

What Services Do Patients Need from Internet Hospitals? An AI-Powered Text Classification Study of Patient Descriptions

Ruoxin Xu, Jing Liu, Tingting Yang, Huizi Ye, Kuanlie Jiang

Submitted to: Journal of Medical Internet Research
on: October 21, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 21

Figures 22

Figure 1..... 23

Figure 2..... 24

Figure 3..... 25

Figure 4..... 26

Figure 5..... 27

Figure 6..... 28

Figure 7..... 29

Figure 8..... 30

What Services Do Patients Need from Internet Hospitals? An AI-Powered Text Classification Study of Patient Descriptions

Ruoxin Xu¹; Jing Liu¹; Tingting Yang¹; Huizi Ye¹; Kuanlie Jiang²

¹ Yue Bei People's Hospital Shaoguan CN

² Hainan Medical University Sanya CN

Corresponding Author:

Jing Liu

Yue Bei People's Hospital

No. 133, Huimin South Road, Wujiang District, Shaoguan City, Guangdong Province, China

Shaoguan

CN

Abstract

Background: Internet hospitals have emerged as a innovative telemedicine model in China, providing online services like consultations, prescription renewals, and report interpretations. These platforms generate a vast amount of authentic, patient-written text describing symptoms and needs. While Natural Language Processing (NLP) and Large Language Models (LLMs) offer powerful tools for text analysis, a significant gap exists in applying these techniques to systematically decode and categorize the specific service intents within real-world clinical texts from internet hospitals. Most existing research relies on data from social media or forums, not from actual clinical practice.

Objective: This study aimed to leverage AI to analyze patient descriptions from an internet hospital platform with three primary objectives:

1. To develop and optimize a specialized large language model for understanding unstructured clinical narratives.
2. To analyze and categorize the specific service intents expressed in patients' medical descriptions.
3. To examine the distribution of these patient intents across different clinical departments to reveal demand patterns for strategic planning.

Methods: The study retrospectively collected 16,457 de-identified patient consultation records from a tertiary hospital's internet hospital in China. A hybrid methodology was employed: First, an unsupervised learning approach was used to discover intent categories. Patient texts were converted into embeddings using the chinese-roberta-wwm-ext-large model, reduced in dimensionality with UMAP, and clustered using K-Means, which revealed eight distinct intent categories. Subsequently, a random sample of 5,000 records was manually annotated by three clinical experts based on these categories (achieving high inter-rater agreement, Fleiss' Kappa = 0.82). This labeled dataset was then used to fine-tune the pre-trained model, and its performance was compared against traditional machine learning models (e.g., SVM with TF-IDF features).

Results: The analysis successfully identified eight coherent patient intent categories. The most prevalent needs were "Online Interpretation of Existing Medical Reports" (25.5%) and "Initial Consultation" (24.4%). The fine-tuned chinese-roberta-wwm-ext-large model achieved the best performance with a weighted F1-score of 0.7030, outperforming traditional models. A highly significant association was found between clinical departments and the types of patient intents ($\chi^2 = 8874.241$, $p < .01$). Based on standardized residual analysis, four distinct operational models for internet hospitals were synthesized:

Centralized Efficiency Departments (e.g., Dentistry, Dermatology): Strongly associated with "Initial Consultation" and "Online Test Booking."

Long-Term Management Departments (e.g., Hematology, Gastroenterology): Primarily linked to "Prescription Renewal for Chronic Diseases."

Balanced Departments (e.g., Nephrology, ENT): Managed a diverse mix of intents like "Report Interpretation" and "Online Follow-up."

Low-Engagement Departments (e.g., Neurosurgery, Oncology): Primarily used for "Report Interpretation" and "Remote Guidance," serving an auxiliary role.

Conclusions: This study demonstrates that a hybrid AI methodology effectively identifies and categorizes patient service intents within internet hospitals. The findings reveal a clear, efficiency-driven allocation of patient demand, with core services being report interpretation and initial consultations, and significant variation across departments. This provides a critical evidence base for hospitals to optimize resource allocation, design specialized operational models for different departments, and strategically develop their telemedicine services. Future work should focus on refining categories, integrating multimodal data (e.g., images), and modeling temporal consultation sequences for a more comprehensive understanding.

(JMIR Preprints 21/10/2025:86216)

DOI: <https://doi.org/10.2196/preprints.86216>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/86216>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/86216>

Original Manuscript

What Services Do Patients Need from Internet Hospitals? An AI-Powered Text Classification Study of Patient' Requests

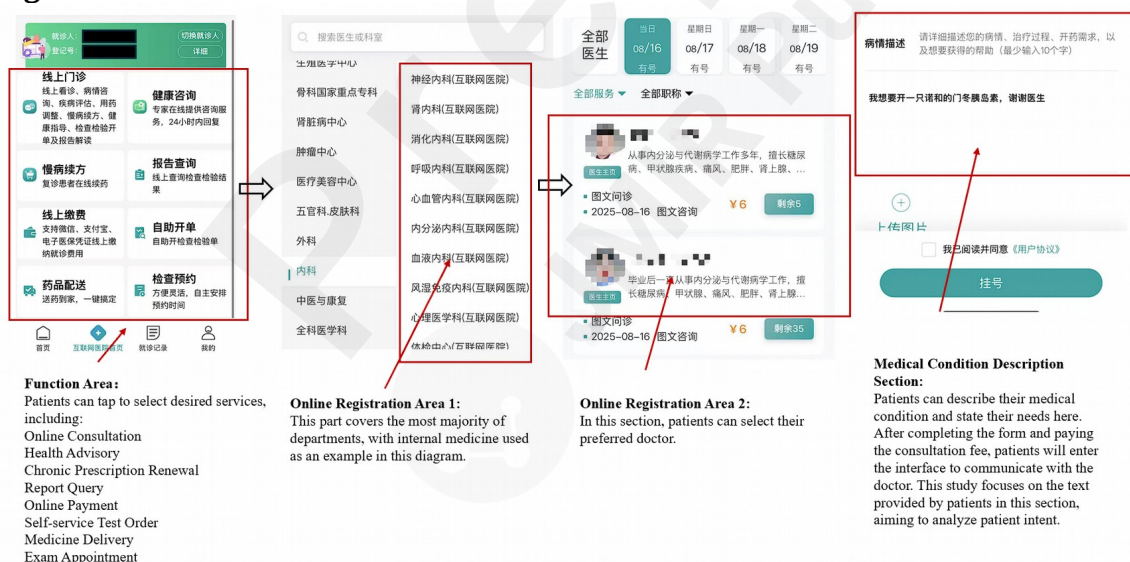
Introduction

The Rise of Internet Hospitals and the Centrality of Patient-Generated Text

In China, internet hospitals have emerged as an innovative telemedicine model that has attracted widespread attention. These platforms leverage in-person medical records and resources to provide online healthcare services for remote patients[1,2], integrating both online and offline resources to establish seamless information channels throughout the entire consultation process[3]. Through online platforms, patients can now complete multiple diagnostic and treatment processes without visiting hospitals in person: real-time access to test reports, online prescription renewals and medication purchases, remote consultations with doctors—traditionally cumbersome aspects of healthcare—can all be efficiently managed via smartphones[4,5]. The Covid-19 pandemic triggered a peak in internet hospital utilization and accelerated policy support, leading to their formal authorization for managing common and chronic diseases and prompting many traditional hospitals to establish their own platforms[6,7]. These internet hospitals, which typically leverage the resources of large tertiary hospitals, can mitigate problems such as long waiting times and limited service accessibility, which is particularly significant given that tertiary hospitals in China still handle a massive volume of outpatient visits.

The data for this study was sourced from a WeChat Mini Program of an internet hospital in Guangdong Province, China. This program provides a suite of characteristic internet hospital services, encompassing eight core functions: online consultations, health advice, chronic disease prescription renewals, report inquiries, online payments, self-service test orders, medication delivery, and exam scheduling. Figure 1 illustrates the typical patient journey within this Mini Program, from login and service selection to final consultation. This operational context generates the primary data for our research: the original textual descriptions written by patients.

Figure 1. Internet-Based Medical Consultation Flowchart



Research Landscape: NLP in Healthcare and the Persisting Gap

Internet hospitals, as a convenient model of care, have proven particularly beneficial for patients with chronic diseases and those with limited mobility[8]. A core function is online consultation, which begins with patients writing self-reports of their symptoms. These firsthand accounts—detailing symptoms, medication history, and health concerns—directly reflect patients' authentic medical needs and constitute a rich corpus for analysis. Current studies on medical texts primarily focus on leveraging Natural Language Processing (NLP) techniques to analyze textual data in the healthcare domain. Key research directions include the development of recommendation systems for online

medical consultations [9] and multi-faceted analysis of digital healthcare interactions by analyzing online reviews[10]. While these studies provide forward-looking insights, the majority rely on data from social media comments, online forums, and similar informal sources. In contrast, our research is grounded in real-world clinical practice, focusing specifically on texts generated by patients throughout the complete clinical pathway within internet hospitals.

Zhang et al. analyzed negative reviews on the Haodf.com platform using manual annotation—a method offering precision but facing significant efficiency challenges when processing large-scale patient data[11]. However, with advancements in artificial intelligence, Large Language Models (LLMs) and machine learning techniques have substantially facilitated text analysis. Despite this, a critical gap remains: the application of these advanced AI techniques to systematically decode and categorize the specific service intents from authentic, clinically-embedded consultation texts from internet hospitals. This study aims to leverage machine learning methodologies to analyze this extensive textual dataset, thereby achieving an optimal balance between accuracy and processing efficiency while addressing this gap.

Study Objectives and Contributions

The objectives of our study are threefold:

1. Developing and optimizing a specialized large language model for medical text to improve its capability in understanding and processing unstructured clinical narratives.
2. Analyzing and categorizing the intents expressed in patients' medical descriptions on online healthcare platforms to determine the specific services they seek.
3. Examining the distribution of these patient intents across different clinical departments to reveal patterns in demand for strategic operational planning.

