International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 7774-7787

https://doi.org/10.70517/ijhsa464669

Perception Recommendation and Trust Cluster Analysis of Patient Trust Tendencies under the "Internet+Medical Health" Model

Jin Yin[1], Boyu Zhang[1,*] and Xiaoqian Huang[1]
[1] College of Economics and Management, Xiamen University of Technology, Xiamen, Fujian, 361024, China
Corresponding authors: (e-mail: 2121021034@stu.xmut.edu.cn).

Abstract In the context of "Internet+ healthcare" services, patient trust has become a key issue of concern. To address this, we propose a study on perception-based recommendation trust clustering using the GWO-K-means algorithm. First, we conduct data collection and preprocessing of perception-based recommendation trust text data. Then, we use the TFIDF_SP algorithm to select features from the perception-based recommendation trust text data. It was found that the K-means clustering algorithm has certain limitations. Therefore, the gray wolf optimization algorithm was used to optimize the K-means clustering algorithm, ultimately designing the GWO-K-means clustering algorithm. Using the GWO-K-means clustering algorithm, a clustering analysis of perceived recommendation trust in patient trust tendencies was conducted from the perspective of "Internet+ healthcare" services. Three categories of features were identified: the first category focuses on search recommendations and patient preferences, the second category emphasizes patient trust tendencies, and the third category primarily addresses patient trust and perceived value.

Index Terms GWO algorithm, K-means algorithm, TFIDF_SP algorithm, perceived recommendation trust

I. Introduction

With the changing demographic structure and evolving demands for healthcare services, "Internet+ healthcare" services based on the emerging "Internet+" model have emerged [1]. The integrated service system formed by the fusion of the two leverages the bridging role of information technology among all parties involved in healthcare services, enabling the application of healthcare services on the internet to facilitate information interaction. This plays a significant role in optimizing the allocation of medical resources, alleviating challenges such as "difficulty in accessing medical care" and "high medical costs," and promoting social development while meeting residents' healthcare needs [2]-[5]. Compared to traditional healthcare models, on one hand, Internet+ healthcare platforms provide users with remote online medical services, addressing spatial constraints [6], [7]. On the other hand, compared to traditional offline hospitals, Internet+ healthcare platforms can provide patients with medical examinations and assistance in a timely manner [8], [9]. Since the concept of "Internet+ healthcare" was first proposed, the service system has begun to take shape, but some issues remain.

Although users are becoming increasingly familiar with Internet-plus healthcare platforms and an increasing number of users are willing to try these platforms for medical assistance, users still harbor doubts about the effectiveness of such platforms, resulting in currently low usage intentions [10]-[13]. Since trust, as a fundamental personal trait, involves a strong mutual dependency between parties, the higher the user's trust in the platform, the more proactive their usage intentions will be [14]-[16]. Therefore, Internet-plus healthcare platforms should focus more on the actual needs and feelings of patients, building a bridge of trust between doctors and patients to enhance users' willingness to use the platform [17]-[19].

Currently, the services provided by Internet-plus healthcare platforms are gradually expanding, primarily including online healthcare education, personal healthcare assessments, professional disease consultation, and healthcare management. Literature [20] points out that the "Internet+" model has reshaped teaching thinking and methods in the medical field, improving the overall level of medical technology training by providing rich teaching resources and services. Literature [21] examined the effectiveness of internet platforms in managing chronic disease patients, leveraging their robust data monitoring and telemedicine capabilities to help patients better manage key health indicators, providing valuable insights for policymakers and healthcare providers. Literature [22] applied a dual medical model combining online support with structured continuous care to the postoperative rehabilitation process for chronic diseases. This method played a significant role in improving postoperative quality of life and mental health while reducing the incidence of postoperative complications. Literature [23] explored an internet-based postpartum healthcare service model, which significantly reduced the incidence of postpartum depression among mothers and provided effective support for improving their quality of life.

Additionally, internet technology provides effective support to medical institutions, further enhancing the level of healthcare services. Literature [24] utilized new technologies such as the internet to design an accurate outpatient triage system, helping patients accurately identify specialized outpatient departments within hospitals, thereby improving patient medical experiences and convenience. Literature [25] indicates that the application of the “Internet+” model in medical institutions will effectively improve the regional social governance system and mechanism construction. By achieving the efficient training of medical management talent and service personnel, it enhances regional medical services. Literature [26] proposed internet-based chronic disease tiered diagnosis and treatment measures, which promote communication between patients and clinical doctors and strengthen data monitoring of health indicators, thereby effectively addressing the prevention, treatment, and management of chronic diseases. Therefore, promoting the development of an “Internet+ healthcare” service system can provide patients with comprehensive medical rehabilitation support.

However, as Internet+ healthcare platforms continue to evolve and improve, related issues have gradually emerged. Literature [27] indicates that online healthcare platforms have expanded the breadth, duration, and depth of interaction between patients and doctors. While this helps doctors attract more patients, excessive interaction depth reduces patients' ability to choose doctors, which is detrimental to the provision of healthcare services. Literature [28] investigated the factors influencing users' willingness to pay on online healthcare service platforms, pointing out that the performance of internet platforms in terms of information quality, system quality, and convenience indirectly influences users' willingness to pay through perceived benefits. Literature [29] elucidated the impact of privacy leakage issues on the efficient operation of internet healthcare platforms, indicating that patients' willingness to disclose private information is strongly correlated with the outcome fairness and procedural fairness of healthcare platforms. Literature [30] analyzed the factors influencing users' use of internet healthcare platforms from four dimensions: perceived quality, perceived value, user trust, and user participation. Among these, online interactive discussions enabled by internet technology, willingness to pay, and privacy disclosure are highly correlated with user trust, making them important focal points. This suggests that in the virtual community of the internet, trust tendencies reflect patients' perception of the extent to which healthcare platforms assist their health. The higher this perception, the greater patients' willingness to use online healthcare services. Therefore, exploring changes in user trust tendencies under the “Internet + healthcare” context is crucial for the development of healthcare service systems.

This paper sets the “Internet + healthcare” service platform as the data source for this study, using Python to collect perceived trust text data related to patient trust tendencies. Additionally, the data undergoes a series of processing steps to enhance the reliability of the research results. A review of relevant literature revealed that using the traditional TFIDF algorithm for feature selection in perceived recommendation trust text data results in issues where the actual meaning of feature words is not clearly expressed. To address this, an improved TFIDF_SP algorithm was designed. The improved TFIDF_SP algorithm was used to perform feature selection in perceived recommendation trust text data. After completing this research, to address the limitations of the traditional K-means clustering algorithm, the GWO algorithm was used to optimize the traditional K-means clustering algorithm, resulting in the GWO-K-means clustering algorithm. Finally, the GWO-K-means clustering algorithm was used to conduct an in-depth analysis of perceived recommendation trust considering patient trust tendencies in the context of “Internet+ healthcare” services.

