

Exploring factors affecting patient satisfaction in online healthcare: A machine learning approach grounded in empathy theory

Junbai Chen^{1,*} , Guoping Wu^{1,*}, Tong Zhang¹, Butian Zhao¹,
Ruoja Wang¹, Xing Zhai¹ and Fengying Guo¹

Abstract

Objective: Empathy between doctors and patients is crucial in enhancing patient satisfaction with medical consultations. This study, grounded in empathy theory, employs natural language processing and machine learning algorithms to explore the factors influencing patient satisfaction in online healthcare services, particularly the impact of doctor–patient empathy.

Methods: Utilizing the three dimensions of the Jefferson Scale of Physician Empathy, seven variables were extracted from patient–doctor dialogs as independent variables, with patient satisfaction as the dependent variable. Employing machine learning algorithms, a classification model was constructed to identify the best-fitting model for exploring the pivotal factors influencing patient satisfaction in online medical services. The optimal model was then chosen to investigate the essential factors impacting patients' satisfaction with online healthcare.

Results: A total of 7586 data points were collected, with 5447 consultation dialogs (71.8%) receiving a satisfactory rating from patients. LightGBM emerged as the best-performing model, achieving an F1 score of 0.78 and an area under the curve value of 0.81. Factors within the Standing in Patient's Shoes and Perspective Taking dimensions were identified as key determinants of patient satisfaction in online healthcare services.

Conclusion: This study broadens the conventional scope of applying empathy theory, signifying its crucial role in cultivating doctor–patient empathy within the realm of online healthcare and elevating the overall quality of medical services. The findings indicate that two pivotal factors influencing patients' satisfaction with online healthcare are doctors' perceived competence and ability to empathize, understanding patients' perspectives, and offering assistance.

Keywords

Empathy, machine learning, online medical services, patient satisfaction, influencing factors

Submission date: 3 July 2024; Acceptance date: 6 December 2024

Introduction

Empathy, synonymous with sympathy, originates from Rogers, the progenitor of humanistic psychology. It entails contemplating and addressing issues from another individual's perspective, manifesting as a sensitive, immediate, and dynamically evolving emotional comprehension of the other person.¹ Mercer and Reynolds delineate clinical empathy as the capacity to grasp a patient's present circumstances, viewpoint, and emotions and to express this comprehension to the patient. This proficiency contributes to improved therapeutic outcomes achieved through precise

treatment.² The Jefferson Scale of Physician Empathy (JSPE) is a psychometric tool developed by Professor Mohammadreza Hojat and colleagues at the Jefferson Medical College in the United States, based on extensive

¹Beijing University of Chinese Medicine, Beijing, China

*Junbai Chen and Guoping Wu are cofirst authors.

Corresponding author:

Fengying Guo, Beijing University of Chinese Medicine, Beijing 102488, China.

Email: guofy@bucm.edu.cn



research literature.³ The scale includes three dimensions: Perspective Taking (PT), Compassionate Care (CC), and Standing in Patient's Shoes (SPS).^{4,5} Scholars have constructed a "PT-CC-SPS" three-dimensional model based on JSPE to analyze empathy in physician–patient communication texts.^{1,6–8} Research has shown that medical professionals' empathy is primarily reflected in patiently listening and deeply understanding patients' emotional experiences, accurately expressing their understanding of the patients, adopting beneficial treatment plans, and reaching a consensus with the patients.⁹ During the medical process, doctors should carefully listen to patient's needs, use an appropriate amount of medical terminology, thoroughly explain the issues of concern to patients,¹⁰ and respond promptly and politely with content closely related to patients' concerns.

Previous research on physician–patient empathy mainly focuses on traditional diagnostic processes. Jane et al. employed semistructured qualitative interviews to study empathy levels in primary care patients during telephone consultations.¹¹ Andrea et al. used questionnaires to demonstrate that chronic pain patients with higher perceived physician empathy have greater treatment satisfaction and fewer symptoms of depression and anxiety.¹² Amirreza et al. surveyed 211 adult patients, proving that perceived empathy helps reduce pain intensity.¹³ The measurement methods in these studies primarily involve questionnaires and interviews, which are flexibly designed based on the target characteristics and provide a rich variety of data.¹⁴ However, these methods may be subject to bias due to the respondents' subjective factors. Compared to post-event questionnaires and interviews, real-world physician–patient interactions can minimize the interference of patients' subjective factors. Additionally, while previous studies often used traditional statistical methods for small-scale data analysis, machine learning methods, with their flexible and efficient data processing and modeling capabilities, are better suited to handling large-scale, highly complex data.¹⁵

The rapid development of network technology has advanced the maturity of Internet interaction models, with online medical platforms leveraging the real-time, free, and open characteristics of the Internet to establish effective communication bridges between doctors and patients. This has gradually formed the typical paradigm of "Internet + Healthcare." As of June 2021, the number of Internet medical users in China reached 239 million, accounting for 23.7% of Internet users, indicating that Internet-based online healthcare has already achieved a significant scale.¹⁶ However, due to the unique nature of online medical consultations, the quality of services provided by different platforms and doctors varies greatly, resulting in generally low patient satisfaction.¹⁷ The factors influencing patient satisfaction in this context are therefore of significant research value. Compared to traditional consultations, online medical users often desire more psychological care¹⁸ and enhancing doctors' empathy in online settings can

improve patient satisfaction.¹⁴ Due to the limitations of the interaction method, empathy in online medical scenarios is often overlooked. Research has shown that inappropriate expressions by doctors during communication can lead to a lack of empathy, causing negative emotions such as anxiety and tension in patients, ultimately affecting their satisfaction with the medical service.¹⁹ In-depth research on the factors influencing patient satisfaction in online healthcare is crucial for improving the quality of online medical services.

While previous studies have demonstrated the significance of empathy in traditional consultations, there is a notable gap in understanding how empathy can be effectively measured and analyzed in online medical consultations, where real-world interactions are less prone to subjective bias and where the scale of data requires more advanced analytical techniques. This study addresses this research gap by exploring the factors influencing satisfaction with online consultations on medical platforms from the perspective of empathy theory. This study is based on the previously proposed "PT-CC-SPS" three-dimensional model, focusing on online healthcare platforms. It extracts relevant variables from doctor–patient dialogue texts and employs machine learning algorithms to replace traditional statistical methods, constructing a classification model to analyze influencing factors. This study aims to more accurately represent physician–patient empathy in the online medical environment and analyze its impact on patient satisfaction. This approach provides a new avenue for improving the quality of online medical services. The overall research approach and methodology are illustrated in Figure 1.

Methods

Data source

The data for this study were sourced from a well-known online medical platform in China. Using a Python-based web scraping algorithm, we collected information such as doctors' details, physician–patient dialogue records, and patient evaluation texts. This included data fields such as doctor ID, dialogue ID, dialogue role, dialogue content, dialogue time, and consultation satisfaction ratings. From September 2016 to September 2021, we obtained a total of 8927 records, encompassing 416,548 doctor–patient dialogue texts. We removed entries from the patient evaluations that lacked either a doctor ID or a dialogue ID. For dialogs with matching doctor IDs and dialogue IDs and occurring within 24 h of each other, we considered them to belong to the same consultation session. The patient consultation texts and the corresponding doctor reply texts from these sessions were merged and stored, resulting in a dataset containing multiple consultation records. Each record in this dataset includes the doctor ID, dialogue ID,