Method

Data Collection

This study retrospectively collected all de-identified, internet-based consultation records from a tertiary hospital in Guangdong Province, China, between November 2024 and March 2025. The use of real-world data from operational digital health platforms is increasingly recognized as crucial for developing pragmatic AI solutions, as it captures the full spectrum of patient communication in a naturalistic setting[12]. Our dataset included structured fields and unstructured texts, primarily consisting of patients' self-reported "symptom descriptions." This analysis focuses exclusively on these descriptions to investigate patient intent, a approach that aligns with the paradigm of leveraging patient-generated text for understanding healthcare needs[13].

Data Preprocessing

We initiated the analysis with 18,254 raw patient descriptions. A rigorous preprocessing pipeline was implemented in Python 3.10 to ensure data quality. This involved the removal of null entries and anomalous texts. Subsequently, we applied regex-based normalization to strip URLs, excessive whitespace, and irrelevant symbols, while carefully preserving clinically critical information such as numerical values and units. This meticulous cleaning process is a critical, often underemphasized step in clinical NLP pipelines, as data quality has been shown to be a primary determinant of final model performance[14]. This process yielded a refined corpus of 16,457 high-quality text entries for modeling.

Unsupervised Clustering and Intent Discovery

To derive a data-driven taxonomy of patient intents without preconceived categories, we employed an unsupervised strategy. The entire analytical workflow is depicted in Figure 2. First, we transformed all texts into contextualized embeddings using the chinese-roberta-wwm-ext-large model. The use of transformer-based embeddings like BERT as a foundation for clustering has been successfully demonstrated in other healthcare text mining tasks, such as phenotyping from electronic health records[15]. The core of this model is the multi-head self-attention mechanism. For each attention $head_{i,j}$, the output is calculated as:

$$head_i = \text{Attention} \quad QW_i^Q \quad KW_i^K \quad VW_i^V \quad (3)$$

The outputs of all heads are then concatenated and projected:

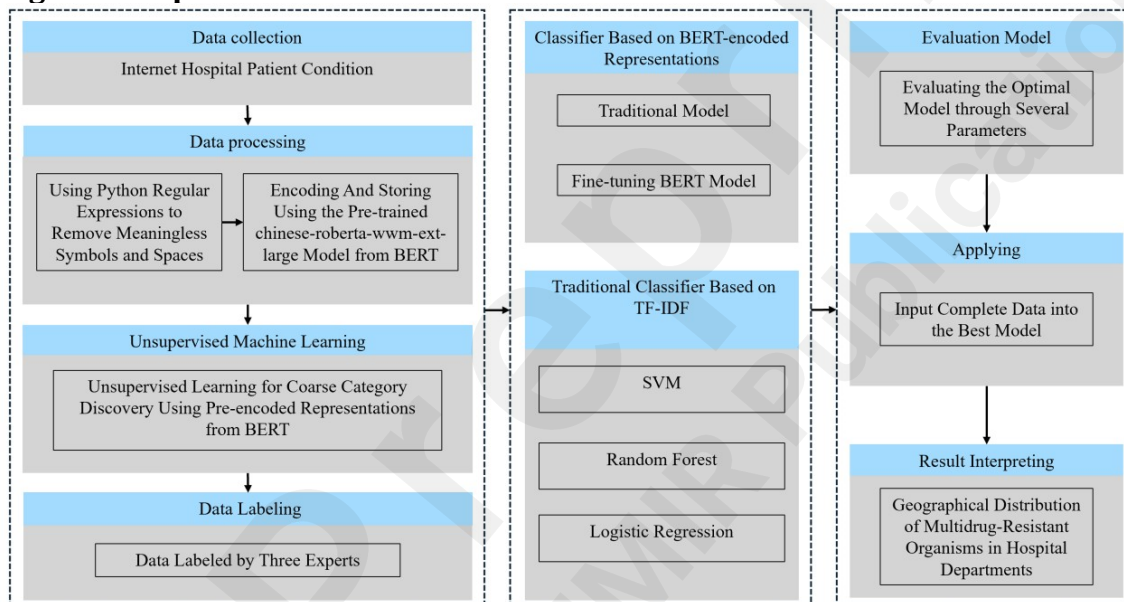
$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1, \dots, head_h) W^O \quad (4)$$

This model was pre-trained using the Whole Word Masking strategy, with the objective of predicting the original vocabulary id of the masked word w_{m_i} based on its context:

$$L_{MLM} = - \sum_{i \in M} \log P(w_i | w_{\setminus i}) \quad (5)$$

To mitigate the "curse of dimensionality" and enhance clustering performance, we projected the high-dimensional BERT embeddings into a lower-dimensional space using Uniform Manifold Approximation and Projection (UMAP). UMAP has gained prominence for its ability to preserve both local and global data structure, making it particularly effective for visualizing and clustering high-dimensional textual data[16]. We then applied the K-Means algorithm to the reduced embeddings. The optimal number of clusters (K=8) was determined by combining the elbow method with a qualitative assessment of clinical interpretability, a mixed-methods approach that ensures statistical patterns are grounded in clinical meaning[17]. A thorough manual review of the resultant clusters allowed us to define eight distinct and meaningful patient intent categories.

Figure 2. Experimental Process



Manual Annotation and Supervised Model Fine-tuning

To facilitate supervised learning, a random sample of 5,000 records was drawn from the preprocessed corpus. Three clinical experts independently annotated this subset according to the eight intent categories. The annotation process demonstrated high reliability, achieving a Fleiss' Kappa of 0.82. The involvement of multiple clinical experts for annotation is considered a gold standard in creating high-quality labeled datasets for medical AI, as it mitigates individual bias and enhances the clinical validity of the categories[18].

This curated dataset was used to fine-tune the chinese-roberta-wwm-ext-large model. The fine-tuning process and its underlying strategy are detailed in Figure 3. To better adapt the model to the medical domain, we experimented with freezing different numbers of layers and ultimately chose to freeze the first 8 layers during fine-tuning. This strategy of progressive layer unfreezing or selective freezing is a recognized technique for effective domain adaptation, preventing catastrophic forgetting of general language patterns while specializing the model for the target domain[19]. The detailed hyperparameters used for this process are listed in Table 2.

Model performance was evaluated on a held-out test set using standard metrics. The calculations for Precision, Recall, F1-score, and the Area Under the ROC Curve (AUC) adhere to the following

formulas. The choice of a comprehensive set of metrics is advised for a nuanced evaluation of multi-class clinical text classification models, as no single metric captures all aspects of performance[20].

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$AUC = \int_0^1 TPR(FPR) dFPR \quad (9)$$

Figure 3. BERT Model Training Strategy

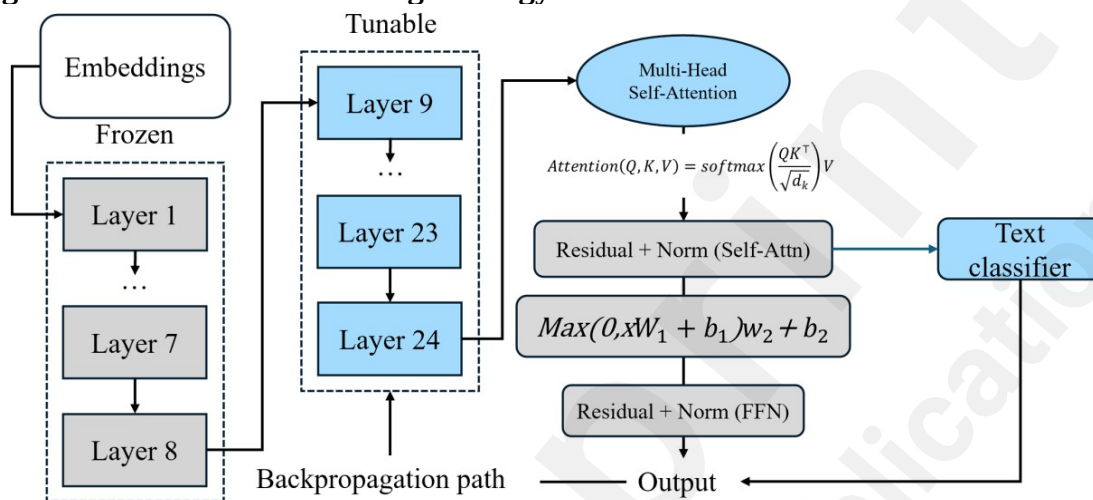


Table 1. Parameter Fine-tuning of BERT Model

Parameter	Value	Description
learning_rate	1e-5	Lower learning rate to adapt to large model
per_device_train_batch_size	8	Smaller batch size to adapt to large model
per_device_eval_batch_size	8	Evaluation batch size
num_train_epochs	6	Number of training epochs (kept unchanged)
weight_decay	0.01	Weight decay for regularization
save_strategy	"epoch"	Save model at the end of each epoch
metric_for_best_model	'f1_weighted'	Metric used to select best model
fp16	True	Use mixed precision training to save memory
warmup_ratio	0.1	Warmup ratio for learning rate scheduler
logging_steps	25	Logging frequency (every 25 steps)
gradient_accumulation_steps	2	Gradient accumulation steps to adapt to small batch size
lr_scheduler_type	"linear"	Learning rate scheduler type

Results

Model Performance and Comparative Analysis

Our investigation began with an unsupervised learning phase to explore the inherent structure of patient requests without preconceived categories. We encoded the entire corpus of 16,457 patient descriptions using the pre-trained chinese-roberta-wwm-ext-large model. The high-dimensional embeddings were then processed using UMAP for dimensionality reduction, followed by the application of the K-Means clustering algorithm. The use of UMAP is particularly advantageous for visualizing and clustering high-dimensional textual data, as it preserves both local and global data structure more effectively than linear methods[21]. Figure 4 depicts the determination of the optimal clustering configuration, which successfully identified eight coherent patient intent categories. The visual representation of these clusters in Figure 5 demonstrates clear separation, validating the distinctness of the semantic groups derived from the patient-generated texts. This data-driven approach to taxonomy discovery mitigates the researcher bias inherent in pre-defined categories and has been successfully applied in other healthcare text mining contexts to uncover latent patterns[22].