II. Perceived recommendation trust clustering design based on patient trust tendencies

II. A. Considerations on the Value of “Internet Plus Healthcare” Services

II. A. 1) Expanding service coverage

The application of internet technology has greatly optimized the health management service process. Functions such as online appointments and remote diagnosis have reduced patient waiting times, enabled more reasonable allocation of medical resources, and significantly improved the efficiency of medical services, while also increasing patient trust. With the help of online platforms, health management services can break through geographical barriers and extend high-quality medical resources to remote areas, effectively expanding the scope of services and allowing more patients to benefit from professional health management. This further strengthens patient trust and provides strong support for the construction of a comprehensive medical service system.

II. A. 2) Promoting the integration of health data

Under the “Internet + Healthcare” model, various types of health data can be aggregated and integrated. Patients' basic physiological indicators and daily health behavior data can be collected and summarized through smart devices and medical information systems. By leveraging big data analytics technology, the rich and comprehensive data can accurately depict an individual's health profile, providing scientific and precise evidence for health management services, thereby reinforcing patients' trust in the system. Based on precise data analysis, health management plans can be customized to meet individual needs, thereby enhancing the effectiveness of health interventions, effectively preventing the onset and progression of diseases, and driving the transformation of health management from traditional experience-based models to precise and scientific approaches. This helps maintain patients' trust in "Internet + Healthcare" services.

II. A. 3) Promoting self-health management

Internet platforms provide patients with a wealth of health knowledge and educational resources, including information on disease prevention, rehabilitation guidance, and promotion of healthy lifestyles. Patients can gain a deeper understanding of health knowledge, thereby enhancing their health awareness, while also increasing their trust in "internet-plus medical and health services." When patients have a high level of trust, they are more likely to actively participate in self-health management, such as maintaining a balanced diet, engaging in regular exercise, and taking medication on time. This shift can improve patients' health status, reduce the burden on the healthcare system, and promote the healthy development of the entire health management ecosystem.

II. A. 4) Optimizing Medical Service Processes

The integration of internet technology enables the redesign and optimization of health management service processes. For example, the addition of online consultation and follow-up sessions facilitates more timely and convenient communication between patients and healthcare providers, reducing the number of trips patients need to make to the hospital and lowering healthcare costs. The sharing and informatization of electronic medical records have improved the accuracy and transmission efficiency of medical information, avoiding issues such as duplicate tests and misdiagnoses caused by information bottlenecks. The optimized service process better meets patient needs, enhances the patient experience, thereby significantly increasing patients' trust in health management services and strengthening their trust in and adherence to the healthcare system.

II. B.Perception Recommendation Trust Text Data Collection and Preprocessing

II. B. 1) Perception Recommendation Trust Text Data Collection

This study uses a "Internet + medical health" service platform as its data source and employs Python to obtain text data on patients' trust tendencies in "Internet + medical health" services as the object of this study. Due to issues such as semantic ambiguity and unscientific and non-standard expressions in the text data on perceived trust in recommendations, this study requires preprocessing of the data to ensure the rigor and validity of its results.

II. B. 2) Perceptual Recommendation Trust Text Data Preprocessing

(1) Data cleaning

Remove meaningless noise data, correct punctuation errors and misuse, and adjust minor inaccuracies and non-standard expressions. Build a synonym list for perceived recommendation trust and standardize professional terminology.

(2) Word segmentation

Use word segmentation software to segment words, laying the theoretical foundation for subsequent research.

(3) Category Annotation

Determine the category system through literature review and expert consultation, ensuring a clear structure and prominent domain characteristics. Classify perception-based recommendation trust information based on relevant literature and combine it with the actual situation of patient perception-based recommendation trust in the "Internet + Healthcare" service platform. Categorize patient perception-based recommendation trust and annotate categories according to the category classification system for annotation and review.

II. C.Perceptual Recommendation Trust Text Feature Selection

In order to improve the feature selection ability of perception recommendation trust text data, a full coverage granularity calculation method is used to analyze the high dimensionality and sparsity of the feature selection algorithm data. In response to the shortcomings of the TFIDF algorithm, an improved TFID_SP algorithm is designed to distinguish the importance of feature words in different parts of a document.

II. C. 1) Traditional TFIDF Algorithm

The TFIDF algorithm is an important algorithm used for feature word analysis in vector space models [31]. IDF refers to inverse document frequency, which tends to decrease as the number of documents increases. Using this feature word cannot achieve the goal of accurately distinguishing document categories. TF refers to term frequency, and as the frequency of feature words increases, the corresponding weight also increases, indicating that the feature word has achieved stronger document distinction performance. The following is the TFIDF calculation formula:

$$w_{i,j}=\frac{\frac{t_j}{\sum_{j=1}^{n}t_j}\times\log\left[\log\left(\frac{N}{n_j}+0.01\right)\right]^2}{\sqrt{\sum_{j=1}^{N}\left\{\frac{t_j}{\sum_{j=1}^{n}t_j}\right\}^2\log\left(\frac{N}{n_j}+0.01\right)}} \tag{1}$$

In the formula, t is the frequency of word t appearing in the m th document, N is the total number of documents, and n is the number of documents containing word t. The TFIDF algorithm can efficiently filter words in a uniform state, but it is important to note that this algorithm has significant limitations that need to be addressed. When calculating contribution, it only analyzes the influence of feature word frequency and does not consider the position or part of speech of feature words. Under different part-of-speech and position conditions, the actual meaning expressed by feature words can vary significantly. Among these, nouns and verbs can achieve stronger thematic expressiveness than adjectives, enabling more precise expression of document content. Additionally, when words appear in the title section, they carry greater significance than when they appear in the main text. For such cases, these types of words must be distinguished and handled separately from regular words.