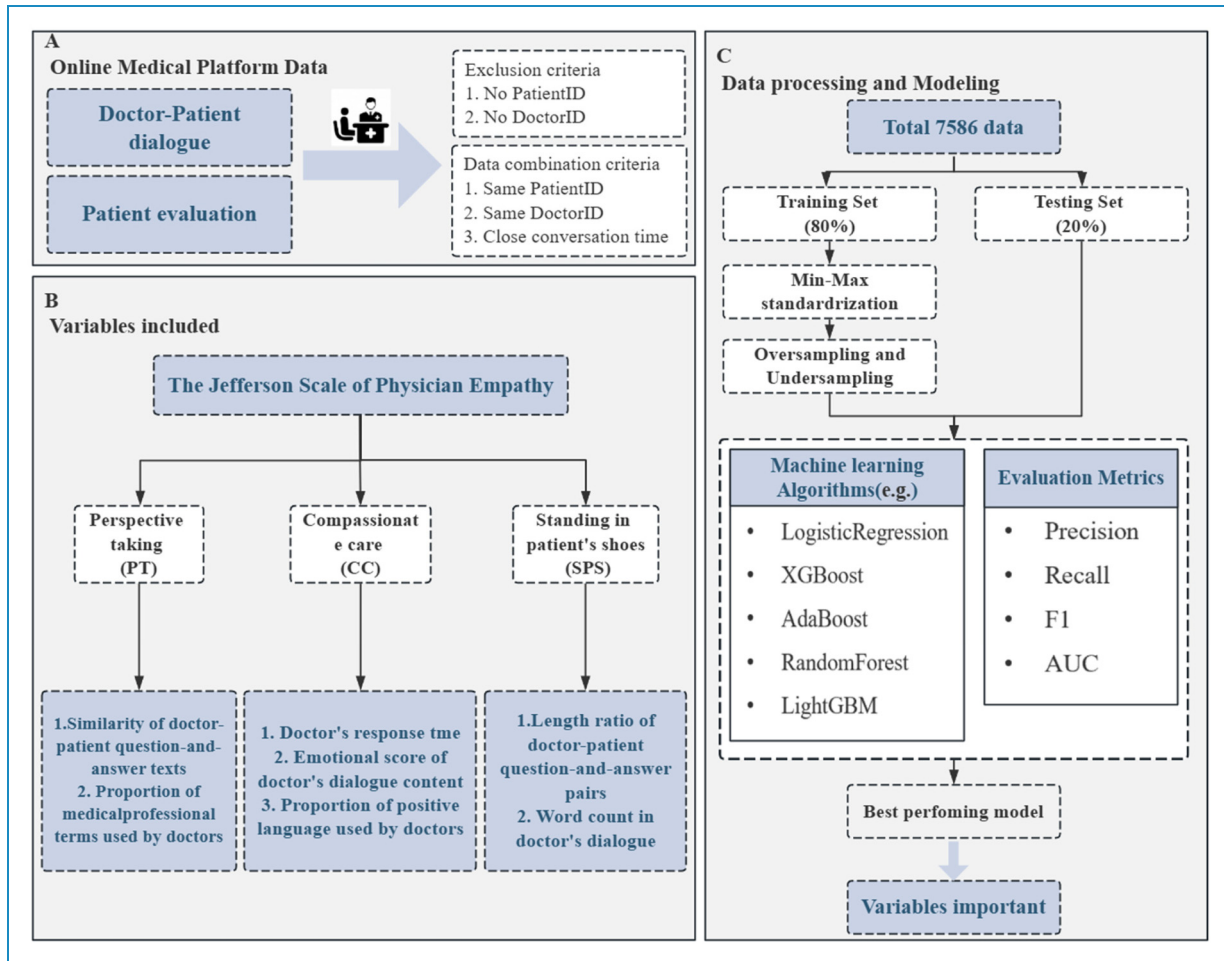


Figure 1. Graphical illustration of research approach and methodology.

patient consultation text, doctor reply text, patient consultation time, doctor reply time, and patient satisfaction rating. The final dataset consists of 7586 records, with 5447 (71.8%) of the consultations rated as satisfactory by patients. Sample data are presented in Table 1.

Feature variable construction

This study employed the “PT-CC-SPS” three-dimensional model from the JSPE to construct an evaluation index system for online consultation satisfaction.^{4,5} In the JSPE, the PT dimension refers to the physician’s ability to view issues from the patient’s perspective, reflecting whether the physician understands and adopts the patient’s viewpoint. Therefore, we selected the Similarity of doctor–patient question-and-answer texts and the Proportion of medical professional terms used by doctors as variables for this dimension. The Similarity of doctor–patient question-and-answer texts measures the extent to which the physician fully understands and addresses the patient’s concerns, while the Proportion of medical professional

terms used by doctors reflects the physician’s ability to adjust their language to match the patient’s level of understanding. The CC dimension refers to the expression of emotions in patient care and understanding of the patient’s experiences, reflecting whether the physician demonstrates care and empathy toward the patient and maintains a positive attitude. For this dimension, we selected the Doctor’s response time, the Emotional score of doctor’s dialogue content, and the Proportion of positive language used by doctors as variables. The Doctor’s response time indicates the physician’s attentiveness to the patient’s concerns, demonstrating care. The Emotional score of doctor’s dialogue content quantifies the intensity of care or empathy expressed in the physician’s words, and the Proportion of positive language used by doctors reflects the physician’s tendency to encourage or comfort the patient. The SPS dimension refers to the physician’s willingness to invest sufficient time and effort to consider the patient’s needs. For this dimension, we chose the Length ratio of doctor–patient question-and-answer pairs and the Word count in doctor’s dialogue as variables. The Length ratio of

Table 1. Example of patient-doctor dialogue data.

Doctor ID	Patient ID	Patient's dialogue	Patient's inquiry time	Doctor's dialogue	Doctor's response time	Patient evaluation
clinic_web_fa2e2aadbe99583	722317092	The director is encountering a new issue. The herbs arrived yesterday, and since they came in the afternoon, I consumed half of the old formula along with the new one. Consequently, my sleep duration has shortened. The first time I woke up, it was almost three and a half hours; today it's two and a half hours! The second time I woke up was at 2 and a half hours, and now it's 2 h! Could this be due to the combination of the old and new remedies, resulting in reduced effectiveness? Or is it possible that the ingredients for nourishing Yin and clearing heat have been diminished?	2021/1/19 9:01	Greetings! Thank you for your trust. Please don't be anxious; insomnia often stems from a yin and yang imbalance. Your reduction in dryness and heat indicates an adjustment, but presently, you are still experiencing liver qi stagnation and spleen and kidney deficiency, contributing to poor sleep quality. Some repetition is normal. In a recent study pressure and mental stress may also be influencing your sleep negatively. Please try to relax.	2021/1/19 9:08	Satisfied

Note. All original texts in this study were in Chinese. For ease of understanding and presentation, the dialogue texts in the samples have been translated into English.

doctor-patient question-and-answer pairs measures whether the physician has invested adequate time and effort in the interaction, providing detailed responses to the patient's questions, while the Word count in doctor's dialogue reflects whether sufficient effective information was provided.