Figure 4. Optimal Configuration Results

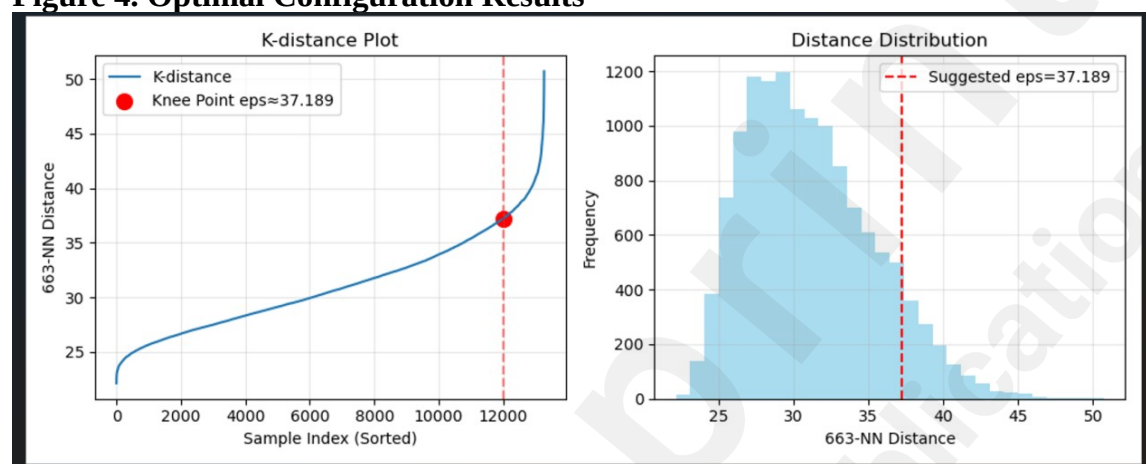
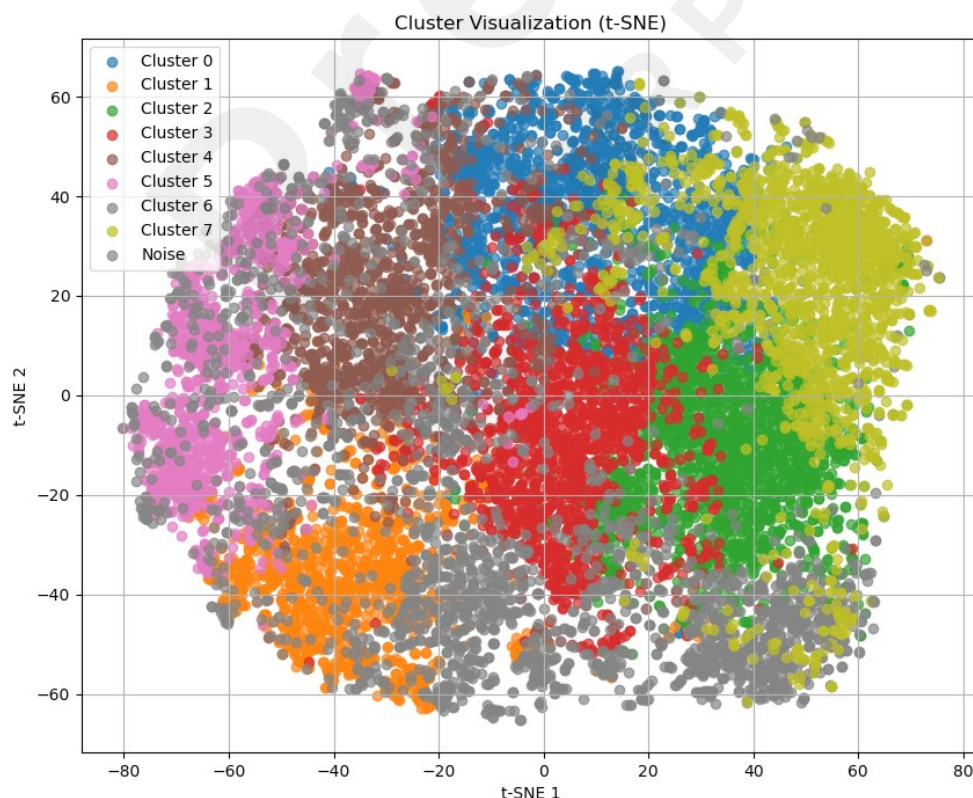


Figure 5. Unsupervised Learning Clustering Results



To transition into a supervised learning framework for precise classification, a random sample of 5,000 records was meticulously annotated by a panel of three clinical experts, achieving an excellent inter-rater agreement with a Fleiss' Kappa of 0.82. The involvement of multiple clinical experts is critical for establishing a reliable ground truth in medical text annotation, as it mitigates individual bias and ensures that the resulting labels reflect a clinical consensus, a practice strongly advocated in clinical NLP research[18]. We employed a comprehensive set of metrics for evaluation—Precision, Recall, F1-Score, and the Area Under the ROC Curve (AUC), as formally defined in Formulas 6-9. The use of a multi-faceted evaluation strategy is essential in healthcare AI, particularly for multi-class problems with potential class imbalance, as it provides a more holistic and realistic assessment of model performance than any single metric alone[23].

A pivotal finding emerged from our benchmark comparison. As detailed in Table 2, traditional machine learning models utilizing TF-IDF features, notably Support Vector Machines (SVM, F1-score=0.6122), outperformed identical models using non-fine-tuned BERT embeddings. This result underscores a critical nuance in medical text analysis: for a broad, category-level classification task like initial intent recognition, explicit keyword-level signals (e.g., "renew prescription," "check report") captured by TF-IDF can provide a strong and immediately effective discriminative baseline. This phenomenon, where simple yet robust traditional methods remain competitive on specific clinical NLP tasks, is consistent with findings from recent comparative studies[24]. The true potential of the deep learning model was unlocked through domain-specific fine-tuning. By strategically freezing the first 8 layers of the chinese-roberta-wwm-ext-large model (a strategy detailed in Figure 3 and parameterized in Table 1), we effectively adapted it to the specialized lexicon and syntax of internet hospital consultations. This process of targeted adaptation is a recognized and efficient strategy for achieving high performance in data-sensitive domains like healthcare, as it allows the model to learn domain-specific representations without overfitting to the typically smaller supervised datasets[25].

Table 2. The accuracy results of traditional models based on TF-IDF feature extraction.

Feature Type	Model	F1-score	Recall	Precision	AUC
TF-IDF	SVM	0.6122	0.6349	0.6121	0.8781
TF-IDF	Logistic Reg.	0.6002	0.6391	0.6025	0.8877
TF-IDF	Random Forest	0.5831	0.6210	0.6032	0.8436
BERT	SVM	0.5774	0.5754	0.5843	0.8652
BERT	Logistic Reg.	0.5923	0.5924	0.5970	0.8633
BERT	Random Forest	0.5557	0.5977	0.5827	0.8180

The fine-tuned model achieved a weighted F1-score of 0.7030. Figure 6, the clustering accuracy bar chart, illustrates the model's proficiency across different intent categories, showing particularly high precision for high-frequency intents. The confusion matrix presented in Figure 7 offers a granular view of the model's error profile. For instance, it reveals occasional misclassifications between "Online Follow-up for Diagnosed Conditions" and "Prescription Renewal for Chronic/Other Diseases"—a clinically understandable overlap, as these intents are often intertwined in the management of chronic illnesses. The achieved performance demonstrates the effectiveness of adapting a large pre-trained language model to a specialized domain, a paradigm that has consistently shown to yield significant gains over generic models across a wide array of text classification tasks[26]. Furthermore, the detailed error analysis afforded by the confusion matrix is a critical step

in clinical AI model validation, as it moves beyond aggregate metrics to uncover specific, clinically meaningful failure modes that must be understood and addressed before real-world deployment[27].