II. C. 2) TFIDF_SP Algorithm

In response to the shortcomings of the TFIDF algorithm, this paper designs an improved TFIDF_SP algorithm. By comprehensively analyzing part-of-speech and position through weight coefficients and establishing composite weights, feature weights are calculated and combined with the TFIDF method to distinguish the importance of feature words in different parts of a document. The corresponding calculation formula is given below:

$$tf_{i,j}=\frac{\lambda_j t_k+u_1 t_{k_1}+u_2 t_{k_2}}{\sum_{j=1}^{l}(\lambda_j t_k+u_1 t_{k_1}+u_2 t_{k_2})} \tag{2}$$

Among them, $t_k=t_{k_1}+t_{k_2}$. That is:

$$w_{i,j}=\frac{tf_{i,j}\times\log\left(\frac{N}{n_j}+0.01\right)}{\sqrt{\sum_{i=1}^{N}(f_{i,j})^2\times\left[\log\left(\frac{N}{n_j}+0.01\right)\right]^2}} \tag{3}$$

In the formula, $tf_{i,j}$ is the word frequency obtained by weighting the feature words, and λ_j is the part-of-speech weighting coefficient. Among these, nouns achieve optimal results when $\lambda=3$, verbs when $\lambda=2$, and other words when $\lambda=1$. t_k is the frequency of word j in document i, and u_1,u_2 are the weight coefficients for the word in the title and body regions, respectively, with optimal results obtained when their values are 4 and 1, respectively. t_{k1},t_{k2} represent the word frequencies of word j under the title and body conditions, respectively, and l is the total number of words contained in the i th document. By applying part-of-speech and position-weighted normalization to feature words and extending the results, feature words can not only reflect high frequency but also more efficiently highlight the structural characteristics of the text.

II. D.Design of Clustering Algorithms for Perception-Based Recommendation Trust

Clustering is an unsupervised machine learning technique. The original purpose of cluster analysis was to group physical objects with similar characteristics in human life, dividing samples in a dataset into clusters such that the similarity among samples within the same cluster is maximized, while the similarity between samples across different clusters is minimized. K-means is a clustering method based on partitioning. It has a simple principle and high computational efficiency. K-means has certain limitations when performing perceptual recommendation trust clustering. K-means requires a high degree of precision in selecting initial cluster centers during algorithm initialization and is prone to getting stuck in local optima. To address this, the gray wolf optimization algorithm is proposed to optimize the clustering algorithm, ultimately designing a gray wolf optimization K-means clustering algorithm tailored for perceptual recommendation trust.

II. D. 1) Traditional K-means clustering algorithm

The core idea of K-means is as follows: first, randomly select k initial cluster centers $c_i(1 \le i \le k)$ from the dataset, calculate the Euclidean distance between the remaining data objects and the cluster centers c_i, identify the cluster center c_i closest to the target data object, and assign the data object to the cluster corresponding to the cluster center c_i [32], [33]. Then, the mean of the data points in each cluster is calculated as the new cluster center, and the process is repeated until the cluster centers no longer change or the maximum number of iterations is reached. The formula for calculating the Euclidean distance between data points and cluster centers in space is:

$$dis(x,c_i)=\sqrt{\sum_{j=1}^{d}(x_j-c_{i,j})^2} \tag{4}$$

In this context, x represents the data sample, c_i denotes the i th cluster center, d is the dimension of the data sample, and x_j and c_{ij} represent the j th attribute values of x and c_i, respectively. The formula for the sum of squared errors (SSE) of the entire dataset is:

$$SSE=\sum_{i=1}^{k}\sum_{x\in C_i}|dis(x,c_i)|^2 \tag{5}$$

Among them, the size of SSE indicates the quality of the clustering results, k is the number of clusters, and C_i is the sample set of the i th cluster. The advantages of K-means clustering include:

(1) It can process perceptual recommendation trust text features, has high stability and scalability, can process numerical data sets, and has good clustering results.

(2) The output results are more intuitive, easier to understand compared to other data analysis methods, and have a clearer geometric meaning.

(3) Fast clustering speed and application in a wide range of fields. The limitations of K-means clustering include: The clustering results and accuracy of the K-means clustering algorithm are closely related to the selection of the initial clustering centers. Therefore, when solving practical problems, the clustering results obtained from the clustering centers are generally locally optimal, so the results will have a certain degree of volatility, high randomness, and significant differences.

II. D. 2) Grey Wolf Optimization Algorithm (GWO)

In GWO, to construct a mathematical model of the gray wolf social hierarchy, the first optimal solution in the wolf pack is referred to as the α wolf, the second optimal solution is referred to as the β wolf, the third optimal solution is referred to as the δ wolf, and the remaining candidate solutions are referred to as ω wolves [34]. During the GWO group hunting (optimization) process, the α, β, and δ wolves guide the ω wolves to search for their optimal regions. Through iterative searches, the α, β, and δ wolves (the three optimal solutions) are used to predict the target of the hunt (the optimal solution), thereby continuously searching for the global optimal solution through guidance. As mentioned above, gray wolf hunting behavior includes tracking and approaching prey, pursuing and surrounding prey, attacking and killing prey.

First, to establish a mathematical model for surrounding prey behavior, equations (6) and (7) are established. A gray wolf can update its position according to equations (6) and (7), and this position may be any random position around the prey:

$$Dis=|C\otimes X_p(t)-X(t)| \tag{6}$$

$$X(t+1)=X_p(t)-A\otimes Dis \tag{7}$$

$$A=2c_1\times r_1-c_1 \tag{8}$$

$$C = c_2 r_2 \tag{9}$$

$$c_1 = 2 - 2t / MaxDT \tag{10}$$

Among these, Dis represents the distance between the gray wolf and the prey, X_p represents the position vector of the prey, X represents the position vector of a gray wolf, and A and C are coefficient vectors, as shown in Equations (8) and (9), where the coefficient c_1 is the parameter for switching between encirclement and attack, i.e., the parameter for switching between exploration and exploitation; it decreases linearly from 2 to 0 during the iteration process, as shown in Equation (10). This parameter determines whether the wolf pack surrounds or attacks the prey. When c_1 is greater than 1, the wolf pack surrounds the prey; When it is less than 1, the wolf pack attacks the prey. c_2 is the difficulty parameter for approaching the prey, whose value is typically set to 2. r_1 and r_2 are random vectors uniformly distributed between [0, 1].