This study utilized Python for indicator extraction and computation. The process involved several steps: First, we obtained a stop word list (<https://github.com/goto456/stopwords>) and a medical word list (<https://github.com/WENGSYX/Chinese-Word2vec-Medicine>). We then employed the Jieba segmentation tool to tokenize both patient's dialogue and doctor's dialogue, removing stop words in the process. Next, we used the xmnlpv0.5.0 natural language processing tool (<https://github.com/SeanLee97/xmnlp>), which includes functions for sentiment analysis and sentence similarity computation, to analyze the

processed dialogs. Details of the process and computation methods are illustrated in Figure 2. The sentiment analysis model in xmnlp is a Naive Bayes model trained on e-commerce review corpora, which provides the probability of a sentence or word having a positive sentiment. Sentence similarity is computed using SentenceBERT, a large-scale pretrained model that represents sentences and calculates cosine similarity between them. For example, in Table 1, we calculated the Doctor's response time by subtracting the Patient's inquiry time from the Doctor's response time, resulting in a response time of 7 min. The Jieba segmentation tool was used to segment both the Patient's Dialogue and Doctor's Dialogue, with a medical word list ensuring that specialized medical terms were correctly segmented. Stop words were removed using a stop word list. Next, the SentenceVector() function from the xmnlp natural language processing tool was employed to vectorize

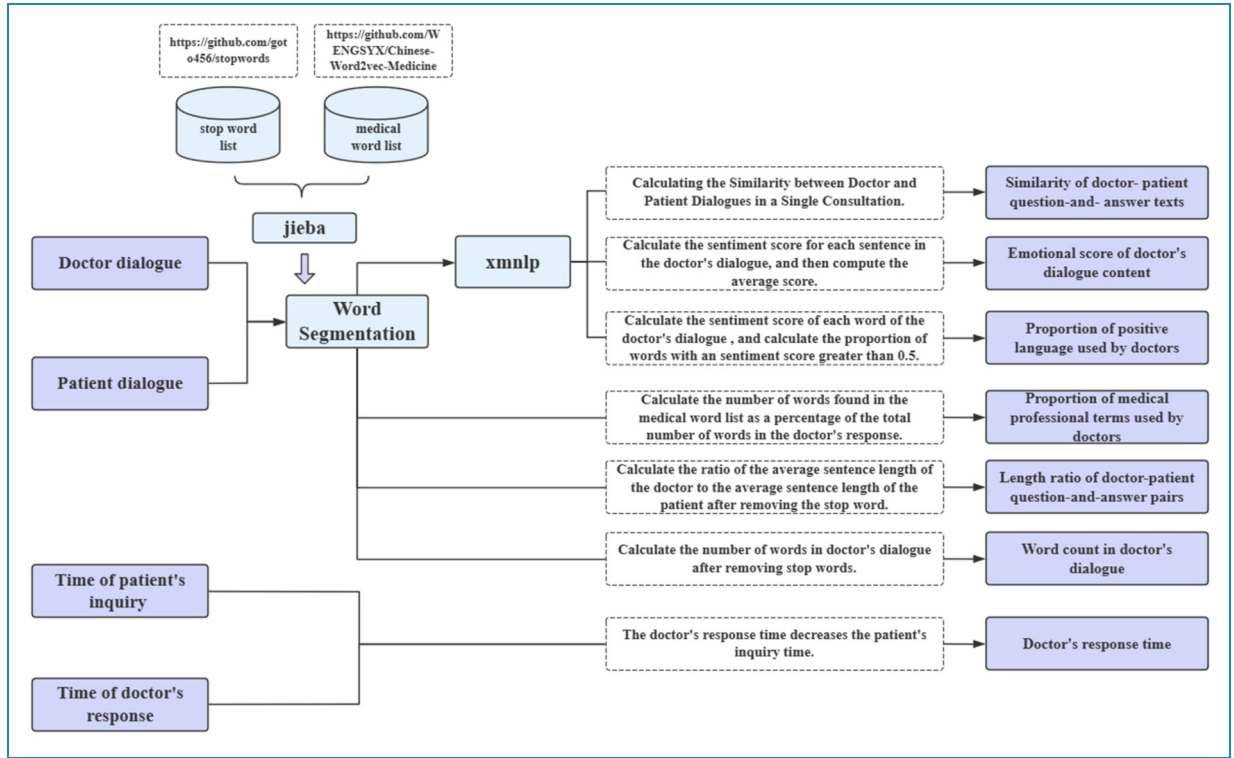


Figure 2. Variable extraction and computation methods.

the Patient's Dialogue and Doctor's Dialogue after stop word removal. The cosine similarity between the two dialogs was calculated, yielding a Similarity of doctor-patient question-and-answer texts of 0.787. After word segmentation and stop word removal, the total number of words in the Doctor's Dialogue was calculated, resulting in a Word count in Doctor's Dialogue of 316. The sentiment() function was then used to compute the emotional score for each word in the Doctor's Dialogue, and words with a score higher than 0.5 (e.g. "thank you," "progress") were counted, totaling 8. Using this with the word count, we calculated the Proportion of positive language used by doctors to be 0.025. Each word in the Doctor's Dialogue (after stop word removal) was then checked against the medical word list. The number of words matching medical terms (e.g. "prescription," "diarrhea") was 16, and this was combined with the total word count to calculate the Proportion of medical professional terms used by doctors, which was 0.051. Finally, the len() function in Python was used to compute the total length of both the Patient's Dialogue and the Doctor's Dialogue. The length ratio of doctor-patient question-and-answer pairs was calculated by dividing the length of the Patient's Dialogue by the length of the Doctor's Dialogue, resulting in a value of 0.731. A sample of the processed data is shown in Table 2.

To avoid the adverse effects of multicollinearity on the predictive performance of the machine learning model

and ensure its validity, a correlation test among the variables was necessary,²⁰ a Pearson correlation analysis was conducted on the cleaned final data, as illustrated in Figure 2. The correlation coefficient between the number of words in the physician's dialogue and the text similarity of the physician-patient dialogue was 0.42, while the Pearson correlation coefficients between other variables were all less than 0.4. This indicates that the correlation between the feature variables is low. Therefore, all features were included in the analysis, with patient satisfaction following online medical consultation as the dependent variable and the seven custom features as independent variables to construct a binary classification prediction model (Figure 3).

Data preprocessing and descriptive statistics

We applied the Min-Max normalization method to the raw data, performing a linear transformation to map the values to the 0 to 1 range. This ensures that the differences between various indicators are not overly large, preventing lower-value indicators from diminishing the accuracy of the results and thus improving the classifier's accuracy. The Min-Max normalization formula is as follows:

$$y_i = \frac{x_i - \min_{1 \leq j \leq n} \{x_j\}}{\max_{1 \leq j \leq n} \{x_j\} - \min_{1 \leq j \leq n} \{x_j\}} \quad (1)$$

Table 2. Example of data after processing.

Doctor ID	Patient ID	Similarity of doctor-patient question-and-answer texts	Doctor's response time	Emotional score of doctor's dialogue content	Proportion of positive language used by doctors	Length ratio of doctor-patient question-and-answer pairs	Proportion of medical professional terms used by doctors	Word count in doctor's dialogue	Satisfaction level
clinic_web_fa2e2aadb99583	722317092	0.787	7	0.276	0.025	0.731	0.051	316	1
clinic_web_fa2e2aadb99583	722171209	0.715	5	0.363	0.029	0.292	0.046	174	1
clinic_web_fa2e2aadb99583	722223360	0.717	5	0.386	0.027	0.542	0.044	183	1
clinic_web_fa2e2aadb99583	722292920	0.785	223	0.489	0.028	1.886	0.060	249	1
clinic_web_fa2e2aadb99583	722230733	0.764	223	0.460	0.034	0.578	0.069	203	1
clinic_web_fa2e2aadb99583	722243867	0.780	19	0.440	0.065	0.396	0.087	321	1
clinic_web_fa2e2aadb99583	722206837	0.570	5	0.447	0.030	32.119	0.059	372	1
clinic_web_fa2e2aadb99583	722037375	0.767	18	0.325	0.042	0.295	0.079	824	1
clinic_web_fa2e2aadb99583	722134128	0.767	4	0.419	0.051	2.140	0.071	673	1

