi:10.1038/s41598-019-42042-y

Lindeman, Neal I., Cagle, P. T., Aisner, D. L., Arcila, M. E., Beasley, M. B., Bernicker, E. H., et al. (2018). Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors: Guideline from the college of American pathologists, the international association for the study of lung

cancer, and the association for molecular pathology. *J. Thorac. Oncol.* 133, 323–358. doi:10.1016/j.jtho.2017.12.001

Nakata, N. (2019). Recent technical development of artificial intelligence for diagnostic medical imaging. *Jpn. J. radiology* 372, 103–108. doi:10.1007/s11604-018-0804-6

Pesapane, F., Codari, M., and Sardanelli, F. (2018). Artificial intelligence in medical imaging: Threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur. Radiol. Exp.* 21, 35–10. doi:10.1186/s41747-018-0061-6

Rubik, B., and Jabs, H. (2018). Artificial intelligence and the biofield: New opportunities and challenges. *Cosmos Hist. J. Nat. Soc. Philosophy* 141, 153–162.

Sankaranarayanan, P., Ragunath, P. K., and Venkatesan, Perumal. (2022). Diagnosis of triple negative breast cancer using expression data with several machine learning tools. *Bioinformation* 184, 325–330. doi:10.6026/97320630018325

Sydorova, E., Pahomova, A., and Halasz, S. (2022). Designing socio-technical systems using the system paradigm in the context of nano-bio-information technology and cognitive science convergence. *Qual. Innov. Prosper.* 262, 170–182. doi:10.12776/qip.v26i2.1637

Wang, R., Jiao, Z., Yang, L., Choi, J. W., Xiong, Z., Halsey, K., et al. (2022). Artificial intelligence for prediction of COVID-19 progression using CT imaging and clinical data. *Eur. Radiol.* 321, 205–212. doi:10.1007/s00330-021-08049-8

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 22 February 2023
DOI 10.3389/fgene.2023.1118831

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Tao Yang,
First Affiliated Hospital of Chongqing Medical University, China
Gongxing Yan,
Chongqing Vocational Institute of Engineering, China

*CORRESPONDENCE
Xingjing Wu,
164903303@stu.cuz.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 08 December 2022
ACCEPTED 13 February 2023
PUBLISHED 22 February 2023

CITATION
Wu X, Tao Z and Cheng W (2023), Microrna-206 induces hypoxic necrosis of femoral head by inhibiting VEGF/PI3K/AKT signaling pathway.
Front. Genet. 14:1118831.
doi: 10.3389/fgene.2023.1118831

Microrna-206 induces hypoxic necrosis of femoral head by inhibiting VEGF/PI3K/AKT signaling pathway

Xingjing Wu[1]*, Zhoushan Tao[1] and Wenjing Cheng[1]

[1]Department of Orthopedics, Yijishan Hospital, Wannan Medical College, Wuhu, Anhui, China

The most common form of non-traumatic necrosis of the femoral head is anoxic necrosis of the femoral head, which is a metabolic disease, mainly involving young and middle-aged people. Apoptosis and its related signal regulation pathway play an important role in the occurrence and development of hypoxic necrosis of the femoral head. In order to investigate the possible pathological manifestations of miR-206 and VEGF/PI3K/AKT signal pathway genes and their interactions in hypoxic necrosis of the femoral head, this paper intended to systematically study the expression and regulation mechanism of miR-206 and VEGF/PI3K/AKT signal pathway genes. The interaction between miR-206 and VEGF/PI3K/AKT signaling pathway and its regulation on apoptosis, differentiation and proliferation of human osteoblast cell line hFOB1.19 (SV40 transfer of human osteoblasts) were studied by double luciferase reporter gene analysis, overexpression and inhibition of miR-206, and gene silencing of VEGF/PI3K/AKT signaling pathway. After 24 h and 48 h of intervention with MicroRNA 206 on osteoblasts, it was found that the fluorescence intensity of caspase-3 was higher than that of 0 h group ($p < 0.05$). This paper has provided an important research basis for the research of femoral head necrosis and the development of new diagnosis and therapeutic drugs for this kind of disease. It also has provided a reference for the further promotion of the chemotherapy drug delivery system.

KEYWORDS

hypoxic necrosis of femoral head, VEGF/PI3K/Akt signal pathway, microRNA-206 molecule, chemotherapy drug delivery system, anoxic necrosis

1 Introduction

MiR-206 is involved in the differentiation of osteoblasts and its expression level is significantly increased in the tissue of induced hypoxic necrosis of the femoral head. However, its specific role and molecular mechanism in the process of steroid induced necrosis of the femoral head remain unknown. Non-traumatic osteonecrosis of the femoral head (NONFH) is the most common orthopedic disease in clinic, most of which are accompanied by a history of corticosteroid use or a long history of alcoholism. The necrosis of femoral head is mostly due to the death of osteoblasts and bone marrow components caused by blood stasis of femoral head vein, damage or interruption of arterial blood supply, which leads to the change and collapse of femoral structure. Pain in hips, buttocks and groins is the main clinical pain, which often occurs in people aged 20–59 years. Due to the impact of COVID-19 in recent years, the clinical use of hormone drugs may further lead to an rise in the proportion of patients, which have a great impact on the quality of life, economic burden and social health undertakings.

As a common difficult and miscellaneous disease in the field of orthopedics, NONFH has attracted extensive attention from researchers in other countries. This study of the functional texture of MicroRNA-206 has set a good precedent for the introduction of chemotherapy drug delivery system. The mechanism of MicroRNA-206 inducing hypoxic necrosis of the femoral head by inhibiting VEGF/PI3K/AKT signal pathway has been explored, which is helpful to understand more treatment methods for necrosis of the femoral head.

Femoral head necrosis is seriously affecting people's physical and mental health. Sheng Zhai believed that hormone-induced necrosis of the femoral head is caused by long-term use of hormones (Sheng et al., 2021). Hua Kun-chi believed that core decompression was an important method to treat femoral head necrosis (Hua et al., 2019). Li Wei believed that in osteonecrosis of the femoral head (ONFH), the blood supply was insufficient to meet the metabolic needs of bone (Li et al., 2017). Zhang Ying believed that bone microcirculation might lead to bone necrosis related to ischemia and hypoxia (Zhang Y. et al., 2017). Leibold Christiane Sylvia performed hip dislocation and varus osteotomy on nine patients with femoral head necrosis (Leibold et al., 2020). However, the treatment of femoral head necrosis proposed by them lacks preciseness, so this paper introduced MicroRNA 206 for comparative optimization.

MicroRNA-206 can enhance cell toxicity. Luo pan believed that the incidence rate of steroid induced ONFH in patients receiving long-term treatment was 9%–40% (Luo et al., 2018). Ma Jian-xiong believed that the mechanism of ONFH was still unclear (Ma et al., 2017). Narayanan Aswath suggested surgical treatment for patients with advanced disease of avascular necrosis of femoral head (Narayanan et al., 2017). Landgraeber Stefan believed that advanced core decompression (ACD) was a new technology for the treatment of ONFH (Landgraeber et al., 2017). Shu Peng believed that ONFH was a common disabling joint disease (Shu et al., 2020). MicroRNA-206 was not combined with femoral head necrosis.

The tumor suppressor PDCD4 (programmed cell death4) plays an significant role in regulating the apoptosis of cardiovascular cells, and the expression of PDCD4 gene is regulated by multiple microRNAs. However, its pathological role and molecular mechanism in the process of human steroid induced avascular necrosis of the femoral head are still unclear. In order to investigate the potential pathological function and interaction of miR-206 and PDCD4 genes in steroid induced avascular necrosis of the femoral head, the functions and mechanisms of miR-206 and PDCD4 expression and their interaction in osteoblast apoptosis and steroid induced avascular necrosis of the femoral head are systematically studied. The correlation coefficient (R value) between miR-206 and mRNA level of VEGF/PI3K/AKT signaling pathway is about 0.48.

2 Method of inducing hypoxic necrosis of femoral head with microRNA-206 depending on chemotherapy drug delivery system

2.1 Mechanism of hypoxic necrosis of femoral head

Modern medicine of avascular necrosis of femoral head is based on precise anatomy and pathology. According to its clinical manifestations, this disease can be attributed to the category of bone arthralgia, bone erosion and hip arthralgia in traditional Chinese medicine. Hip joint is the main mechanical conversion mechanism and key load-bearing part of the human body. It is mainly composed of acetabulum and femoral head. The pathogenesis of the disease is complex. It is widely believed that alcohol, hormone, pregnancy and immune diseases are important inducements, and osteocyte necrosis and cystic degeneration in the femoral head are its pathological manifestations. However, the specific mechanism is not yet clear, so there is a series of theoretical arguments. In China, due to the limitation of economic level, education level and medical conditions, the consequences of NONFH (The results of bone biopsy are consistent with ischemic necrosis, but other examinations are normal) are disastrous. Most patients have entered the collapse period and peri collapse period when they go to hospital. Although most scholars have done a lot of research in the field of hip protection and achieved some results, they often miss opportunities. Most patients need to undergo hip replacement surgery, and the economic burden and complications brought by the surgery are too much for patients. In order to achieve early diagnosis and treatment, many researchers have done a lot of work in the field of molecular biology. Through basic and experimental research, the aim is to find relevant molecular biological markers for early detection, diagnosis and effective intervention Q^{N+1}. However, they are all in the bottleneck period and there is no important breakthrough. However, due to the limitations of the research methods of traditional Chinese medicine and the difficulties K^N in integrating it with modern molecular biology, the scientific basis provided by most of the studies is not convincing (Calori et al., 2017). The incidence of infection after replacement of femoral head necrosis was 2.1%–10.3%. Because of the huge number of artificial joint replacement operations carried out every year, the number of infected cases is also countless. The infection caused by any reason will have serious consequences, eventually leading to prosthesis loosening and surgical failure. Severe infection can cause paralysis.

$$Q^{N+1} = f\left(Y, M^N\right) \tag{1}$$

$$\mathrm{K^N} = \mathrm{T} + \mathrm{V} \tag{2}$$

$$\mathrm{x_M} = \mathrm{N^{-1}}\left(\mathrm{M_1}\right)(\tau - \mathrm{Y} - \mathrm{d}) \tag{3}$$

The relative expression level of PI3K/AKT signal pathway is positively correlated with total cholesterol (it is a major steroid compound in mammals and plays an important role in basic cell life activities), low density lipoprotein and other indicators. The more obese patients are, the higher the relative expression level of PI3K/AKT signal pathway is. Clinical research shows that hyperlipidemia and lipid metabolism disorder are one of the high risk factors of femoral head necrosis. Other studies also found that after the application of PI3K inhibitor, the expression of SREBP (Steel Regulatory Element Binding Protein) decreases significantly, which leads to a significant decrease in thyroglobulin indicators. It leads to lipid metabolism disorder and excessive obesity, and even leads to the formation of atherosclerosis (the symptoms of atherosclerosis mainly depend on vascular disease and the degree of ischemia of the affected organs), thus affecting the pathogenesis of femoral head necrosis J (Peng et al., 2021).

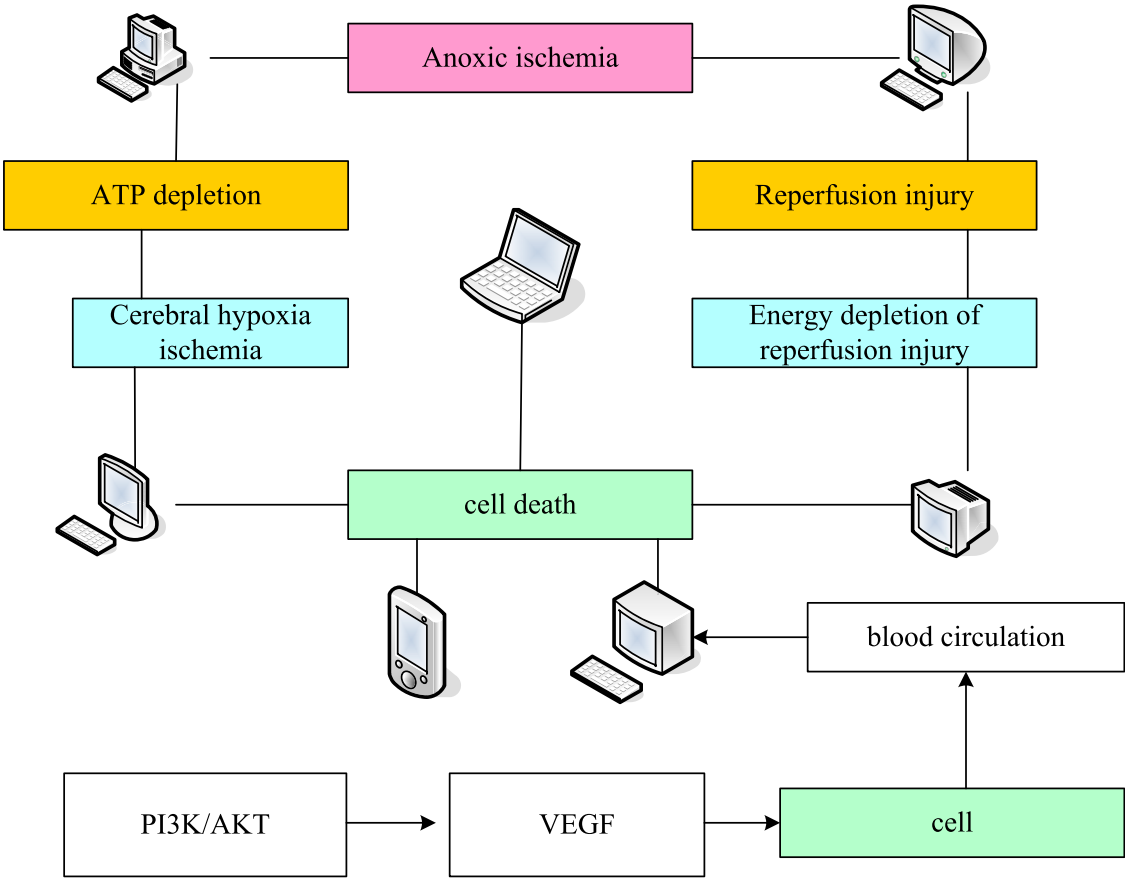


FIGURE 1
Pathogenesis of NONFH.

$$J = \sum_{i=0}^{n} J_i \tag{4}$$

$$N = KS - \frac{CS}{\|S + C\|} \tag{5}$$

$$bg = vgS + \frac{k}{\|l + m\|} \tag{6}$$

AKT combines with eNOS (Endothalial Nitric Oxide Synthase) phosphorylation to promote the production of NO (nitrous oxide). Its upstream factor PTEN (photosynthesis and tensin homolog deleted on chromosomal ten) protein can improve the vascular permeability by affecting the cell growth cycle, so as to adjust the blood circulation around the necrosis of femoral head, thus achieving the role of preventing and treating NONFH. The improvement of vascular permeability (The contraction of endothelial cells is called recurrent transient response because the half-life of the inflammatory mediator of endothelial cell contraction is short and the endothelial cell contraction caused by it is reversible) around the femoral head can provide a certain direction for the treatment of early and middle femoral head necrosis. However, most of them stay in the animal experiment stage, and further research is needed to obtain more intuitive clinical data. The pathogenesis of NONFH is shown in Figure 1.

Studies have shown that PI3K/AKT can regulate the phosphorylation of subunits when activated by external stimulation, thus starting VEGF gene transcription to promote angiogenesis. Moreover, by activating and supplementing VEGF (The gene structure of VEGF family proteins is complex, among which VEGF-A protein is the most studied. The precursor mRNA of VEGF-A gene transcription can form different fragments of VEGF-A protein through alternative splicing), it can promote the high expression of PI3K/AKT, stimulate the production of more VEGF and miRNA protein expression related to angiogenesis in the body, so as to better promote angiogenesis.

PI3K/AKT can directly induce the generation and expression of VEGF. VEGF can promote cell generation and regulate peripheral blood supply, which can help vascular repair and regeneration, and can nourish peripheral bone cells to alleviate the early development of femoral head necrosis. It is possible for VEGF to improve the blood circulation around necrosis of femoral head. The theory is relatively perfect. However, it has not been further applied to clinical treatment. It is worth further discussion and research to better clarify its mechanism of action (Wang et al., 2017; Wen et al., 2020).

The diagnostic criteria is that when the patient meets one of the following three points, the femoral head necrosis can be diagnosed, regardless of whether there are clinical symptoms or signs.

X-ray film: necrotic focus surrounded by sclerotic band and segmental collapse in the femoral head can be observed; Crescent sign and other specific manifestations.

Magnetic resonance imaging (MRI) examination:

T1W1 band low signal, T2W "double line" sign, the lateral low signal band is hyperplastic sclerotic bone, and the medial high signal band is granulation fiber tissue repair.

CT: necrotic focus with clear contour; Subchondral fracture.

2.2 MicroRNA-206 induces anoxic necrosis of femoral head

2.2.1 Clinical bone tissue sample

The clinical bone tissue samples of the experimental group used in this paper are from 15 patients with steroid induced avascular necrosis of the femoral head, including nine male patients and six female patients. Another 15 femoral head tissue samples from patients undergoing total hip replacement (The artificial prosthesis, including the femur and acetabulum, is fixed on the normal bone by bone cement and screws to replace the diseased joint) due to femoral neck fracture are used as the control group in this article, including eight male and seven female patients with femoral neck fracture. All patients participating in the study signs the informed consent form, and obtains detailed clinical pathology and prognosis data of all patients in the experimental group and the control group. The study covering relevant patient screening, clinical examination, clinical data acquisition and subsequent experimental analysis. The collected femoral head bone tissue samples are frozen in the prepared liquid nitrogen immediately after the operation for subsequent experimental analysis of total RNA and protein sample extraction.

Cell line: The product classification numbers of hFOB1.19 and human renal epithelial cell line 293 T are CRL-11372 and CRL-3216 respectively.

Main reagents and antibodies: conventional chemical reagents: disodium hydrogen phosphate, methanol, anhydrous ethanol, glacial acetic acid, isopropanol and other conventional reagents used in this paper are analytical pure. Total RNA extraction reagent (Trizol solution) is purchased from Bioengineering Co., Ltd. (Shanghai), and the product number is B511311. Diethyl pyrocarbonate is purchased from Sigma Aldrich (China) Co., Ltd. (Shanghai, China), and the product number is 472565. Agarose is purchased from Yisheng Biotechnology Co., Ltd. (Shanghai, China), and the product number is 10208ES60.

DNA marker is purchased from Beyotime Biotechnology Company (Nantong, China), and the product number is D0107. Nucleotide fluorescent dye (SYBR GreenI) is purchased from Thermo Fisher Scientific (China) Co., Ltd. (Shanghai, China), and the product number is S7563. Taq DNA polymerase is purchased from Fullgold Biotechnology Co., Ltd. (Beijing, China), and the product number is AP101-12. The basic idea of clinical experiment is shown in Figure 2.

2.2.2 Total RNA extraction process of bone tissue and cell line

In this paper, all patients with hypoxic avascular necrosis of the femoral head and the contrast group are used to extract total RNA from bone tissue, human osteoblasts and renal epithelial cells using a standardized process. In order to prevent sample contamination and affect the accuracy of experimental results, consumables such as gun heads and centrifuge tubes used in the extraction process are soaked in DEPC water (which is used for dissolution of RNA precipitation and various reaction systems containing RNA, such as reverse transcription, annealing of siRNA, etc.) in advance. The specific extraction process is as follows:

(1) Tissue and cell homogenate treatment: For bone tissue samples, the bone tissue samples stored in liquid nitrogen tank shall be ground in liquid nitrogen immediately after being taken out until uniform fine particles are formed. Then, pre cooled TRIzol (which is an RNA extract that can be directly extracted from cells to RNA) solution (1 mL of TRIzol solution is added to every 0.1 g of bone tissue) is added to fully mix, so that RNA samples can be fully dissolved in TRIzol solution. For cultured human osteoblasts and renal epithelial cells, 800 g of them are centrifuged for 5 min to collect cells, and then an appropriate amount of pre cooled PBS solution is added to resuspension cells. PBS solution is removed by centrifugation, and cells are washed twice in this way. Finally, about 1 mL TRIzol solution is added to resuspension cells and repeated blowing to promote cell lysis. The volume of reagent and solution used in the following experimental process is calculated according to 1 mL of homogenate.

(2) The centrifuge tube is inverted onto a clean absorbent paper, and then it is left to dry at room temperature. Finally, about 40 ul double distilled water treated by DEPC is added, and the sample is shaken on a vortex oscillator for a short time to promote the dissolution of the sample. It is stored at −80°C for standby.

(3) Agarose gel electrophoresis: 2 μL of the obtained RNA sample is taken out and separated by 2% agarose gel electrophoresis and photographed as v. The quality a and integrity of the obtained sample dP are analyzed (Zhu et al., 2020).

$$a = \sin \frac{2mt}{c} \tag{7}$$

$$v = \frac{at}{2\pi} \cos \frac{2vt}{t} \tag{8}$$

$$dP = dq_x + dq_y \tag{9}$$

(4) Determination of RNA sample solubility: Micropipette is used to suck 1 μL extracted RNA sample, and measure the RNA sample solubility and A260/A280 ratio through micro spectrophotometer. It is RNA sample lp meeting the purity requirements (Zhang M. et al., 2017; Yan et al., 2018).

$$K = d + y_m \tag{10}$$

$$o_{total} = \frac{o_F}{c} \tag{11}$$

$$lp = \frac{k_{VS}}{k_{max}} = \frac{v_M}{b_{VS}} \tag{12}$$

2.2.3 Extraction of total protein from bone tissues and cell lines

In order to analyze the changes of protein levels of related genes, the animal whole protein extraction kit (product number: C510003)

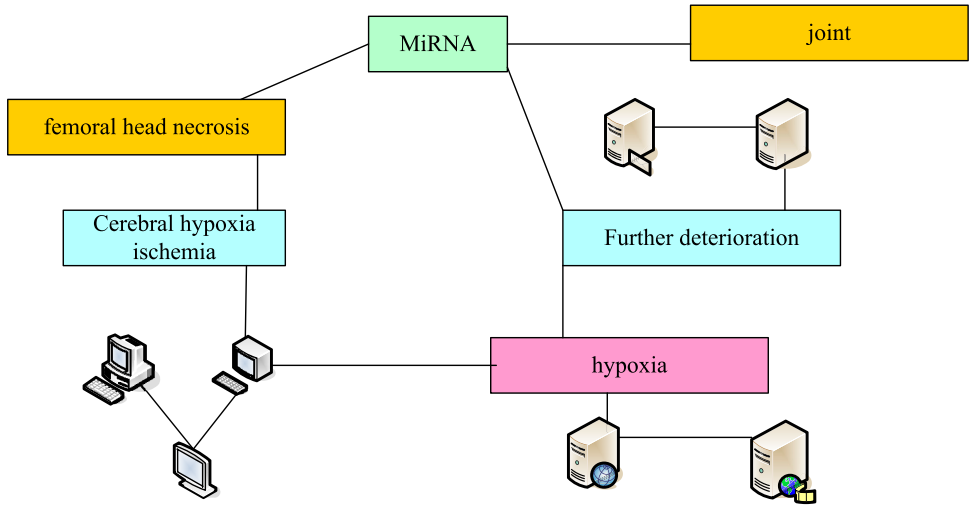


FIGURE 2
Basic idea of clinical experiment.

produced by Shanghai Sangong Biotechnology Co., Ltd. is used to extract the total protein of femoral head tissue, human osteoblasts and renal epithelial cells. The specific extraction process is described below. Extraction process of total protein from femoral head is:

(1) Reagent preparation: Lysis buffer provided in the kit is stored at 4°C. Protease inhibitor, phosphatase inhibitor and benzylsulfonyl fluoride solution are stored at—20°C. Before protein extraction, the relevant reagents are placed on ice for several minutes.
(2) In the precooled Lysis Buffer, phosphatase inhibitor (5 μL/mL), protease inhibitor (5 μL/mL) and benzylsulfonyl fluoride solution (10 μL/mL) are added, fully mixed and evenly placed on ice for several minutes for standby.
(3) The femoral head tissue stored in liquid nitrogen tank is taken out. In liquid nitrogen, all bone tissue blocks are broken and ground into fine powder. After the liquid nitrogen volatilizes completely, the precooled Lysis buffer is added to continue grinding for several minutes to form tissue homogenate.
(4) The tissue homogenate obtained is transferred to a precooled 1.5 mL Eppendorf centrifuge tube, left on ice for 15 min. It is subjected to severe eddy vibration for 3–5 times to fully extract the protein.
(5) After centrifugation at 4°C for 10 min, the supernatant is transferred to a new precooled 1.5 mL Eppendorf centrifuge tube. The obtained sample is the total protein of the tissue. After the protein solubility is determined, it is packaged and stored at—70°C for standby.

Necrosis of the femoral head can cause the increase of C-reactive protein, suggesting that infection: the important clinical significance of the increase of C-reactive protein is to prompt infection, especially bacterial infection, which will lead to a significant increase of C-reactive protein. Virus infection usually has a slight increase or normal level of C-reactive protein, so C-reactive protein can be used as a differential indicator of infection.