In GWO, it is assumed that the first three optimal wolves α,β and δ are the optimal solutions, and it is required that other wolves (including ω wolves) update their positions based on the positions of α,β and δ wolves to establish a mathematical model of hunting behavior, as shown in Equations (11) to (17). In a 2D search space, gray wolves use the guidance of the three optimal wolves α,β and δ to determine the distance between the optimal wolves and the prey, thereby updating their own positions and continuously moving toward the prey to reduce the distance between themselves and the prey. For:

$$Dis_\alpha = | C_1 \otimes X_\alpha(t) - X(t) | \tag{11}$$

$$Dis_\beta = | C_2 \otimes X_\beta(t) - X(t) | \tag{12}$$

$$Dis_\delta = | C_3 \otimes X_\delta(t) - X(t) | \tag{13}$$

$$X_1 = X_\alpha(t) - A_1 \otimes Dis_\alpha \tag{14}$$

$$X_2 = X_\beta(t) - A_2 \otimes Dis_\beta \tag{15}$$

$$X_3 = X_8(t) - A_3 \otimes Dis_\delta \tag{16}$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{17}$$

When gray wolves surround their prey within a small area, they end the hunting behavior by attacking the prey. The mathematical simulation of approaching the prey is based on the decrease in the c_1 value, and $|A|$ decreases as the c_1 value decreases. When each component value of A randomly varies within the range [-1, 1], the gray wolves update their positions to approach and gather around the prey at any location between their current position and the prey, based on α, β, and δ. This enables GWO to achieve better search accuracy and convergence quality.

II. D. 3) Improvements to the K-means clustering algorithm

This section uses GWO to optimize the K-means clustering algorithm for perception-based recommendation trust. First, assuming that each coyote consists of k cluster centers, the dimension of the solution vector should be equal to $k \times$ the number of perception-based recommendation trust text features in the data sample. The objective function is given by Equation (5). Next, GWO is executed until the algorithm's termination condition is met, and the optimal perception-based recommendation trust cluster centers are output.

III. Empirical validation analysis

III. A. Perceptual recommendation text feature selection analysis

III. A. 1) Evaluation Indicators

The effectiveness of feature selection in perceptual recommendation text is often evaluated using precision and recall rates. Precision refers to the ratio of perceptual recommendation text (assumed to be M) that matches the manually selected results among the text judged by the algorithm (assumed to be N), expressed as: P = M/N. Recall rate refers to the ratio of the perceived recommended text (O) that matches the algorithm's selection to the total perceived recommended text (N) that should be selected manually, expressed as R = O/N. Accuracy and recall rate reflect the quality of a text feature selection from two perspectives. However, neither of them fully represents overall performance. Therefore, combining them is necessary for better results. This is why we introduce the F1 score. The formula is as follows:

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \tag{18}$$

III. A. 2) Algorithm Verification Analysis

Since the traditional TFIDF algorithm is applied to the field of text features, the experiments we designed also focus on the selection of text features related to perceived recommendation trust, as this ensures fairness. The data used in the experiment was sourced from text data on the “Internet + Healthcare” service platform, with a total of 8,000 patient trust-oriented perceptual recommendation text feature data points crawled. Based on this, evaluation metrics were employed to explore the performance of the TFIDF_SP algorithm in perceptual recommendation text feature selection. To ensure the validity of the results, the experiment was conducted 20 times. The algorithm validation analysis results are shown in Table 1. Based on the data in the table, the performance metrics of the TFIDF algorithm in text feature selection for perceived recommendation trust were 0.714, 0.687, and 0.699, respectively. After improving the TFIDF algorithm, its performance improved, with metrics of 0.888, 0.881, 0.884, indicating that introducing feature weighting on the basis of the traditional TFIDF algorithm is more beneficial for text feature selection in perceptual recommendation.

Table 1: Algorithm verification and analysis results

N	TFIDF			TFIDF_SP		
	Precision	Recall	F1	Precision	Recall	F1
1	0.634	0.653	0.643	0.939	0.937	0.938
2	0.631	0.628	0.629	0.884	0.901	0.892
3	0.732	0.795	0.762	0.915	0.863	0.888
4	0.758	0.633	0.690	0.906	0.805	0.853
5	0.798	0.660	0.722	0.893	0.916	0.904
6	0.727	0.649	0.686	0.935	0.932	0.933
7	0.739	0.747	0.743	0.876	0.911	0.893
8	0.647	0.611	0.628	0.864	0.915	0.889
9	0.757	0.709	0.732	0.833	0.884	0.858
10	0.605	0.674	0.638	0.937	0.904	0.920
11	0.799	0.783	0.791	0.896	0.943	0.919
12	0.623	0.669	0.645	0.833	0.895	0.863
13	0.796	0.743	0.769	0.899	0.823	0.859
14	0.799	0.772	0.785	0.888	0.892	0.890
15	0.785	0.742	0.763	0.813	0.884	0.847
16	0.722	0.654	0.686	0.868	0.816	0.841
17	0.671	0.642	0.656	0.899	0.832	0.864
18	0.656	0.653	0.654	0.914	0.853	0.882
19	0.723	0.617	0.666	0.874	0.898	0.886
20	0.684	0.701	0.692	0.898	0.824	0.859
Total	0.714	0.687	0.699	0.888	0.881	0.884

III. A. 3) Algorithm Application Analysis

After verifying the performance of the TFIDF_SP algorithm in selecting perceptual recommendation trust text features, the TFIDF_SP algorithm was then applied to select perceptual recommendation trust text features that consider patient trust tendencies in the context of “Internet + medical health” services. The results of the algorithm application analysis are shown in Table 2. The data in the table shows that the TFIDF_SP algorithm selected a total of 30 keywords for perceptual recommendation trust text features related to patient trust tendencies. Rehabilitation Guidance < Internet < Online Appointment < Smart Devices < Medical Resources < Information Management < Medical Costs < Patient Needs < Online Consultation < Remote Diagnosis < Satisfaction < Electronic Medical Records < Health Awareness < Hidden Trust < Medical System < Patient Information < Family Relationships < Experience < Health Behavior < Direct Trust < Service Attitude < Medical Health < Number of Visits <

Recommendation Quality < Patient Intent < Perceived Value < Search Recommendation < Trust Tendency < Patient Trust < Perceived Trust.

Table 2: Algorithm application analysis

No.	Key words	Word frequency	Weight	Rank	No.	Key words	Word frequency	Weight	Rank
1	The Internet	6	0.0105	28	16	Remote diagnosis	12	0.0211	21
2	Medical and health care	22	0.0387	8	17	Medical resources	8	0.0141	25
3	Service attitude	21	0.0369	10	18	Intelligent device	7	0.0123	27
4	Patient trust	42	0.0738	2	19	Healthy behavior	19	0.0334	12
5	Kinship	17	0.0299	14	20	Health awareness	15	0.0264	17
6	Medical treatment cost	10	0.0176	23	21	Health awareness	16	0.0281	15
7	Experience	18	0.0316	13	22	Online consultation	11	0.0193	22
8	Satisfaction	14	0.0246	19	23	Electronic medical record	14	0.0246	19
9	Rehabilitation guidance	4	0.0070	30	24	Information-based management	8	0.0141	25
10	Search recommendation	37	0.0650	4	25	Patient's needs	10	0.0176	23
11	Perceived trust	54	0.0949	1	26	Direct trust	20	0.0351	11
12	Perceived value	33	0.0580	5	27	Hide trust	15	0.0264	17
13	Patient's will	29	0.0510	6	28	Follow-up visit	22	0.0387	8
14	Trust tendency	40	0.0703	3	29	Recommended quality	23	0.0404	7
15	Online reservation	6	0.0105	28	30	Patient Information	16	0.0281	15