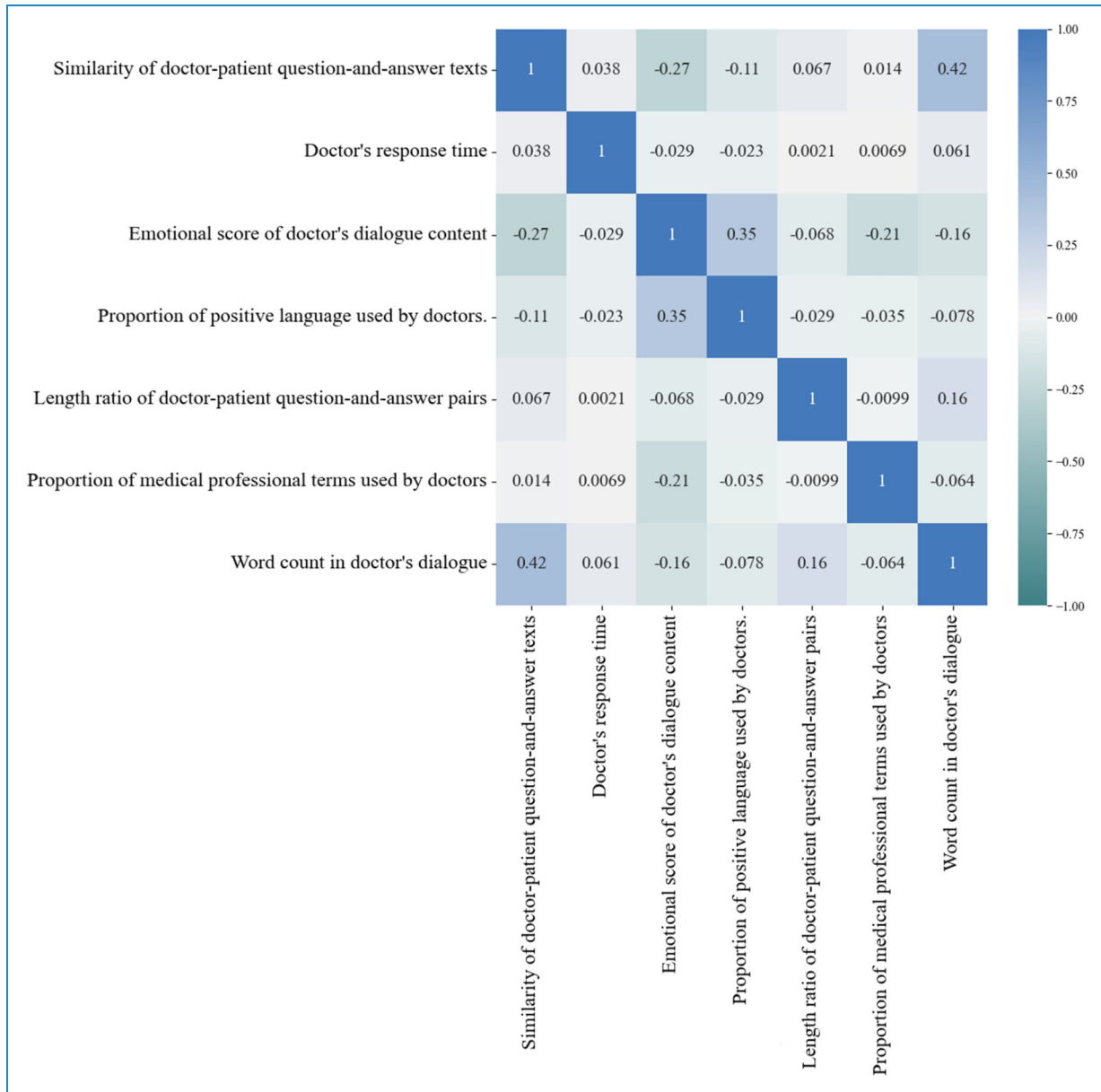


Figure 3. Correlation heatmap for variables used in the machine learning model.

The dataset was randomly split into training and test sets with an 8:2 ratio, aiming to maintain consistent label distributions within each subset. This was achieved by using a random seed for reproducibility and ensuring stratification of labels (implemented using the `train_test_split()` function with the `stratify` parameter). The training set was used for model training, while the test set was used for model evaluation and selection. Considering platforms serving patients from different age groups or regions may exhibit distinct satisfaction trends, the imbalance in the distribution of positive and negative samples (approximately 7:3) may affect the generalizability of the model across platforms with different patient demographics, so it was necessary to address

this issue through advanced resampling techniques. The SMOTE method synthesizes new minority class samples through linear interpolation between minority class samples and their k -nearest neighbors. While this method alleviates information redundancy caused by random oversampling to some extent, it also tends to produce overlapping and noisy samples.²¹ The SMOTEENN algorithm, a combination of SMOTE and ENN algorithms, employs the ENN method for deep cleaning of the data generated by SMOTE. It has been shown to perform better than other classical sampling methods on several standard datasets.²² In this study, the SMOTEENN algorithm was used to perform both undersampling and oversampling on the

training set data—repeatedly selecting the minority class and randomly selecting a small portion of the majority class. Specific parameter settings are provided in Table 3. Consequently, the processed training data had a more balanced distribution of positive and negative samples (approximately 3:2), which reduced the impact of data imbalance on the classifier's prediction results.

Descriptive statistical analysis was performed on the preliminarily cleaned data. All independent variables were continuous quantitative variables and followed a normal distribution. However, the test for homogeneity of variances between groups failed. Therefore, the data were described using means and standard deviations, and non-parametric tests were used for group comparisons.

Model construction, training, evaluation, and analysis

Firstly, multiple independent classification models were constructed using Python 3.7, including Logistic Regression, XGBoost, AdaBoost, RandomForest, and LightGBM, with all models utilizing default parameter settings. Considering the SMOTEENN could produce noisy or less distinctive samples, potentially affecting model precision. To counterbalance this risk, these models were trained using 10-fold cross-validation, where the training set was further divided into 10 subsets. In each iteration, nine subsets were used for training, and the remaining one subset was used as the validation set. This process was repeated 10 times, with a different subset used as the validation set in each iteration. The final result was obtained by averaging the outcomes of these 10 validations. Subsequently, precision, recall, and F1 scores were used as evaluation metrics (formulas 1–3), where TP denotes the number of true positive instances, FP denotes the number of false positive instances, and FN denotes the number of false negative instances. Additionally, receiver operating characteristic (ROC) curves were plotted for

each model, and the area under the ROC curve (AUC-ROC) was computed to further assess the accuracy of different models. Finally, the best-performing model was selected, and the SHAP (Shapley Additive Explanation) algorithm was employed for model interpretation and feature importance analysis. Shapley Additive Explanation is a machine learning explanation algorithm based on cooperative game theory, calculating additive contributions of each feature to represent its impact on the model's predictions²³:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Results

Descriptive statistics

Based on the descriptive statistics in Table 4, all variables exhibit a *p*-value below 0.05, signifying the statistical significance of the difference in patient satisfaction. Descriptive statistics were further performed for selected outcome indicators related to patient satisfaction. The results revealed that out of the total, the patients rated 5447 questioning dialogs (71.8%) as satisfactory.

Table 4. Descriptive statistics of each variable.

Variables	Mean	Standard deviation	<i>p</i> value
Similarity of doctor–patient question-and-answer texts	0.713	0.112	<0.001
Doctor's response time	15.997	59.251	0.012
Emotional score of doctor's dialogue content	0.389	0.106	<0.001
Proportion of positive language used by doctors	0.034	0.023	<0.001
Length ratio of doctor–patient question-and-answer pairs	3.012	7.758	<0.001
Proportion of medical professional terms used by doctors	0.063	0.036	<0.001
Word count in doctor's dialogue	290.433	267.327	<0.001

Table 3. SMOTEENN algorithm parameter settings.