Figure 6. Clustering Accuracy Bar Chart

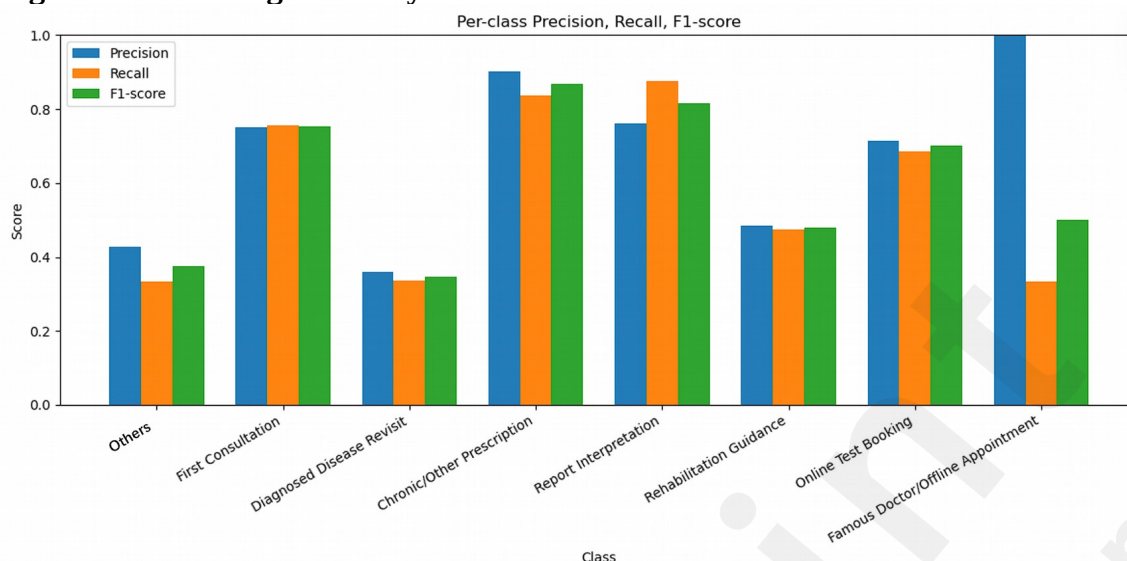
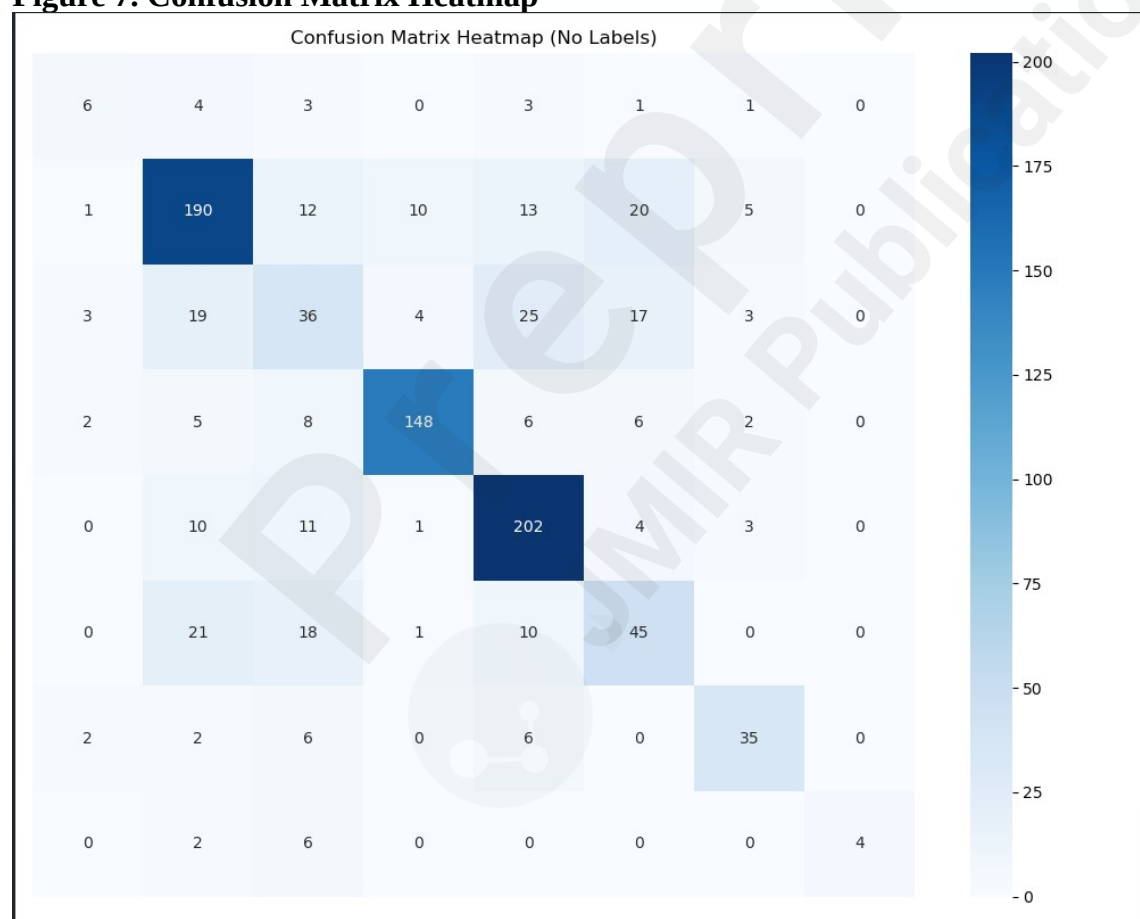


Figure 7. Confusion Matrix Heatmap



Distribution of Patient Intents and Cross-Departmental Analysis

Deploying the optimized model on the full dataset (N=16,457) provided a macroscopic view of patient demand within the internet hospital ecosystem. The distribution of patient intents, cataloged in Table 3, reveals a compelling narrative. The two most predominant needs were "Online Interpretation of Existing Medical Reports" (25.5%, 4199/16457) and "Initial Consultation" (24.4%, 4019/16457). This finding powerfully underscores a core utility of the platform: it primarily functions as an accessible and efficient channel for patients seeking to understand existing clinical

data and to initiate the diagnostic process for new health concerns. This pattern aligns with established models of telehealth utilization, where efficiency gains in information access and initial triage are among the most immediate and valued benefits for patients[28].

Table 3. Model prediction result

Department	Others	First Consultation	Diagnosed Disease Revisit	Chronic/Other Prescription	Report Interpretation	Rehabilitation Guidance	Online Test Booking	Famous Doctor/Offline Appointment	Total
Traditional Chinese Medicine	1	125	8	58	6	27	2	0	227
Ophthalmology and Otorhinolaryngology and Stomatology	29	425	224	90	547	195	61	12	1583
Obstetrics/Gynecology & Pediatrics	33	1038	295	347	978	569	163	17	3440
Surgical	62	587	766	395	1148	363	239	42	3592
Non-surgical	70	1305	509	438	905	492	197	32	3948
Total	257	4019	1929	3307	4199	1803	850	93	16457

Our analysis further uncovered significant heterogeneity in how different clinical specialties engage with the platform. A multivariate Chi-square test confirmed that consultation volumes were disproportionately distributed across departments ($\chi^2 = 27917.301$, $p < .01$). More profoundly, a Chi-square test of independence established a robust and statistically significant association between the clinical department and the types of patient intents expressed ($\chi^2 = 8874.241$, $p < .01$, $df = 312$). The application of such statistical tests is fundamental to health services research for rigorously quantifying variations and associations in patient care patterns[29]. The considerable magnitude of this χ^2 value strongly suggests that this relationship is not coincidental but reflects deeply embedded, systematic patterns in patient behavior and service utilization across medical specialties.

To decode these specific patterns, we conducted a standardized residual analysis, the results of which are visualized in the heatmap in Figure 8. This analytical technique allows us to move beyond confirming a general association to pinpointing the exact department-intent pairs that were significantly over- or under-represented ($|\text{standardized residual}| > 1.96$). The use of residual analysis to uncover the underlying structure in cross-tabulated data is a powerful and recommended method in categorical data analysis for health research[30]. Guided by these data-driven insights from Figure 8, we synthesized four distinct operational models for internet hospital services:

Centralized Departments (e.g., Dentistry, Dermatology, Rehabilitation): These departments exhibited strong associations with "Initial Consultation" and "Online Test Booking." Their operational model is characterized by high immediacy and efficiency, often enabling a complete virtual care loop—from initial consultation and visual diagnosis to e-prescription and medication delivery—for a well-defined set of conditions. The high suitability of such specialties for end-to-end telemedicine models has been documented in prior research focusing on service integration and efficiency[31];

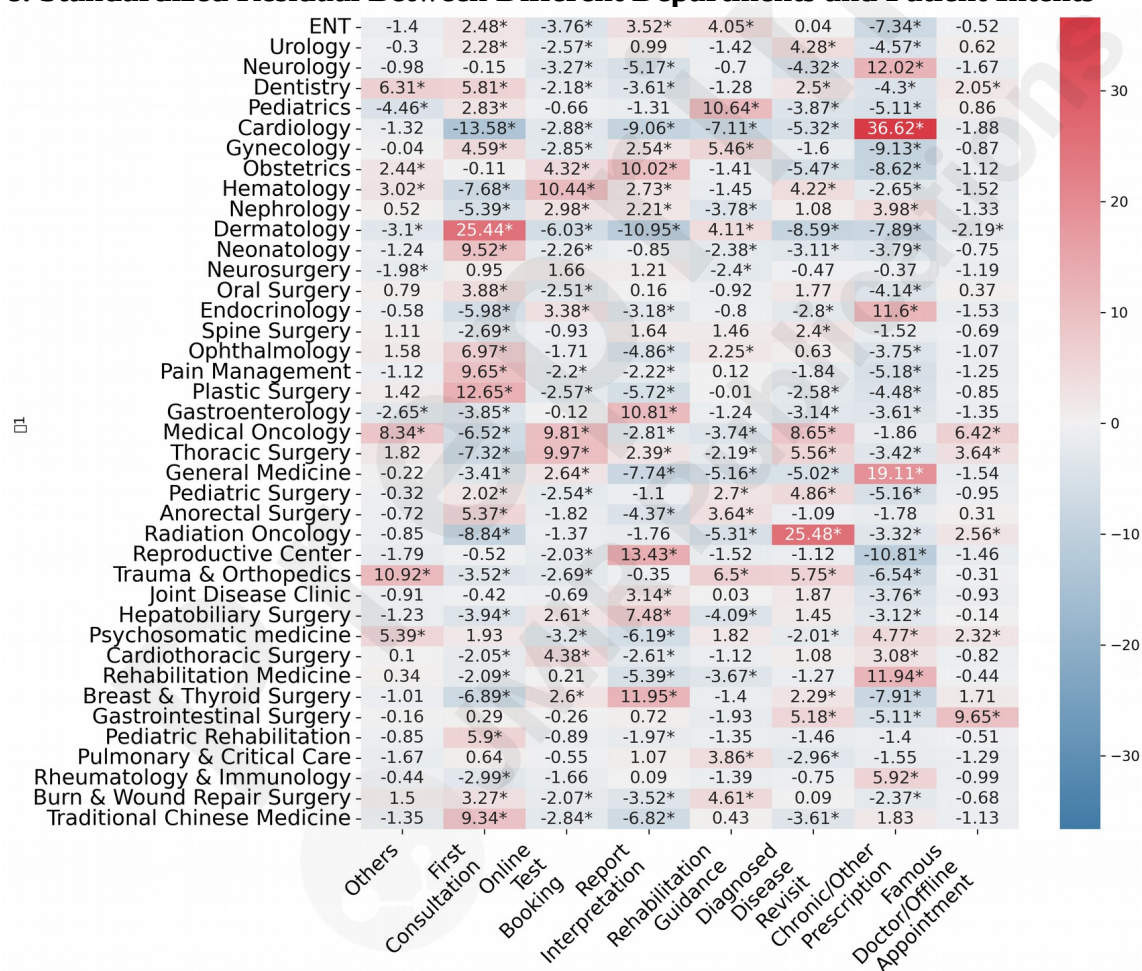
Long-Term Management Departments (e.g., Hematology, Gastroenterology, Breast & Thyroid Surgery): As anticipated, these specialties were significantly linked to "Prescription Renewal for Chronic/Other Diseases." This pattern highlights the internet hospital's pivotal role in facilitating continuous, long-term care for patients with stable chronic conditions. This finding resonates strongly with a large body of evidence demonstrating that telehealth interventions are particularly effective for improving the continuity and convenience of chronic disease management[32];