III. B. Clustering Case Analysis

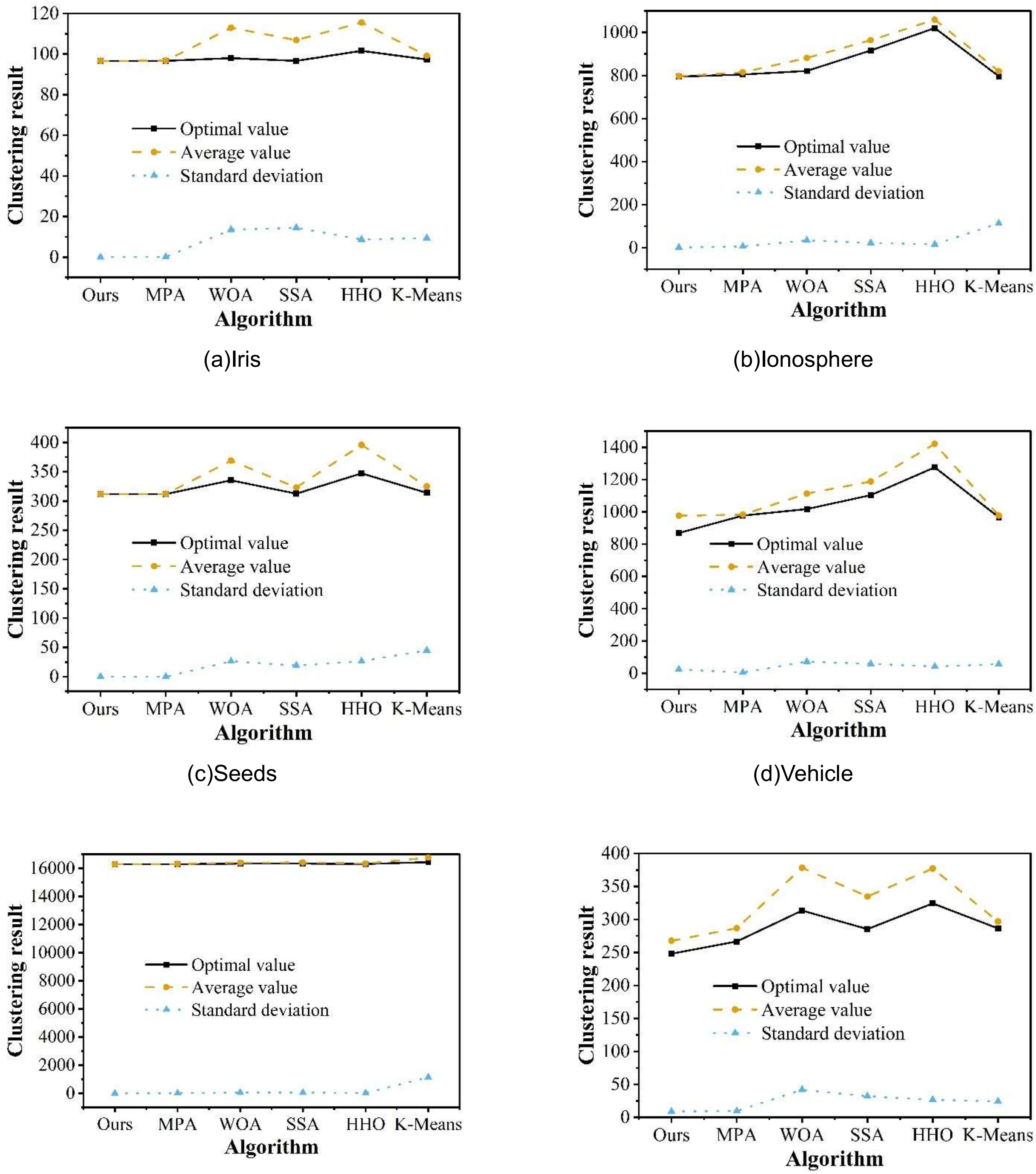
III. B. 1) Analysis of the Optimization Effect of the Gray Wolf Algorithm

This chapter selects six commonly used datasets from the UCI dataset to test the optimization capabilities of the gray wolf algorithm for global optimization. The test datasets are shown in Table 3.

Table 3: Test dataset

Dataset name	Sample quantity	Number of features	Number of clusters
Iris	150	4	3
Ionosphere	351	34	2
Seeds	210	7	3
Vehicle	846	18	4
Wine	178	13	3
Glass	214	9	6

The simulation environment for this paper is as follows: Windows 8, Intel(R) Core(TM) i7-8265UCPU@2.4GHz, with simulation software Matlab 2016b. Figure 1 shows the experimental results of the improved algorithm and comparison algorithms (MPA: basic marine predator algorithm, WOA: Whale Algorithm, SSA: Sparrow Optimization Algorithm, HHO Harris Eagle Optimization Algorithm, K-Means) on six clustering datasets. (a) to (d) represent the Iris dataset, Ionosphere dataset, Seeds dataset, Vehicle dataset, Wine dataset, and Glass dataset, respectively. For the Iris dataset, the Gray Wolf Optimization Clustering Algorithm achieves the best optimal value (93.332), mean (93.124), and standard deviation (0.0192), with its optimal value matching the theoretical optimal predicted value. Its optimization performance significantly outperforms other algorithms, including the original K-Means algorithm. The Gray Wolf Optimization Clustering Algorithm also demonstrates the best performance across the remaining five datasets. The original K-Means algorithm has poor optimization efficiency across all clustering datasets. However, since K-Means clustering calculations are easily influenced by initial cluster centers, it has reached a local optimum. After introducing the Gray Wolf optimization algorithm into the K-Means clustering framework, the optimal value, mean, and standard deviation values all improved to some extent, demonstrating the optimization effect of the Gray Wolf algorithm on clustering algorithms.