2.2.4 HFOB1.19 osteoblast transfection method

First, the cells inoculated with hFOB1. 19 cells are placed on a 24 well cell culture plate. The cells with a confluence of about 50%–60% are washed twice with PBS solution, and 300 μL serum free medium is added to each cell culture hole to continue to culture (5% CO_2, 37°C). The basic culture medium without serum is diluted with Lipofectamine 2000 reagent at a ratio of 50:1 (to ensure the demand of each cell culture hole 50 μL), and diluted with the basic culture medium without serum. The target DNA fragments to be transfected are diluted at an appropriate ratio (to ensure each cell culture hole 50 μL), and left at room temperature for several minutes. Subsequently, the two diluted solutions are mixed by the same volume and left at room temperature for 20 min. Finally, each cell culture well is dripped with the mixed dilute solution obtained from the above 100 μL and gently mixed. It is continuously cultured for about 6 h and replaced with fresh medium for further culture. Cells are collected for subsequent cytological and molecular biological correlation analysis for about 24 h.

2.2.5 Cell activity test method (CCK8 method)

Cell Counting Kit-8 (CCK8) method is used to detect the cell viability of cultured human hFOB1. 19 osteoblasts. The experimental operation process is carried out according to the technical methods provided in the manual. The simple experimental process is as follows. After transfection, hFOBl. 19 osteoblasts are cultured for 24–27 h and used for CCK-8 analysis of cell activity. Cells in each group are collected at three different time points, namely, 24, 48 and 72 h. First, pipettes are used to gently remove old media. Then, according to the ratio of 100:1, the fresh complete culture medium is diluted with CCK-8 reagent. Finally, the cell culture plate n (j) is put back into the incubator to continue to culture for about 1 h, and then detected by the microplate reader for statistical analysis of the difference c_y in cell activity among groups. At least three biological repeats shall be set for the above analysis (Wang et al., 2019).

TABLE 1 Reverse transcription system.

Reaction system	Unit (μL)
RNA samples	1
OligodT primers (18)	1
dNTP mixed solution	0.5
Reverse transcriptase	0.5
DEPC water	2
Reverse transcription buffer	2

$$k_g = g_{FS} - g \tag{13}$$

$$\mathrm{n(j)} = \frac{\Delta \mathrm{l}}{\Delta \mathrm{z} \star \mathrm{n}} \tag{14}$$

$$\mathrm{a_{as}} = \frac{1}{\mathrm{d}(\lambda)} \tag{15}$$

$$\mathrm{c_y} = \mathrm{h} + \frac{\mathrm{g}}{\mathrm{P_{max}} \cdot \mathrm{t_0}} \tag{16}$$

During the miR-206 expression level difference experiment, the femoral head tissue samples of 15 patients with femoral neck fracture who underwent total hip arthroplasty were collected as the control group.

2.3 Statistical treatment

Pearson correlation method (which can measure the correlation between vectors) is used to analyze the correlation between the parameters between groups. In this paper, $p < 0.05$ and $p < 0.01$ are used as the criteria for evaluating statistically significant and extremely significant differences (Wu et al., 2018; Petek et al., 2019).

$$q_D = \frac{2\varepsilon_s}{j \star h} \sqrt{\frac{m_A m_D}{m_A + m_D}} \tag{17}$$

$$\mathrm{f_{bi}} = \ln \frac{\mathrm{kbj_D}}{\mathrm{m_i}} \tag{18}$$

$$\eta = \frac{\mathrm{v}}{\mathrm{ls/m_0}} \tag{19}$$

$$\mathrm{d} = \sum_{i=0}^{n} (\mathrm{g_i}\omega_i + \mathrm{c_i}) \tag{20}$$

3 Results of microRNA-206 induced hypoxic necrosis of femoral head

In this paper, the reverse transcription of mRNA is carried out using EasyScripte reverse transcriptase produced by Beijing Quanshijin Biotechnology Co., Ltd. The extracted RNA sample 1 μg is used as the template, and the reaction system is 7 μL. The prepared reaction system is fully mixed evenly. Ordinary PCR (Polymerase Chain Reaction, which can amplify DNA, making a small amount of DNA become more) instrument is used for cDNA synthesis. The specific reaction procedure is: 70°C, 4 min; 37°C, 55 min; 95°C for 4 min. The synthesized cDNA samples can be used for quantitative PCR analysis or stored at—15°C for future use. The reverse transcription system is shown in Table 1.

At the same time, the femoral head tissue samples of 15 patients with femoral neck fracture undergoing total hip replacement are collected as the control group. Total RNA samples from femoral head tissues of two groups of patients are extracted. The total cDNA samples of each sample are obtained by reverse transcription. The expression level of miR-206 between the experimental group and the control group is compared by real-time fluorescent quantitative PCR. The difference of miR-206 expression level is shown in Figure 3. The expression level of the experimental group was 4.1, and that of the control group was 3.2.

The difference of VEGF/PI3K/AKT signal pathway gene expression level between the two groups of bone tissues by fluorescence quantitative PCR and western blotting is compared (the mRNA level of VEGF/PI3K/AKT signal pathway is shown in Figure 4A. The average data of the experimental group is 1.2, and the average data of the control group is 1.5. The regulation of this gene expression may be an important molecular pathological mechanism for the occurrence and development of steroid induced avascular necrosis of the femoral head (the level of VEGF/PI3K/AKT signal pathway protein is shown in Figure 4B). The mRNA and protein levels of VEGF/PI3K/AKT signaling pathway are shown in Figure 4.

Results are analyzed by real-time fluorescent quantitative PCR and western blotting. In order to verify this, SPSS software (SPSS is the first statistical software in the world to use graphical menu driven interface. Its most prominent feature is that the operation interface is extremely friendly and the output results are beautiful) is used to analyze the correlation between miR-206 and VEGF/PI3K/AKT signal pathway mRNA levels in steroid induced avascular necrosis of the femoral head. There is a significant negative correlation between miR-206 and VEGF/PI3K/AKT signal pathway mRNA levels in femoral tissue of steroid induced avascular necrosis of the femoral head. The correlation coefficient (R value) is about 0.48, which is statistically significant. There may be some interaction between miR-206 and VEGF/PI3K/AKT signaling pathway genes. The interaction between miR-206 and VEGF/PI3K/AKT signal pathway genes is shown in Figure 5.

Based on the previous research on miR-206, and the specific binding ability of miR-206 and VEGF/PI3K/AKT signal pathway genes to start specific regions, two kinds of osteoblasts with different miR-206 expression levels are constructed in human hFOB1. 19 osteoblasts through the Lipofectamine TM2000 transfection system to study the regulatory role of miR-206 in the cellular processes of osteoblast activity, proliferation and apoptosis in detail. At the same time, the downstream regulatory mechanism of miR-206 regulating osteoblast biological function is studied with the above cell materials, especially the possible role of VEGF/PI3K/AKT signaling pathway in the regulation of osteoblast function by miR-206. First, the effect of miR-206 expression level on hFOB1. 19 osteoblast activity is analyzed by Cell Counting Kit-8 (CCK8) experiment. After human hFOB1. 19 osteoblasts are transfected with miR-206 mimics, miR-206 inhibitor and control fragments, the cell viability of hFOB1.19 osteoblasts is detected by CCK-8 at 24, 48 and 72 h, respectively. The expression level of miR-206 in human osteoblasts is negatively correlated with its cell activity. It shows that miR-206 has the effect of inhibiting osteoblast activity (the activity of 24 H osteoblast is shown in Figure 6A, 48H–72H

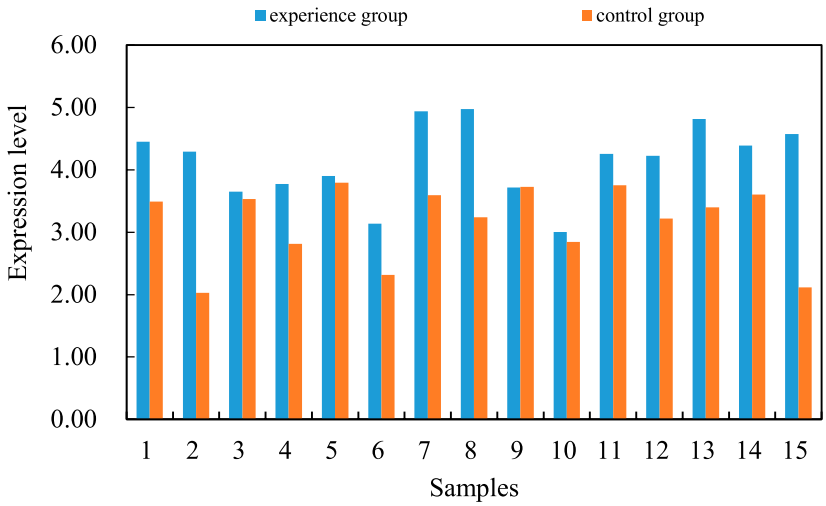


FIGURE 3
Differences in miR-206 expression levels.

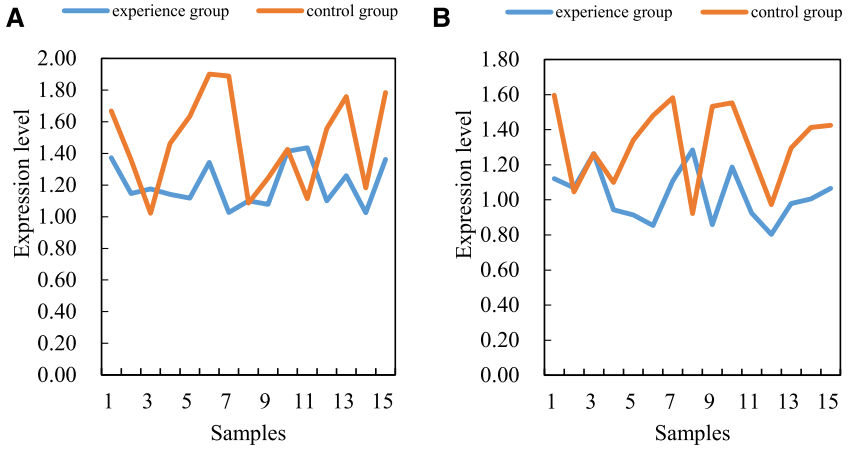


FIGURE 4
mRNA and protein levels of VEGF/PI3K/AKT signaling pathway. **(A)** MRNA **(B)** Protein.

osteoblast activity is shown in Figure 6B. The regulation of miR-206 expression level on osteoblast activity is shown in Figure 6.

The dynamic balance between osteoblasts and osteoclasts is the main mechanism to maintain the normal function of human bones. In order to further understand the cytological mode of miR-206 involved in the pathological process of steroid induced avascular necrosis of the femoral head, the apoptosis of human hFOB1. 19 osteoblasts overexpressed and silenced miR-206 are further analyzed. First, the apoptosis of osteoblasts transfected with MicroRNA-206 mimics and MicroRNA-206 inhibitor is compared and analyzed by Hoechst 33258 staining method. In particular, miR-206 mimics transfected human hFOB1. 19 osteoblasts leads to extremely significant apoptosis. Apoptosis is hardly observed in osteoblasts of miR-206 inhibitor transfection group and normal control group. This result shows that overexpression of miR-206 can promote apoptosis of human hFOB1. 19 osteoblasts. At the same time, the apoptosis of the three groups of human osteoblasts is detected by flow cytometry. After Annexin V and propidium iodide staining, flow cytometry is used to quantitatively analyze the proportion of apoptotic cells in the three groups. The expression level of MicroRNA-206 has a strong promoting effect on osteoblast apoptosis. The regulatory effect of miR-206 expression level on osteoblast apoptosis is shown in Figure 7.

The expression level of VEGF/PI3K/AKT signal pathway gene in osteoblasts transfected with MicroRNA-206 mimics and MicroRNA-206 inhibitor is first analyzed (VEGF/PI3K/AKT signal pathway is shown in Figure 8A). Quantitative RTPCR (reverse transcription polymerase chain reaction) analysis shows that the mRNA level of VEGF/PI3K/AKT signal pathway gene in osteoblasts transfected with MicroRNA-206 mimics is significantly downregulated. On the contrary, the mRNA level of VEGF/PI3K/

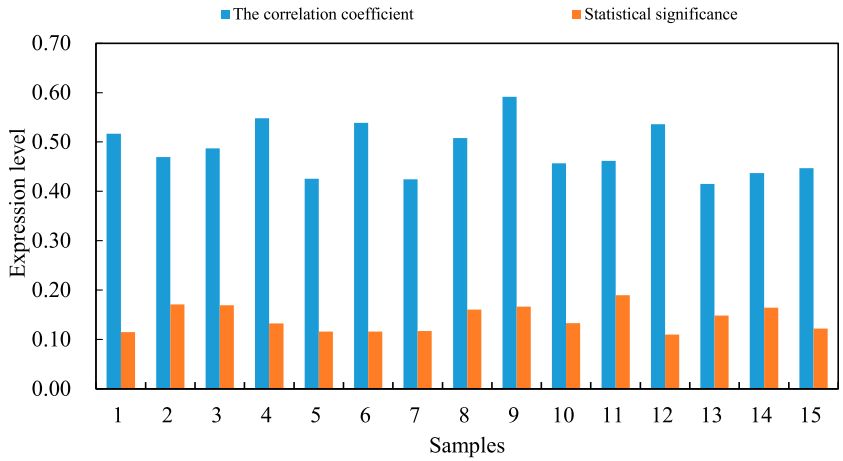


FIGURE 5
Interaction between miR-206 and VEGF/PI3K/AKT signal pathway genes.

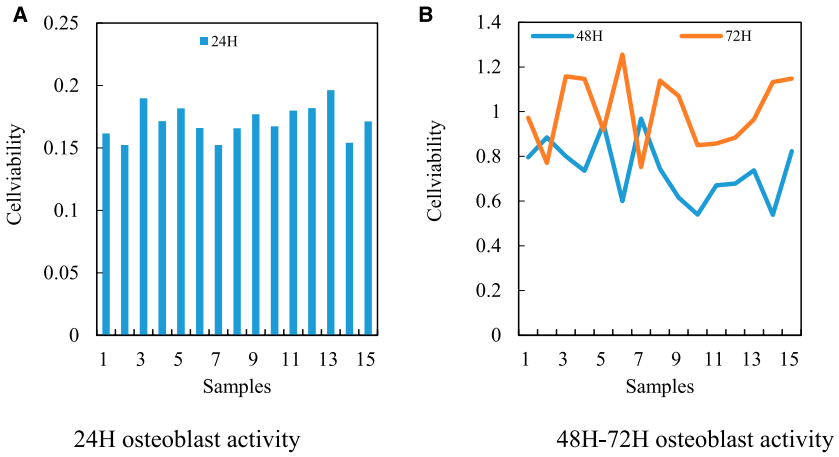


FIGURE 6
Regulation of miR-206 expression level on osteoblast activity. **(A)** 24H osteoblast activity **(B)** 48H–72H osteoblast activity.

AKT signal pathway gene in osteoblasts transfected with MicroRNA-206 inhibitor increases dramatically. These results indicate that MicroRNA-206 has a strong negative regulatory effect on the mRNA level of VEGF/PI3K/AKT signaling pathway genes. Alkaline phosphatase (ALP) is a major marker gene for osteogenic differentiation of bone marrow mesenchymal stem cells. The study also finds that the expression level of this gene in osteoblasts is also significantly regulated by MicroRNA-206, and MicroRNA-206 mimics significantly inhibits the expression of ALP gene. MicroRNA-206 inhibitor significantly promotes the expression level of ALP gene (ALP is shown in Figure 8B). The regulation of hypoxic avascular necrosis of the femoral head is shown in Figure 8.

ALP is marked with green fluorescence. The osteoblasts are intervened with 10–6 mol/L MicroRNA 206 for 0, 12, 24 and 48 h, and then immunofluorescence staining is performed. The

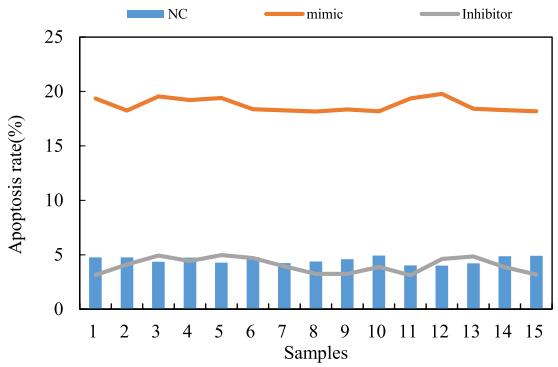


FIGURE 7
Regulation of miR-206 expression level on osteoblast apoptosis.

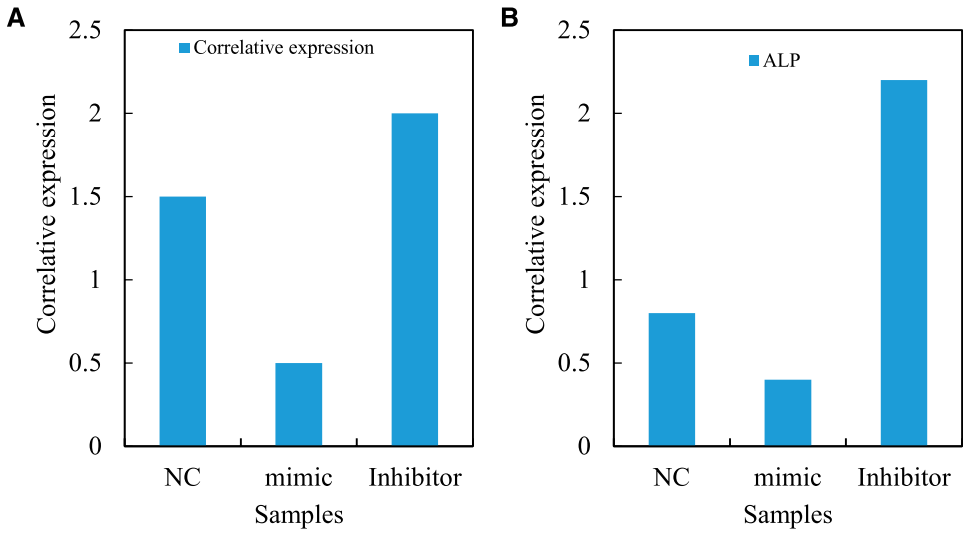


FIGURE 8
Regulation of hypoxic avascular necrosis of femoral head. **(A)** VEGF/PI3K/AKT signal pathway **(B)** ALP.

TABLE 2 Comparison of absorbance.

Time (h)	ALP (mol/L)	*p*
48	8	<0.05
24	11	
12	15	
0	20	

expression of ALP fluorescence decreases significantly at 12, 24 and 48 h. There is only a small amount of immunofluorescence expression in 0 h group. The fluorescence intensity of caspase-3 is higher in the treatment group treated with MicroRNA 206 for 24 and 48 h than that in the control group ($p < 0.05$). MicroRNA-206 can upregulate caspase-3 expression and induce osteoblast apoptosis in a time dependent manner. The absorbance contrast is shown in Table 2.

Then, in the logarithmic growth period, the necrosis cells of the femoral head are taken after hypoxia in good growth condition. The cell density to 2X 104/well is adjusted with the third time medium, and is connected to the 96 well plate. Then the MTT detection method is used. The medium is sucked out and 150 ul DMSO (Dimethyl sulfide) is added to shake for 10 min. The absorbance value OD568 of each hole is measured by microplate reader. Compared with miR-206 + VEGF/PI3K/AKT group, the average cell proliferation rate of VEGF/PI3K/AKT pretreatment group increases significantly ($p < 0.05$). Compared with the control group, the average cell increment rate of the pretreatment group has no significant change. The relative increment rate of cells under different experimental conditions is shown in Table 3.

The chemotherapy drug delivery system has laid the foundation for personalized therapy. In view of the high accuracy of this method, it can determine the optimal dose of drugs for patients, thereby reducing side effects and reducing treatment costs. This paper also considers the targeted drug delivery of the chemotherapy drug delivery system. The first antibody is diluted with sealing solution, so that PVDF (polyvinylidene fluoride) film is immersed in the first antibody incubation solution, and incubated overnight in a 4°C incubator. The dilution concentration of antibody is shown in Table 4.

In the process of regulating the apoptosis, proliferation and differentiation of human osteoblasts by MicroRNA 206, VEGF/PI3K/AKT signaling pathway and other important regulatory factors related to apoptosis and differentiation have undergone significant changes. A specific siRNA (Small interfering RNA) targeting VEGF/PI3K/AKT signaling pathway gene is designed. At the same time, based on the previous research results about the significant increase of gene expression level of VEGF/PI3K/AKT signal pathway in human osteoblasts transfected with MicroRNA 206 inhibitor, MicroRNA 206 inhibitor and VEGF/PI3K/AKT signal pathway are co transfected into human hFOB1. 19 osteoblasts, with a view to building a human osteoblast strain co silenced with MicroRNA 206 and VEGF/PI3K/AKT signal pathway genes. It is used to compare with the osteoblasts transfected with MicroRNA 206 inhibitor, and to evaluate the regulatory role of VEGF/PI3K/AKT signaling pathway in the process of apoptosis of osteoblasts mediated by MicroRNA 206 and its molecular mechanism. Quantitative RT-PCR analysis shows that the mRNA level of VEGF/PI3K/AKT signal pathway gene in hFOB1.19 osteoblasts co transfected with MicroRNA 206 inhibitor and VEGF/PI3K/AKT signal pathway decreases significantly compared with human

TABLE 3 Relative cell growth rate under different experimental conditions.

Group	For the first time	The second time	Third time
Blank	0.07	0.069	0.068
Control	1.4	1.3	1.4
miR-206	1.1	1.2	1.1
VEGF/PI3K/AKT	1.6	1.5	1.7
miR-206 + VEGF/PI3K/AKT	1.3	1.3	1.2

TABLE 4 Diluted concentration of antibody.

Name	Dilution ratio
LC3	1:1,000
Beclin1	1:1,000
P62	1:1,000
AKT	1:1,000
P-AKT	1:1,000

osteoblasts transfected with MicroRNA 206 inhibitor alone. It indicates that this cell line can be used for subsequent research. At the same time, quantitative RT-PCR analysis confirms the expression of ALP gene in osteoblasts transfected with MicroRNA-206 inhibitor. (VEGF/PI3K/AKT signal expression is shown in Figure 9A. The ALP gene in the osteoblasts transfected by MicroRNA-206 inhibitor is shown in Figure 9B. The comparison of regulatory effects of MicroRNA-206 is shown in Figure 9.

The osteoblasts are intervened with 10–8 mol/L, 10–7 mol/L, 10–6 mol/L MicroRNA 206 for 12, 24 and 48 h, and then CCK8 is detected. The results shows that after 12 h of intervention, 10–7 mol/L, 10–6 mol/L MicroRNA-206 can significantly reduce the cell proliferation activity. 10 * mol/LMicroRNA-206 has no significant inhibitory effect on osteoblast proliferation ($p > 0.05$). Compared with 10–8 mol/L group, the proliferation activity of 10–7 mol/L and 10–6 mol/L MicroRNA 206 groups is significantly decreased ($p < 0.05$). Compared with 10–7 mol/L group, the proliferation activity of 10–6 mol/L MicroRNA-206 osteoblasts decreases significantly ($p < 0.05$). It is suggested that MicroRNA 206 inhibits osteoblast proliferation in a dose-dependent manner. At the same concentration of MicroRNA 206, the relative proliferation rate of osteoblasts decreases gradually with the prolongation of the time of MicroRNA 206 intervention for 12, 24 and 48 h. The results shows that MicroRNA 206 inhibits the proliferation of osteoblasts. The longer the action time is, the more obvious the inhibition is. The influence difference of different time and concentration is shown in Table 5. The femoral head gradually healed after necrosis, as shown in Figure 10.

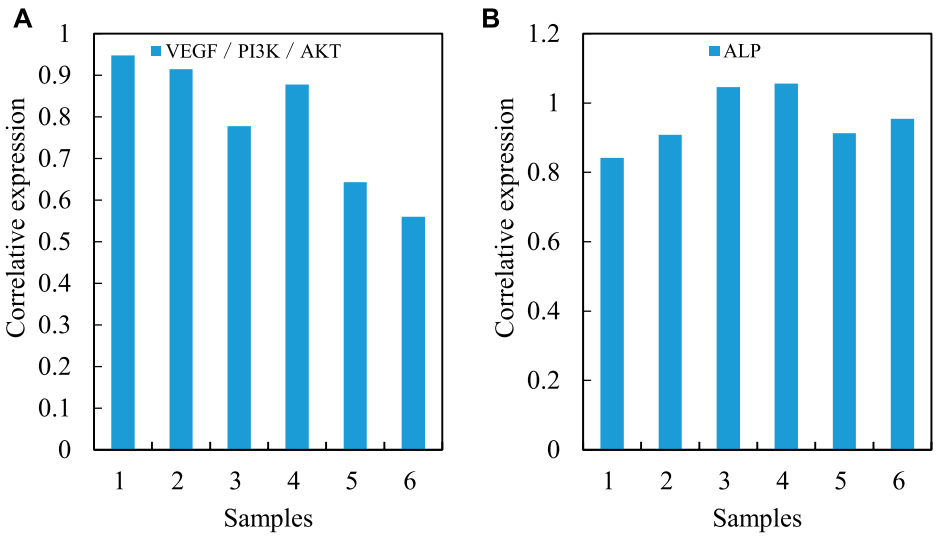


FIGURE 9
Comparison of regulatory effects of MicroRNA-206. **(A)** VEGF/PI3K/AKT signal expression **(B)** MicroRNA-206 inhibitor transfection of ALP gene in osteoblasts.