Hyperparameters	Implication	Value
n_classes	Number of categories	2
class_sep	Degree of separation between categories	2
weights	Sample weights for each category	0.7,0.3
n_informative	Number of informative features	2
random_state	Seeds for random number generators	1

Performance of models

Table 5 presents the classification results of the machine learning algorithms. Among the models, LightGBM stands out as the top performer, achieving an F1 value of 0.78 and an AUC value of 0.81. The ROC curves for the five machine learning models are illustrated in Figure 4. Compared to traditional statistical models, machine learning algorithms can more comprehensively capture complex and nonlinear relationships.¹⁵ However, due to differences in their structures, different models may exhibit varying performance when addressing the same problem. As shown in Table 5, the LightGBM model achieves a prediction accuracy of 78%, a recall rate of 79%, an F1 score greater than 0.78, and an AUC greater than 0.81, all of which are the highest values. The four ensemble learning models—XGBoost, AdaBoost, RandomForest, and LightGBM—demonstrate performance improvements over the single model LogisticRegression, indicating that integrating multiple weak classifiers helps enhance classifier accuracy.^{24,25} Among these models, LightGBM performed the best. While XGBoost uses a layer-wise tree growth strategy that may be slower for large-scale data, LightGBM adopts a leaf-wise growth approach, which can lead to faster training times and potentially better performance in handling large datasets.²⁶ AdaBoost, while useful in focusing on difficult-to-classify instances, is more sensitive to noise and data imbalance, which may have led to slightly lower performance in this dataset.²⁷ RandomForest, despite being robust in reducing overfitting through multiple decision trees, lacks the boosting mechanism present in LightGBM and is less efficient in capturing the finer patterns in complex, nonlinear data.^{26–28} LightGBM's leaf-wise growth strategy, histogram-based algorithm, and ability to handle sparse features allowed it to efficiently model intricate relationships without overfitting, making it particularly well-suited to the dataset's complexity and scale.

Table 5. Summary performance measures of predictive models in the testing dataset.

Model	Precision	Recall	F1	AUC
LogisticRegression	0.74	0.76	0.72	0.78
XGBoost	0.76	0.77	0.76	0.79
AdaBoost	0.77	0.78	0.77	0.80
RandomForest	0.78	0.79	0.77	0.80
LightGBM	0.78	0.79	0.78	0.81

AUC: area under the curve.

Shapley Additive Explanation-based model interpretation analysis

The best-performing classification model (LightGBM) was interpreted using SHAP values to analyze the importance of seven predictor variables. The results are depicted in Figure 5, where the horizontal axis represents SHAP values. A SHAP value less than 0 indicates a negative impact of the feature on patient satisfaction, while a value greater than 0 indicates a positive impact.²⁹ Each data point represents a sample, with color denoting the magnitude of the feature value—ranging from blue to red indicating low to high feature values. The width of each point reflects the impact of the feature on the result: wider points indicate a greater influence.

The analysis reveals that the top three most important features contributing to patient satisfaction are Word count in doctor's dialogue, Length ratio of doctor–patient question-and-answer pairs, and Similarity of doctor–patient question-and-answer texts. These features demonstrate a significant positive correlation with patient satisfaction, as indicated by their concentration of red points in the region where SHAP values are greater than 0. Specifically, a higher Word count in doctor's dialogue and Length ratio of doctor–patient question-and-answer pairs may reflect a doctor's effort to provide detailed explanations and address patient concerns thoroughly, leading to higher patient satisfaction. Similarly, a higher similarity of doctor–patient question-and-answer texts might indicate that the doctor effectively tailors responses to match patient queries, fostering a sense of understanding and empathy. Conversely, the two least important features are the proportion of positive language used by doctors and the Proportion of medical professional terms used by doctors. These variables show a negative correlation with patient satisfaction, albeit with a lower impact, as evidenced by their concentration of red points in the region where SHAP values are less than 0. This lower impact may be due to the nuanced nature of doctor–patient communication.³⁰ For example, an excessive use of positive language, while well-intentioned, could be perceived as insincere or overly simplistic, especially if not paired with substantive answers to patient concerns. Similarly, the use of medical terminology, while important for accuracy, might overwhelm patients without sufficient contextual explanation, thereby diminishing its effect.

Discussion

Doctor's perceptual ability as a key factor influencing patient satisfaction in online medical consultations

In this study, the predictive model for patient satisfaction highlights two key features: Word count in doctor's dialogue and Length ratio of doctor–patient question-and-answer

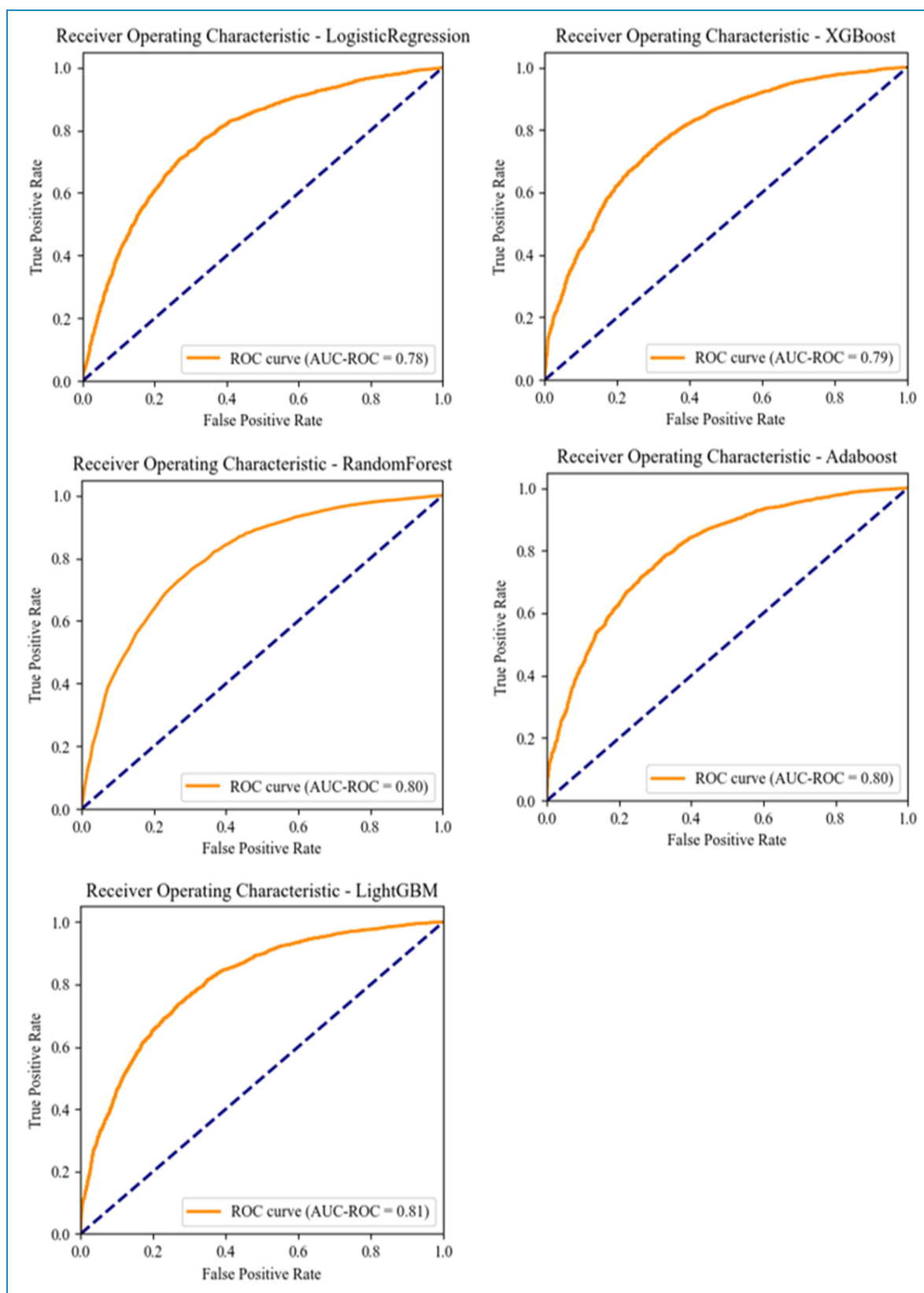


Figure 4. Receiver operating characteristic (ROC) curves for the individual predictive models (A) LogisticRegression (AUC: 0.78). (B) XGBoost (AUC: 0.79). (C) AdaBoost (AUC: 0.80) (D) RandomForest (AUC: 0.80). (E) LightGBM (AUC: 0.81). AUC: area under the curve.