Balanced Departments (e.g., Nephrology, Thoracic Surgery, ENT): These departments managed a

diverse portfolio of patient intents, including "Report Interpretation," "Online Follow-up," and "Initial Consultation." This reflects a hybrid, integrated care model where the online platform effectively complements in-person services for complex conditions that require sustained monitoring and patient education. The value of such blended care models, which leverage the strengths of both digital and physical touchpoints, is increasingly recognized as a sustainable approach for managing a wide range of conditions[33];

Low-Engagement Departments (e.g., Urology, Neurosurgery, Oncology): These high-acuity specialties demonstrated relatively lower overall engagement on the platform, with their presence primarily associated with "Report Interpretation" and "Remote Guidance for Disease Recovery." This suggests a focused and highly valuable role for telemedicine in these fields: it serves as an auxiliary tool for perioperative counseling and post-discharge symptom monitoring. This supportive, rather than replacement, role for telehealth in complex surgical and oncological care is consistent with emerging best practices that aim to enhance patient support without attempting to virtualize core, high-risk interventions[34].

Figure 8. Standardized Residual Between Different Departments and Patient Intents



Discussion

Finding the Most Suitable Model Is the Key to Analyzing These Texts.

This research developed a multi-label consultation intent recognition system using real-world data from an internet hospital platform, employing a hybrid methodology of "unsupervised semantic discovery → category system construction → manual annotation → model training/validation" to enhance structured interpretation of patient inputs for treatment prediction and resource allocation. Medical experts conducted annotations with cross-review mechanisms, significantly improving inter-annotator agreement (IAA) and reducing bias, while unsupervised clustering of BERT embeddings

validated annotation quality through distinct semantic clusters in t-SNE visualizations—particularly effective in small-sample categories like "chronic disease follow-up." The fine-tuned chinese-RoBERTa-wwm-ext-large model outperformed traditional methods in accuracy and weighted F1-score, especially for high-frequency intents, demonstrating robustness against ambiguous semantics and imbalanced data. However, when applying machine learning for text classification, insufficient sample size often serves as a primary constraint. For example, in this study, departments with lower utilization rates may demonstrate reduced model accuracy and recall compared to those with higher utilization. Nevertheless, BERT has established a leading position in small-sample machine learning among natural language processing techniques, exhibiting superior robustness and resistance to overfitting when handling limited data. Particularly in specialized domains such as medicine, by leveraging pre-trained models with minimal fine-tuning and capitalizing on the inherent strengths of the Transformer architecture, it can effectively extract distinctive features of specific categories. The study addresses constraints in Chinese medical NLP—notably annotation costs and corpus complexity—by systematically combining supervised/unsupervised methods, validating BERT's efficacy per prior research[35], and establishing scalable foundations for intelligent consultation systems in healthcare settings.

The Mechanisms and Implications of Demand Allocation in Internet Hospital Services

The 16,457 internet-hospital consultations in dataset reveals a robust efficiency-driven allocation of demand that mirrors patterns observed in international telehealth literature: high concentration in chronic-care and maternal/child services, mid-level use in hybrid specialties, and relatively lower direct-treatment demand in surgically intensive fields.

Mechanistically, two interacting forces explain this pattern and carry operational implications. First, task substitutability—routine, repeatable activities with low incremental diagnostic risk (prescription renewals, test-interpretation, brief follow-ups) are highly amenable to remote management, a finding consistent with systematic evidence that telemedicine delivers measurable value for chronic disease continuity of care and reduces utilization burdens on in-person services[32,36]. Second, time-cost sensitivity and life-stage priorities induce behavioral sorting: parents and chronic patients weight travel and waiting costs heavily and thus prefer online channels for screening, reassurance, and medication continuity, a dynamic also documented in studies of Chinese internet-hospital adoption and operational models[37,38]. From a policy and platform design perspective, this implies prioritizing digital capacity and algorithmic triage for departments with both high demand and high online substitutability to maximize marginal returns; investing in hybrid workflows such as tele-triage followed by fast in-person slots for balanced specialties; and treating low-engagement but high-acuity departments as candidates for auxiliary tele-services—such as perioperative counselling and symptom surveillance—rather than full virtualization. These prescriptions are aligned with empirical findings that targeted telehealth deployment, rather than blanket expansion, yields better access, safety, and cost-effectiveness[39].

Modular Design and Implementation Strategies for Optimizing Online Medical Consultation Services

An effective way to optimize online consultation services is to break them down into several "functional modules"—such as triage/screening, structured medical history and imaging collection, treatment decision-making, long-term monitoring, and administrative processes. This modular analysis method can more clearly reveal the heterogeneity among different departments and provide a clear basis for developing executable implementation strategies. In departments highly reliant on imaging and visual judgment (such as dermatology, some oral and ENT specialties), telemedicine has shown high consistency with offline diagnosis for various common diseases. With standardized photography procedures and asynchronous image reading standard operating procedures (SOPs), these departments can establish a safe and reliable triage-treatment pathway[40].

For departments requiring long-term management (e.g., cardiology, endocrinology,

gastroenterology), key success factors could utilize bidirectional integration with Electronic Health Record (EHR) systems, automated alert mechanisms based on laboratory thresholds, and a performance indicator system centered on continuous health outcomes (such as improvement in test indicators and prescription renewal delay rates). The remote management model for these departments is highly consistent with the effectiveness observed in chronic disease management research[41]. In perioperative and oncology scenarios, existing literature indicates that online follow-ups can serve as an auxiliary risk management tool, reducing unnecessary face-to-face visits while ensuring patient safety and without increasing readmission rates[32,42]. Therefore, the design of implementation pathways should match the functional modules: imaging-related modules should follow a progression from retrospective consistency validation to prospective piloting, and then to routine auditing; chronic disease management modules need to establish continuous outcome dashboards and threshold-triggered escalation rules; perioperative modules should set clear warning criteria and open rapid offline referral channels. Accordingly, success evaluation should also adopt a modular approach: imaging modules should use diagnostic consistency rate as the core metric, chronic disease modules should focus on biomarker improvement and medication adherence, and triage/screening modules should evaluate their effectiveness in avoiding unnecessary visits and the timeliness of reaching final decisions. Internet hospital practices and related implementation studies show that this function-centric, differentiated deployment strategy, compared to a "one-size-fits-all" approach to telemedicine promotion, is more conducive to ensuring medical safety while achieving sustainable expansion and operation of services[38,43].

Limitation and future research

This study has several limitations. First, the classification of patient conditions is limited to eight broad categories, whereas real-world medical scenarios often involve more complex patient needs. Future research could refine the current coarse-grained categories by introducing a more detailed hierarchical classification based on disease types (e.g., common diseases, chronic conditions, complex disorders) or human physiological systems, thereby enabling more precise allocation of medical resources. Second, in internet-based hospital consultation platforms, users can upload images to assist in diagnosis, yet this study relies solely on textual information. Future work could extend the model's multimodal information fusion capability by integrating visual data from user-uploaded images, supplementing textual semantics with visual cues to enhance both discriminative power and generalization performance. Third, the current approach treats each medical description as an independent instance, disregarding the temporal sequence of consultations from the same patient. This results in incomplete utilization of available data. Subsequent studies could incorporate sequence modeling mechanisms to jointly model a user's historical consultation records as temporal context, enabling a more comprehensive understanding of disease progression and dynamic changes in user intent. Finally, future efforts should continue to strengthen the model's interpretability and robustness, improving its tolerance to noise such as non-standard expressions and spelling errors, thereby ensuring stable and reliable performance in real-world clinical settings.