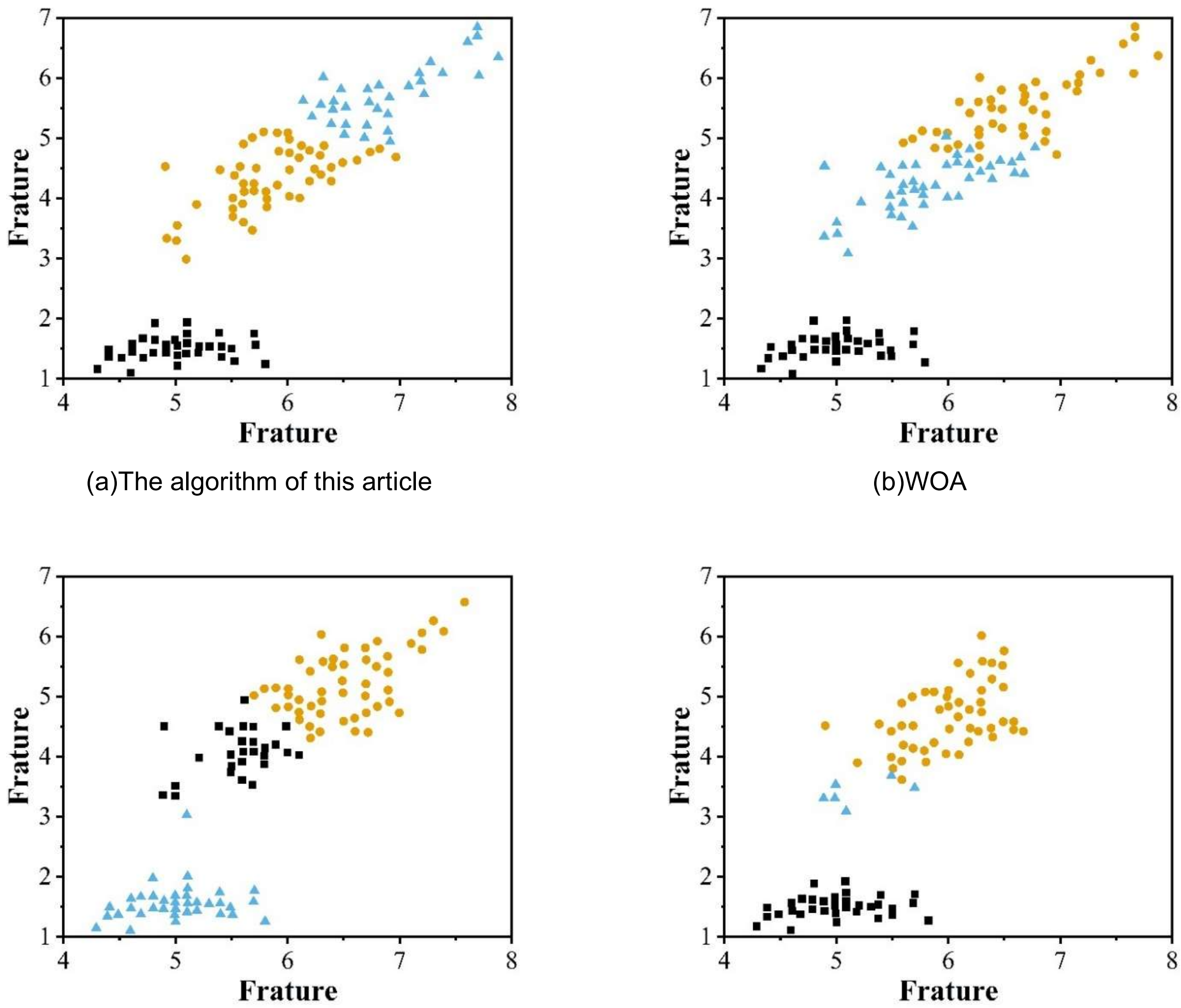
(a)Iris (b)Ionosphere

(c)Seeds (d)Vehicle

(e)Wine (f)Glass

Figure 1: Comparison of clustering results

Figure 2 shows the clustering results of the Iris dataset, where (a) to (d) represent the algorithms proposed in this paper, WOA, SSA, and HHO, respectively. Based on the feature clustering results, it can be seen that the K-means clustering algorithm optimized by the gray wolf algorithm yields the best clustering results, demonstrating the scientific validity and effectiveness of the improvements proposed in this paper.



(a)The algorithm of this article (b)WOA

(c)SSA (d)HHO

Figure 2: The clustering effect of the Iris dataset

III. B. 2) Analysis of algorithm application effectiveness

The effectiveness of the Gray Wolf Optimization Clustering Algorithm has been demonstrated in the preceding section. The following section will explore the application effectiveness of the algorithm. As indicated in Section 3.1.2, this study utilizes the TFIDF_SP algorithm to select 30 text feature data points related to patient trust inclination and perceived recommendation trust. Considering the magnitude of the weights, this study focuses on seven text features: perceived trust, patient trust, trust inclination, search recommendation, perceived value, patient willingness, and recommendation quality. Cluster analysis requires weak correlations between features, so it is necessary to analyze the correlations between text features before implementing cluster analysis. The results of the correlation analysis are shown in Figure 3, In the figure, X1 to X7 represent perceived trust, patient trust, trust tendency, search recommendation, perceived value, patient willingness, and recommendation quality, respectively. Based on the data in the figure, the correlation coefficient between recommendation quality and perceived trust is greater than 0.6, indicating a high correlation between the two, which may influence the clustering analysis results. Therefore, variable dimension reduction is necessary. The remaining features do not exhibit high correlations.

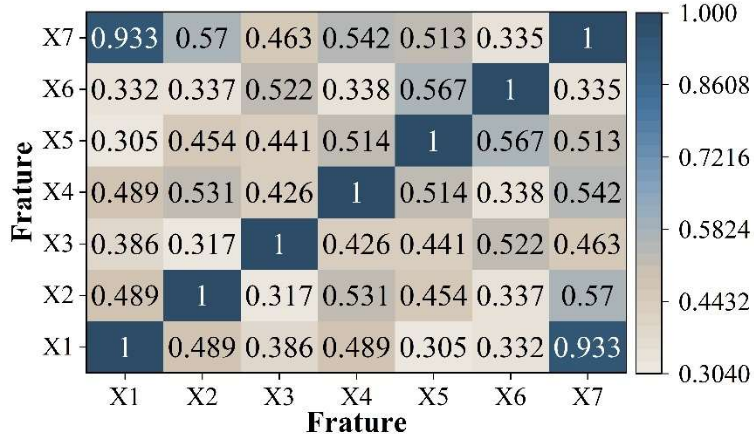


Figure 3: Results of correlation analysis

Since recommendation quality and perceived trust are highly correlated, principal component analysis (PCA) was used in this study. The results of PCA are shown in Table 4. It can be seen that the first principal component explains 65.1% of the variance. Therefore, the first principal component was extracted as the principal component, resulting in minimal information loss and an optimal outcome. The first principal component was then used as a new indicator: perceived recommendation trust (Y).