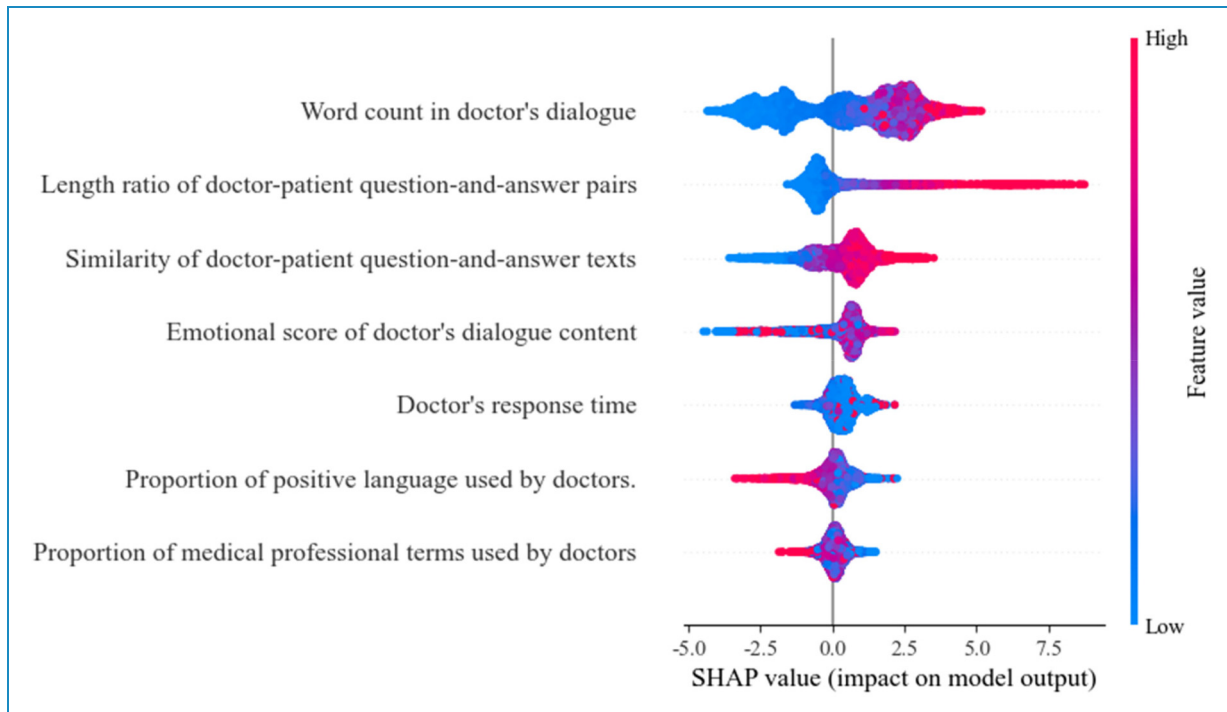


Figure 5. Shapley Additive Explanation (SHAP) variable importance plots by LightGBM model.

pairs. These features belong to the “SPS” dimension of empathy. In the context of online medical services, the perception of emotional attitudes by doctors and the application of empathy heavily rely on the expression of language and words.^{30,31} For instance, if a patient expresses negative emotions, a doctor who can perceive this and respond with comforting language such as “Please don’t worry” or “Rest assured, it’s not a major issue” demonstrates empathy effectively, thereby alleviating patient anxiety. The study results indicate a positive correlation between the Length ratio of doctor–patient question-and-answer pairs, Word count in doctor’s dialogue, and patient satisfaction. This suggests that longer dialogue by doctors contributes to higher patient satisfaction rates. However, the quality of the content, such as the doctor’s attentiveness, choice of words, and engagement with the patient’s concerns also plays a crucial role.

In one example, a patient inquired about the combination of medications for folliculitis and Liuwei Dihuang Pills. The doctor responded by advising against using Liuwei Dihuang Pills immediately and provided dietary and lifestyle recommendations, demonstrating an understanding of the patient’s concerns and providing concise advice. Despite the Length ratio of doctor–patient question-and-answer pairs being below average, the doctor successfully addressed the patient’s primary concerns, explained the reasons for discontinuing Liuwei Dihuang Pills, and offered suggestions to meet the patient’s needs, thus displaying empathy and enhancing the patient’s confidence in the treatment plan.³²

Conversely, some doctors, due to factors like busy schedules, may provide comprehensive answers and recommendations but fail to express care or greetings, thus neglecting the emotional needs of patients and lacking empathy. This could lead to patient distrust and lower satisfaction with online medical consultations.³³ For instance, in a doctor–patient interaction, a consultant detailed her mother’s history of hyperlipidemia, and recent symptoms such as dizziness, headache, and angina, seeking the doctor’s advice on the urgency of the situation. “The doctor replied: Hello, to be actively treated, first of all, it is recommended to take musk cardio-protection pills or nitroglycerin tablets sublingual, to relieve angina pectoris, take aspirin 100 mg once a day, to rest, cannot work. Secondly, you can choose the treatment of proprietary Chinese medicine, each time, four capsules, three times a day. After treatment, the pain can be relieved, and you can elective treatment, if the relief is not obvious, it is best to timely treatment to prevent the occurrence of a heart attack.” In this doctor–patient dialogue, the consultant expressed her urgency and concern about her mother’s condition at the same time as asking for a consultation, but the doctor focused only on proposing solutions, ignored the consultant’s emotional needs, and did not provide sufficient emotional comfort, failing to achieve full empathy, which ultimately led to low patient satisfaction with online healthcare.

Regarding another variable in the “SPS” dimension—Proportion of medical professional terms used by doctors

—this study found a negative correlation with patient satisfaction. Clinical medical terms are often concise and information-rich but may be challenging for ordinary patients to understand. Overuse of medical terminology can create communication barriers and hinder patient understanding.^{34,35}

To improve patient satisfaction in online healthcare, doctors should enhance their communication skills by incorporating empathetic language and actively acknowledging patient concerns, balancing time constraints by expressing care even in brief interactions. Simplifying medical jargon into layman's terms can also foster better understanding and trust. Online medical platforms can support these efforts by developing empathy training programs for doctors, implementing AI-powered tools to analyze and improve doctor–patient dialogs in real time, and providing patient-friendly resources, such as glossaries for common medical terms. Encouraging doctors to address patient-specific concerns thoughtfully, especially in complex or emotionally charged cases, can further ensure that patients feel heard and supported, ultimately translating into higher satisfaction and better healthcare outcomes.

Perspective-taking behavior enhances patient satisfaction

In terms of feature importance, Similarity of doctor–patient question-and-answer texts ranks third, falling under the “Perspective Taking” dimension of empathy. Perspective taking involves accurately understanding the thoughts of communication partners by overcoming one's self-centered viewpoint and attempting to understand a specific communication scenario from their perspective,³⁶ often reflected in linguistic similarity.³⁷ The study findings indicate that lower doctor–patient dialogue text similarity correlates with lower patient satisfaction rates. Conversely, higher similarity complicates determining patient satisfaction. Typically, when faced with various patient inquiries or consultations, doctors need to succinctly state and address each issue raised by the patient.³⁸ For example, in a particular doctor–patient dialogue, the patient primarily consulted on four symptoms: (a) Increased warmth in hands and feet during nighttime anxiety. (b) Waking up during sleep with frequent dreams towards waking. (c) Reduced tinnitus compared to previous experiences. (d) Yellowish urine, occasionally with heat. “The doctor responded: The warmth in hands and feet indicates Yin deficiency and excessive internal heat, related to anxiety and emotional instability. Difficulty sleeping, easy waking, and vivid dreams are signs of liver Qi stagnation disturbing the heart's spirit. Please don't worry about recurring issues; everything progresses in a spiral. Please maintain confidence in your current treatment. Good night.” In this interaction, the doctor addressed each symptom point by point,

emphasizing accurate understanding and providing explanations and care, making the patient more receptive to the doctor's concern and advice, thereby enhancing patient satisfaction.