1. Li L, Liu G, Xu W, Zhang Y, He M. Effects of Internet Hospital Consultations on Psychological Burdens and Disease Knowledge During the Early Outbreak of COVID-19 in China: Cross-Sectional Survey Study. *J Med Internet Res* 2020 Aug 4;22(8):e19551. doi: 10.2196/19551
2. Lai Y, Chen S, Li M, Ung COL, Hu H. Policy Interventions, Development Trends, and Service Innovations of Internet Hospitals in China: Documentary Analysis and Qualitative Interview Study. *J Med Internet Res* 2021 July 20;23(7):e22330. doi: 10.2196/22330

3. Kulikowski CA. Ethics in the History of Medical Informatics for Decision-Making: Early Challenges to Digital Health Goals. *Yearb Med Inform* 2022 Aug;31(01):317–322. doi: 10.1055/s-0042-1742491
4. Han Y, Lie RK, Guo R. The Internet Hospital as a Telehealth Model in China: Systematic Search and Content Analysis. *J Med Internet Res* 2020 July 29;22(7):e17995. doi: 10.2196/17995
5. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. *The Lancet Global Health* 2015 Aug;3(8):e445–e446. doi: 10.1016/S2214-109X(15)00042-X
6. Zhi L, Yin P, Ren J, Wei G, Zhou J, Wu J, Shen Q. Running an Internet Hospital in China: Perspective Based on a Case Study. *J Med Internet Res* 2021 Sept 16;23(9):e18307. doi: 10.2196/18307
7. Ding L, She Q, Chen F, Chen Z, Jiang M, Huang H, Li Y, Liao C. The Internet Hospital Plus Drug Delivery Platform for Health Management During the COVID-19 Pandemic: Observational Study. *J Med Internet Res* 2020 Aug 6;22(8):e19678. doi: 10.2196/19678
8. Liu X, Ben Liu Q. Superior medical resources or geographic proximity? The joint effects of regional medical resource disparity, geographic distance, and cultural differences on online medical consultation. *Social Science & Medicine* 2024 June;350:116911. doi: 10.1016/j.socscimed.2024.116911
9. Journal of Medical Internet Research - Online Medical Consultation Service–Oriented Recommendations: Systematic Review. Available from: <https://www.jmir.org/2024/1/e46073> [accessed Oct 20, 2025]
10. Bellary S, Kumar Bala P, Chakraborty S. Utilizing online reviews for analyzing digital healthcare consultation services: Examining perspectives of both healthcare customers and healthcare professionals. *International Journal of Medical Informatics* 2024 Nov 1;191:105587. doi: 10.1016/j.ijmedinf.2024.105587
11. Journal of Medical Internet Research - Unhappy Patients Are Not Alike: Content Analysis of the Negative Comments from China's Good Doctor Website. Available from: <https://www.jmir.org/2018/1/e35/> [accessed Oct 20, 2025]
12. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, Liu Y, Topol E, Dean J, Socher R. Deep learning-enabled medical computer vision. *npj Digit Med* Nature Publishing Group; 2021 Jan 8;4(1):5. doi: 10.1038/s41746-020-00376-2
13. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review - ScienceDirect. Available from: <https://www.sciencedirect.com/science/article/pii/S1532046417301685> [accessed Oct 21, 2025]
14. Wang Y, Sohn S, Liu S, Shen F, Wang L, Atkinson EJ, Amin S, Liu H. A clinical text classification paradigm using weak supervision and deep representation. *BMC Medical Informatics and Decision Making* 2019 Jan 7;19(1):1. doi: 10.1186/s12911-018-0723-6
15. Beaulieu-Jones BK, Greene CS, Pooled Resource Open-Access ALS Clinical Trials Consortium. Semi-supervised learning of the electronic health record for phenotype

- stratification. *J Biomed Inform* 2016 Dec;64:168–178. PMID:27744022
16. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*; 2020. doi: 10.48550/arXiv.1802.03426
 17. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv*; 2022. doi: 10.48550/arXiv.2203.05794
 18. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc* 2019 Apr 1;26(4):364–379. PMID:30726935
 19. Ruder S. Universal Language Model Fine-tuning for Text Classification.
 20. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 2009 July 1;45(4):427–437. doi: 10.1016/j.ipm.2009.03.002
 21. Dadu A, Satone VK, Kaur R, Koretsky MJ, Iwaki H, Qi YA, Ramos DM, Avants B, Hesterman J, Gunn R, Cookson MR, Ward ME, Singleton AB, Campbell RH, Nalls MA, Faghri F. Application of Aligned-UMAP to longitudinal biomedical studies. *Patterns* 2023 June 9;4(6):100741. doi: 10.1016/j.patter.2023.100741
 22. Si Y, Du J, Li Z, Jiang X, Miller T, Wang F, Jim Zheng W, Roberts K. Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review. *Journal of Biomedical Informatics* 2021 Mar 1;115:103671. doi: 10.1016/j.jbi.2020.103671
 23. Gaetani M, Mazwi M, Balaci H, Greer R, Maratta C. Artificial intelligence in medicine and the pursuit of environmentally responsible science. *The Lancet Digital Health Elsevier*; 2024 July 1;6(7):e438–e440. PMID:38816299
 24. Lu H, Ehwerhemuepha L, Rakovski C. A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. *BMC Medical Research Methodology* 2022 July 2;22(1):181. doi: 10.1186/s12874-022-01665-y
 25. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *arXiv*; 2023. doi: 10.48550/arXiv.1910.10683
 26. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
 27. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine* 2019 Oct 29;17(1):195. doi: 10.1186/s12916-019-1426-2
 28. Lu W, Zhai Y. Self-Adaptive Telemedicine Specialist Recommendation Considering Specialist Activity and Patient Feedback. *Int J Environ Res Public Health MDPI*; 2022 May;19(9). doi: 10.3390/ijerph19095594
 29. Hayat MJ, Powell A, Johnson T, Cadwell BL. Statistical methods used in the public health literature and implications for training of public health professionals. *PLoS One* 2017 June

7;12(6):e0179032. PMID:28591190

30. Sharpe D. Your Chi-Square Test is Statistically Significant: Now What?
31. Lokken TG, Blegen RN, Hoff MD, Demaerschalk BM. Overview for Implementation of Telemedicine Services in a Large Integrated Multispecialty Health Care System. *Telemed J E Health* 2020 Apr;26(4):382–387. PMID:31433261
32. Xiao Z, Han X. Evaluation of the Effectiveness of Telehealth Chronic Disease Management System: Systematic Review and Meta-analysis. *J Med Internet Res* 2023 Apr 27;25:e44256. PMID:37103993
33. Eze ND, Mateus C, Cravo Oliveira Hashiguchi T. Telemedicine in the OECD: An umbrella review of clinical and cost-effectiveness, patient experience and implementation. *PLoS One* 2020;15(8):e0237585. PMID:32790752
34. Ferrari-Light D, Geraci TC, Chang SH, Cerfolio RJ. Novel Pre- and Postoperative Care Using Telemedicine. *Front Surg* 2020 Dec 1;7:596970. PMID:33335911
35. Keloth VK, Hu Y, Xie Q, Peng X, Wang Y, Zheng A, Selek M, Raja K, Wei CH, Jin Q, Lu Z, Chen Q, Xu H. Advancing entity recognition in biomedicine via instruction tuning of large language models. Wren J, editor. *Bioinformatics* 2024 Mar 29;40(4):btae163. doi: 10.1093/bioinformatics/btae163
36. Ranjith N, Kanna Baskaran SN, R S, Ganesan A, Kanakaraj R, S G, Shafy SM, Nair AS. Current trends in telemedicine for chronic disease management. *Bioinformation* 2025;21(4):587–593. PMID:40636162
37. Zhong Y, Hahne J, Wang X, Wang X, Wu Y, Zhang X, Liu X. Telehealth Care Through Internet Hospitals in China: Qualitative Interview Study of Physicians' Views on Access, Expectations, and Communication. *Journal of Medical Internet Research* 2024 Mar 29;26(1):e47523. doi: 10.2196/47523
38. Chen X, Wu X, Zhang Q, Jing R, Cheng W, Tian J, Jin C. The construction and operational models of internet hospitals in China: a hospital-based survey study. *BMC Health Serv Res* 2023 June 21;23(1):669. PMID:37344831
39. Wang L, Liang D, HuangFu H, Ke C, Wu S, Lai Y. Enterprise-led internet healthcare provision in China: insights from a leading platform. *Front Digit Health Frontiers*; 2025 Mar 3;7. doi: 10.3389/fdgth.2025.1491183
40. Bashshur RL, Shannon GW, Tejasvi T, Kvedar JC, Gates M. The Empirical Foundations of Teledermatology: A Review of the Research Evidence. *Telemed J E Health* 2015 Dec;21(12):953–979. PMID:26394022
41. Xiao Z, Han X. Evaluation of the Effectiveness of Telehealth Chronic Disease Management System: Systematic Review and Meta-analysis. *J Med Internet Res* 2023 Apr 27;25:e44256. PMID:37103993
42. Uppal A, Kothari AN, Scally CP, Roland CL, Bednarski BK, Katz MHG, Vauthey J-N, Chang GJ, D3CODE Team. Adoption of Telemedicine for Postoperative Follow-Up After Inpatient

Cancer-Related Surgery. JCO Oncol Pract 2022 July;18(7):e1091–e1099. PMID:35263166

43. Kheterpal MK, Borre ED, Cheema U, Nicholas MW, Cooner EW, Phinney D, Gagnon K, Zullig LL, King HA, Malcolm EJ, Chen SC. Implementation evaluation of a teledermatology virtual clinic at an academic medical center. Implement Sci Commun 2023 Oct 27;4(1):130. PMID:37891695