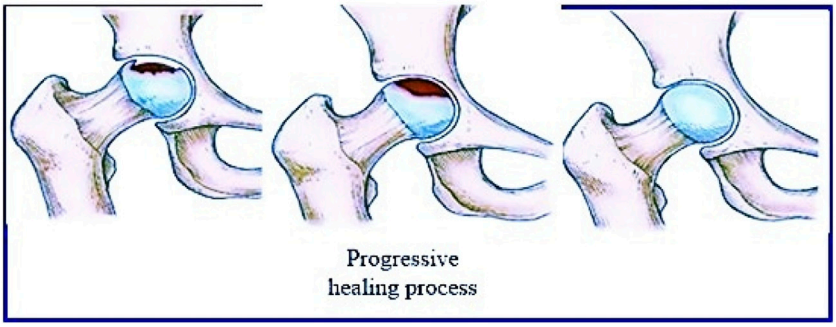


FIGURE 10
The femoral head gradually healed after necrosis (https://xn7p.cn/0f75e).

TABLE 5 Difference of influence of different time and concentration.

Group	12 h	24 h	48 h
10^{-6} mol/L	0.46	0.61	0.76
10^{-7} mol/L	0.48	0.62	0.79
10^{-8} mol/L	0.49	0.72	0.99
Contrast	0.51	0.71	1.12

4 Conclusion

This article has explained the correlation between constitution and NONFH at the molecular level, which is conducive to the play of the unique advantages of the theory of "treating diseases before occuring" in exploring the prevention and revealing the causes of NONFH. It is the direction of continuous development of medicine to explore the mechanism of NONFH blood circulation disorder from the perspective of molecular biology, understand the causes of the disease from a microscopic perspective, and obtain treatment methods. PI3K/AKT can delay and prevent femoral head necrosis by correcting abnormal lipid metabolism, enhancing vascular permeability and promoting vascular regeneration. In this paper, the research on PI3K/AKT promoting vascular regeneration by enhancing VEGF was relatively more complete, and the relevant evidence was also more sufficient. This paper fully discussed the safety and feasibility of the chemotherapy drug delivery system through local tissue absorption of chemotherapy drugs. However, it is still in the laboratory research stage, and the effect of PI3K/AKT on NONFH bone tissue needs to be improved. At the same time, obesity, genetics and environmental factors can also be studied from a microscopic perspective through signal pathways, so as to more clearly understand the mechanism and treatment of NONFH. It is the future direction to know and understand the pathological changes and pathogenesis of NONFH more accurately to help prevent and treat it in advance.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/ participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This study was supported by Science and Technology Project Fund of Wuhu City, Anhui Province, China (2021yf62) and grant KY20260018 from Doctor Research Foundation for Wannan Medical College Yijishan Hospital of China.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Calori, G. M., Mazza, E., Colombo, A., Mazzola, S., and Colombo, M. (2017). Core decompression and biotechnologies in the treatment of avascular necrosis of the femoral head. *EFORT open Rev.* 2 (2), 41–50. doi:10.1302/2058-5241.2.150006

Hua, K-C., Yang, X. G., Feng, J. T., Wang, F., Yang, L., Zhang, H., et al. (2019). The efficacy and safety of core decompression for the treatment of femoral head necrosis: A systematic review and meta-analysis. *J. Orthop. Surg. Res.* 14 (1), 306–314. doi:10.1186/s13018-019-1359-7

Landgraeber, S., Warwas, S., Clasen, T., and Jager, M. (2017). Modifications to advanced core decompression for treatment of avascular necrosis of the femoral head. *BMC Musculoskelet. Disord.* 18 (1), 479. doi:10.1186/s12891-017-1811-y

Leibold, C. S., Schmaranzer, F., Siebenrock, K. A., and Steppacher, S. D. (2020). Femoral osteotomies for the treatment of avascular necrosis of the femoral head. *Oper. Orthop. Traumatol.* 32 (2), 116–126. doi:10.1007/s00064-019-00642-x

Li, W., Ye, Z., Wang, W., Wang, K., Li, L., and Zhao, D. (2017). Clinical effect of hyperbaric oxygen therapy in the treatment of femoral head necrosis: A systematic review and meta-analysis. *Der Orthop.* 46 (5), 440–446. doi:10.1007/s00132-016-3360-8

Luo, P., Gao, F., Han, J., Sun, W., and Li, Z. (2018). The role of autophagy in steroid necrosis of the femoral head: A comprehensive research review. *Int. Orthop.* 42 (7), 1747–1753. doi:10.1007/s00264-018-3994-8

Ma, J., He, W., Zhao, J., Kuang, M., Bai, H., Sun, L., et al. (2017). Bone microarchitecture and biomechanics of the necrotic femoral head. *Sci. Rep.* 7 (1), 13345. doi:10.1038/s41598-017-13643-2

Narayanan, A., Khanchandani, P., Borkar, R. M., Ambati, C. R., Roy, A., Han, X., et al. (2017). Avascular necrosis of femoral head: A metabolomic, biophysical, biochemical, electron microscopic and histopathological characterization. *Sci. Rep.* 7 (1), 10721. doi:10.1038/s41598-017-10817-w

Peng, P., Nie, Z., Sun, F., and Peng, H. (2021). Glucocorticoids induce femoral head necrosis in rats through the ROS/JNK/c-Jun pathway. *FEBS Open bio* 11 (1), 312–321. doi:10.1002/2211-5463.13037

Petek, D., Hannouche, D., and Suva, D. (2019). Osteonecrosis of the femoral head: Pathophysiology and current concepts of treatment. *EFORT open Rev.* 43, 85–97. doi:10.1302/2058-5241.4.180036

Sheng, Z., Xiaoping, H., Lu, D., XiZhe, W., Jie, Z., Qing, L., et al. (2021). Identification of key non-coding RNAs and transcription factors regulators and their potential drugs for steroid-induced femoral head necrosis. *Genomics* 113 (2), 490–496. doi:10.1016/j.ygeno.2020.12.034

Shu, P., Sun, D. L., Shu, Z. X., Tian, S., Pan, Q., Wen, C. J., et al. (2020). Therapeutic applications of genes and gene-engineered mesenchymal stem cells for femoral head necrosis. *Hum. Gene Ther.* 31 (5-6), 286–296. doi:10.1089/hum.2019.306

Wang, T., Teng, S., Zhang, Y., Wang, F., Ding, H., and Guo, L. (2017). Role of mesenchymal stem cells on differentiation in steroid-induced avascular necrosis of the femoral head. *Exp. Ther. Med.* 13 (2), 669–675. doi:10.3892/etm.2016.3991

Wang, Y., Zhu, W., Xiao, K., Li, Z., Ma, Q., Li, W., et al. (2019). Self-healing and injectable hybrid hydrogel for bone regeneration of femoral head necrosis and defect. *Biochem. biophysical Res. Commun.* 508 (1), 25–30. doi:10.1016/j.bbrc.2018.11.097

Wen, P., Zhang, Y., Hao, L., Yue, J., Wang, J., Wang, T., et al. (2020). The effect of the necrotic area on the biomechanics of the femoral head-a finite element study. *BMC Musculoskelet. Disord.* 21 (1), 211–218. doi:10.1186/s12891-020-03242-0

Wu, W., He, W., Wei, Q. S., Chen, Z. Q., Gao, D. W., Chen, P., et al. (2018). Prognostic analysis of different morphology of the necrotic-viable interface in osteonecrosis of the femoral head. *Int. Orthop.* 42 (1), 133–139. doi:10.1007/s00264-017-3679-8

Yan, Y-Q., Pang, Q-J., and Xu, R-J. (2018). Effects of erythropoietin for precaution of steroid-induced femoral head necrosis in rats. *BMC Musculoskelet. Disord.* 19 (1), 282–287. doi:10.1186/s12891-018-2208-2

Zhang, M., Shi, C. Y., Zhou, Z. L., and Hou, J. F. (2017). Bone characteristics, histopathology, and chondrocyte apoptosis in femoral head necrosis induced by glucocorticoid in broilers. *Poult. Sci.* 96 (6), 1609–1614. doi:10.3382/ps/pew466

Zhang, Y., Sun, R., Zhang, L., Feng, L., and Liu, Y. (2017). Effect of blood biochemical factors on nontraumatic necrosis of the femoral head: Logistic regression analysis. *Der Orthop.* 46 (9), 737–743. doi:10.1007/s00132-017-3408-4

Zhu, W., Guo, M., Yang, W., Tang, M., Chen, T., Gan, D., et al. (2020). CD41-deficient exosomes from non-traumatic femoral head necrosis tissues impair osteogenic differentiation and migration of mesenchymal stem cells. *Cell death Dis.* 11 (4), 293. doi:10.1038/s41419-020-2496-y

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 23 February 2023
DOI 10.3389/fgene.2023.1134994

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Defei Liu,
Southwest University, China
Youbao Ma,
Zhejiang University of Technology, China

*CORRESPONDENCE
Yajun Ji,
20200154@ntit.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 31 December 2022
ACCEPTED 08 February 2023
PUBLISHED 23 February 2023

CITATION
Yan L and Ji Y (2023), Correlation of GDFT combined with rehabilitation therapy in DNA damage repair of esophageal cancer cells.
Front. Genet. 14:1134994.
doi: 10.3389/fgene.2023.1134994

Correlation of GDFT combined with rehabilitation therapy in DNA damage repair of esophageal cancer cells

Lihua Yan[1] and Yajun Ji[2]*

[1]Rehabilitation Physiotherapy Department, Central Hospital, Cangzhou, Hebei, China, [2]No. 1 Anesthesia Department, Central Hospital, Cangzhou, Hebei, China

Esophageal cancer is a common malignant tumor with a high incidence and a serious threat to human health. The treatment of esophageal cancer is a complex process, which requires the comprehensive use of a variety of treatment methods. At present, the treatment of esophageal cancer mainly includes surgery, radiotherapy, chemotherapy and immunotherapy. The research on the treatment of cancer cells based on Goal directed fluid therapy (GDFT) combined with rehabilitation therapy is the focus of the current society. This paper proposed a study on DNA damage repair of cancer cells based on goal directed fluid therapy combined with rehabilitation therapy, aiming to optimize the traditional treatment of esophageal cancer by using goal directed fluid therapy technology. The algorithm proposed in this paper was an electroencephalogram (EEG) signal optimization algorithm based on combined rehabilitation therapy. Through this algorithm, the electroencephalogram signal could be optimized. The algorithm could speed up signal processing, and improve signal reliability and stability by reducing the influence of interference signals and improving the signal to noise ratio. These optimization measures could better help researchers analyze and understand electroencephalogram signals, so as to help better study brain functions and diseases. Through the test and investigation on the treatment of cancer cells based on goal directed fluid therapy combined with rehabilitation therapy, the results showed that the blood transfusion volume of goal directed fluid therapy treatment and conventional treatment was 251.5 mL and 288.3 mL respectively. This showed that after goal directed fluid therapy treatment, the input amount of various medical fluids was relatively reduced, and the use of medical fluids was more economical. In addition, their bleeding volumes were 295.2 mL and 324.4 mL, respectively. Urine volume was 382.3 mL and 418.1 mL respectively. This showed that after goal directed fluid therapy treatment, the patient's blood loss and urine volume were relatively reduced, which has improved the patient's health. This experiment has proved the excellent ability of goal directed fluid therapy combined with rehabilitation therapy in the treatment of esophageal cancer, and this research result has also proved the excellent medical effect of goal directed fluid therapy technology. Similarly, this paper also provided valuable reference information for the treatment of esophageal cancer.

KEYWORDS

esophageal cancer, cellular DNA damage repair, goal directed fluid therapy, combined rehabilitation therapy, EEG signal optimization algorithm

1 Introduction

In the medical field, esophageal cancer is a common malignant tumor, which usually occurs on the inner wall of the esophagus. In recent years, great progress has been made in the treatment of esophageal cancer, creating opportunities for improving the survival rate of patients. However, there are still many difficulties in the treatment of esophageal cancer, such as the standardization of the scope of surgical resection and lymph node resection, and the early detection rate of esophageal mucosal and submucosal lesions. Therefore, the treatment of esophageal cancer needs to be further optimized and improved. The research direction of this topic is to use GDFT technology to improve the efficacy of cancer cell therapy. This study not only optimizes the use of medical fluids in esophageal cancer data, but also improves the repair effect of DNA damage, which has made a great contribution to the optimal treatment of cancer. Therefore, with the help of the current advanced GDFT technology, the treatment experience and effect of esophageal cancer can be greatly optimized.

Cancer is a common malignant tumor with high incidence and poor therapeutic effect. Nowadays, many scholars have conducted more scientific research on cancer. Strickler John H pointed out that the "liquid biopsy" method of cell-free DNA (cfDNA) in the blood of cancer patients was increasingly used in clinical practice. His research showed that the frequency of genomic changes detected in cfDNA was comparable to that observed in three independent tissue based colorectal cancer sequencing pharmacopoeias. His analysis also found a group of new extracellular domain mutations that mediated drug resistance by blocking the binding of anti cellular antibodies (Strickler et al., 2018). Cai et al. (2017)'s research pointed out that reversing abnormal DNA methylation and related gene silencing by inhibiting DNA methyltransferases (DNMTs) was an important potential cancer treatment method. He not only defined the leading role of DNA methyltransferase DNMT1, but also defined its different roles in the maintenance of genome wide DNA methylation. Irianto et al. (2017) found that the migration of micron scale narrowing broke the nucleus. By releasing the green fluorescent protein (GFP) located in the nucleus, he observed that for wild type and subclonal U2OS cells, the nuclear damage induced by migration was reversible. In addition to the persistent genomic differences between stable clones revealed by DNA array and sequencing, the acquisition and loss of hundreds of megabases in many chromosomes were typical changes and heterogeneity in bone cancer. Su et al. (2020) believed that accurate cancer cell recognition and effective treatment were very necessary and challenging in clinical practice. He reported the first example of DNA tetrahedral nanostructures (DTNs), real-time monitoring and imaging of three kinds of intracellular miRNAs based on fluorescence "off" to "on" mode, and cancer treatment induced by miRNA silencing. DTNs were self assembled from seven customized single stranded nucleic acid chains containing three target miRNAs recognition sequences. In addition, he studied the correlation between inhibiting migration and invasion of cancer cells and inducing apoptosis of cancer cells, and predicted that the development of intelligent nano platforms would open a door for cancer diagnosis and treatment. The above research topics have studied cancer research from multiple directions, which has provided great help for the research work of this paper. However, the contents of these research papers do not combine cancer research with GDFT technology. Therefore, further research is needed.

Nowadays, with the rapid development of science and technology, a variety of intelligent technologies have been applied to multiple cross cutting research topics. Therefore, GDFT technology can be applied to the treatment research, which provides more advanced research for cancer treatment. At present, many scholars have studied the correlation between liquid therapy technology and cancer. Yin et al. (2021) studied the correlation between GDFT and the treatment of adenocarcinoma and tumor. He evaluated somatic mutations, circulating tumor cells (CTC) and circulating tumor DNA (ctDNA) in patients with pancreatic ductal adenocarcinoma (PDAC) who had pathological complete response (pCR) after adjuvant therapy, and found that they were related to prognosis. In his study, he pointed out the existence of somatic mutation, CTC and ctDNA even in patients with pCR to PDAC, which might predict the early recurrence probability of cancer. Sivakumar et al. (2022) pointed out that GDFT technology and genome mapping had changed breast cancer care. His research showed that cancer usually evolved and evaded therapeutic intervention by acquiring genomic mutations. He examined tissue and fluid biopsy patients as part of routine clinical care to characterize the evolution of the tumor and to identify potential vulnerabilities in the event of recurrence. In general, his research had provided insights into treatment and selection driven tumor evolution, and identified a potential combination treatment scheme for advanced breast cancer. It can be seen from the above research briefs of various scholars that their research results are of great reference value and provide a good direction for this study. However, their research also has some shortcomings. It can be seen that most of their research on knowledge technology theory, which leads to the lack of practicality and authenticity of their research results.

Esophageal cancer is a relatively common malignant tumor, and its incidence rate and mortality increase with age. At present, the treatment of esophageal cancer is a field that has been widely concerned by the society, and its development prospects are good. In addition to surgical treatment, esophageal cancer also has chemotherapy, radiotherapy and other means, but these can not cure the tumor, just to prolong the life of patients, and reduce their pain, thus improving the quality of life. In addition, because the early symptoms of esophageal cancer are not obvious, many patients are not found until the disease develops to the late stage. Therefore, early screening is very important. It is suggested that physical examination should be carried out in time to make early diagnosis, early treatment and control of the disease, so as to avoid serious consequences in the late stage. The innovation of this paper is to apply GDFT technology to the treatment of esophageal cancer, which can quickly detect tumor cells, quickly locate and accurately kill them, thus improving the survival rate and treatment experience of patients.

2 Treatment of esophageal cancer

2.1 Concept and pathology of esophageal cancer

Esophageal cancer refers to the cancer of the esophagus, and refers to the cancerous changes in the tissues of the esophagus. Esophageal cancer is a fatal cancer. It occurs in the esophagus, which is a long tube from the throat to the stomach (Adjiri, 2017; Chang

et al., 2019). Esophageal cancer can affect people's swallowing of food and drink, leading to dysphagia, chest pain and vomiting. Symptoms of esophageal cancer may also include coughing, dry throat, hoarseness, and throat discomfort. The cause of esophageal cancer is not completely clear, but there are some factors that may increase the risk, including smoking, excessive drinking, long-term exposure to smoke or toxins, eating many pickled foods, suffering from Egyptian peduncle (an oral cancer), throat infection or asthma. The treatment of esophageal cancer depends on the extent of cancer spread and health status. Common treatment methods include surgery, chemotherapy, radiotherapy and immunotherapy. Early detection and treatment of esophageal cancer is very important to improve the survival rate. The pathology of esophageal cancer has the following categories. Esophageal cancer can be divided into three types: esophageal adenocarcinoma, esophageal sarcoma and esophageal neuroendocrine carcinoma.

2.1.1 Esophageal adenocarcinoma

Esophageal adenocarcinoma is a common malignant cancer, which usually occurs in the middle and lower part of the esophagus (Sato et al., 2017; Song et al., 2019). It is a cancer caused by gland cells (cells with secretory function) in the esophagus. The early symptoms of esophageal adenocarcinoma are often not obvious. Common symptoms include: throat discomfort, dysphagia, anorexia, weight loss, cough, dry throat, hoarseness, etc. If these symptoms occur, patients must seek medical attention as soon as possible. In order to determine the condition, the doctor may suggest some examinations, including gastroscopy, CT scanning, PET (positron emission tomography) scanning, etc., which can help the doctor understand the condition and decide on the treatment plan. The treatment of esophageal adenocarcinoma depends on the severity of the disease. Common treatment methods include surgery, chemotherapy and radiotherapy. If the condition is mild, only surgery may be required. If the disease is serious, chemotherapy and radiotherapy may be combined. The purpose of the operation is to remove the tumor in order to cure it. Chemotherapy and radiotherapy are carried out after surgery, aiming to eliminate cancer cells left in the body to prevent recurrence. The prognosis of esophageal adenocarcinoma depends on the severity of the disease. Early detection and treatment can improve the survival rate. It is recommended that regular physical examination be carried out to detect and treat esophageal cancer in time.

During the treatment of esophageal adenocarcinoma, patients may need to receive long-term chemotherapy and radiotherapy, which may cause a greater burden on the body. In order to reduce the burden, patients may need to receive nutritional support treatment to maintain health. In addition, during treatment, the patient may have difficulty swallowing, which can lead to malnutrition and weight loss. To solve this problem, the doctor may suggest that patients use enteral nutrition or gastrointestinal infusion, or use swallowing aids. In general, the treatment of esophageal adenocarcinoma is a complex process, which requires comprehensive consideration of many factors. During the treatment, the patient needs to closely cooperate with the doctor and actively cooperate with the treatment. At the same time, they should also pay attention to their physical health and maintain a good psychological state to help them cope with the challenges posed by treatment (Savoli and Bhatt, 2022).

2.1.2 Esophageal sarcoma

Esophageal carcinoma, also known as esophageal sarcoma, is a relatively common malignant tumor, which occurs in the inner wall of the esophagus (Short et al., 2017; Ma et al., 2019). The esophagus is the tube connecting the throat and stomach in the human body, and food is transported to the stomach through the esophagus during eating. When esophageal cancer occurs, the tumor grows in the inner wall of the esophagus and may spread to surrounding tissues and organs. There are many symptoms of esophageal cancer. Common symptoms include sore throat, food blockage, cough, vomiting, anorexia, and significant weight loss. If these symptoms are found, patients should go to the hospital immediately for diagnosis and treatment. There are many ways to treat esophageal cancer, including surgery, chemotherapy, radiotherapy, and immunotherapy. The specific treatment plan depends on the specific situation of the patient and the severity of the tumor. If found and treated in time, the cure rate of esophageal cancer is high. Therefore, if the above symptoms are found, patients should go to the hospital immediately for diagnosis and treatment. In addition to surgery, chemotherapy, radiotherapy and immunotherapy, nutritional support for patients with esophageal cancer is also very important. Since patients with esophageal cancer may suffer from anorexia and significant weight loss, nutritional support should be paid attention to. Nutritional support includes oral nutrition liquid, enteral nutrition liquid, parenteral nutrition liquid, etc. These nutrient solutions are composed of various nutrients, which can help patients supplement nutrition and maintain health.

In addition, for patients with esophageal cancer, smoking and drinking should be avoided as far as possible. Smoking and drinking are both risk factors of esophageal cancer, which can increase the risk of esophageal cancer. Therefore, patients should avoid smoking and drinking as much as possible to reduce the risk of esophageal cancer. In a word, esophageal cancer is a serious disease. If it is found and treated in time, the cure rate is high. Therefore, if there are relevant symptoms, patients should go to the hospital immediately for diagnosis and treatment.

2.1.3 Esophageal neuroendocrine carcinoma

Esophageal neuroendocrine carcinoma is a rare cancer, which occurs in the neuroendocrine cells in the esophagus (Sohda and Kuwano, 2017; Yang et al., 2020). These cells are secreted by hormones, which can control and regulate digestion, endocrine and physiological functions. Esophageal neuroendocrine carcinoma is rare, but when it occurs, it can cause serious problems. The incidence rate of this cancer is very low, but if not treated in time, it may lead to death. Symptoms of esophageal neuroendocrine carcinoma may include chest pain, dyspnea, sore throat, pharyngeal discomfort, dysphagia, neck mass, weight loss, and weight loss. The diagnosis of this cancer is usually made by chest X-ray, computed tomography, magnetic resonance imaging and radionuclide scanning. The treatment of esophageal neuroendocrine carcinoma may include surgery, chemotherapy, radiotherapy and targeted therapy. The choice of these treatments depends on the location, size and expansion of the cancer and the health of the patient. If esophageal neuroendocrine carcinoma is diagnosed and treated early, the survival rate of patients may be higher. Therefore, if one has the above symptoms, please consult doctors for help as soon as possible.

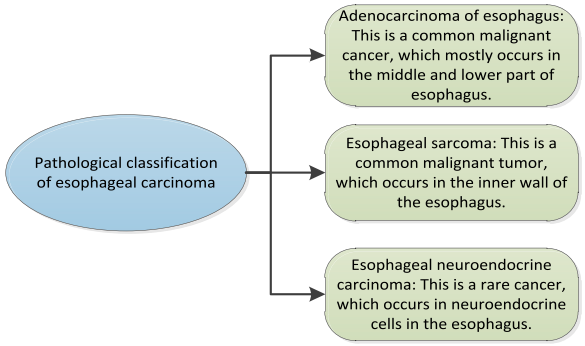


FIGURE 1
Pathological classification of esophageal cancer.

During the treatment of esophageal neuroendocrine carcinoma, patients may need to receive supportive treatment to help them cope with the side effects of treatment and improve their quality of life. These supportive treatments may include nutritional support, pain management, psychological support and lifestyle intervention. It is an important part of the treatment of esophageal neuroendocrine carcinoma to keep close communication with doctors. Patients should discuss treatment options and expected results with doctors, and report any problems or concerns during treatment. In general, esophageal neuroendocrine carcinoma is a rare but serious disease. Early diagnosis and treatment can improve the survival rate and prognosis of patients. It is recommended to consult doctors for more information and treatment options on esophageal neuroendocrine carcinoma.

The pathological classification of esophageal cancer is shown in Figure 1.

2.2 DNA damage

DNA damage refers to the damage to DNA molecules. DNA molecule is the genetic material of cells, responsible for storing and transmitting genetic information. When DNA is damaged, it may lead to the loss or change of genetic information, leading to cell function defects or cell canceration. DNA damage may be caused by many different factors, including chemical damage, physical damage and biological damage. Chemical damage may come from natural or man-made chemical substances. Physical damage may come from radiation or high temperatures. Biological damage may result from bacterial or viral infection. If DNA damage cannot be repaired, the following consequences may result.

2.2.1 Physiological function disorder

Cell proliferation refers to the process of cell division and is the basis of cell growth and organism growth. When cell proliferation decreases, it may lead to genomic instability. Genome refers to the combination of all genes of an organism, which endows the organism with characteristics, morphology and functions. When the genome is unstable, it may lead to changes in genes in cells, which may lead to physiological dysfunction. These physiological disorders may include physical development disorder, immune function disorder, metabolic function disorder, etc. In addition, when cell proliferation decreases, leading to genomic instability and physiological dysfunction, many different symptoms may occur. These symptoms may include weight loss, fatigue, anemia, skin itching, immune dysfunction, etc. If treatment cannot be carried out in time, this situation may lead to more serious consequences, even life threatening (Zhao et al., 2021). In order to avoid this, it is recommended to maintain a healthy lifestyle, including a balanced diet, proper physical exercise and avoiding harmful factors in the natural environment, such as radiation and pollution. If people find any abnormal symptoms, they should go to a doctor as soon as possible so that they can get treatment in time.