Table 4: The results of principal component analysis

Ingredients	Initial eigenvalue			Extract the sum of squares and load them		
	Total	variance%	Cumulative %	Total	variance%	Cumulative %
1	1.302	65.1	65.1	1.302	65.1	65.1
2	0.698	34.9	34.9			

Next, we calculate the descriptive statistics for each text feature, as shown in Table 5. From the means in Table 5, we can see that the magnitudes of the six indicators vary greatly. For example, the maximum value for trust tendency is 943, while the maximum value for patient trust is only 72. Therefore, in order to eliminate the influence of the scale and ensure that all indicators are treated equally in the analysis, it is necessary to standardize the data.

Table 5: Describe the statistics

Text features	Valid data	Minimum value	Maximum value	Mean value	Standard deviation
Y	100	43	943	444.55	7.18
X2	100	3	72	42.62	5.24
X3	100	5	423	262.47	8.49
X4	100	24	414	258.91	7.96
X5	100	14	627	361.52	9.87
X6	100	6	366	164.42	6.08

After the data has undergone standardization processing, to establish a K-means clustering model, it is necessary to first determine the K value for modeling. In this paper, the contour coefficient is used to evaluate the K value of the K-means model, and the contour coefficients corresponding to different K values are calculated, as shown in Figure 4. A larger contour coefficient indicates better clustering efficiency at that K value. As shown in Figure 4, the contour coefficient is largest when K is 3, indicating that clustering patients' trust tendencies in the “Internet+ healthcare” context into three categories yields the best clustering results, with a value of 0.357. Therefore, we set the number of clusters K to 3 and performed GWO-K-means clustering on the data.

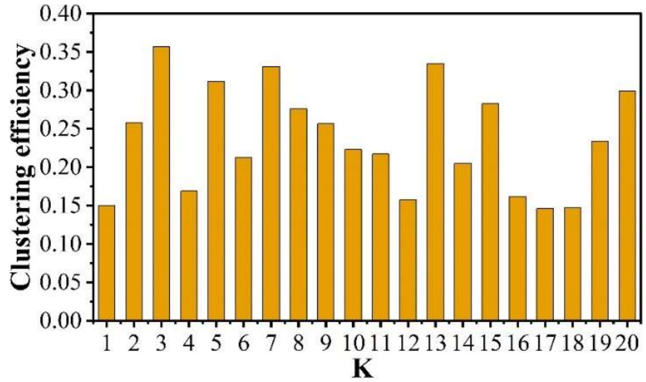


Figure 4: Contour coefficient variation graph

The GWO-K-means clustering algorithm was used to investigate the perception and recommendation of patient trust tendencies in the context of “Internet+ healthcare.” The results of the cluster analysis are shown in Table 6. The cluster results are analyzed as follows:

The first category of features includes: patient trust (22.45), trust tendency (181.53), search recommendations (165.23), perceived value (125.96), and patient willingness (152.87). It can be concluded that the perceived recommendation trust of patient trust tendencies primarily manifests in search recommendations and patient willingness.

The second category of features includes: patient trust (34.17), trust inclination (191.84), search recommendations (112.19), perceived value (143.47), and patient willingness (140.53). This category of features primarily focuses on enhancing patient trust inclination.

The second category of features includes: patient trust (41.33), trust tendency (133.65), search recommendations (145.05), perceived value (153.25), and patient preference (111.38). This category of features focuses on patient trust and perceived value.

Table 6: Cluster analysis results

Category	X2	X3	X4	X5	X6
1	22.45	181.53	165.23	125.96	152.87
2	34.17	191.84	112.19	143.47	140.53
3	41.33	133.65	145.05	153.25	111.38

IV. Conclusion

This paper first uses the Python software package to obtain text data on patients' perceived trust in recommendation trust from the “Internet + Healthcare” service platform, and then preprocesses the data. Using the TFIDF_SP algorithm, text feature selection is conducted on the perceived recommendation trust text data, resulting in 30 text features. Finally, the features with higher weights are selected as the research objects, and the GWO-K-means clustering algorithm is applied to perform clustering analysis on the perceived recommendation trust related to patient trust tendencies under the “Internet+Healthcare” platform. The results show that the first category of features focuses on search recommendations (165.23) and patient preferences (152.87), the second category emphasizes patient trust tendencies (191.84), and the third category primarily addresses enhancing patient trust (41.33) and perceived value (153.25). These findings provide valuable guidance for advancing the development of Internet+ healthcare.

Funding

This work was supported by National Natural Science Foundation of China Youth Project (72001184): Perceptual Trust transfer model of Internet Medical Service based on Intuitionistic Fuzzy sets. And also supported by Youth Project of Fujian Provincial Natural Science Foundation (2021J05264): Research on the perceived Trust transfer model of the elderly on Internet + Medical and health Services.

References

[1] Yang, F., Shu, H., & Zhang, X. (2021). Understanding “internet plus healthcare” in china: policy text analysis. Journal of Medical Internet Research, 23(7), e23779.

[2] Jiang, X., Xie, H., Tang, R., Du, Y., Li, T., Gao, J., ... & Xie, G. (2021). Characteristics of online health care services from China’s largest online medical platform: cross-sectional survey study. Journal of medical Internet research, 23(4), e25817.

[3] Feng, X., Cao, Y., & Li, Y. (2019). Medical Service Platform and Internet+ Medicine Innovation Construction. International Journal of Advanced Nursing Education and Research, 4(1), 25-30.

[4] Lee, H. Y., Lee, K. H., Lee, K. H., Erdenbayar, U., Hwang, S., Lee, E. Y., ... & Youk, H. (2023). Internet of medical things-based real-time digital health service for precision medicine: Empirical studies using MEDBIZ platform. Digital health, 9, 20552076221149659.

[5] Wang, R., Huang, Y., Zhang, X., & Yao, Y. (2023). Online dialogue with medical professionals: An empirical study of an online “Ask the Doctor” platform. International Journal of Medical Informatics, 177, 105123.

[6] XU, W. L., WANG, W. T., & WANG, X. D. (2020). Effect of of Internet plus medical assistance on quality of life among residents in rural poverty-stricken regions in Hainan province. Chinese Journal of Public Health, 36(5), 693-696.

[7] Dong, J., Pan, H., Su, H., & Xie, J. (2022, December). “Internet Plus” Community Home-Based Elderly Care Services Combining Medicine and Sports. In 2022 2nd International Conference on Public Management and Intelligent Society (PMIS 2022) (pp. 853-863). Atlantis Press.