However, some doctors achieve high dialogue text similarity by simply copying the patient's wording to describe symptoms without delving into explanations or providing concise summaries, failing to fully consider the patient's actual situation. For instance, a patient inquired about leg numbness following an abortion, whether they could consume Ginseng Bolus to invigorate the spleen, and treatments for pelvic inflammation, constipation, heavy dampness, and acne. The doctor replied: “Just had an abortion, Qi, and blood deficiency. You can eat some ginseng! Not recommended to eat more. You can take Shiquan Dabu Pills, they're not very harsh, take for half a month.” This response, although high in text similarity, was overly simplistic, lacked attention to the patient's specific questions, and failed to provide explanations, potentially leading the patient to perceive the doctor's attitude as perfunctory,⁸ undermining patient trust and resulting in a negative medical consultation experience.

To address this, doctors should focus on achieving meaningful dialogue text similarity by actively engaging with the patient's specific concerns and offering detailed explanations rather than mirroring the patient's wording without added value. Training programs could emphasize perspective-taking skills, guiding doctors to identify and address underlying patient needs effectively. Online platforms can assist by integrating tools that provide feedback on the quality of doctor–patient interactions, highlighting areas needing more thorough responses.

Doctors need to further enhance empathetic care during diagnosis and treatment

In our study, while three variables in the “CC” dimension showed a relatively low impact on patient satisfaction, they remain crucial and should not be overlooked.³⁹ Existing research indicates that using positive language and emotions effectively enhances patients' psychological care, promptly responding to patient queries, and helping alleviate negative emotions, thereby improving doctor–patient empathy and patient satisfaction.⁴⁰ This is because positive and supportive language conveys the doctor's concern, understanding, and support, and providing comfort and confidence to patients facing illness. Timely responses reduce patient wait times, avoiding unnecessary anxiety during the waiting period, aligning with our findings that emotional scores in doctor–patient dialogue and response times positively correlate with patient satisfaction. However, our study also found that excessively positive language does not necessarily enhance patient satisfaction; rather, it may lead patients to perceive the doctor's words as

insincere or overly formal, thereby reducing trust.⁴¹ Furthermore, doctors often focus more on showing concern for the medical condition rather than for the individual patient,¹⁰ potentially increasing the emotional distance between doctors and patients, which also hampers empathy development and patient satisfaction.

Therefore, doctors should be mindful of their language use during diagnosis and treatment, maintaining sincerity and patience while avoiding excessive positive language. Additionally, doctors should further enhance their comprehensive understanding and application of empathy theory and skills to improve patient satisfaction and healthcare service quality.

Conclusion

This study employed machine learning algorithms from the perspective of empathy theory to analyze factors influencing patient satisfaction in online healthcare settings, exploring how doctor–patient empathy connections are established. The findings suggest that beyond basic medical care, doctors should enhance their ability to perceive patients' needs, maintain a proactive and positive attitude, and provide thorough responses to questions. Additionally, doctors should strive to think from the patient's perspective and use language that resonates with patients to better meet their empathy needs, thereby improving patient satisfaction.

The main contributions of this study are as follows: first, it analyzes the impact of physician–patient empathy on patient satisfaction through the examination of dialogue texts from online medical platforms. Second, it applies machine learning models to the study of empathy in the online healthcare domain. Finally, it offers strategic recommendations to enhance physician–patient empathy in online healthcare, thereby improving patient satisfaction.

However, the study has limitations. One key limitation is the generalizability of the findings, as the data were sourced from a single online platform, which may not be representative of other healthcare platforms, geographic regions, or medical specialties. The absence of personal patient data, doctor-specific professional information, and audiovisual materials from doctor–patient interactions further limits the study's ability to capture the full spectrum of empathy-related dynamics. Additionally, the study could not fully address issues such as a class imbalance in the dataset, specifically the uneven distribution between satisfied and dissatisfied patient evaluations, potentially biasing positive predictions. Therefore, future research will aim to collect and integrate data from platforms that serve different geographic regions, socioeconomic backgrounds, or medical specialties, construct a more balanced dataset, and enhance the depth and breadth of the study. This will allow for more accurate capture of patient feedback, promoting fairer, more transparent healthcare services, and improved patient satisfaction.

Acknowledgments: This work is supported by, the Funds for the Industry-University-Research Innovation of Chinese Universities (grant numbers 2021LDA12004), Beijing University of Chinese Medicine Basic Research Fund (Open Bidding Leadership) Project (grant numbers 2023-JYB-JBZD-068), 2022 Educational Science Research Project of Beijing University of Chinese Medicine (grant numbers XJY22045), 2022 Basic Research Business Fund 'Top-Down Leadership' Project (grant numbers 2022JYBJBRW12), and 2024 Independent Research Projects for Postgraduate Students of Beijing University of Chinese Medicine (ZJKT2024033).

Contributorship: JC and GW participated in the method design, analyzed data, and drafted the initial manuscript. JC, GW, TZ, BZ, RW, XZ, and FG participated in text checking correction and helped to draft the manuscript. BZ, FG, and RW oversaw and provided input on all aspects of manuscript writing and the final analytical plan. All the authors read and approved the final manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process: During the preparation of this work, the author(s) used chatGPT in order to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: We would like to clarify that the data used in this study were obtained from publicly available sources on the Internet (e.g. https://www.chunyuyisheng.com/pc/qa/u1ZScDhXZhiPKuUA1ohn_A/). Patients who used this online healthcare platform provided consent for their consultation records to be publicly accessible as part of the platform's terms of use. Furthermore, the platform had removed all personal identifiers from the data prior to our use, ensuring patient anonymity and privacy. We confirm that the content of the paper submitted by us does not involve any human or animal experiments. Therefore, we are not required to provide any ethical review documents or approvals related to human or animal experiments.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Industry-University-Research Innovation of Chinese Universities, 2024 Independent Research Projects for Postgraduate Students of Beijing University of Chinese Medicine, Beijing University of Chinese Medicine Basic Research Fund (Open Bidding Leadership) Project, Ministry of Education Industry-University Cooperative Education Program of China, (grant number 2021LDA12004, ZJKT2024033, 2023-JYB-JBZD-068, 202102001001).

Guarantor: JC, RW, XZ, FG.

Statement of human and animal rights: Our research primarily focuses on theoretical exploration, literature review, and data analysis and does not involve direct intervention or research on biological entities, whether human or animal.

Statement of informed consent: The data used in this study were obtained from publicly available sources on the Internet. Patients who used this online healthcare platform provided consent for their consultation records to be publicly accessible as part of the platform's terms of use. There is no human subject in this article and informed consent is not applicable.