2.2.2 Cell death

DNA is a kind of information carrier in organism, which stores the genetic information of organism. DNA damage is very common in cells, which can be caused by many factors, including radiation, toxins and other environmental pressures. When DNA is damaged, it may not be able to perform its function normally, which may lead to abnormal signal transmission. Signal transduction is a kind of communication mechanism in organism, which enables information in cells to be transmitted to cells through signal transduction, so as to coordinate cell activities. When DNA is damaged, it may lead to abnormal signal transmission, which means that information in cells cannot be transmitted normally. This abnormal signal transduction may lead to cell death. Cell death refers to the process in which cells stop living and gradually decompose. Generally speaking, cell death is a normal physiological process. However, if it occurs excessively, it may lead to tissue damage or disease. In the case of abnormal signal transduction caused by DNA damage, cells may die because they cannot receive the necessary signals. This cell death may lead to a decline in tissue function and may, in some cases, lead to disease. On the other hand, if the DNA damage response pathway is overactive or dysfunctional, it may also lead to cell death or other negative results. In general, signal pathways involved in DNA damage response play a crucial role in maintaining genome integrity and ensuring cell survival. The maladjustment of these pathways has serious consequences for the health of organisms.

2.2.3 Abnormal cell proliferation and metabolism

DNA is a macromolecule, which stores the blueprint of biological genetic information. This information is contained in DNA sequences and guides protein synthesis through transcription and transformation. Cells usually have the ability to repair DNA damage. However, if the DNA damage is too much or the repair mechanism is damaged, this may lead to abnormal gene expression, and then lead to abnormal cell proliferation and metabolic abnormalities. When DNA is damaged, these information may be changed or destroyed, which may lead to abnormal gene expression. Abnormal gene expression refers to changes in the level of gene activity, which may lead to abnormal cell proliferation and metabolic abnormalities. For example, if the expression level of a gene is too high or too low, it may lead to abnormal cell proliferation, leading to tumor formation. Similarly, if abnormal gene expression leads to changes in the activity of some enzymes, this may lead to abnormal metabolism and have a negative impact on the health of organisms.

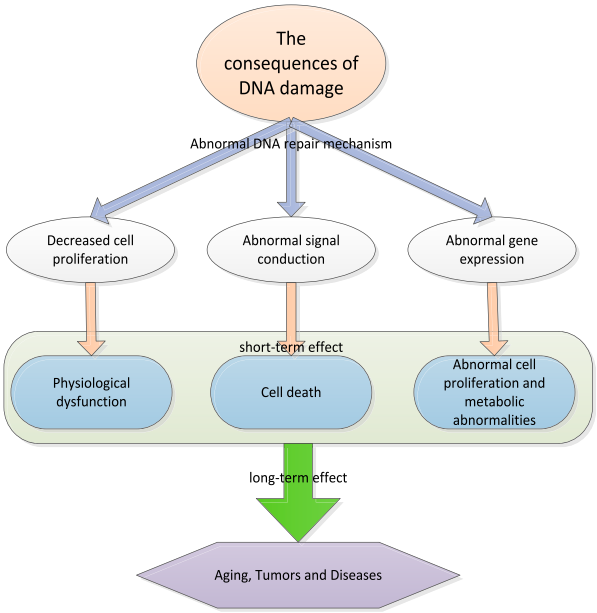


FIGURE 2
Consequences of DNA damage.

In general, DNA damage has a huge impact on human health. As the genetic material of cells, DNA stores the genetic information of all cells and guides how cells divide, grow and function. Finally, in the long run, when DNA is damaged, cells may not function normally, which may lead to a series of problems, including aging, tumors and diseases.

The consequences of DNA damage are shown in Figure 2.

2.3 Cancer treatment and DNA damage repair

Cancer treatment refers to the treatment of patients with cancer, aiming to inhibit or eliminate the growth and reproduction of cancer cells (Kopinski et al., 2021). Cancer is a disease caused by DNA damage. Therefore, one of the ways to cure cancer is to eliminate cancer cells by repairing DNA damage. DNA damage repair means that after DNA damage occurs, the repair mechanism in the cell can repair and restore DNA to its original state. This repair process can be implemented in many ways, such as direct repair, alternative repair, and deletion after repair. These repair mechanisms are completed by special repair enzymes, such as multifunctional repair enzymes and transferases. DNA damage repair is also widely used in cancer treatment. For example, chemotherapy drugs damage DNA. Therefore, in the process of chemotherapy, the DNA damage repair mechanism in cells plays an important role to protect cells from damage. In addition, some drugs can also directly act on the DNA damage repair mechanism to make it more effective. DNA damage repair is an important biological process, which is carried out in cells to repair DNA damage, so as to protect the integrity of the genome and the normal function of cells. In the treatment of cancer, DNA damage repair can start from the following aspects.

2.3.1 Blocking repair pathway

Blocking of repair pathway refers to the process of blocking DNA repair pathway with drugs or other methods. This method allows drugs to kill cancer cells more effectively, because cancer cells usually have higher DNA repair capacity than normal cells. For example, PARP (Poly ADP-ribose Polymerase) inhibitors can block the activity of PARP repair pathway, which is a method commonly used to treat breast cancer and ovarian cancer. PARP is a protein that plays an important role in DNA repair process and is usually used by cancer cells to repair DNA damage. Therefore, the blocking of PARP repair pathway can make drugs more effective in killing cancer cells. Another commonly used method to block the repair pathway is to use TOPO II inhibitors, which is a method used to treat leukemia and tumor. TOPO II is a protein that plays an important role in DNA repair. Therefore, the blocking of TOPO II repair pathway can make the drug more effective in killing cancer cells. In general, the blocking of the repair pathway is an effective way to treat cancer, which can make drugs more effective in killing cancer cells, but may also lead to side effects, so it needs to be used with caution (Arslan and Koyuncu, 2021).

2.3.2 Activating repair path

The activation of repair pathway refers to the process of activating DNA repair pathway by drugs or other methods. This method can help normal cells better repair DNA damage, thereby reducing the risk of cancer. For example, ATM (Ataxia telangiectasia mutated) activator can activate ATM repair pathway, which is a method commonly used to prevent cancer and promote DNA repair. ATM is a protein that plays an important role in DNA repair and is usually used by normal cells to repair DNA damage. Therefore, the activation of ATM repair pathway can help normal cells better repair DNA damage, thereby reducing the risk of cancer. Another common way to activate the repair pathway is to use DNA repair enzymes to directly repair DNA damage, which in some cases can help reduce the risk of cancer. In general, activation of the repair pathway is an effective way to prevent cancer, which can help normal cells better repair DNA damage. However, there are not enough studies to prove its effectiveness in the treatment of cancer.

2.3.3 Taking advantage of weaknesses in the repair approach

The weakness of the repair approach refers to the method of treating cancer by using the loopholes or defects of the repair approach. This method is based on the fact that cancer cells usually have higher DNA repair capacity than normal cells. Therefore, in some cases, the use of weaknesses in the repair pathway can be exploited to make drugs more effective in killing cancer cells. For example, BRCA (breakthrough cancer susceptibility gene) mutation leads to defects in BRCA repair pathway. Therefore, BRCA mutated cancers may be sensitive to PARP inhibitors. PARP inhibitors are drugs that can block the repair pathway of PARP. Therefore, the use of the weakness of BRCA repair pathway can enable PARP inhibitors to kill cancer cells more effectively. Another common way to exploit the weaknesses of the repair pathway is to use DNA repair enzyme mutants. These enzyme mutants have higher efficiency than normal enzyme mutants in repairing DNA damage, so they can be used to treat certain types of cancer. In general, the use of the vulnerability of the repair pathway is an

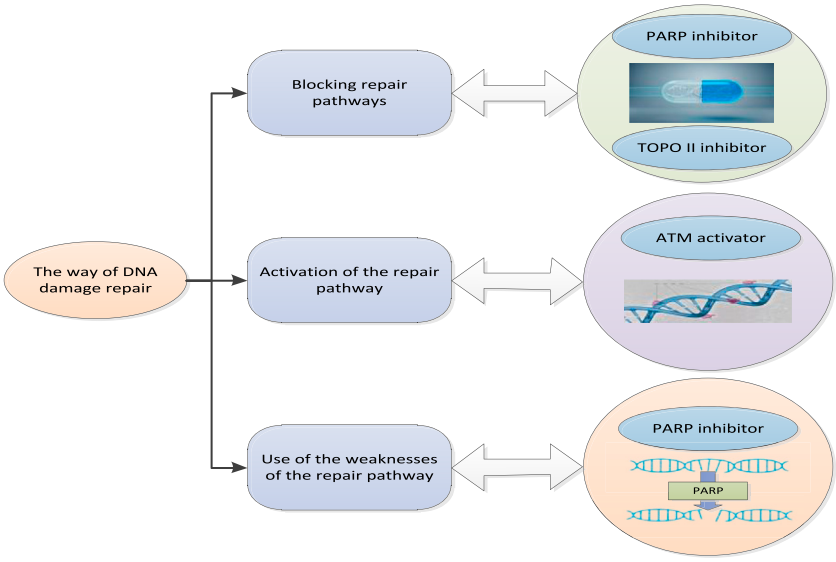


FIGURE 3
Pattern of DNA damage repair.

effective way to treat cancer, but further research is needed to determine its effect in different situations. In general, there is an inseparable relationship between cancer treatment and DNA damage repair, and the effectiveness of DNA damage repair is one of the important factors determining the effect of cancer treatment. Through in-depth study of the mechanism of DNA damage repair, more effective drugs and therapies can be developed to improve the efficiency of cancer treatment. In addition, DNA damage repair may also be related to some other diseases, such as hereditary diseases, senile diseases, etc. Therefore, the research on DNA damage repair is not only helpful to improve the efficiency of cancer treatment, but also may provide new therapeutic methods and ideas.

The DNA damage repair method is shown in Figure 3.

3 GDFT combined with rehabilitation therapy technology

3.1 Target oriented liquid therapy

Targeted liquid therapy is a medical technology, which aims to use liquid or gas to directly inject drugs or other substances into specific target areas of the body to treat certain diseases or restore body functions. This technique is often used to treat diseases in organs or tissues, such as tumors, inflammation or infections.

In target oriented liquid therapy, doctors use imaging technology (such as ultrasound, magnetic resonance imaging or computed tomography) to help determine the location of the treatment target area. Then, the doctor makes a small incision in the skin and uses a guide to inject drugs or other substances into the target area. In some cases, it may be necessary to use multiple imaging techniques to help doctors locate the target area more accurately. Other equipment, such as a liver separator or liver cancer treatment device, may also be needed to help inject drugs into the target area. Targeted liquid therapy may be used to treat various diseases, including tumor, inflammation, and infection. It may also be used to restore body functions, such as improving liver function through injections of drugs or promoting blood circulation through injections. The advantages of goal oriented liquid therapy include accuracy and pertinence. Because the drug is directly injected into the target area, it can reduce the impact on other parts and reduce the risk of side effects. However, this technology also has some limitations. For example, it may not be suitable for some diseases, or it needs to use special equipment for treatment.

In the goal oriented liquid therapy, doctors use certain liquid drugs, such as biological agents with precise formula and use, to directly treat specific diseases of patients. This method can reduce the damage to other healthy tissues and make the drug more effective. In addition, goal oriented liquid therapy mainly guides liquid therapy through "fluid response." Liquid reaction refers to the chemical reaction between drugs and cells, tissues or organs in the human body. These reactions may lead to changes in the mechanism of action, bioavailability or toxicity of the drug. For example, when a drug enters the human body, it may react with some enzymes or proteins to change the mechanism of action of the drug. These liquid reactions can be studied through chemical analysis, biological experiments or pharmacological experiments. In addition, the fluid response is to observe the changes of cardiac output and other indicators after a certain volume of fluid is given to the body. If cardiac output increases, fluid infusion continues, otherwise, it stops.

The Frank Starling curve of liquid reaction is shown in Figure 4.

3.2 EEG signal optimization algorithm based on combined rehabilitation therapy

EEG signal is the physiological signal of human brain, which can reflect human cognitive activity, emotional state, memory, attention

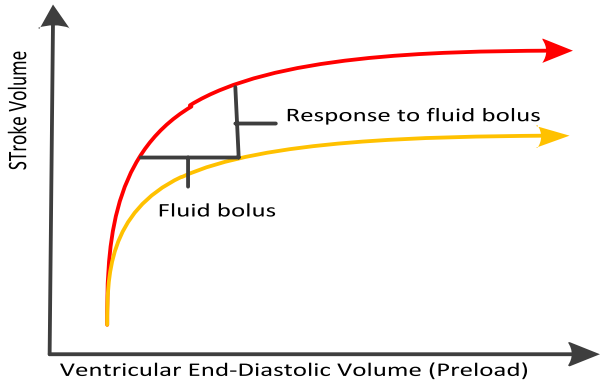


FIGURE 4
Frank starling curve of liquid reaction.

level, and other information. The optimization algorithm of EEG signal usually refers to the method of processing, analyzing and visualizing EEG signal to improve signal quality, and enhance information content and signal interpretability.

In order to ensure the accuracy of EEG signals, it is necessary to denoise EEG signals. Since the signal has been processed by power frequency denoising, low-pass filtering and baseline removal, this section directly removes the adaptive eye artifacts.

First, three channel signals (o_1, o_2, and o_3) are used for image artifact removal. The calculation formula is as follows:

$$a(t) = j(t) + \{o_1(t), o_2(t), o_3(t)\}b \quad (1)$$

Among them, a(t) represents a certain EEG channel signal. j(t) represents the filtered value of this channel. b represents the weight of EEG artifacts on this channel. The filtering model of multi-channel EEG signal can be expressed as:

$$A_{T*M} = J_{T*M} + O_{T*N}*B \quad (2)$$

Among them, EEG signal A and filtered EEG signal J have M channels respectively, and the length of each channel is T. The ophthalmic noise O has N channels, and B represents the channel weight.

Therefore, the filtered target signal can be expressed as:

$$J = A - OB \quad (3)$$

At the same time, the left and right sides calculate the covariance of the noise matrix to obtain the formula:

$$Cov(O, J) = Cov(O, (A - OB)) \quad (4)$$

Since signal J and noise O are independent, their cross covariance is zero. Therefore, the weight B can be expressed as:

$$B = D_{NN}*D_{NY} \quad (5)$$

Among them, D_{NN} is the autocovariance matrix of EEG signal, and D_{NY} is the cross covariance matrix of EEG signal.

Through the above algorithm formulas, EEG signals can be optimized. EEG signal optimization algorithm is a technology used to improve the quality of EEG signal. It improves the accuracy and clarity of EEG signals by reducing the influence of interference signals, improving the signal to noise ratio, speeding up signal processing, improving the reliability and stability of signals, and reducing the noise of EEG signals. These optimization measures can better help researchers analyze and understand EEG signals, so as to help better study brain functions and diseases.

4 Implementation and testing of esophageal cancer treatment based on GDFT combined with rehabilitation therapy technology

4.1 Testing of EEG signal optimization algorithm based on combined rehabilitation therapy

After the EEG signal optimization algorithm based on combined rehabilitation therapy is proposed, it can not only rely on theory, but also need to test the calculation effect of the algorithm in practical use. This experiment tests the traditional EEG signal algorithm and the EEG signal optimization algorithm based on the combined rehabilitation treatment.

First, the measurement accuracy of these two algorithms for common EEG waves need to be counted: Delta, Theta, Alpha, Beta and Gamma. The test results are shown in Table 1.

It can be seen from the test results in Table 1 that the measurement accuracy of EEG signal optimization algorithm based on combined rehabilitation treatment for various EEG signals is higher than that of traditional EEG signal algorithm. Among them, the biggest difference in measurement accuracy is the beta EEG measurement. The measurement accuracy of beta EEG by traditional EEG signal algorithm and EEG signal optimization algorithm based on combined rehabilitation treatment is 88.2% and 94.8% respectively, with a difference of 6.6%.

In addition, it is also necessary to investigate the measurement error rate of these two algorithms in a long running environment. The results are shown in Figure 5.

It can be seen that the measurement error rates of these two algorithms are increasing when running for a long time. However, it can be seen that the measurement error rate of the traditional EEG signal algorithm in each run is greater than that of the EEG signal optimization algorithm based on combined rehabilitation therapy. When the number of runs is 150, the measurement error rates of these two algorithms are 5.74% and 4.72% respectively. Through these tests, it can be shown that the EEG signal optimization algorithm based on combined rehabilitation therapy has better performance.

4.2 Experiment based on GDFT combined with rehabilitation technology in the treatment of esophageal cancer

4.2.1 Experimental direction

The research direction selected in this paper is to test the effect of GDFT combined with rehabilitation technology in the treatment of esophageal cancer, and its therapeutic effect on esophageal cancer patients is explored.

TABLE 1 Measurement accuracy of EEG.

	Traditional brain-wave signal algorithm (%)	Optimization algorithm of brainwave signal based on combined rehabilitation therapy (%)	Differential value (%)
Delta	88.6	94.8	6.2
Theta	89.4	94.1	4.7
Alpha	90.8	95.2	4.4
Beta	88.2	94.8	6.6
Gamma	90.2	95.2	5.0

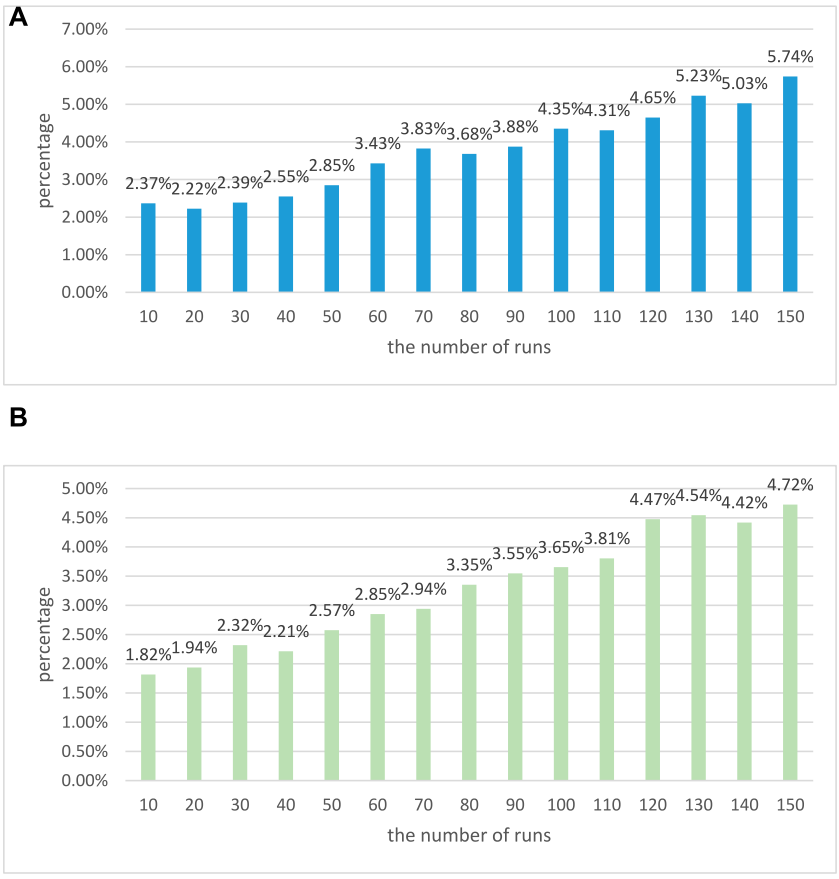


FIGURE 5
Operation error diagram of two algorithms. **(A)** Operation error diagram of traditional EEG signal algorithm. **(B)** Operation error diagram of EEG signal optimization algorithm based on combined rehabilitation therapy.

4.2.2 Experiment content

Adults who are selected for esophageal cancer surgery are randomly divided into two groups. During the operation, the patients are treated with GDFT and conventional therapy, and their physical conditions are observed.

4.2.3 Experimental methods

In this survey, the comparative survey mode is adopted to ensure the scientific and effective experiment.

4.2.4 Experimental results

First, two groups of patients with esophageal cancer are selected and their basic information is collected for statistics, as shown in Table 2.

From the basic information of esophageal cancer in Table 2, it can be seen that the gender ratio, average age, average body mass index and average EEG bispectral index of the GDFT treatment group and the conventional treatment group are very close. This shows that the basic information gap between the two groups of

TABLE 2 Comparison of basic information of two groups of patients before treatment.

	GDFT treatment group	Conventional treatment group
Gender ratio (male/female)	52.8%	53.2%
Average age(year)	52.7	52.5
Average BMI(body mass index) (kg/m^2)	22.32	22.66
Average BIS(Bispectral index)	97.56	97.84

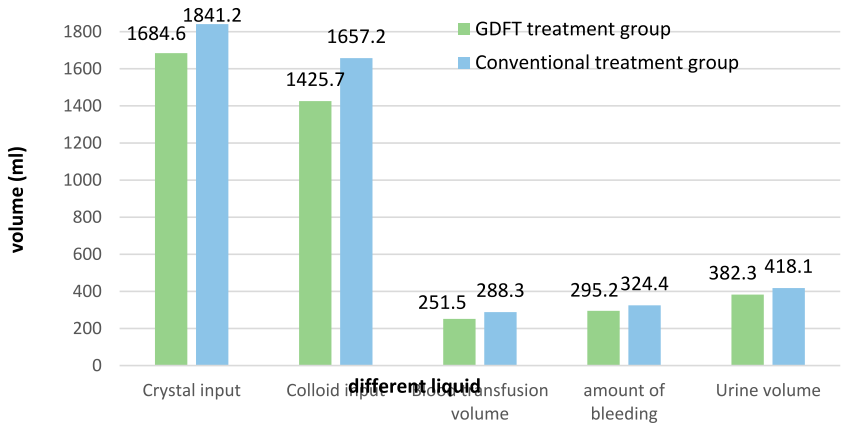


FIGURE 6
Fluid input and output of two groups of patients during treatment.

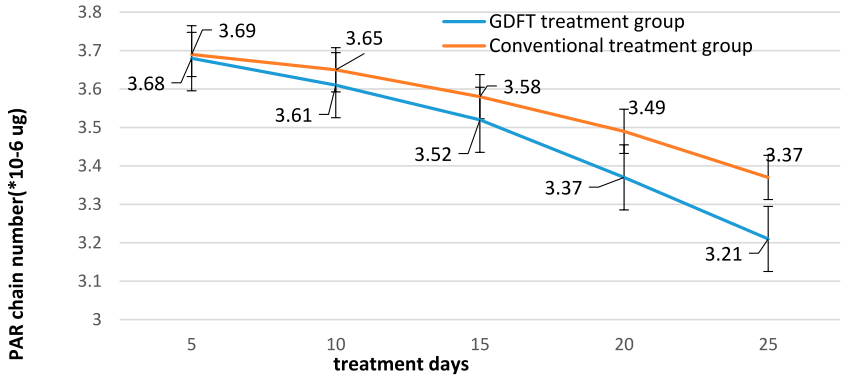


FIGURE 7
DNA damage repair of two groups of patients.

patients with esophageal cancer is small, ensuring the authenticity of the follow-up experiment.

During the investigation, first of all, the liquid intake and output of these two treatment technologies during treatment are investigated. The results are shown in Figure 6.

It can be seen from Figure 6 that in the treatment of esophageal cancer patients, the crystal input volume of GDFT treatment and conventional treatment is 1,684.6 mL and 1841.2 mL respectively, and their colloid input volume is 1,425.7 mL and 1,657.2 mL respectively. In addition, their blood transfusions are 251.5 mL and 288.3 mL, respectively. This shows that after GDFT treatment, the input amount of various medical fluids is relatively reduced, and the use of medical fluids is more economical. In addition, there is a comparison of liquid output. Their bleeding volume is 295.2 mL and 324.4 mL respectively, and their urine volume is 382.3 mL and 418.1 mL respectively. This shows that after GDFT treatment, the patient's blood loss and urine volume are relatively reduced, which improves the patient's health.

After that, the DNA damage repair of the patient needs to be investigated after using the two treatment schemes. poly adp-ribose

polymerase (PARP) can detect single strand breaks in DNA and initiate DNA damage repair reaction. Once single strand breaks are detected, PARP combines with DNA and starts to generate PAR chains as signals for other DNA repair enzymes. The test results are shown in Figure 7.

According to the data in Figure 7, it can be seen that the number of PAR chains in these two groups of patients decreases with the increase of treatment days after the use of GDFT treatment and conventional treatment. Their PAR chains have decreased from 3.68 * 10-6ug and 3.69 * 10-6ug to 3.21 * 10-6ug and 3.37 * 10-6ug, respectively. This shows that these two groups of treatment technologies can effectively treat patients' DNA damage. In addition, it can be seen from the figure that the number of PAR chains in the GDFT treatment group is significantly reduced, which proves that GDFT treatment is more effective in repairing DNA damage.