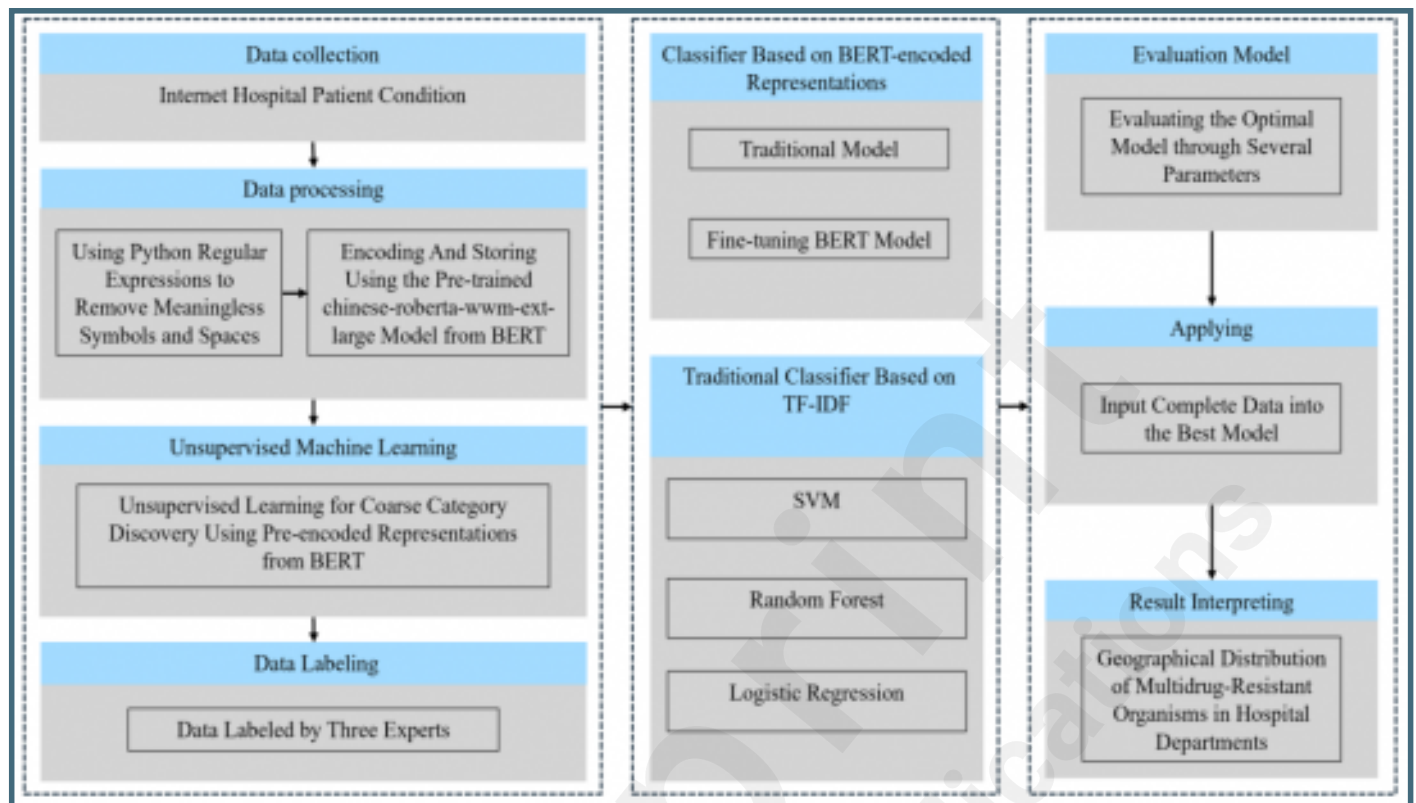
Supplementary Files

Figures

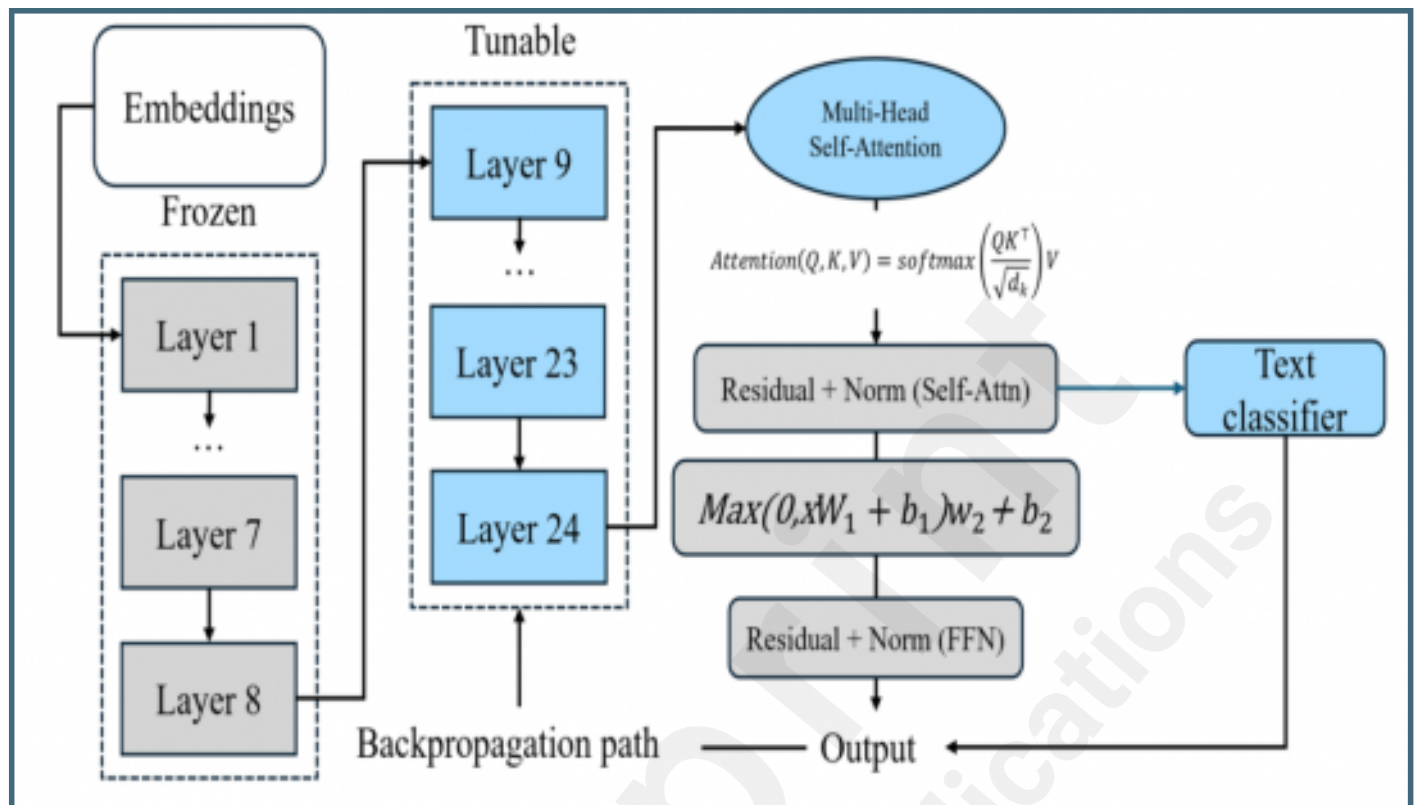
Internet-based medical consultation flowchart.



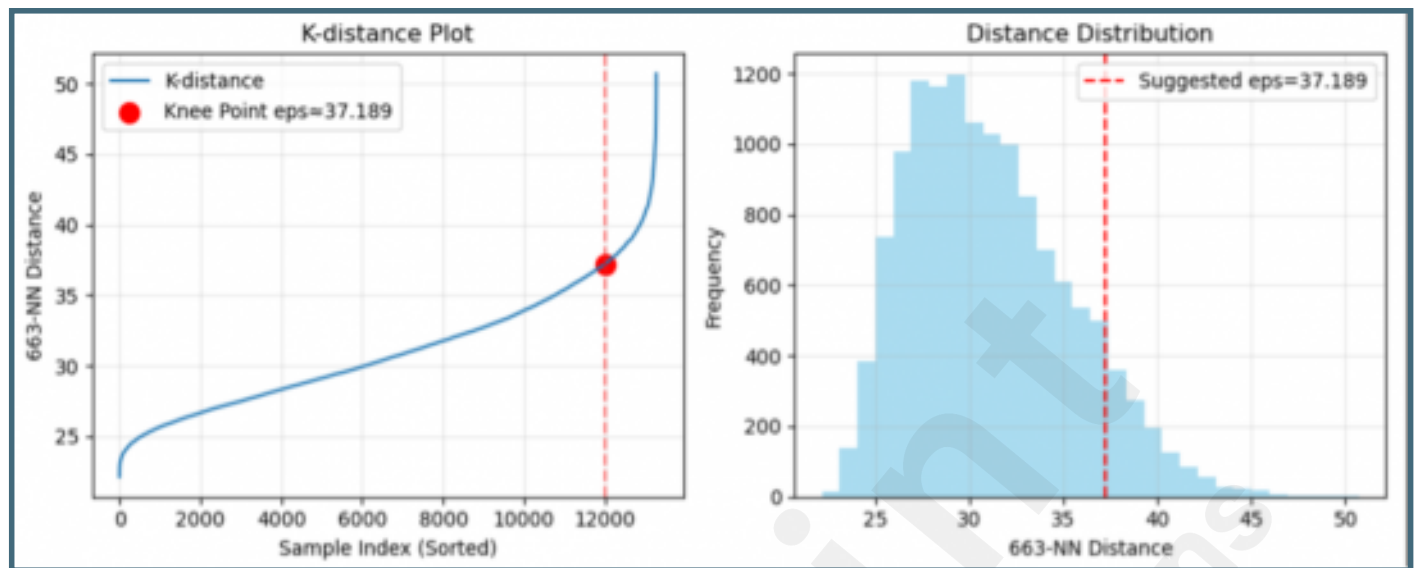
Experimental process.