ORCID ID: Junbai Chen  <https://orcid.org/0009-0004-3536-7811>

References

- McNulty JP and Politis Y. Empathy, emotional intelligence and interprofessional skills in healthcare education. *J Med Imaging Radiat Sci* 2023; 54: 238–246.
- Wang X, Wang R, Sheng F, et al. The effects of empathy by caregivers on healthcare service satisfaction. *Front Psychol* 2022; 13: 912076.
- Mallory L, Floyed R, Doughty C, et al. Validation of a modified Jefferson scale of empathy for observers to assess trainees. *Acad Pediatr* 2021; 21: 165–169.
- Diaz Valentin MJ, Garrido Abejar M, Fuentes Chacon RM, et al. Validation to the Spanish of the Jefferson Empathy Scale health professions students version and its psychometric properties in nursing students. *Nurse Educ Pract* 2019; 40: 102629.
- Fashami FM, Nili M, Mottaghi M, et al. Measuring empathy in Iranian pharmacy students using the Jefferson Scale of Empathy-health profession student version. *Am J Pharm Educ* 2023; 87: ajpe8687.
- Huang C-W, Wu BCY, Nguyen PA, et al. Emotion recognition in doctor-patient interactions from real-world clinical video database: Initial development of artificial empathy. *Comput Methods Programs Biomed* 2023; 233: 107480.
- Schwartz R, Dubey M, Blanch-Hartigan D, et al. Physician empathy according to physicians: a multi-specialty qualitative analysis. *Patient Edu Couns* 2021; 104: 2425–2431.
- Surchat C, Carrard V, Gaume J, et al. Impact of physician empathy on patient outcomes: a gender analysis. *Br J Gen Pract* 2022; 72: e99–e107.
- Gutiérrez-Puertas L, Gutiérrez-Puertas V, Ortiz-Rodríguez B, et al. Communication and empathy of nursing students in patient care through telenursing: a comparative cross-sectional study. *Nurs Educ Today* 2024; 133: 106048.
- Cunico L, Sartori R, Marognoli O, et al. Developing empathy in nursing students: a cohort longitudinal study. *J Clin Nurs* 2012; 21: 2016–2025.
- Vennik J, Hughes S, Lyness E, et al. Patient perceptions of empathy in primary care telephone consultations: a mixed methods study. *Patient Edu Couns* 2023; 113: 107748.
- Too A, Gatien C and Cormier S. Treatment satisfaction mediates the association between perceived physician empathy and psychological distress in a community sample of individuals with chronic pain. *Patient Edu Couns* 2021; 104: 1213–1221.
- Fatehi A, Brown LE, Versluis Y, et al. The relationship of perceived empathy with levels of pain intensity and incapability among patients visiting a musculoskeletal specialist. *Patient Edu Couns* 2023; 115: 107900.
- Martikainen S, Falcon M, Wikstrom V, et al. Perceptions of doctors' empathy and patients' subjective health status at an online clinic: development of an empathic anamnesis questionnaire. *Psychosom Med* 2022; 84: 513–521.
- Deimazar G and Sheikhtaheri A. Machine learning models to detect and predict patient safety events using electronic health records: a systematic review. *Int J Med Inform* 2023; 180: 105246.
- Ren HY, Wang LA and Wu JJ. The faster or richer the response, the better performance? An empirical analysis of online healthcare platforms from a competitive perspective. *Decis Support Syst* 2024; 184: 114274.
- Huang Z and Yuan L. Enhancing learning and exploratory search with concept semantics in online healthcare knowledge management systems: an interactive knowledge visualization approach. *Expert Syst Appl* 2024; 237: 121558.
- Xiong Q, Luo F, Chen Y, et al. Factors influencing fatigue, mental workload and burnout among Chinese health care workers during public emergencies: an online cross-sectional study. *BMC Nurs* 2024; 23: 428.
- Terry C and Cain J. The emerging issue of digital empathy. *Am J Pharm Educ* 2016; 80: 58.
- Arora A, Bojko L, Kumar S, et al. Assessment of machine learning algorithms in national data to classify the risk of self-harm among young adults in hospital: a retrospective study. *Int J Med Inform* 2023; 177: 105164.
- Tang J, Wang X, Wan H, et al. Joint modeling strategy for using electronic medical records data to build machine learning models: an example of intracerebral hemorrhage. *BMC Med Inf Decis* 2022; 22: 278.
- Xing M, Zhang Y, Yu H, et al. Predict DLBCL patients' recurrence within two years with Gaussian mixture model cluster oversampling and multi-kernel learning. *Comput Methods Programs Biomed* 2022; 226: 107103.
- Park J, Kim J, Ryu D, et al. Factors related to steroid treatment responsiveness in thyroid eye disease patients and application of SHAP for feature analysis with XGBoost. *Front Endocrinol* 2023; 14: 1079628.
- Caruana A, Bandara M, Musial K, et al. Machine learning for administrative health records: a systematic review of techniques and applications. *Artif Intell Med* 2023; 144: 102642. 2023/10/03.
- Leiser F, Rank S, Schmidt-Kraepelin M, et al. Medical informed machine learning: a scoping review and future research directions. *Artif Intell Med* 2023; 145: 102676. 2023.
- Shi JC, Jia SL, Bai CX, et al. A novel approach for asphalt pavement invisible distress classification prediction by machine learning algorithms. *Int J Pavement Eng* 2024; 25: 2343087.
- Song YLQ, Yuan Q, Liu HQ, et al. Machine learning algorithms to predict mild cognitive impairment in older adults in China: a cross-sectional study. *J Affect Disorders* 2025; 368: 117–126.
- Ji X, Chang W, Zhang Y, et al. Prediction model of hypertension complications based on GBDT and LightGBM. *J Phys Conf Ser* 2021; 1813: 012008 (012008pp).

29. Tong H-J, Huang Z-M, Li Y-L, et al. Machine learning to analyze the factors influencing myopia in students of different school periods. *Front Public Health* 2023; 11: 1169128.
 30. Dwamena F, Holmes-Rovner M, Gaulden CM, et al. Interventions for providers to promote a patient-centred approach in clinical consultations. *Cochrane Database Syst Rev* 2013; 12: CD003267.
 31. Martikainen S, Falcon M, Wikstrom V, et al. Perceptions of doctors' empathy and patients' subjective health status at an online clinic: development of an empathic anamnesis questionnaire. *Psychosom Med* 2022; 84: 513–521. 2022.
 32. Barriere V and Balahur A. Multilingual multi-target stance recognition in online public consultations. *Mathematics* 2023; 11: 2161.
 33. Bernardi R and Wu PF. Online health communities and the patient-doctor relationship: an institutional logics perspective. *Soc Sci Med* 2022; 314: 115494.
 34. Meng FB, Liu YX, Zhang XF, et al. General knowledge-sharing and patient engagement in online health communities: an inverted U-shaped relationship. *J Knowl* 2024; 28: 763–788.
 35. Yang YB, Guo XT, Wu TS, et al. The effects of social media use and consumer engagement on physician online return: evidence from Weibo. *Internet Res* 2024; 34: 371–397.
 36. Kong L, Chen Y, Wang L, et al. Effect of perspective-taking on trust between doctors and patients: a randomized controlled trial. *J Clin Psychol Med Settings* 2023; 30: 708–715. 2023/02/07.
 37. Khedr S, D'Angelo D, Santos AT, et al. Identification of clinical risk factors affecting patient-physician communication. *J Surg Res* 2023; 282: 246–253.
 38. Chen S, Guo X, Wu T, et al. Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis. *Inf Process Manag* 2020; 57: 102253.
 39. Liu C, Li Y, Fang M, et al. Using machine learning to explore the determinants of service satisfaction with online healthcare platforms during the COVID-19 pandemic. *Service Business* 2023; 17: 449–476.
 40. Shaw SC, Carravallah L, Johnson M, et al. Barriers to health-care and a 'triple empathy problem' may lead to adverse outcomes for autistic adults: a qualitative study. *Autism* 2024; 28: 1746–1757.
 41. Blease C and Torous J. ChatGPT and mental healthcare: balancing benefits with risks of harms. *BMJ Mental Health* 2023; 26: e300884.
-