5 Conclusion

Combined rehabilitation therapy is a comprehensive rehabilitation treatment method that integrates behavior correction, drug therapy and physical therapy to help patients overcome bad behavior habits and improve the quality of life. The combined rehabilitation treatment includes several aspects: exercise, sleep, stress management, diet management and emotion management, which help to improve the quality of life, and eliminate psychological stress to improve social support, so that patients can better cope with the challenges brought by the disease, and eventually recover to a good state of life. The rehabilitation therapy based on GDFT has a very excellent effect in the treatment of esophageal cancer, and its experimental effect is very significant in the real medical environment. In addition, the EEG signal optimization algorithm proposed in this paper based on the combined rehabilitation treatment also had a good effect. It has been proved through experiments that after using the EEG signal optimization algorithm based on combined rehabilitation therapy, the accuracy of various EEG measurements has been effectively improved. According to a variety of experimental results, the combination of GDFT and rehabilitation technology had excellent medical effect in the treatment of esophageal cancer. After using the combined rehabilitation technology based on GDFT, the fluid intake and output of the patients were optimized. In addition, it also had excellent repair effect on DNA damage. In conclusion, this paper clearly discussed the good therapeutic effect of GDFT combined with rehabilitation therapy in esophageal cancer, which has provided a favorable research direction for cancer treatment.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LY and YJ: Writing.

Funding

This work was supported by the comparison of different targeted liquid therapy in thoracic surgery (Subject No. 20211257).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adjiri, A. (2017). DNA Mutations may not be the Cause of Cancer. *Oncol. Ther.* 5 (1), 85–101. doi:10.1007/s40487-017-0047-1

Arslan, E., and Koyuncu, I. (2021). Comparison of amino acid metabolisms in normal prostate (PNT-1A) and cancer cells (PC-3). *Oncologie* 23 (1), 105–117. doi:10.32604/oncologie.2021.014764

Cai, Y., Tsai, H. C., Yen, R. W. C., Zhang, Y. W., Kong, X., Wang, W., et al. (2017). Critical threshold levels of DNA methyltransferase 1 are required to maintain DNA methylation across the genome in human cancer cells. *Genome Res.* 27 (4), 533–544. doi:10.1101/gr.208108.116

Chang, X., Zhang, C., Lv, C., Sun, Y., Zhang, M., Zhao, Y., et al. (2019). Construction of A Multiple-Aptamer-Based DNA Logic Device on Live Cell Membranes via associative toehold activation for accurate cancer cell identification. *J. Am. Chem. Soc.* 141 (32), 12738–12743. doi:10.1021/jacs.9b05470

Irianto, J., Xia, Y., Pfeifer, C. R., Athirasala, A., Ji, J., Alvey, C., et al. (2017). DNA damage follows repair factor depletion and portends genome variation in cancer cells after pore migration. *Curr. Biol.* 27 (2), 210–223. doi:10.1016/j.cub.2016.11.049

Kopinski, P. K., Singh, L. N., Zhang, S., Lott, M. T., and Wallace, D. C. (2021). Mitochondrial DNA variation and cancer. *Nat. Rev. Cancer* 21 (7), 431–445. doi:10.1038/s41568-021-00358-w

Ma, F., Wei, S., and Zhang, C-Y. (2019). Construction of a robust entropy-driven DNA nanomachine for single-molecule detection of rare cancer cells. *Anal. Chem.* 91 (12), 7505–7509. doi:10.1021/acs.analchem.9b01617

Sato, H., Niimi, A., Yasuhara, T., Permata, T. B. M., Hagiwara, Y., Isono, M., et al. (2017). DNA double-strand break repair pathway regulates PD-L1 expression in cancer cells. *Nat. Commun.* 8 (1), 1751. doi:10.1038/s41467-017-01883-9

Savoli, A., and Bhatt, M. (2022). Chronic patients' emotions toward self-managing care IT: The role of health centrality and dependence on IT. *J. Organ. End User Comput.* 34, 14. doi:10.4018/JOEUC.288550

Short, M. W., Burgers, K., and Vincent, F. (2017). Esophageal cancer. *Am. Fam. physician* 95 (1), 22–28.

Sivakumar, S., Jin, D. X., Tukachinsky, H., Murugesan, K., McGregor, K., Danziger, N., et al. (2022). Tissue and liquid biopsy profiling reveal convergent tumor evolution and therapy evasion in breast cancer. *Nat. Commun.* 13 (1), 7495. doi:10.1038/s41467-022-35245-x

Sohda, M., and Kuwano, H. (2017). Current status and future prospects for esophageal cancer treatment. *Ann. Thorac. Cardiovasc. Surg.* 23 (1), 1–11. doi:10.5761/atcs.ra.16-00162

Song, T., Shah, S., Bui, H., Garg, S., Eshra, A., Fu, D., et al. (2019). Programming DNA-based biomolecular reaction networks on cancer cell membranes. *J. Am. Chem. Soc.* 141 (42), 16539–16543. doi:10.1021/jacs.9b05598

Strickler, J. H., Loree, J. M., Ahronian, L. G., Parikh, A. R., Niedzwiecki, D., Pereira, A. A. L., et al. (2018). Genomic landscape of cell-free DNA in patients with colorectal cancer. *Cancer Discov.* 8 (2), 164–173. doi:10.1158/2159-8290.CD-17-1009

Su, J., Wu, F., Xia, H., Wu, Y., and Liu, S. (2020). Accurate cancer cell identification and microrna silencing induced therapy using tailored DNA tetrahedron nanostructures. *Chem. Sci.* 11 (1), 80–86. doi:10.1039/c9sc04823e

Yang, Y-M., Hong, P., Xu, W. W., He, Q. Y., and Li, B. (2020). Advances in targeted therapy for esophageal cancer. *Signal Transduct. Target. Ther.* 5 (1), 229. doi:10.1038/s41392-020-00323-3

Yin, L., Pu, N., Thompson, E., Miao, Y., Wolfgang, C., and Yu, J. (2021). Improved assessment of response status in patients with pancreatic cancer treated with neoadjuvant therapy using somatic mutations and liquid biopsy analysis. *Clin. cancer Res.* 27 (3), 740–748. doi:10.1158/1078-0432.ccr-20-1746

Zhao, P., Xu, Y., Wu, A., Hu, Y., Cai, A., Wang, X., et al. (2021). Circular RNA hsa_circ_0016863 regulates the proliferation, migration, invasion and apoptosis of hepatocellular carcinoma. *Oncologie* 23 (4), 589–601. doi:10.32604/oncologie.2021.018713

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 02 March 2023
DOI 10.3389/fgene.2023.1139412

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Tao Liu,
Zhongnan Hospital, Wuhan University, China
Defei Liu,
Southwest University, China
Zhe Song,
Taylor's University, Malaysia

*CORRESPONDENCE
Shuo Liu,
lxjandcha@163.com

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 07 January 2023
ACCEPTED 15 February 2023
PUBLISHED 02 March 2023

CITATION
Chen H, Liu S, Li X, Wang Z and Zhang C (2023), Prognostic analysis of inflammatory response-related genes and biomarkers in patients with urothelial carcinoma of ureter.
Front. Genet. 14:1139412.
doi: 10.3389/fgene.2023.1139412

Prognostic analysis of inflammatory response-related genes and biomarkers in patients with urothelial carcinoma of ureter

Huaian Chen, Shuo Liu*, Xiujun Li, Zhe Wang and Chao Zhang

First Affiliated Hospital of Hebei North University, Zhangjiakou, Hebei, China

Ureteral urothelial carcinoma is a common urinary system tumor, accounting for 40% to 60% of all ureteral diseases. This study attempted to analyze the prognosis of patients with urothelial carcinoma, judging ureteral urothelial carcinoma by genes and biomarkers of inflammatory response. In this paper, co-expression network analysis and gene-based image fusion evaluation methods were proposed to obtain the prognosis results of patients with ureteral urothelial carcinoma. The experimental results showed that the levels of PLR and NLR increased, and the levels of HGB and HCT decreased; high PLR and high NLR levels, low HGB and low HCT levels were all risk factors affecting bladder urothelial carcinoma, and their ratios (OR) were 1.023, 1.611, 0.961, 0.859, 1.015, 1.072, 0.979, and 0.951, respectively. However, high PLR and high NLR levels were independent risk factors for bladder urothelial carcinoma, and their OR values were 1.497 and 1.071, respectively. Through biomarker diagnosis, the area under the curve, sensitivity, specificity and Youden index of hsa-mir-17, hsa-mir-93, hsa-mir-429 and hsa-mir-20a all exceeded 0.9, indicating that this is a potential diagnostic indicators. All in all, during the treatment of ureteral cancer, in order to reduce tumor recurrence, systemic therapy should be combined with ureteral cancer. In addition, this study also analyzed the prognosis of chemotherapy patients, and the results showed that immunotherapy may increase the risk of tumor cell reperfusion during chemotherapy.

KEYWORDS

patients with urothelial carcinoma of ureter, inflammatory reaction, related genes, prognostic analysis, image fusion

1 Introduction

Epithelial carcinoma of the urinary tract is a cancer that occurs in some epithelium covering the urinary tract. If all covered epithelial tissues are exposed to urine for a long time, it is very likely to be affected by carcinogenic factors in urine to form cancer. Although urothelial carcinoma and cystourethropathy have the same cell origin, the two have completely different biological manifestations and prognosis. Upper tract urothelial carcinoma is a relatively rare heterogeneous cancer, but the incidence of upper tract urothelial carcinoma in China has increased rapidly in recent years, and the number of related reports has gradually increased. In the nearly 30 years from 1973 to 2005, the

overall incidence of upper urinary tract urothelial carcinoma in China has increased steadily, which is mainly related to the increase of ureteral carcinoma. Ureteral carcinoma used to be one of the rarest urological neoplasms, accounting for only about one percent of all urothelial malignancies. In recent years, due to the improvement of detection methods and the growth of people's lives, the incidence rate is also increasing. The prognosis of ureteral cancer is poor, and the five-year survival rate is about 41%–67%. This is mainly related to the thinner wall of the ureter, the richer lymphatic drainage near it, the easy penetration of cancer cells through the muscle layer, and the invasion and movement that can be caused. Although there are various treatment methods for urothelial carcinoma of the ureter, in general, the 10-year survival rate of most patients after diagnosis is still very low, and only about 20%–30% of patients can survive for more than 10 years. Most of these patients have a poor prognosis, and their 5-year survival rate is only 15%. Among tumor-associated inflammatory response pathway genes, about 50% of the data were concentrated in patients with urothelial carcinoma. Therefore, studying and analyzing the expression of inflammation-related genes associated with ureteral carcinoma is of great significance for determining the prognosis of patients. This article analyzes the prognosis of patients with urothelial carcinoma of the ureter through inflammatory response genes and biomarkers in order to make certain contributions to the treatment of urothelial carcinoma of the ureter.

Based on the existing research results, various scholars have carried out relevant research on Urothelial Carcinoma of Ureter: Ureteral fibroepithelial polyp is a rare benign lesion, and it is a good diagnostic index. Bansal Devanshu reported a case of severe hematuria in a 56-year-old woman. After a preliminary examination, she was diagnosed with lower ureteral carcinoma and underwent surgery (Bansal et al., 2022). Small cell carcinoma is a common type of neuroendocrine cancer, mostly found in the lung, accounting for about 20%–30% of all lung cancers. A few are outside the lung, accounting for about 0.1%–0.4% of all small cell carcinomas. Extrapulmonary small cell carcinoma may appear in multiple organs of the body, while small cell carcinoma of the urinary tract is not uncommon in clinical practice. However, sporadic case reports have also been reported in recent years, and primary small cell carcinoma of the ureter is even rarer. Yang et al. (2019) studied the cases of patients with primary ureteral small cell carcinoma after kidney transplantation, and then the other side of the ureter and bladder with primary urothelial carcinoma. The purpose of Alessandro et al. (2020) is to explore the clinical diagnosis and prognostic value of upper tract urothelial carcinoma. A total of 6,619 cases were identified in the surveillance, epidemiology and final result prevalence database, including 3,719 cases of renal pelvis involvement and 2,971 cases of ureteral involvement. Alessandro et al. (2020) evaluated predictors of surgical technique. However, these scholars lacked certain technical demonstrations on the exploration of urothelial ureteral carcinoma. After research, it was found that genes related to inflammatory response have certain research value on the prognosis of urothelial ureteral carcinoma. For this, relevant literature on inflammatory response was consulted.

Some scholars also have some research on inflammatory response: Mori et al. (2021) explored the prognostic role of preoperative systemic immune inflammatory index in patients with upper tract urothelial carcinoma undergoing radical nephroureterectomy. Itami et al. (2019) evaluated the clinical relevance of a comprehensive preoperative assessment of inflammatory, nutritional, and muscle markers in patients with upper tract urothelial carcinoma undergoing therapeutic nephroureterectomy. Chen et al. (2021). evaluated the significance of indicators related to inflammation in predicting the prognosis of ureteral carcinoma. However, these scholars did not analyze the prognosis of patients with Urothelial Carcinoma of Ureter based on inflammatory response-related genes and biomarkers, but only discussed it from a superficial level.

In this paper, the following conclusions are drawn by analyzing the prognosis of patients with Urothelial Carcinoma of Ureter: Ureteral cancer should be combined with systemic therapy to reduce its recurrence rate. On this basis, it is necessary to conduct an in-depth discussion on its related genes and its influence on the recurrence rate and prognosis of ureteral tumors.

The innovations of this paper are as follows: 1) To explore the detection methods of ureteral cancer, study the gene co-expression network analysis and propose an image fusion evaluation method; 2) Based on this, the simulation analysis of the prognosis experiment of patients with urothelial carcinoma of the ureter was carried out.

2 Prognostic analysis method of inflammatory response-related genes in patients with urothelial carcinoma of ureter

2.1 Detection and prognostic analysis of ureteral carcinoma

The preoperative diagnosis of ureteral carcinoma mainly relies on imaging, endoscopy and histopathological examination. At present, there are mainly B-ultrasound, intravenous urography (IVU), retrograde radiography, spiral CT urography (CTU), fluorescence *in situ* hybridization (FISH), ureteroscopy, etc. (Mounsif et al., 2019; Petros et al., 2021). Ureteral tumor refers to the tumor occurring in the ureter, which can be divided into benign tumor and malignant tumor according to its nature. Figure 1 shows the auxiliary examination for ureteral cancer. In the past, the auxiliary examination of ureteral cancer mostly relied on IVU and retrograde angiography, but due to the development of imaging and imaging techniques, such as CTU, urologists can more directly observe ureteral tumors (Wang et al., 2018; Takashi et al., 2020). CTU has advantages that other auxiliary examinations do not have: 1) In coronal and sagittal three-dimensional reconstruction of the urinary tract, it can comprehensively and intuitively display the entire tortuous course of the upper urinary tract, greatly increasing the diagnostic accuracy of renal and urothelial malignant tumors. 2) Compared with IVU, CTU has a diameter of less than 15 mm and a mass that is covered by the superposition of gas. Therefore, the diagnostic sensitivity and accuracy of CTU for ureteral transitional cell carcinoma can be

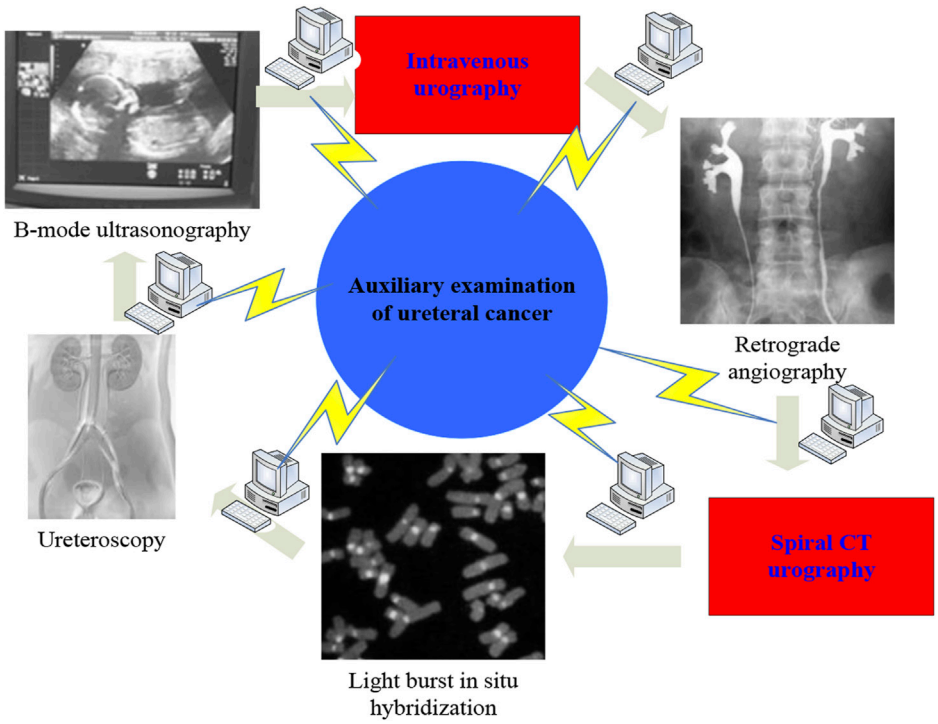


FIGURE 1
Auxiliary examination for ureteral cancer.

as high as 89%–100%. 3) CTU can not only show tumors growing on the surface of the ureter, but also infiltrating and growing tumors, and even lead to stenosis and atresia of the ureteral lumen, which is not available in B-ultrasound and IVU (Wang et al., 2019; Ren et al., 2021). 4) CTU can display fat tissue around the renal pelvis and ureter, enlarged lymph nodes, distant metastasis, etc., and can make a correct diagnosis of clinical staging, laying the foundation for preoperative evaluation. 5) CTU is divided into arterial phase, venous phase, and excretion phase. Compared with other auxiliary examinations, more imaging data can be obtained, which is helpful to improve the accuracy of clinical diagnosis. Spiral CT urography is a continuous volume scan performed during the peak period of contrast agent excretion by injecting contrast agent. The obtained image is processed by computer to obtain a three-dimensional representation of the urinary tract. CTU has become an important means of diagnosing ureteral tumors.

Changes in lifestyle such as smoking and drinking are the main factors affecting the prognosis of patients with ureteral cancer (Chang et al., 2021; Zhou et al., 2022). The survival analysis of patients with ureteral cancer found that the prognosis of smokers was worse than that of women (Sebastiano et al., 2019; Teruo et al., 2020). The study also found that there were certain differences in lifestyle changes of patients with ureteral cancer at different stages. In conclusion, we can see from the influence of the above factors that we should pay attention to the prognostic evaluation of ureteral cancer in clinical practice.

Because the age of the incidence of ureteral cancer is younger in clinical practice, and the location of the incidence is mostly between the bladder and the kidney, early diagnosis and treatment effect are very important (Nir et al., 2020). Clinically, it is necessary to comprehensively analyze and evaluate the tumor stage and pathological classification, the relationship between stage and clinical prognosis, and the prognosis between different stages, so as to obtain the best treatment plan suitable for the patient's condition. In addition, for the elderly patients with ureteral cancer or those with family history and early onset of ureteral cancer, surgery or chemotherapy should be carried out as far as possible to avoid death due to deterioration of the condition. Since the incidence rate of ureteral cancer ranks first among the malignant tumors of the urinary system, it is very important to choose the treatment strategy for this disease.

2.2 Gene co-expression network analysis

Since the data of the network module is easily disturbed by outlier data, it is important to ensure the reliability of the results of the subsequent gene co-expression system network analysis. First of all, the information must be preprocessed to control the quality of the data, so all outlier sample information must be cleared before the establishment of the gene co-expression system. The inter-array correlation (IAC) coefficient plot can be used to assess the spread of chip data, the sample clustering dendrogram, the IAC distribution line graph, and the average IAC scatter point curve to evaluate the degree of data outlier can also be used, and the data that is obviously scattered should be deleted.

WGCNA (Weighted Gene Co-Expression Network Analysis) is an Internet technical analysis package for weighted genome co-expression, which can be used to discover template resources and module hub genomes in chip resources. In this package, concepts such as module, eigenvector gene, module identity, gene significance, module significance, hub gene, etc. are defined. The template refers to a group of genomes with similar expression profiles, which can indicate that their genomes are highly related in various functions. The characteristic vector genome refers to the first principal component genome that can represent the expression levels of all genes in a module. The template identity refers to the similarity between the expression profile of each gene and a characteristic vector gene expression profile. The significance of the genome refers to the relative value of the *p*-value of the significance test of the differential expression of the genome between a genome and different groups of clinicopathological characteristics, which is used to reflect the degree of correlation between the genome and external signals. The significance of the template refers to the average of the gene significance of each gene group involved in the module, which can indicate the high correlation between the module and some clinical application signals. The central genome refers to a group of key genes with the largest connection in a certain model, and it can also be called the most important gene group in a certain model. It determines the important characteristics of the template to a certain extent, and the hub genome in establishing and describing the network model is usually more biologically meaningful.

2.3 Image fusion evaluation

Medical imaging, also known as image acquisition, image conversion, image visualization, etc., refers to the use of image scanners to obtain image data non-invasively, and to provide information such as contrast, brightness, and details of images based on image technology. Current medical imaging acquisition modes include: ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), positron emission computed tomography (PET), single photon emission computed tomography (SPECT), etc.

CT images use attenuation coefficients to measure tissue density, which is an important anatomical data. CT images show that the bone image is clear, so the bones and contrast-enhanced blood vessels can be clearly distinguished. However, its performance on soft tissue lesions is limited due to the high cost of high-density *in vivo* detection, which may adversely affect its molecular imaging. MRI is the rotational imbalance produced by the hydrogen protons in the human body in a strong magnetic field. This technology can accurately display the blood perfusion of different tissues, and can clearly reflect the imaging features of soft tissues. However, its gray level distribution is still non-uniform. At the same time, due to the different types of scanning equipment in different regions, there are certain differences in their intensity values. Both PET and SPECT are imaging techniques for visualizing radioactive material. PET imaging can be performed over a dynamic range, for example, in temporal sequences of two-dimensional images of active regions of the brain. SPECT uses a narrow-span collimator, and its imaging has high spatial resolution, but it lacks in sensitivity compared with PET.

In order to further treat patients with urothelial carcinoma of the ureter, it is necessary to analyze the images of the patient's site. The subjective evaluation refers to the analysis of the advantages and disadvantages of the fusion algorithm, such as image distortion, internal details, etc., using the researcher's reasoning. In terms of image fusion, the subjective evaluation method is simple and reliable. However, due to the influence of the surrounding environment, there are great differences in people's mental state and vision level, so the evaluation accuracy of the image is not high. For this reason, researchers judge the quality of image fusion on the basis of comprehensive evaluation and objective evaluation. Figure 2 shows the application of medical image fusion in Urothelial Carcinoma of Ureter.

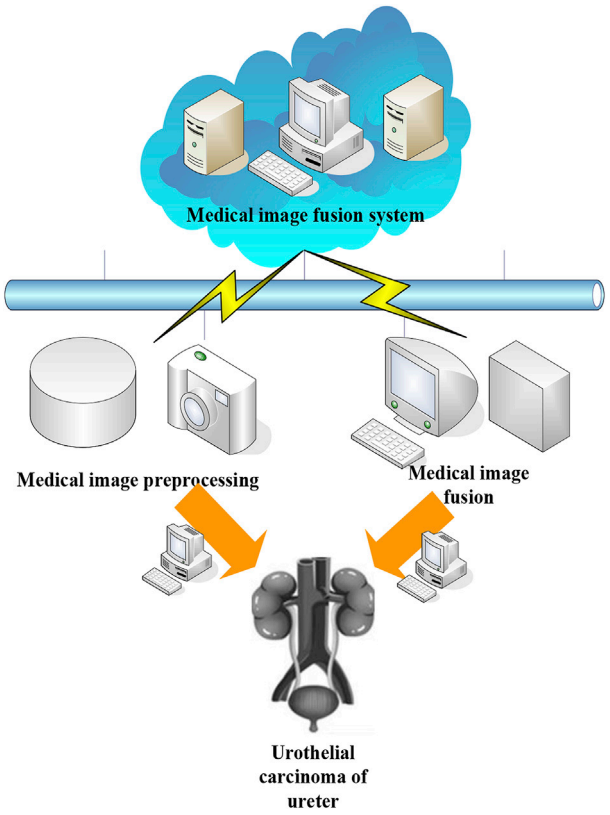


FIGURE 2
Application of medical image fusion in Urothelial Carcinoma of Ureter.

At present, in order to test the clarity and validity of images output by urothelial carcinoma of the ureter, scholars have proposed many indicators of image parameters. For the above evaluation indicators, some representative evaluation indicators can be selected. The public variables included in the following formulas are as follows: O_G is the fused image, and O_T is the reference image, and Z and M are the width and height of the image respectively.

2.3.1 Entropy (EN)

As EN increases, the information in the fused image also increases. EN can be represented by the following formula.

$$EN = -\sum_{o=0}^{A-1} q(o)\log_2 q(o) \quad (1)$$

where o is the gray level and q(o) is the probability associated with o.

TABLE 1 Experimental object data.

Gender	Number of people	Proportion
Male	120	60%
Female	80	40%
Age	Number of people	Proportion
50–59 years old	48	24%
60–69 years old	65	32.5%
Over 70	87	43.5%

2.3.2 Gradient index (W)

The output range of W is between [0,1]. When the output is close to 1, clear image boundary information can be obtained. W can be represented by the following formula.

$$W = \frac{\sum_{o=1}^{Z}\sum_{k=1}^{M}\left(W_{O_1O_G}(o,k)E_{O_1}(o,k) + W_{O_2O_G}(o,k)E_{O_2}(o,k)\right)}{\sum_{o=1}^{Z}\sum_{k=1}^{M}\left(E_{O_1}(o,k) + E_{O_2}(o,k)\right)} \quad (2)$$

where E is a sliding window with a fixed size.