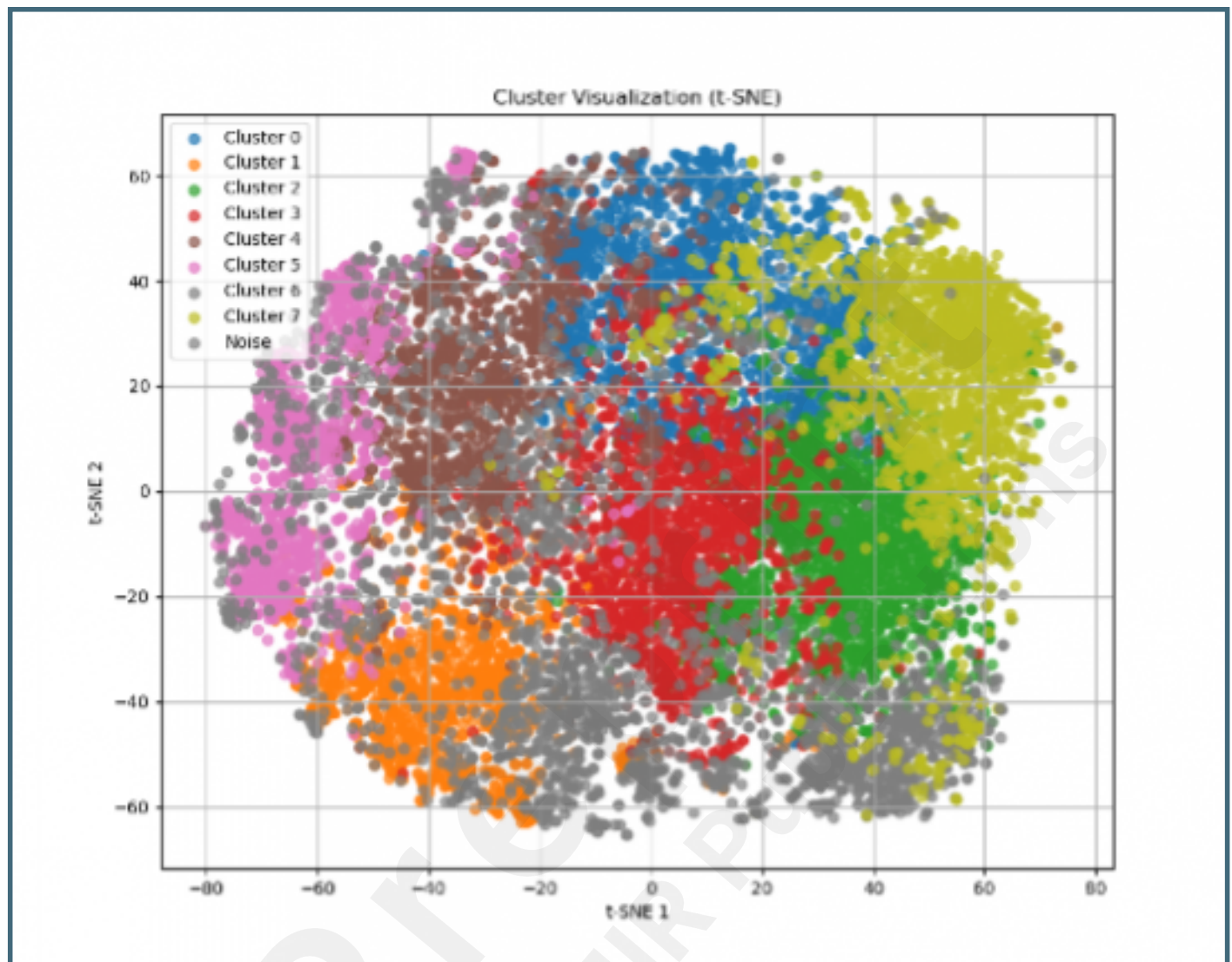
BERT model training strategy.



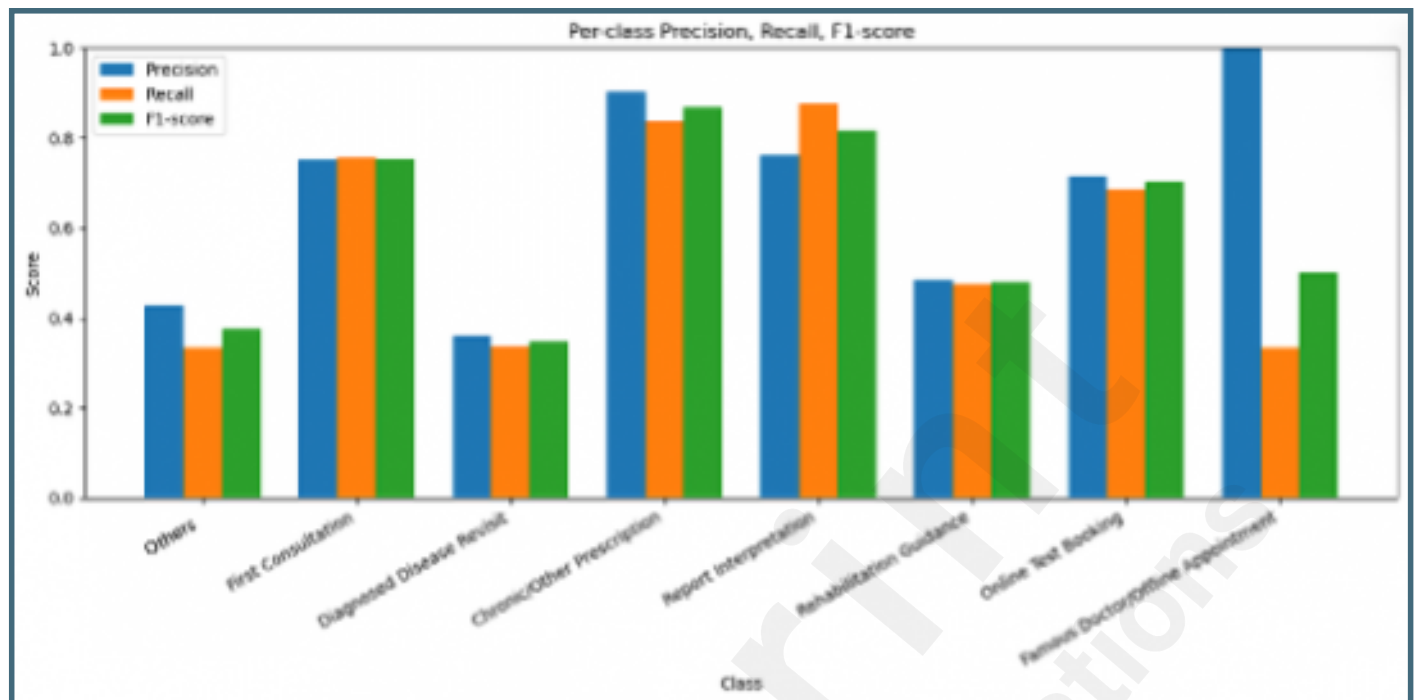
Optimal configuration results.



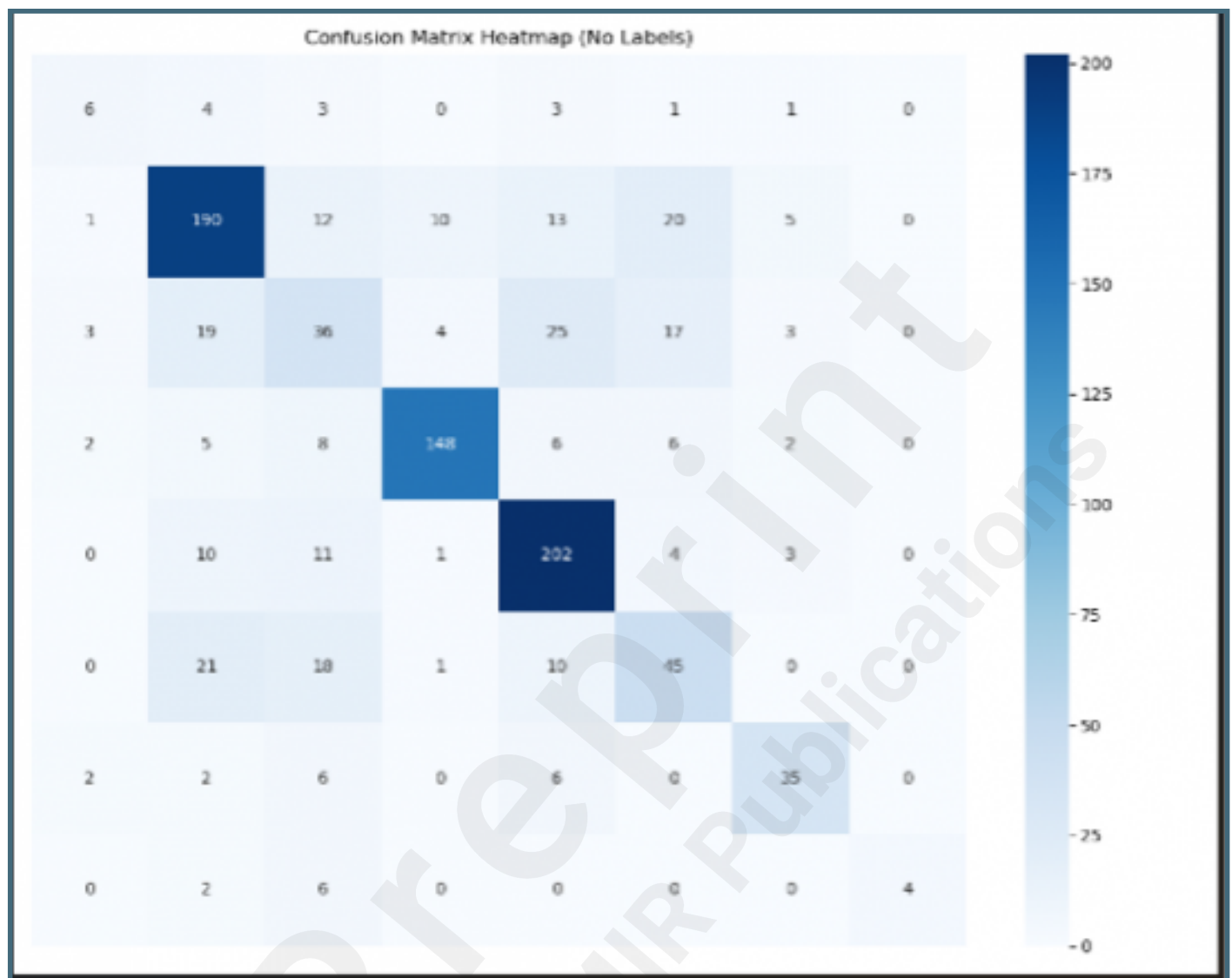
Unsupervised learning clustering results.



Model clustering accuracy bar chart.



Confusion matrix heatmap.



Standardized residual between different departments and patient Intents.