[8] Lijuan, Y. A. N. G., Yanju, T. A. O., Jiali, H. A. O., Lijun, Y. A. N. G., & Huifang, H. U. A. N. G. (2025). Application value of medical care combined with Internet plus+ continuing care in continuing care of early preterm infants. Journal of Bengbu Medical University, 50(3), 381-384.

[9] Hu, L., Chen, H., Mo, W., Han, Y., & Sun, H. (2024). Best evidence summary for management of older people with type 2 diabetes Mellitus using ‘Internet plus nursing services’. International journal of older people nursing, 19(6), e12657.

[10] He, J. (2023). The impact of users' trust on intention to use the mobile medical platform: Evidence from China. Frontiers in Public Health, 11, 1076367.

[11] Lai, P., & Jiang, J. (2024, January). Analysis of intention to use the Internet medical care platform from the perspective of digital economy. In Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence (pp. 214-220).

[12] Roettl, J., Bidmon, S., & Terlutter, R. (2016). What predicts patients’ willingness to undergo online treatment and pay for online treatment? Results from a web-based survey to investigate the changing patient-physician relationship. Journal of medical Internet research, 18(2), e32.

[13] You, J., Chen, Y., & Huang, Y. (2022, October). Analysis of the impact of smart medical platform construction on medical willingness: ——Based on quantitative analysis of urban residents in Wuhan. In Proceedings of the 3rd International Symposium on Artificial Intelligence for Medicine Sciences (pp. 185-189).

[14] Wang, H., Zhang, J., Luximon, Y., Qin, M., Geng, P., & Tao, D. (2022). The determinants of user acceptance of mobile medical platforms: An investigation integrating the TPB, TAM, and patient-centered factors. International Journal of Environmental Research and Public Health, 19(17), 10758.

[15] Lin, S. H., & Lin, T. M. (2018). Demand for online platforms for medical word-of-mouth. Journal of International Medical Research, 46(5), 1910-1918.

[16] Gong, Y., Wang, H., Xia, Q., Zheng, L., & Shi, Y. (2021). Factors that determine a patient's willingness to physician selection in online healthcare communities: a trust theory perspective. Technology in Society, 64, 101510.

[17] Bidmon, S., Terlutter, R., & Röttl, J. (2014). What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. Journal of medical Internet research, 16(6), e3122.

[18] Menghan, W. A. N. G., & Weihua, F. A. N. G. (2020). Influence Factors of Public Willingness to Use Internet Medical Service Platform. Journal of Beijing University of Aeronautics and Astronautics Social Sciences Edition, 33(3), 150-156.

[19] Xiong, J., & Zuo, M. (2023). Factors influencing health care professionals’ adoption of mobile platform of medical and senior care in China. Information & Management, 60(5), 103798.

[20] Zhu, J., Jiang, D. W., & Wu, Y. (2020). Application of Internet Plus Education in Medical Technical Training. TMR Theory and Hypothesis, 3(1), 285-289.

[21] Ma, X., Qin, L., He, C., & Huang, Y. (2025). Long-Term Evaluation of the Effects of An'Internet Plus' Continuous Care Platform in Chronic Disease Management. Computational Molecular Biology, 15.

[22] Fu, H., Zhou, G., Wang, M., & Chen, Y. (2025). Research on Internet plus continuing nursing under dual heart medical model after percutaneous coronary intervention. Medicine, 104(24), e42778.

[23] Jinfang, Y. U. A. N., Jun, L. Y. U., Jinmei, Q. I. N., Jian, Y. A. N. G., & Xiaofang, X. U. (2024). Intervention effect of" Internet plus" postpartum health care service on postpartum depression. Shanghai Journal of Preventive Medicine, 36(9), 899-903.

[24] Yang, L., Pang, J., Zuo, S., Xu, J., Jin, W., Zuo, F., ... & Sun, X. (2024). Evolution of the “Internet Plus Health Care” Mode Enabled by Artificial Intelligence: Development and Application of an Outpatient Triage System. Journal of Medical Internet Research, 26, e51711.

[25] Tian, Z., Li, X., Wang, S., Xu, B., Ding, W., & Sun, Y. (2020, April). The Mechanism and Role of “Internet plus” in Promoting Medical Institutions to Achieve Social Governance--Taking Medical Colleges in Qigihar as an Example. In Journal of Physics: Conference Series (Vol. 1533, No. 4, p. 042060). IOP Publishing.

[26] Liang, G., Jiang, H., Huang, C., Que, X., Tang, J., Lu, J., & Gao, J. (2020). Diabetes health management strategy based on internet plus graded diagnosis and treatment strategy. Annals of Palliative Medicine, 9(6), 3915922-3913922.

[27] Liu, H., Zhang, Y., Li, Y., & Albright, K. (2023). Better interaction performance attracts more chronic patients? Evidence from an online health platform. Information Processing & Management, 60(4), 103413.

[28] Ye, P., & Liu, L. (2021). Factors affecting user intention to pay via online medical service platform: Role of misdiagnosis risk and timeliness of response. International Journal of Healthcare Information Systems and Informatics (IJHISI), 16(4), 1-26.

[29] Feng, C. L., Cheng, Z. C., & Huang, L. J. (2019). An investigation into patient privacy disclosure in online medical platforms. Ieee Access, 7, 29085-29095.

[30] Guo, Y., Zu, L., Chen, D., & Zhang, H. (2023). Research on influencing factors of satisfaction with the use of public health internet platform: evidence from Ding Xiang Doctor (DXY) internet medical platform. International Journal of Environmental Research and Public Health, 20(3), 2276.
[31] Guixian Xu,Xu Gao,Xin Zhang & Xiaobing Zhao. (2011). Improved TFIDF weighting for imbalanced biomedical text classification. Energy Procedia,11,2360-2367.
[32] Jaya Mabel Rani A. & Pravin A. (2023). Clustering by hybrid K-Means and black hole entropic fuzzy clustering algorithm for medical data. International Journal of Modeling, Simulation, and Scientific Computing,14(04),
[33] Yastrebov Alexander,Kubuś Łukasz & Poczeta Katarzyna. (2022). Multiobjective evolutionary algorithm IDEA and k-means clustering for modeling multidimenional medical data based on fuzzy cognitive maps. Natural Computing,22(3),601-611.
[34] Meenakshi,H. Sirmathi & J. Anitha Ruth. (2019). Cloud computing-based resource provisioning using k-means clustering and GWO prioritization. Soft Computing,23(21),10781-10791.