2.3.3 Average gradient (AVG)

The larger the AVG value, the better the quality of the fused image.

$$AVG = \frac{1}{ZM}\sum_{o=1}^{Z}\sum_{k=1}^{M}\sqrt{\frac{\nabla O_{G_c}^2(o,k) + \nabla O_{G_u}^2(o,k)}{2}} \quad (3)$$

where $\nabla O_{G_c}^2(o,k)$ is the first-order difference of O_G in the c direction, and $\nabla O_{G_u}^2$ is the first-order difference of O_G in the u direction.

2.3.4 Mutual information (MI)

MI refers to the further extraction of information such as image details and textures based on the intensity similarity between the O_T image and the O_G image. MI can be represented by the following formula.

$$MI = \sum_{o=1}^{Z}\sum_{k=1}^{M} j_{O_TO_G}(o,k) \times \log_2\left(\frac{j_{O_TO_G}(o,k)}{j_{O_T}(o,k)j_{O_G}(o,k)}\right) \quad (4)$$

where $j_{O_TO_G}$ is the joint gray level histogram of O_T and O_G.

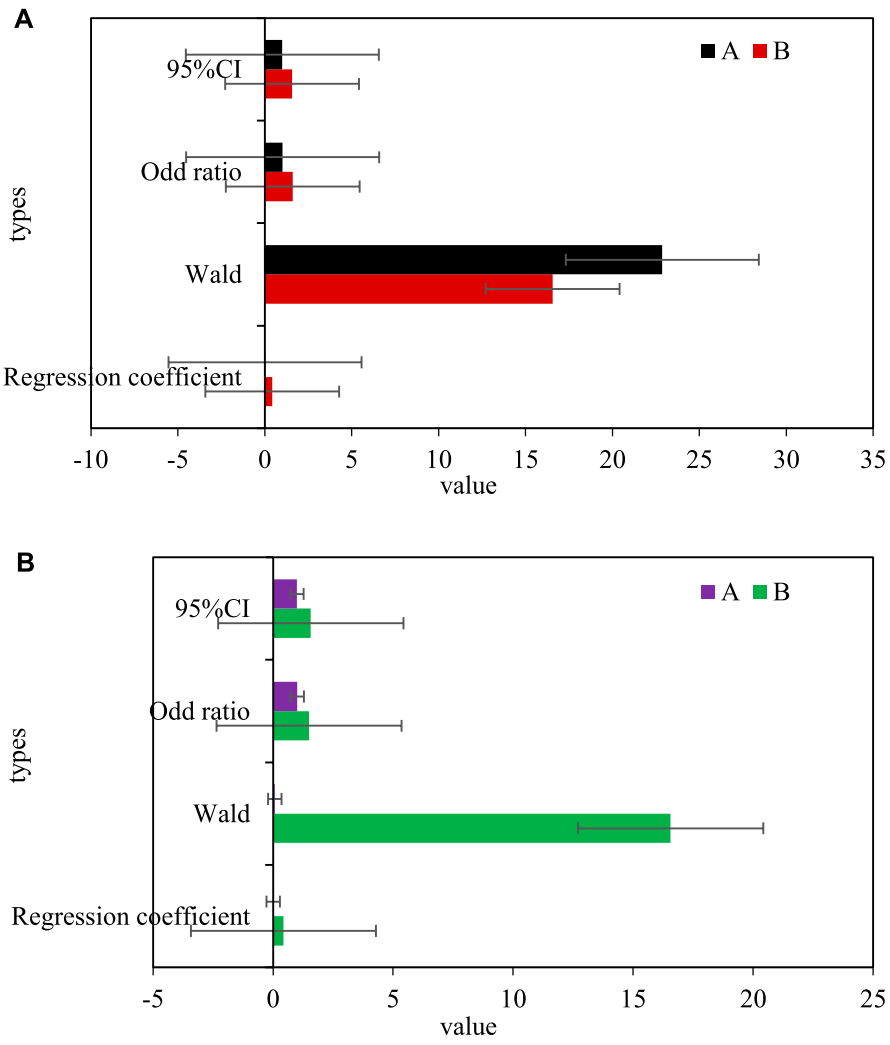


FIGURE 3
(A) Univariate logistic risk factor analysis of inflammatory markers and ureteral urothelial carcinoma. **(B)** Multivariate logistic risk factor analysis of inflammatory indicators and ureteral urothelial carcinoma.

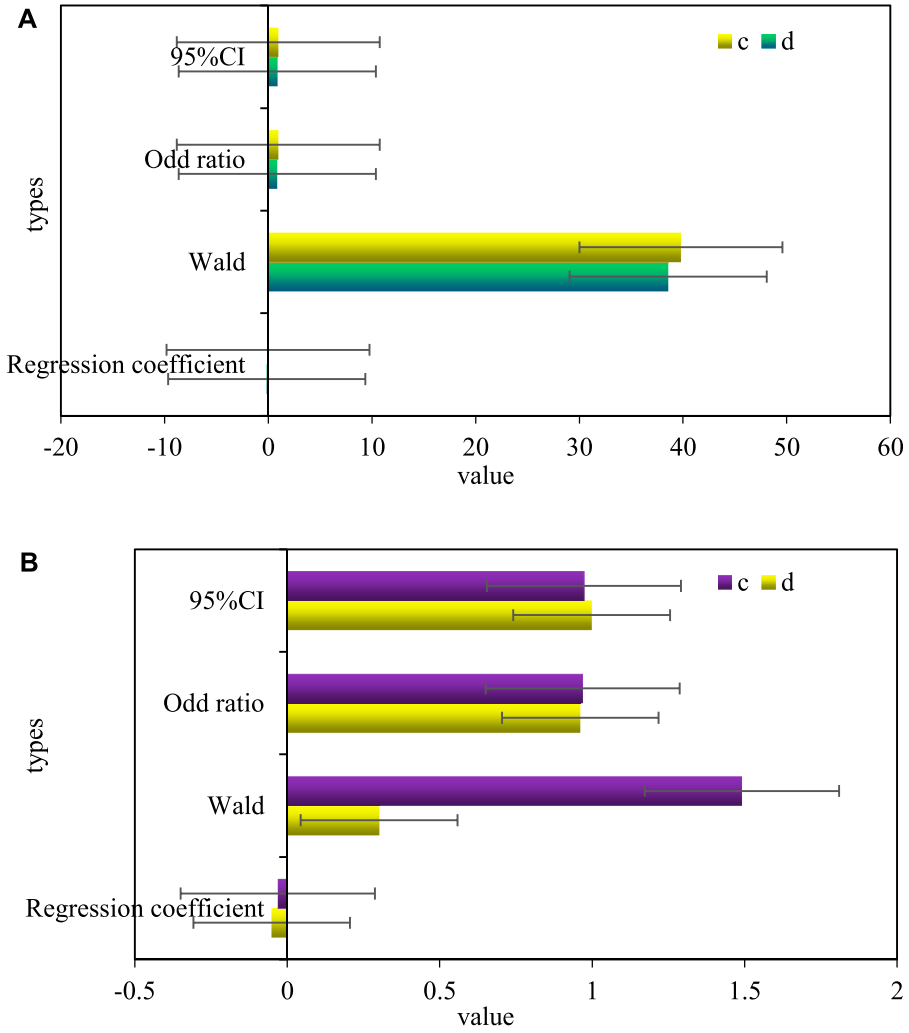


FIGURE 4
(A) Univariate logistic risk factor analysis of hemoglobin index and ureteral urothelial cancer. **(B)** Multivariate logistic risk factor analysis of hemoglobin index and ureteral urothelial carcinoma.

2.3.5 Standard deviation (SD)

The value of SD is proportional to the contrast of the fused image. SD can be represented by the following formula.

$$SD = \sqrt{\frac{1}{ZM}\sum_{o=1}^{Z}\sum_{k=1}^{M}\left(O_G(o,k) - i_{O_G}\right)^2} \quad (5)$$

where, i is the average value of the image.

2.3.6 Spatial frequency (SF)

The SF value is proportional to the overall clarity of the fused image, and is calculated from the row frequency and column frequency. SF can be represented by the following formulas.

$$SF = \sqrt{RF^2 + CF^2} \quad (6)$$

$$RF = \sqrt{\frac{1}{ZM}\sum_{o=1}^{Z}\sum_{k=1}^{M}\left(O_G(o,k) - O_G(o,k-1)\right)^2} \quad (7)$$

$$CF = \sqrt{\frac{1}{ZM}\sum_{o=1}^{Z}\sum_{k=1}^{M}\left(O_G(o,k) - O_G(o-1,k)\right)^2} \quad (8)$$

Therefore, the above indicators can be used to analyze the effect of Urothelial Carcinoma of Ureter images very well.

3 Prognosis results of patients with urothelial carcinoma of ureter

3.1 Experimental materials and methods

This article collected a total of 200 patients with Urothelial Carcinoma of Ureter who were treated in the Department of Urology of Hospital, First Affiliated Hospital of Hebei North University from June 2018 to December 2021 and were hospitalized for surgical treatment and had clear pathological diagnosis after surgery. Among them, 120 were male, accounting for 60% of the total number of cases, and 80 were female patients, accounting for 40% of the total number of cases. The age of onset of patients included in the case data was 38–87 years old, 48 patients (24%) were 50–59 years old, 65 patients (32.5%) were 60–69 years old,

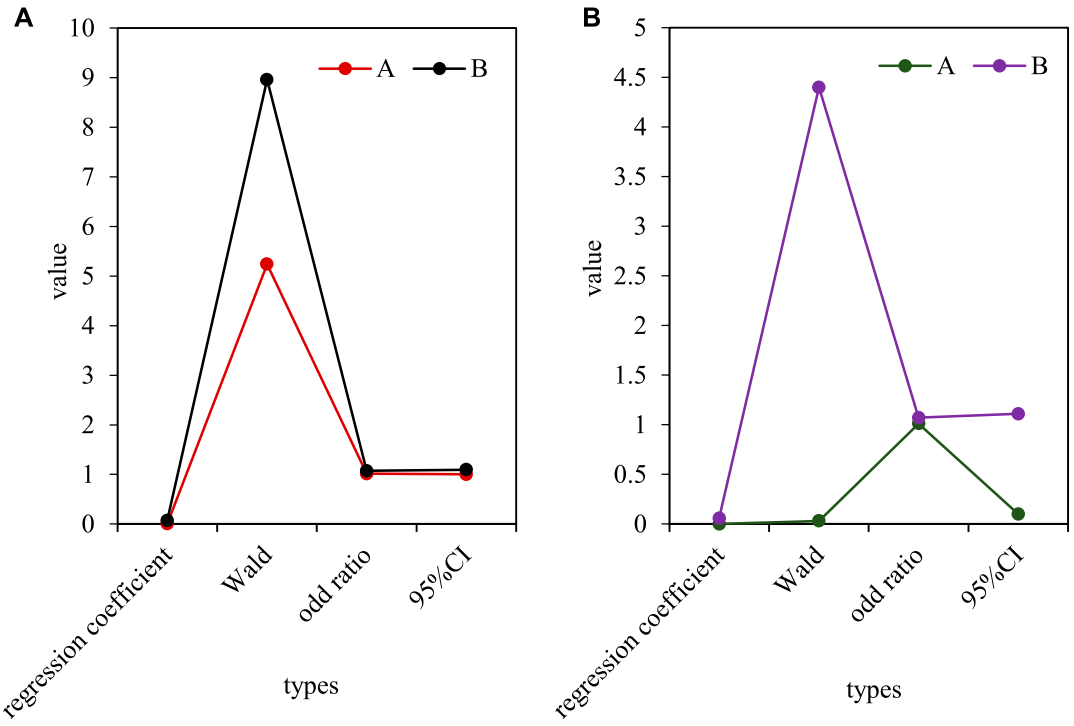


FIGURE 5
(A) Univariate logistic risk factor analysis of inflammatory indicators and pathological grades in ureteral urothelial carcinoma. **(B)** Multivariate logistic risk factor analysis of inflammatory index and pathological grade in ureteral urothelial carcinoma.

and 87 patients (43.5%) were over 70 years old. The results showed that upper urothelial carcinoma increased with age, as shown in Table 1.

This paper studied the changes of platelet-lymphoid ratio (PLR), neutrophil-lymphocyte ratio (NLR), hemoglobin (HGB), and hematocrit (HCT) in patients with bladder urothelial carcinoma before operation, so as to study the clinical significance of relevant indicators. Their impact on bladder urothelial carcinoma, its direct relationship between clinicopathological grade and stage, and its clinical utility were analyzed.

Statistical analysis tools: All statistical analysis uses SPSS19.0 statistical analysis software package. $p < 0.05$ means significant difference, and the difference is statistically significant.

3.2 Inflammatory indicators and risk factors associated with urothelial carcinoma of ureter

Using different dependent variables of serum PLR (A) and NLR (B) levels, the incidence rate of bladder urothelial carcinoma was used as a dependent variable, and the univariate Logistic risk factor study was first used. The results showed that: Increased levels of PLR and NLR are the main risk factors for bladder urothelial carcinoma, with odd ratio (OR) values of 1.023 and 1.611, and 95% confidence intervals (CI) of 1.009 and 1.569, respectively. Then, based on the conclusion of the single-factor Logistic risk factor study, the indicators with statistical value (PLR, NLR) in the single-factor analysis were introduced into the multi-factor Logistic risk factor study, and the risk factors of bladder urothelial carcinoma with independent value was screened out. The results showed that high NLR content was an independent risk factor for bladder urothelial carcinoma, with an OR value of 1.497 and a 95% CI of 1.573. As shown in Figure 3, Figure 3A is the univariate Logistic risk factor analysis of inflammatory indicators and Urothelial Carcinoma of Ureter, and Figure 3B is the multivariate Logistic risk factor analysis of inflammatory indicators and Urothelial Carcinoma of Ureter.

3.3 Hemoglobin index and related risk factors of urothelial carcinoma of ureter

The different dependent variables of HGB (c) and HCT (d) levels in serum and the incidence rate of Urothelial Carcinoma of Ureter were used as dependent variables, and single-factor Logistic risk factors were used for research. The results showed that the decline of HGB and HCT levels were the main risk factors affecting the carcinogenesis of ureteral urothelium, the OR values were 0.961 and 0.859, and the 95% CI were 0.961 and 0.873. Based on the conclusions of the single-factor Logistic risk factor study, the factors with statistical value (HGB, HCT) in the single-factor analysis were introduced into the multi-factor Logistic risk factor study, and the risk factors with independent value that caused ureteral urothelial canceration was screened out. As shown in Figure 4, Figure 4A is a univariate Logistic risk factor analysis of hemoglobin index and Urothelial Carcinoma of Ureter, and Figure 4B is a multivariate Logistic risk factor analysis of hemoglobin index and Urothelial Carcinoma of Ureter.

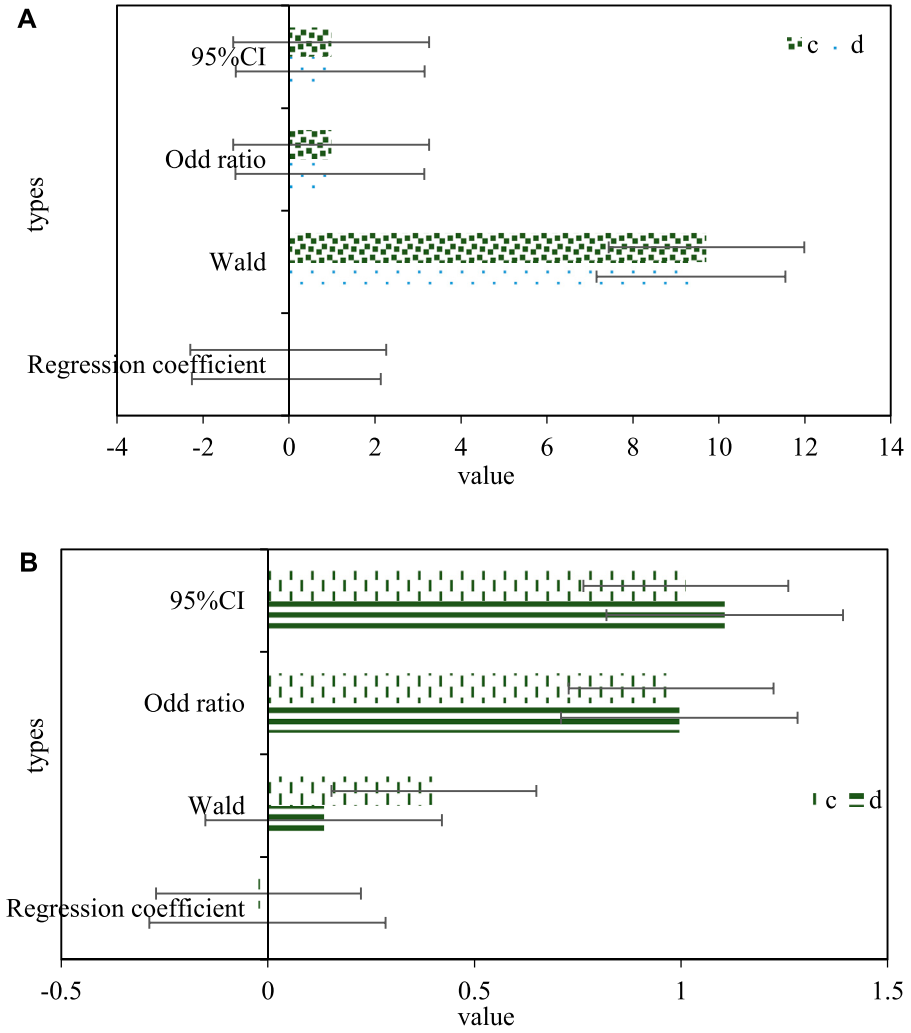


FIGURE 6
(A) Univariate logistic risk factor analysis of hemoglobin index and pathological classification for ureteral urothelial carcinoma. **(B)** Multivariate logistic risk factor analysis method for hemoglobin index and pathological type of ureteral urothelial carcinoma.

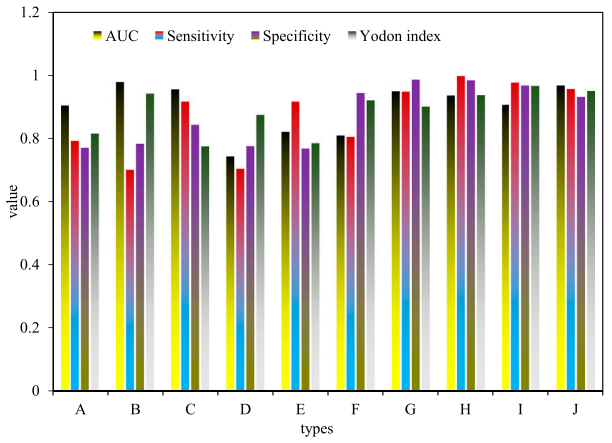


FIGURE 7
Results of ROC curve analysis of expressed genes.

3.4 Risk factors associated with inflammatory indicators and pathological grading of urothelial carcinoma of ureter

Taking the difference in serum PLR (A) and NLR (B) levels as the dependent variable, and the pathological classification of Urothelial Carcinoma of Ureter as the dependent variable, this paper first used a single-factor Logistic risk factor study. The results showed that high PLR and high NLR content were the main risk factors leading to the pathological classification of Urothelial Carcinoma of Ureter, with OR values of 1.015 and 1.072, and 95% CI of 1.003 and 1.096, respectively. Based on the results of single factor Logistic risk factor analysis, the indexes (PLR, NLR) with statistical significance in single factor analysis were included in the multivariate Logistic risk factor analysis. Risk factors with independent meanings that were closely related to the classification of ureteral and urothelial carcinoma lesions were selected. The results showed that high NLR level was an

independent risk factor affecting the risk of pathological classification of bladder urothelial carcinoma, with an OR value of 1.071 and a 95% CI of 1.111. As shown in Figure 5, Figure 5A is a single factor Logistic risk factor analysis of the inflammation index and the pathological grade of Urothelial Carcinoma of Ureter, and Figure 5B shows the multifactor Logistic risk factor analysis of inflammation index and pathological grade of Urothelial Carcinoma of Ureter.

3.5 Risk factors related to hemoglobin index and pathological grade of urothelial carcinoma of ureter

Serum low HGB (c) and low HCT (d) levels were designed as independent variables, and the pathological grade of Urothelial Carcinoma of Ureter was designed as dependent variables. First, through the single factor Logistic risk factor analysis, the results showed that: Low HGB and low HCT levels were risk factors affecting the pathological grade of Urothelial Carcinoma of Ureter, with OR values of 0.979 and 0.951, and 95%CI of 0.983 and 0.956, respectively. Based on the research results of the single factor Logistic risk factor analysis method, the factors (HGB, HCT) with statistical significance in the single factor analysis were introduced into the multifactor Logistic risk factor analysis method. Thus, risk factors with independent meanings that are closely related to the classification of ureteral and urothelial carcinoma lesions were screened out. The results showed that there was no single risk factor in HGB and HCT that reflected the risk of pathological classification of urothelial carcinoma of the ureter. As shown in Figure 6, Figure 6A is a univariate Logistic risk factor analysis of hemoglobin index and pathological classification of Urothelial Carcinoma of Ureter. Figure 6B shows the multifactor Logistic risk factor analysis method of hemoglobin index and pathological classification of urothelial carcinoma of the ureter.

3.6 Evaluation of the diagnostic effect of biomarkers

A better marker for tumor detection, the closer the area under the receiver operating characteristic curve (ROC), sensitivity, specificity, and Youden index are to one, the stronger it is. It is generally considered that the detection ability of the test markers with the area under the ROC curve, sensitivity, specificity and Youden index all exceeding 0.9 is strong, and it is suitable for clinical diagnosis. In this paper, ROC curve research was carried out on ten newly screened expressed genes, and the capacity under the curve (AUC), sensitivity, specificity and Youden index were calculated, and the results are shown in Figure 7. At the same time, taking into account the above technical indicators, it was found that the area under the ROC curve, sensitivity, specificity and Youden index are higher than 0.7, suggesting that it may be of general diagnostic ability as a biomarker for urothelial carcinoma of the ureter, and has certain reference value. The range, sensitivity, specificity and Youden index of Hsa-mIr-17 (G), hsa-mir-93 (H), hsa-mir-429 (I), hsa-mir-20a (J) curves are all higher than 0.9, indicating that its detection function is very good, and it has become a potential detection marker for bladder urothelial carcinoma.

4 Conclusion

Urothelial Carcinoma of Ureter, as the most common malignant tumor in the urinary system, has attracted more and more attention. Its occurrence and development is a complex pathological change process, which is affected by many factors. The molecular mechanism and specific therapeutic targets related to its occurrence and development have always been the hotspots in the field of medical research. With the development of molecular ecology of cancer genes, a series of studies on large-scale proteomics and genomics have shown that the malignant development, invasion, transformation and sociobiological processes of bladder cancer are closely related to the different expressions of mRNA and miRNA. The theoretical research results played a very critical role in finding biomarkers of bladder urothelial carcinoma and important gene molecular targets that directly affect the pathological process of bladder urothelial carcinoma. During the treatment of ureteral cancer, it is necessary to consider the use of targeted and immunotherapy in combination with systemic therapy to reduce cancer recurrence and also need to take into account the prognosis of chemotherapy patients. It showed that the use of immunotherapy during chemotherapy can increase the risk of recurrent infusion of tumor cells when chemotherapy drugs are infused. This paper also found that inflammation-related genes have higher expression levels in patients with RCC, indicating that inflammatory response-related pathways may be an important factor in the development of RCC. In the future, this paper needs to further study the genes involved in the inflammatory response pathway and the significance of predicting the postoperative recurrence rate and prognosis of ureteral cancer. However, due to the limitations of time and technology, this paper has not carried out a detailed study of the problems encountered in the study of ureterourethral epithelial carcinoma, which will be further discussed in the future.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This study was supported by medical science research program of Hebei Health Commission, No.20211529.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alessandro, V., Antonelli, A., Martini, A., Falagario, U., Carrieri, G., Grob, M. B., et al. (2020). Ureteral location is associated with survival outcomes in upper tract urothelial carcinoma: A population-based analysis. *Int. J. Urology* 2711, 966–972. doi:10.1111/iju.14336

Bansal, D., Chaturvedi, S. R. M., Samit, C., and Anant, K. (2022). Giant ureteral fibro-epithelial polyp: A rare but important differential of ureteral urothelial cell carcinoma. *J. Clin. Urology* 156, 515–518. doi:10.1177/2051415820931629

Chang, C-W., Ou, C-H., Yu, C-C., Lo, C-W., Tsai, C-Y., Cheng, P-Y., et al. (2021). Comparative analysis of patients with upper urinary tract urothelial carcinoma in black-foot disease endemic and non-endemic area. *BMC cancer* 211, 80–89. doi:10.1186/s12885-021-07799-4

Chen, H. A., Liu, S., Li, X. J., Wang, Z., Zhang, C., and Li, F. Q. (2021). Clinical value of inflammatory biomarkers in predicting prognosis of patients with ureteral urothelial carcinoma. *Beijing xue xue bao* 2, 302–307. Yi xue ban= Journal of Peking University. Health Sciences 53.

Itami, Y., Miyake, M., Tatsumi, Y., Gotoh, D., Hori, S., Morizawa, Y., et al. (2019). Preoperative predictive factors focused on inflammation-nutrition-and muscle-status in patients with upper urinary tract urothelial carcinoma undergoing nephroureterectomy. *Int. J. Clin. Oncol.* 245, 533–545. doi:10.1007/s10147-018-01381-y

Mori, K., Resch, I., Miura, N., Laukhtina, E., Schuettfort, V. M., Benjamin, P., et al. (2021). Prognostic role of the systemic immune–inflammation index in upper tract urothelial carcinoma treated with radical nephroureterectomy: Results from a large multicenter international collaboration. *Cancer Immunol.* 9, 2641–2650. Immunotherapy 70. doi:10.1007/s00262-021-02884-w

Mounsif, A., Cheriyan, S. K., Foerster, B., Shariat, S. F. S. K. C., Charles, C. P., Beat, F., et al. (2019). Optimal management of upper tract urothelial carcinoma: An unmet need. *Curr. Treat. Options Oncol. 20* 5, 40–11. doi:10.1007/s11864-019-0637-2

Nir, K., F Matin, P. S., Pierorazio, P. M., Gore, P. J. L., Ahmad, S., Hu, B., et al. (2020). Primary chemoablation of low-grade upper tract urothelial carcinoma using UGN-101, a mitomycin-containing reverse thermal gel (OLYMPUS): An open-label, single-arm, phase 3 trial. *lancet Oncol.* 216, 776–785. doi:10.1016/S1470-2045(20)30147-9

Petros, S., Pyrgidis, N., Brookman-May, S., Mykoniatis, I., Karasavvidis, T., and Hatzichristou, D. (2021). Does ureteral stenting increase the risk of metachronous upper tract urothelial carcinoma in patients with bladder tumors? A systematic review and meta-analysis. *J. Urology* 2054 (2021), 956–966. doi:10.1097/JU.0000000000001548

Ren, S., Feng, H., Bao, Y., Yi, W., Ou, Y., Wang, Y., et al. (2021). Ureteral urothelial carcinoma with squamous cell carcinoma and sarcomatoid carcinoma differentiation: A case report. *BMC Surg.* 211, 96–6. doi:10.1186/s12893-021-01099-1

Sebastiano, N., Mazzone, E., Preisser, F., Tian, Z., Mistretta, F. A., Shariat, S. F., et al. (2019). Rates of lymph node invasion and their impact on cancer specific mortality in upper urinary tract urothelial carcinoma. *Eur. J. Surg. Oncol.* 45 (7), 1238–1245. doi:10.1016/j.ejso.2018.12.004

Takashi, Y., Sato, S., Kimura, T., Iwatani, K., Onuma, H., Yanagisawa, T., et al. (2020). HER2 status in molecular subtypes of urothelial carcinoma of the renal pelvis and ureter. *Clin. Genitourin. Cancer* 184, e443–e449. doi:10.1016/j.clgc.2019.12.003

Teruo, I., Matsuyama, H., Komura, K., Ibuki, N., Fujimoto, K., Shiina, H., et al. (2020). Tumor location based segmentation in upper-tract urothelial carcinoma impacts on the urothelial recurrence-free survival: A multi-institutional database study. *Curr. Urol.* 144, 183–190. doi:10.1159/000499240

Wang, J., Wang, G., Shan, H., Wang, X., Wang, C., Zhuang, X., et al. (2019). Gradiently degraded electrospun polyester scaffolds with cytostatic for urothelial carcinoma therapy. *Biomater. Sci.* 3, 963–974. Biomaterials. doi:10.1039/c8bm01317a

Wang, Y., Liu, H., and Wang, P. (2018). Primary sarcomatoid urothelial carcinoma of the ureter: A case report and review of the literature. *World J. Surg. Oncol.* 161, 77. doi:10.1186/s12957-018-1383-9

Yang, X. G., Jin, M. S., Guo, L., Zhu, L., and Qu, L. M. (2019). Homolateral ureter primary small cell carcinoma and urinary tract epithelial carcinoma of the contralateral ureter and bladder after renal transplantation: Report of a case. *Zhonghua bing li xue za zhi= Chin. J. pathology* 489, 735–736. doi:10.3760/cma.j.issn.0529-5807.2019.09.018

Zhou, W., Dang, L., Chong, C., Cheng, X., Qin, Y., and Su, W. (2022). Ureteral tumor with morphological features analogous to phyllodes tumor: A unique case with concomitant urothelial carcinoma. *Diagn. Pathol.* 171, 94–96. doi:10.1186/s13000-022-01277-6

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 March 2023
DOI 10.3389/fgene.2023.1147047

OPEN ACCESS

EDITED BY
Deepak Kumar Jain,
Chongqing University of Posts and Telecommunications, China

REVIEWED BY
Tiefeng Wu,
Qingdao University of Technology, China
Peibo Bao,
Tongji University, China

*CORRESPONDENCE
Qun Li,
Y15010901417@163.com

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 18 January 2023
ACCEPTED 01 March 2023
PUBLISHED 16 March 2023

CITATION
Yin H and Li Q (2023), Characteristics of malignant tumors of digestive system and autoimmune diseases.
Front. Genet. 14:1147047.
doi: 10.3389/fgene.2023.1147047

Characteristics of malignant tumors of digestive system and autoimmune diseases

Hang Yin[1] and Qun Li[2]*

[1]Day Clinic, The Hainan Hospital of Chinese PLA General Hospital, Sanya, Hainan, China, [2]Department of Thoracic Surgery, The Hainan Hospital of Chinese PLA General Hospital, Sanya, Hainan, China

With the development of preoperative examination technology, endoscopic ultrasonography-guided fine-needle aspiration biopsy (EUS-FNA) has been widely used in preoperative pathological diagnosis. However, challenges remain in obtaining appropriate tissue samples and obtaining accurate pathological results to predict disease risk. Therefore, this study aimed to analyze the characteristics of digestive system malignancies and their autoimmune diseases, and to analyze the clinicopathological features, preoperative CT imaging features and pathological grades of pNENs of different pathological degrees of pNENs on the prognosis of pNENs. Experimental results showed that non-functioning pancreatic neuroendocrine tumors showed prominent surrounding hypervascular lesions on multiphase CT examination. Among them, arterial phase and portal venous phase were most clearly imaged at the end, and the degree of local vascular invasion could be used as an index to evaluate its resectability. The sensitivity of CT examination was 63% to 82%, and the specificity was 83% to 100%, depending on the size.

KEYWORDS

digestive system, malignant tumor, characteristics analysis, D-dimer, autoimmune diseases

1 Introduction

Currently, Stroke has been found to be the most common neurological complication in cancer patients other than tumor brain metastases. It has been reported that up to about 15% of cancer patient autopsies show thromboembolic complications, and 50% of these patients have clinical manifestations of stroke during their lifetime. Retrospective studies have shown that malignancy is diagnosed in 6%–12% of patients with ischemic stroke. In actual clinical work, however, these numbers are much smaller because the association between stroke and cancer is often overlooked by clinicians. This is often because cancer-related strokes are difficult to diagnose, especially in patients with no known cancer history or strokes of unknown origin. Patients with cancer may experience a stroke in a different way than others without the disease. It is well accepted that blood hypercoagulability caused by cancer plays a major role in the underlying mechanism of stroke in cancer patients. At present, there are more and more relevant studies, but the sample size of the studies is generally small, and there is still a lack of systematic studies with large sample sizes.

The survival rate of cancer patients is rising, and more cancer patients are also developing strokes, thanks to the quick development of more efficient cancer treatments. Therefore, research into the connection between cancer and acute ischemic stroke is crucial. In patients with malignant tumors complicated by acute ischemic cerebral

infarction, this study conducted a retrospective case-control investigation on the clinical characteristics, vascular risk factors, and imaging characteristics of conventional stroke. This article discusses the related risk factors of CAIS, aiming to improve the clinical understanding of the disease and provide a theoretical basis for clinical diagnosis and treatment.

The innovation point of this study: 1) The CT imaging data (including outpatient and inpatient CT imaging data) of 83 patients were investigated. This article retrospectively analyzed preoperative CT images and pNEN classification. And this paper compared the differences in CT imaging features between G1 and G2. 2) the data of 56 patients were examined and analyzed to obtain CT values of arterial phase, portal venous phase, advanced tumors and normal tissues, as well as CT imaging features of G1pNEN and G2pNEN patients. In this paper, the data of 56 patients were examined and analyzed to obtain the CT values of arterial phase, portal venous phase, advanced tumor and normal tissues, as well as the CT imaging characteristics of G1pNEN and G2pNEN patients. Thus, the difference and ROC curve were compared to analyze the diagnostic value of absolute and relative CT magnification of tumor lesions in distinguishing G1 and G2 patients.

2 Related work

Numerous autoimmune disorders seem to affect women more often, which raises the possibility of a hormonal cause. There are minimal research on tumor estrogen receptor (ER) and progesterone receptor (PR) status, as well as breast cancer risk in systemic and organ-specific autoimmune disorders. Schairer evaluated the relationship between several organ- and system-specific autoimmune disorders, as well as tumor ER and PR status, and overall breast cancer risk in this study. They discovered the first case of breast cancer in a woman under the age of 66 using Linked Surveillance, Epidemiology and End Results (SEER)-Medicare data. They used unconditional and multinomial logistic regression to estimate odds ratios (ORs) and 99.9% confidence intervals (CIs) (Schairer et al., 2018). Immune checkpoint inhibitors (ICIs) are being used more frequently in the metastatic situation or as a supplement to other therapies to treat a wide variety of cancer types. Due to worries about increased toxicity, patients with autoimmune illnesses have been mostly excluded from clinical trials utilizing ICIs. Kennedy et al. (2019) LC reviews current case series describing patients with solid tumors who already have autoimmune disease and examines similarities between clinical autoimmune illness and ICI-induced immunotherapy-related side effects. Which autoimmune patients are likely candidates for ICI therapy is also included in his review (Kennedy et al., 2019). However, no trials have examined its usage, either as a monotherapy or as a combination medication, in the treatment of cancer. According to the Hong J study, Y-320 caused significant harm to MDR tumor cells when it was given concurrently with other chemotherapeutic drugs, although having little effect on multidrug-resistant (MDR) cell lines. Further investigation indicated that P-glycoprotein is a substrate of Y-320 (P-gp). When paclitaxel and Y-320 were administered together, tumor development was dramatically suppressed by 77.1%, as opposed to 6.5% in the group receiving paclitaxel alone (Hong et al., 2020). Numerous autoimmune disorders seem to affect women more often, which raises the possibility of a hormonal cause. Kennedy LC evaluated the relationship between several systemic and organ-specific autoimmune disorders, overall breast cancer risk, and tumor ER and PR status in this study. Despite the fact that these samples were Medicare recipients in good health and free of breast cancer, there was some experimental risk (Kennedy et al., 2019). Immune checkpoint inhibitors (ICIs) are being used more frequently in the metastatic situation or as a supplement to other therapies to treat a wide variety of cancer types. The similarities between clinical autoimmune disorders and the unfavorable effects of ICI-induced immunotherapy are discussed by Radovich et al. (2018). Additionally, he reviews case studies that depict individuals who have solid tumors and pre-existing autoimmune disorders. The prevalence of autoimmune reactions brought on by immune checkpoint blockage during cancer treatment emphasizes the value of interdisciplinary research on autoimmune illnesses and malignancies. In lymphoid malignancies, Shao et al. (2019) points out the connection between particular IL-7R signaling and steroid resistance. Additionally, he showed that after receiving steroid treatment, IL-7R-positive cells that are resistant to steroids may be found in mouse bone marrow, spleen, or in a mouse model of autoimmune arthritis. Everolimus, a mTOR inhibitor, may be used in patients with advanced relapsed and refractory thymic epithelial malignancies, according to newly available research (TET). The identification of patients and the discovery of predictive biomarkers of response are still difficult tasks. Here, Yan et al. (2017) gives a complete genetic characterization of TET patients' thymic tumors and describes a single-center experience with everolimus in those patients. There were still mistakes in the calculations of time to treatment failure and overall survival even though patients had had significant pretreatment with an average of three prior lines of treatment.

3 Medical image enhancement algorithms

3.1 Histogram equalization

The histogram equalization method has a relatively simple theoretical principle, a short running time and the maximum possible preservation of the original image information. It is often used in medical image enhancement (Hellyer et al., 2020). This method remaps the pixel value distribution range of the image to the entire gray level range to enhance the image contrast.

Histogram equalization (HE) expands the gray-level distribution range of the original input image by counting the pixel values, distribute the original image histogram to the global gray level range through the remapping function [the gray level range of the 8-bit image is in (0,255)], and avoids the concentration of pixel values in a certain gray-scale range (AlKindi, 2017; Li et al., 2017). The original input image is defined as X, and its probability density distribution function is shown in Formula (1);

$$p(X_k) = \frac{n_k}{n} \tag{1}$$

$$c_{(x)} = \sum_{i=0}^{k} p(X_i) \tag{2}$$

Among them, $x = X_k$, $k = 0, 1, \ldots, L-1$, it can be deduced that c (L-1) = 1. Histogram equalization is to perform a function transformation operation on the pixel distribution histogram of the image. The process is to stretch and redistribute the histogram distribution of the input image through a certain function transformation (Nozaka et al., 2020; Ohkubo et al., 2020). For an image with a dynamic range of (X_k; X_{L-1}), its mapping function can be defined as:

$$f_{(x)} = X_0 + (X_0 + X_{L-1})c_{(x)} \tag{3}$$

The advantages of the histogram equalization algorithm are mainly reflected in that it regards the histogram of the input image as a whole operation unit, performs global mapping function f(x) transformation on the image pixels, and expands the distribution range of dynamic gray levels (Agra et al., 2020; Ryabov et al., 2020). In addition, according to the information entropy theory, it is easy to know that when the probability distribution of pixel values is uniform, the value of image information entropy is the largest, which makes the enhanced image outline clear and colorful (Zhao et al., 2020).

3.2 Bipartite histogram equalization

The bisection histogram equalization algorithm first calculates the gray mean value of the image, and then divides the histogram of the original image into two parts according to this value. Finally, the two sub-histograms are equalized respectively, so as to better maintain the original brightness of the image. Let X_m be the mean of the image X, from which we know $X_m \in (X_0, X_1, \ldots, X_{L-1})$. Taking the image brightness mean point X_m as the boundary, the histogram is divided into two parts: X_L, and X_U (Lukianchenko et al., 2020). The dynamic range of the pixels in the lower half of the histogram is set to $X_L \in (X_0, X_1, \ldots, X_m)$. The dynamic range of the pixels in the upper half of the histogram is in $X_U \in (X_{m+1}, X_{m+2}, \ldots X_{L-1})$, and their respective probability density distribution functions are shown in Formula (4) and Formula (5), and the expressions are:

$$p_L(X_k) = \frac{n_L^k}{n_L} \tag{4}$$

$$p_U(X_k) = \frac{n_U^k}{n_U} \tag{5}$$

Among them, the value of the density distribution function k of the upper and lower parts is taken as $k = 0, 1, \ldots m$ and $k = m+1, m+2, \ldots L-1$ respectively. n_L represents the sum of pixels in the lower half $n_L = \sum_{k=0}^{m} n_l^k$, and n_U represents the sum of pixels in the upper half $n_U = \sum_{k=m+1}^{L-1} n_u^k$. It can be seen that $n = n_L + n_U$.

According to the conditions, the average brightness of the image enhanced by the He algorithm is equal to the median of the pixel display range (Li et al., 2020; Makhdoomi et al., 2020). Therefore, such methods are not applicable in many fields. The average brightness of the image enhanced by the BBHE algorithm will be limited by the average brightness of the histogram of the high and low parts, thus playing the role of brightness preservation.

3.3 Histogram equalization based on motion blur

The histogram equalization method based on dynamic blur (BPDFHE) can improve the problem of detail loss to a certain extent (Manikandan, 2020). The dynamic blur histogram equalization method mainly uses the process of fuzzy statistics to map the input image to the fuzzy domain through the transformation function. This method effectively solves the roughness of the grayscale range of the image, which is beneficial to improve the performance of the algorithm.

First, the input image histogram needs to be transformed into a blur histogram through a function. A fuzzy histogram is a sequence of real numbers consisting of $h(X)\ X \in \{0, 1, \ldots, L-1\}$, where h(X) represents the distribution probability of gray levels in the neighborhood of X. $\tilde{I}$ (i,j) is set as the grayscale blur value of I (i,j), and the blur histogram is shown in Formula (6):

$$h(X) \leftarrow h(X) + \sum_i \sum_j \mu_{(i,j)X}, k \in [a, b] \tag{6}$$

Among them, $\mu_{I(i,j)i}$ is a triangular fuzzy membership function. Next, take the local maximum value of the histogram as the limit point to obtain a plurality of sub-histograms. Denote the first derivative and second derivative of the fuzzy histogram as h′(X), h"X), respectively, and their derivative expressions are:

$$h'(X) = \frac{dh(X)}{dX} = \frac{h(X+1) - h(X-1)}{2} \tag{7}$$

$$h''(X) = \frac{d^2h(X)}{dX^2} = h(X+1) - 2h(X) + h(X-1) \tag{8}$$

The second derivative h”(X) directly uses the second central difference operator to calculate the second derivative of the fuzzy histogram. The purpose of this is to minimize the approximation error accumulated from the first derivative calculation. After the histogram is divided into multiple sub-histograms, each sub-histogram is remapped to the gray level dynamic range, and finally the brightness of the image is normalized.

$$\frac{E}{S} = T\left(\frac{p2}{p1}\right) * T\left(\frac{p3}{p2}\right) *, , , * T\left(\frac{pn}{pn-1}\right) \tag{9}$$

$$\log\left(\frac{E}{S}\right) = \log\left[T\left(\frac{p2}{p1}\right)\right] + \log\left(T\left(\frac{p3}{p2}\right)\right) +, , , + \log\left(T\left(\frac{pn}{p_{n-1}}\right)\right) = \sum_{i=2}^{n} \log\left[\left(T\left(\frac{pi}{p_{i-1}}\right)\right)\right] \tag{10}$$

$$\log\left(\frac{E}{S}\right) = \sum_{i=2}^{n} \log\left[\left(T'\left(\frac{pi}{p_{i-1}}\right)\right)\right] = \sum_{i=2}^{n} T'(\log p_i - \log p_{i-1}) \tag{11}$$

$$D = 2^{\log 2 \min (rows, cols) - 1)} \tag{12}$$

After one comparison, rotate 90° clockwise, the distance is halved, convert Maiko D = -D/2 and continue the comparison until |p|<1. The iterative Formula is as follows:

$$r_{n+1}(x, y) = \frac{r_n(x, y) + r_n'(x, y)}{2} \tag{13}$$

$$r_n'(x,y) = \begin{cases} r_n(x,y) + \Delta L & r_n(x,y) + \Delta L \le \max \\ \max & r_n(x,y) + \Delta L \le \max \end{cases} \quad (14)$$

The single-scale Retinex algorithm expression is

$$\log R(x,y) = \log I(x,y) - \log(F(x,y) * I(x,y)) \quad (15)$$

$$F(x,y) = K \cdot e^{\frac{-(x^2+y^2)}{\sigma^2}} \quad (16)$$

$$\iint F(x,y)dxdy = 1 \quad (17)$$

4 Experimental characteristics

Through previous studies, we found that in the comparison of the pathological characteristics of GI grade and G2 grade pNENs, G2 grade pNENs and patients' tumors were larger in maximum diameter, more prone to invasive growth, more prone to lymph node metastasis, and more prone to distant metastasis. Preoperative invasive EUS + FNA examination cannot obtain 100% accurate pathological results. So whether these differences can be reflected by simple and easy CT imaging examination, we combined CT, the most commonly used preoperative examination method in 266 patients, to screen out 83 cases of G1 and G2 grades with CT images. The CT imaging features were used to analyze the differences in imaging features between G1 and G2 patients, and to observe whether the pathological grade of the patients could be judged by CT imaging. Through the further quantified CT indexes of 56 patients with pancreatic neuroendocrine tumors, we measured the CT values of arterial phase, delayed phase, portal venous phase, plain scan tumor and normal tissue. To analyze the significance of CT imaging features on the pathological grading of pancreatic neuroendocrine tumors, so as to predict its possible prognosis and then adopt more appropriate surgical methods.

4.1 Materials and methods

4.1.1 Research objects

Inclusion criteria for patients with acute cerebral infarction with active cancer (cancer group, n = 28) were as follows. 1) Acute ischemic cerebral infarction is defined as sudden focal or generalized neurological damage that persists for >24 h and has imaging-related symptoms, or symptoms of neurological damage <24 h. However, the patient's acute ischemic cerebral infarction was verified by MRI diffusion-weighted imaging (DWI). 2) Active cancer was defined as a malignant tumor that had been diagnosed or untreated within 1 year before the stroke, as confirmed by the patient's previous medical records. Either a previously inactive malignancy that recurred within 1 year before the stroke, or a newly diagnosed malignancy during the stroke hospitalization. The diagnosis of tumor is confirmed by tissue or liquid pathological biopsy. Inclusion standards for the cancer group were as follows: 1) Patients without diffusion-weighted imaging. 2) Patients with clinically diagnosed tumors but not confirmed by pathological biopsy; 3) Patients with benign tumors, secondary intracranial metastatic tumors and primary intracranial tumors. 4) Inactive cancer is defined as a patient with a history of cancer who has undergone surgery more than 5 years prior to the onset of this acute stroke and has no symptoms or signs of recurrence. 5) Patients with cerebral hemorrhage, encephalitis, multiple sclerosis or other central nervous system diseases. 6) Patients with severe heart, liver, lung and renal insufficiency.

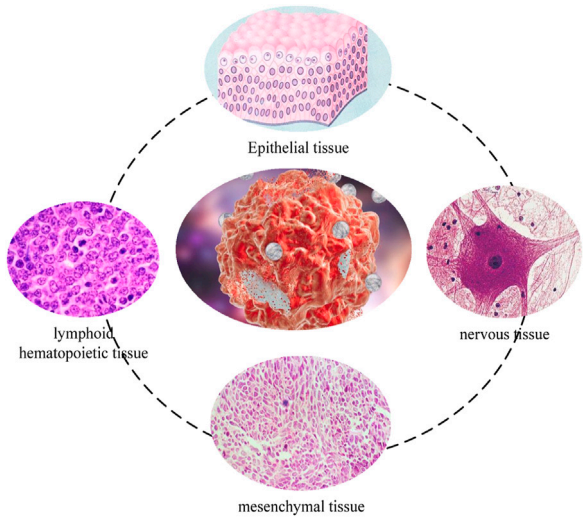


FIGURE 1
Classification of malignant tumors.

4.1.2 Data collection

Clinical baseline data were collected for all patients, including patient age, sex, and time from cancer diagnosis to stroke event. This article documents traditional vascular risk factors for stroke, including hypertension, diabetes, hyperlipidemia, atrial fibrillation, smoking, and history of ischemic stroke. The patient's laboratory tests have coagulation parameters D-dimer, fibrinogen degradation products, and fibrinogen values. We collected data on imaging studies of patients. Generally speaking, malignant tumors include many types, and the main sources include carcinoma of epithelial tissue, mesenchymal tissue source, lymphoid hematopoietic tissue source, neural tissue source and other tissue sources. The classification of malignant tumors is shown in Figure 1.

4.1.3 Statistical analysis

Statistical analysis software SPSS22.0 was used for statistical analysis in this study. Individual factors of $p < 0.1$ have been incorporated into stepwise multiple accounting regression to obtain a complete predictive model. Then, the ROC curve analysis was used to compare the prediction performance of D_D, FDP, Fib and the comprehensive prediction model, and Kappa test was carried out between D-D, the comprehensive prediction model and the real results, and the test level was a = 0.05.

FIGURE 2
Location and sex ratio of primary cancer.

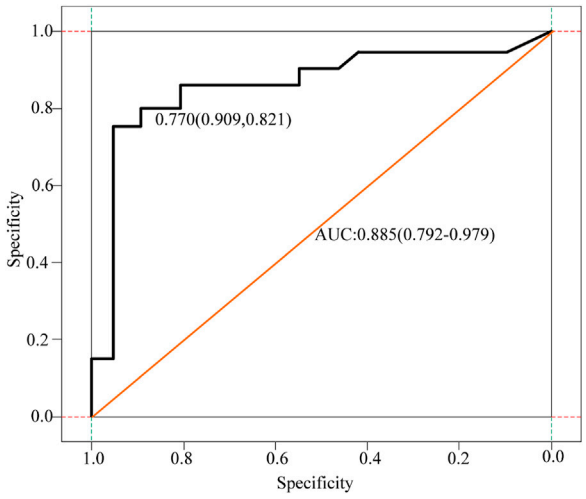


FIGURE 3
Receiver operating curves for cancer and non-cancer patients with stroke.

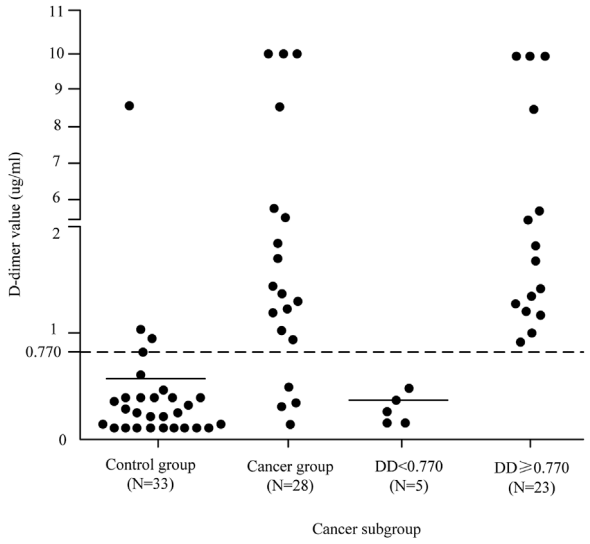


FIGURE 4
D-dimer value distribution patterns in each patient subgroup.

4.2 Clinical characteristics of patients with cancer-related acute ischemic cerebral infarction

Lung cancer, gastrointestinal cancers, and cancers of the female reproductive system are the three most prevalent cancers. The median interval between the diagnosis of malignancy and the beginning of the stroke was 8.5 months. Acute cerebral infarction was the first presentation in four patients (14.3%), and cancer (ovarian cancer, lung cancer) was subsequently discovered during the neurology hospitalization. From the time of the cancer diagnosis until the commencement of the ischemia infarction, the distribution of time was as follows: 14.3% in 3, 14.3% in 3–6, 25.0% in 6–12, 25.0% in 12–24, and 7.1% in >24 months. The prevalence of risk factors, such as hypertension, diabetes, atrial fibrillation, hyperlipidemia, and smoking, is lower in the experimental group compared to the control group. While the value of fibrinogen was not substantially different, the D-dimer and FDP values of the cancer group were greater than those of the control group. The location and sex ratio of primary cancer are shown in Figure 2.

The D-dimer test results of a few patients (3 cases in the cancer group) exceeded the upper limit of hospital's detection system. The receiver operating characteristic curve (ROC) analysis showed that the cut-off value of D-dimer to distinguish cerebral infarction

TABLE 1 Stepwise forward multivariate logistic regression analysis.

Variable	Estimate	Se	z	Wald	p	Or (95 CI%)
(Intercept)	−4.107	1.122	−3.661	13.404	0	0.016 (0.002, 0.148)
DD (u g/ml)	0.408	0.21	1.941	3.767	0.052	1.504 (0.996, 2.272)
FDP (ug/ml)	1.303	0.414	3.148	9.909	0.002	3.680 (1.635, 8.281)
AMBIs	2.573	1.083	2.376	5.645	0.018	13.099 (1569,109.364)

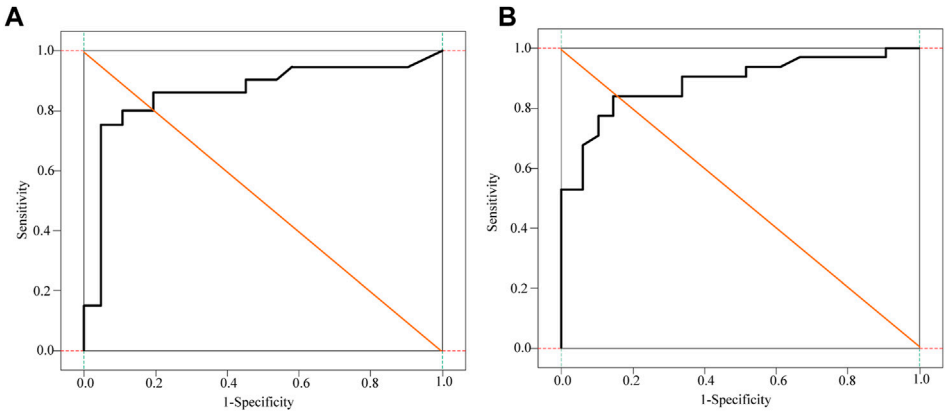


FIGURE 5
Subject D-D dimer and FDP curve analysis of patients. **(A)** D-D dimer curve **(B)** FDP curve.

patients with or without malignancy was 0.770 μg/ml. In the control group, 12.1% had values higher than the upper limit of the normal reference value of DD, and only two patients had DD values > 0.770 μg/ml. Besides, 17.9% of the patients in the cancer group had DD values < 0.770 μg/ml, and these 17.9% of patients also had DD values within the normal reference range. Figure 3 displays receiver operating curves for stroke patients with and without malignancy.

As can be seen from the analysis in Figure 3, the cancer group was divided into two subgroups with the value of 0.770 μg/ml as the critical value, D-dimer <0.770 μg/ml and D-dimer ≥0.770 μg/ml. The pattern of D-dimer values distribution in each subgroup of individuals with acute ischemic cerebral infarction is shown in Figure 4.

As can be seen from Figure 4, the D-dimer values in the cancer group were scattered, and when D-dimer was <0.770 μg/ml, the distribution matched that of the control group quite well.

4.3 Related risk factors in patients with malignant tumor-related cerebral infarction

The analysis results showed that plasma D-dimer (OR = 1.504, 95%CI:0.996–2.272, p = 0.052) and fibrin (ogen) degradation products (OR = 3.680, 95%CI:1.635–8.281, p = 0.002) level, and acute multiple cerebral infarction (OR = 13.099, 95%CI: 1.569–109.364, p = 0.018) were independent risk factors for acute ischemic cerebral infarction in patients with malignant tumors. The stepwise forward multivariate logistic regression analysis is shown in Table 1.

It can be seen from Table 1 that a comprehensive prediction model was obtained based on the CAIS-related factors screened out in multivariate logistic regression, and the risk index of CAIS in the patients in this study was calculated according to the prediction model. The DD, FDP, Fib and comprehensive prediction models are analyzed and compared by drawing the ROC curve, as shown in Figure 5 and Figure 6.

It can be seen from Figure 5 and Figure 6 that DD, FDP and comprehensive prediction models all have high diagnostic value for malignant tumor-related cerebral infarction. However, the AUC of the comprehensive prediction model was 0.925, which was significantly higher than the AUC area predicted by other risk factors alone, and the predicted cut-off value was 0.636. The Kappa test was used to analyze the consistency of D-dimer levels, the comprehensive prediction model and the real results. The results of the Kappa consistency test are shown in Table 2.

It can be seen from Table 2 that D-D≥0.770 μg/ml (Kappa = 0.734, p < 0.001) and the comprehensive prediction model (Kappa = 0.799, p < 0.001) are in good agreement with the real results. However, the comprehensive prediction model has a high degree of consistency. Therefore, compared with DD ≥ 0.770 μg/ml for the diagnosis of malignant tumor-related cerebral infarction, the prediction effect of the comprehensive prediction model may be better.

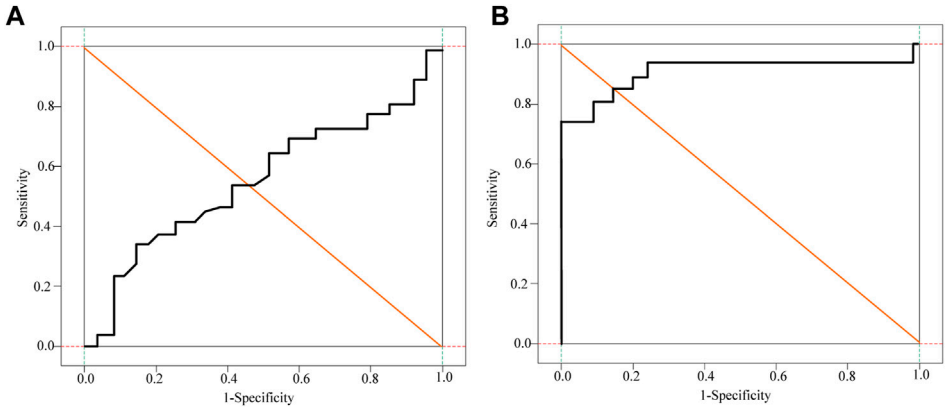


FIGURE 6
Subject Fib and Logistic Regression curve analysis of patients. **(A)** Fib curve **(B)** Logistic Regression curve.

TABLE 2 Kappa consistency test.

	Total (n = 61)	Control group (n = 33)	Cancer group (n = 28)	Kappa (95%C)	P
D-dimer value					
<0.770	35 (57.38)	30 (90.91)	5 (17.86)	0.734 (0.554–0.899)	<0.001
20.770	26 (42.62)	3 (9.09)	23 (82.14)		
Comprehensive forecast				0.799 (0.632–0.932)	<0.001
<0.636	39 (63.93)	33 (100.00)	6 (21.43)		
20.636	22 (36.07)	0 (0.00)	22 (78.57)		

TABLE 3 ROC analysis results of CT absolute and relative enhancement in clinical diagnosis of G1 and G2 pNENs.

Degree of reinforcement (CT value)	AUC	95%CI	Specificity	Sensitivity	Cut-off
Absolute arterial phase	0.775	0.648–0.901	0.621	0.926	69.55
Portal phase absolute	0.811	0.700–0.922	0.655	0.889	84.20
Absolute delay	0.803	0.687–0.918	0.926	0.586	55.85
Arterial phase relative	0.692	0.553–0.831	0.483	0.889	12.85
Portal phase relative	0.760	0.633–0.887	0.517	0.926	11.15

5 CT imaging features

5.1 CT imaging features of 83 patients with G1 pNENs and G2 pNENs

The tumor size of G2 pancreatic neuroendocrine tumor was significantly different from that of G1 pancreatic neuroendocrine tumor (47.4 + 28.9 VS. 28.6 + 17.7, $p < 0.01$). The CT imaging capabilities of G1 and G2pNEN showed that the largest tumor diameter was larger in G2, and the rest were size, number, border, shape, cystic degeneration, homogeneous enlargement, pancreatic duct enlargement, calcification, peripheral lymph node metastasis, and distant metastasis. The difference was not statistically significant.

5.2 The clinical value of CT absolute and relative enhancement values in predicting G1 and G2 pathological grades

To further clarify the clinical value of CT-related imaging functions in determining abnormal G1 and G2 scores, ROC curves were used to identify and analyze the absolute and relative CT magnification values of patients with G1 and G2 pNETs. The ROC analysis results of CT absolute and relative enhancement in the clinical diagnosis of G1 and G2 pNENs are shown in Table 3.

The ROC curve analysis of different absolute enhancement degrees and the ROC curve analysis of two different relative enhancement degrees are shown in Figure 7 and Figure 8.

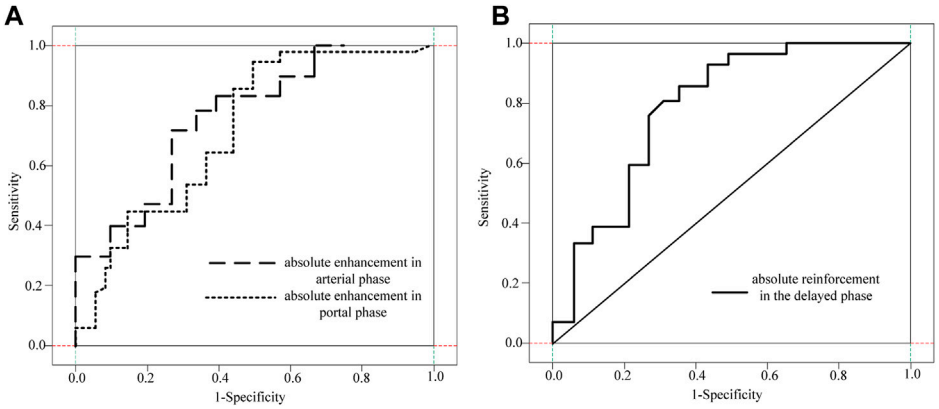


FIGURE 7
ROC curve analysis of three different absolute enhancement degrees. **(A)** Absolute enhancement in arterial and portal phases **(B)** Absolute enhancement in delayed phase.

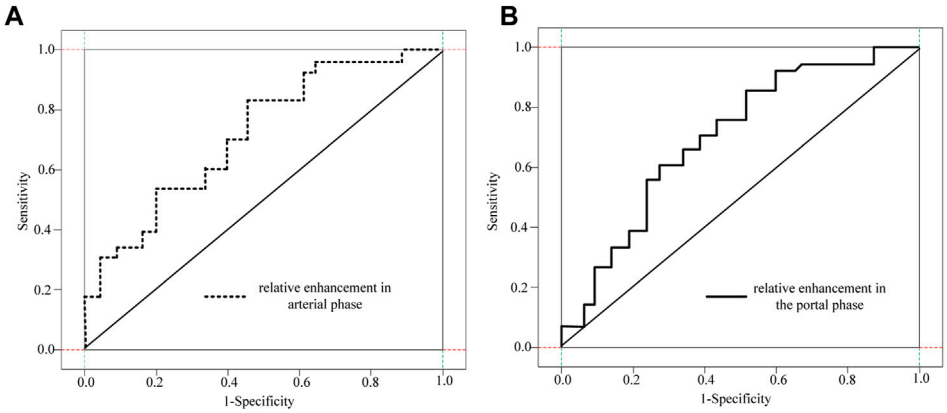


FIGURE 8
ROC curve analysis of two different relative enhancement degrees. **(A)** Relative enhancement degree in arterial phase **(B)** Relative enhancement degree in portal venous phase.

As shown in Figure 7 and Figure 8, the absolute and relative CT magnification values can effectively distinguish the G1 and G2 spin pNEN patients from the area under the absolute magnification ROC (AUC) curve. The portal phase value was 0.811 (95% CI 0.700–0.922), the specificity was 0.811 (95% CI 0.700–0.922) 0.655, and the sensitivity was 0.889.

Multi-slice computed tomography (CT) or magnetic resonance imaging (MRI) of the pancreas is still the preferred method for evaluating the imaging manifestations of neuroendocrine tumors and the extent of tumor invasion. On multiphase CT, non-functional pancreatic neuroendocrine tumors appear as well-defined peripheral hypervascular lesions, which are most clearly visualized in the end arterial and portal phases. The extent of local vascular invasion can be used as an indicator for assessing resectability. The sensitivity of CT examination is 63%–82%, and the specificity is 83%–100%, which varies according to its size. The sensitivity for detecting liver metastases was 82% and the specificity was 92%. About 10% of non-functional pancreatic neuroendocrine tumors are found as cystic lesions of the pancreas, with a misdiagnosis rate of about 43%. MRI improves tissue contrast in assessing the pancreas and liver. Its sensitivity is 85%–100%, the specificity is 75%–100%, the average detection rate of non-functioning neuroendocrine tumors is 73%, and the average detection rate of liver metastases is 82%. Preoperative examinations also included magnetic resonance cholangiopancreatography (MRCP). MRI is most useful when monitoring or confirming liver metastases. The comparison of the patients before and after surgery is shown in Figure 9.

As shown in Figure 9A, a 36-year-old female patient, who was unfit for contrast-enhanced CT examination because of abdominal pain, showed typical imaging manifestations of pancreatic neuroendocrine tumors with marked enhancement in the arterial phase. The hypodense foci in the right hepatic lobe were not visualized on octreotide scanning, and were assessed as hepatic cysts by MRI. Pancreatic lesions were removed by laparoscopic pancreatic body and tail + splenectomy. There was no recurrence or

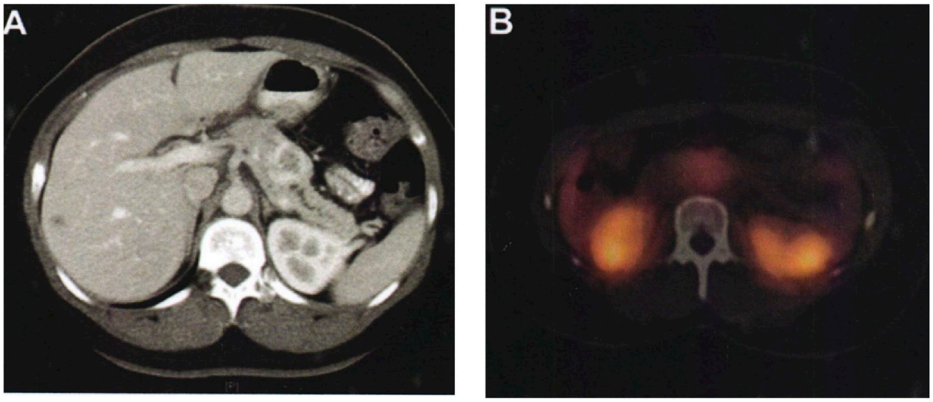


FIGURE 9
Comparison of patients before and after surgery. **(A)** The pancreas tissue of the patient before surgery. **(B)** The pancreas tissue of the patient after surgery.

metastasis in the follow-up 2 years after operation. Figure 9B shows a preoperative octreotide scan of the same patient showing typical features of a non-functional pancreatic neuroendocrine tumor with mildly elevated octreotide uptake.

In addition to traditional imaging, there are two nuclear medicine examination methods: somatostatin receptor scintigraphy (SRS) and positron emission tomography (PET). The somatostatin receptor family (SSTRs) includes five subtypes. That is, SSTR1 is highly expressed in many pancreatic neuroendocrine tumors, especially SSTR-2. It can therefore be visualized by radionuclide-labeled somatostatin receptor analogs (octreotides). SRS is often used to locate, stage, check for metastases, monitor recurrence, and evaluate response to systemic therapy in non-functional pancreatic neuroendocrine tumors. There are no data to support these tests to provide more valuable information than CT, MRI. In addition, SRS examination is expensive, which will increase the financial burden of patients.

6 Conclusion

In this study, the clinical and pathological characteristics of 266 patients with pancreatic neuroendocrine tumors were compared and analyzed, and it was found that patients with GI grade pancreatic neuroendocrine tumors and G2 and G3 pancreatic neuroendocrine tumors had obvious lymphadenopathy. The difference in distant metastases was statistically significant, but the extent of gastrointestinal and pancreatic neuroendocrine tumors showed only statistically significant differences in tumor diameter on computed tomography. This analysis was performed for the following reasons: computed tomography scans are more reliable and objective for larger diameter tumors, but standard CT scans are sufficiently objective for distant metastases, lymph node metastases, and invasive tumor growth. At the same time, 83 cases are relatively small, and reducing the sample size will have a certain impact on the results. A retrospective analysis of the preoperative CT image data of 56 patients with G1 and G2pNEN showed that the CT image features of GI and G2 were significantly different in CT values. The absolute and relative amplitudes of the arterial and portal phases were significantly higher than those of G2-grade patients and had important diagnostic implications in distinguishing G1-grade and G2pNEN patients. Of course, this study also has some shortcomings. The first study was designed as a retrospective analysis and had some selection bias. Second, because this was a retrospective study, the criteria for determining feature expression on specific CT images were inconsistent and biased. Third, due to the small number of G3 pNEN cases, this study should only include imaging features of G1 and G2 patients, and further studies are needed to analyze G3 patients.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Agra, G., Formiga, N. S., Oliveira, S., Sousa, A. T. O., Soares, M. J. G. O., and Lopes Costa, M. M. (2020). Instrument validation on nurses' knowledge and practice in palliative care for people with cutaneous malignant tumor wound. *Aquichan* 20 (1), 1–17. doi:10.5294/aqui.2020.20.1.2

AlKindi, G. M . (2017). A rare case of inflammatory myofibroblastic tumor of the mandible mimicking a malignant tumor. *Saudi Dent. J.* 29 (1), 36–40. doi:10.1016/j.sdentj.2016.10.001

Anehosur, V., Monis, P. L., Joshi, A., Bhat, A., and Kumar, N. (2020). Enigma of CD56 and S100 positive-high-grade undifferentiated malignant tumor of uncertain histogenesis of the mandible - a rare case report. *Int. J. Med. Dent. Case Rep.* 7 (1), 1–4. doi:10.15713/ins.ijmdcr.150

Hellyer, J. A., Ouseph, M. M., Padda, S. K., and Wakelee, H. A. (2020). Everolimus in the treatment of metastatic thymic epithelial tumors. *Lung Cancer* 149, 97–102. doi:10.1016/j.lungcan.2020.09.006

Hong, J., Jing, S., Zhang, Y., Owusu-Ansah, K. G., Chen, B., Chen, R., et al. (2020). Y-320, a novel immune-modulator, sensitizes multidrug-resistant tumors to chemotherapy. *Am. J. Transl. Res.* 12 (2), 551–562.

Kennedy, L. C., Bhatia, S., Thompson, J. A., and Grivas, P. (2019). Preexisting autoimmune disease: Implications for immune checkpoint inhibitor therapy in solid tumors. *J. Natl. Compr. Cancer Netw. JNCCN* 17 (6), 750–757. doi:10.6004/jnccn.2019.7310

Li, C., Ma, Y., Fei, F., Zheng, M., Li, Z., Zhao, Q., et al. (2020). Critical role and its underlying molecular events of the plasminogen receptor, S100A10 in malignant tumor and non-tumor diseases. *J. Cancer* 11 (4), 826–836. doi:10.7150/jca.36203

Li, P. P., Wang, Z. H., Huang, G., Li, Y., Ni, J. S., Liu, H., et al. (2017). Application of liver three-dimensional visualization technologies in the treatment planning of hepatic malignant tumor. *Zhonghua wai ke za zhi Chin. J. Surg.* 55 (12), 916–922. doi:10.3760/cma.j.issn.0529-5815.2017.12.008

Lukianchenko, A. B., Valiev, R. K., Romanova, K. A., Medvedeva, B. M., Nurberdiev, M. B., Stashkiv, V. I., et al. (2020). Primary hydatid cyst of pancreas mimicking malignant tumor. *Med. Vis.* 24 (4), 51–63. doi:10.24835/1607-0763-2020-4-51-63

Makhdoomi, M. A., Almotlaq, A., and Tawfiq, N. A. (2020). Small bowel malignant tumor: A diagnostic dilemma, a case report and literature review. *Int. Surg. J.* 7 (11), 3785–3789. doi:10.18203/2349-2902.isj20204692

Manikandan, T. (2020). An automated computer-aided diagnosis system for malignant tumor localization from lung CT images for surgical planning. *Int. J. Psychosoc. Rehabilitation* 24 (5), 2724–2729. doi:10.37200/ijpr/v24i5/pr201975

Nozaka, M., Yokoyama, H., Kitayama, K., Nagawa, D., Hamadate, M., Miura, N., et al. (2020). Clinical outcomes of acute myocardial infarction patients with a history of malignant tumor. *vivo (Athens, Greece)* 34 (6), 3589–3595. doi:10.21873/invivo.12203

Ohkubo, J-I., Ohkubo, S., Takahashi, A., Kawaguchi, R., Hasegawa, S., and Suzuki, H. (2020). Microwave energy device-assisted endoscopic resection of a parapharyngeal malignant tumor. *Pract. Oto-Rhino-Laryngologica* 113 (2), 123–128. doi:10.5631/jibirin.113.123

Radovich, M., Pickering, C. R., Felau, I., Ha, G., Zhang, H., Jo, H., et al. (2018). The integrated genomic landscape of thymic epithelial tumors. *Cancer Cell* 33 (2), 244–258.e10. doi:10.1016/j.ccell.2018.01.003

Ryabov, A. B., Pikin, O. V., and Abu-Khaidar, O. B. (2020). Solitary fibrous malignant tumor of the pleura with associated hypoglycemia: A case report. *Sib. J. Oncol.* 19 (2), 147–152. doi:10.21294/1814-4861-2020-19-2-147-152

Schairer, C., Catherine, S., and Gadalla, S. M. (2018). Autoimmune diseases and breast cancer risk by tumor hormone-receptor status among elderly women. *Int. J. Cancer* 142 (6), 1202–1208. doi:10.1002/ijc.31148

Shao, C. Y., Zhong, F., Wang, X. F., and Wang, H. (2019). Mucosal-associated invariant T cells and tumors. *Chin. Pharmacol. Bull.* 35 (12), 1650–1653.

Yan, B., Liu, H., You, J., and Fang, L. (2017). Clinicopathological characteristics of patients with gastrointestinal malignant tumor in some parts of Hainan province. *Cancer Res. Clin.* 29 (8), 556–560.

Zhao, M. M., Xie, Y. M., Liu, H., Zhang, Y., Lu, Q., and Zhuang, Y. (2020). Complex network analysis of combination medication of patients with kidney malignant tumor based in real world. *Zhongguo Zhong yao za zhi = Zhongguo zhongyao zazhi = China J. Chin. materia medica* 45 (14), 3299–3306. doi:10.19540/j.cnki.cjcmm.20200314.501

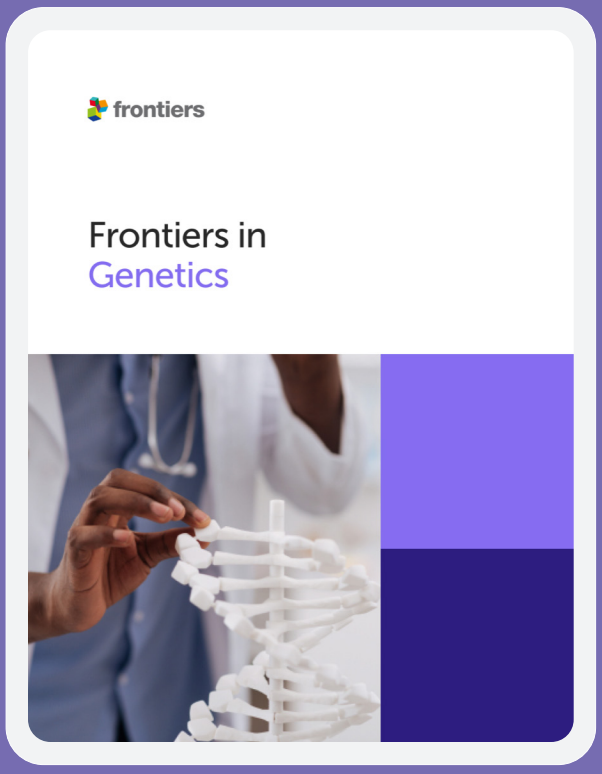
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact