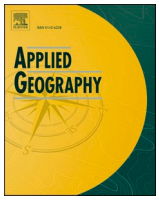
Applied Geography 189 (2026) 103954

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog



Digital grievances: A geospatial analysis of public discourse on doctor-patient conflict in China

Pu Zhang [a], Yiliang Li [b], Zheng Wei [c], Pan Hui [a,c,d], Na Jiang [a,*]

[a] The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
[b] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China
[c] Emerging Interdisciplinary Areas, Academy of Interdisciplinary Studies, The Hong Kong University of Science and Technology, Hong Kong
[d] Department of Computer Science, University of Helsinki, Helsinki, Finland

ARTICLE INFO

Handling Editor: J Peng

Keywords:
Doctor-patient relationship
Social media
Public opinion
Health geography
Regional disparities
China

ABSTRACT

The escalating conflict between doctors and patients in China has become a critical societal issue, with social media emerging as a primary arena for public grievance. However, the thematic content, spatial heterogeneity, and underlying drivers of this vast online discourse remain poorly understood. This study addresses this gap by analyzing over 300,000 user comments from the social media platform Douyin, linked to provincial-level socioeconomic and healthcare statistics. Employing a hybrid method combining manual open coding with a Large Language Model (LLM), we identify core discussion themes and use a normalized "Topic Index" with Lasso-selected Multiple Linear Regression (MLR) and SHAP-GIS analysis to explain their geographical distribution. Our findings reveal a discourse dominated by grievances, yet a counter-narrative of empathy for professionals is highly endorsed by users. Geographically, the analysis uncovers a distinct "Modernization Paradox." Nationally, Hospital Density serves as a significant stabilizer, showing a negative correlation with grievance levels. However, SHAP visualization reveals that in modernized coastal regions, this buffering effect is offset by the "Modernization Penalty", where higher Life Expectancy actively drives critical systemic discourse. We conclude that online public grievance is not a simple function of scarcity but is shaped by a structural tension between the stabilizing effects of infrastructure and the escalating pressures of patient complexity and rising expectations in modernized regions. This suggests that solutions must move beyond simply expanding physical infrastructure to focus on optimizing patient flow and humanizing service interactions in high-pressure medical centers.

1. Introduction

A striking geographical disparity defines China's healthcare landscape: while modernization hubs like Beijing and Shanghai concentrate the nation's best medical resources yet grapple with overcrowding and intense conflicts, vast inland regions struggle with basic accessibility and resource scarcity (Jia et al., 2022; Pan et al., 2016; Yan et al., 2022). This spatial inequality sets the stage for a severe crisis in the relationship between patients and physicians. Once a profession of high esteem, doctors are now frequently targets of public anger in a system characterized by endemic mistrust, soaring costs, and systemic pressures (Luo et al., 2022; Q. Wu, Jin, & Wang, 2021). This erosion of trust has manifested in uniquely disruptive and often violent forms, from highly publicized protests and organized disturbances known as *yinao* (医闹) to physical assaults against medical staff, making Chinese hospitals some of the most dangerous workplaces in the world (Chen et al., 2022; Sun et al., 2022). This conflict is not merely a medical dispute but a profound social issue with distinct spatial dimensions, reflecting broader regional anxieties regarding equity, access, and institutional accountability in a rapidly modernizing society (Meng et al., 2021).

To understand the gravity of this issue, one must look at the specific context of China's healthcare transition (Liu et al., 2020). The shift from state-welfare to market-oriented healthcare has created a "profit-driven" public hospital system that often relies on high-volume services to survive, structurally eroding trust (Y. Wu, Jin, & Wang, 2021). The consequences are severe and tangible. A lot of physicians have reported experiencing conflicts with patients (He & Qian, 2016). High-profile tragedies, such as the brutal killing of Dr. Yang Wen in Beijing in 2019, shocked the nation and triggered a "chilling effect," leading to widespread defensive medicine and a growing attrition of medical

* Corresponding author. The Hong Kong University of Science and Technology (Guangzhou), China.
E-mail address: najiang@hkust-gz.edu.cn (N. Jiang).

https://doi.org/10.1016/j.apgeog.2026.103954
Received 22 September 2025; Received in revised form 17 February 2026; Accepted 23 February 2026
Available online 26 February 2026

talent. In response, the Chinese government has elevated hospital security to a national priority, implementing "zero-tolerance" laws and mandatory security checkpoints in hospitals (Hu et al., 2024). However, these top-down security measures often fail to address the root causes of mistrust, leaving the underlying social grievances to ferment and explode in the digital sphere.

In recent years, these long-simmering grievances have found a powerful new arena: social media. Platforms like Douyin (the Chinese version of TikTok) have evolved from entertainment channels into crucial spaces where personal medical experiences are shared and collective sentiments are spatially aggregated (Fung & Hu, 2022; Zhang et al., 2024). Unlike traditional surveys or news media reports, these platforms provide an unprecedented, real-time window into the unsolicited, unfiltered, and deeply personal sentiments of millions (Dong & Xie, 2024). Crucially, this digital discourse acts as a "virtual mirror" to the physical world, reflecting the uneven distribution of healthcare resources across the country. The viral spread of a single negative encounter can shape the perceptions of a vast audience, while the aggregation of countless individual comments can form a potent, data-rich barometer of public opinion (Dong & Xie, 2024). Yet, despite the evident importance of this digital discourse, it remains a "black box", its thematic structure, spatial heterogeneity, and the underlying geographical drivers are poorly understood.

Scholarship on China's doctor-patient relationship has extensively documented its challenges through valuable, yet traditional, methodologies (Wang & Du, 2023). Researchers have used national surveys to track declining levels of trust (Du et al., 2023), conducted ethnographic studies to reveal the daily frictions within hospital walls (He & Qin, 2023), and analyzed state-controlled news media to understand official narratives (Wei & Mao, 2023). While essential, these approaches often miss the raw, spontaneous, and user-generated nature of discourse that now thrives online (Chen et al., 2021). Parallelly, a vast body of literature has explored China's profound regional disparities in economic development and healthcare access, consistently highlighting a deep cleavage between affluent coastal provinces and the less developed interior (Fu et al., 2022; Peng et al., 2024). However, few studies have bridged these two domains to investigate how these physical, macro-level structural realities fundamentally shape the substance and tone of public discourse in the digital sphere.

This study stands at the confluence of these research streams. While we know that doctor-patient trust is low, that social media is a key forum for debate, and that China is regionally unequal, no research to our knowledge has systematically connected these phenomena. We lack a large-scale, data-driven map that shows not just *that* people are dissatisfied, but *what* specifically they are concerned about, *where* these concerns are most intense, and *why* they differ from one region to another.

Therefore, this paper aims to fill this gap by addressing three central research questions that progress from descriptive mapping to explanatory modeling:

RQ1: What are the primary themes of public discourse regarding doctor-patient conflict on Chinese social media?
RQ2: How does the salience of these themes vary across China's provinces?
RQ3: To what extent do regional healthcare resources and socioeconomic modernization explain these spatial variations, and are these effects uniform across different development contexts?

To answer these questions, we employ a novel, three-stage research design. We first collect and clean up over 300,000 user comments from Douyin and compile a corresponding set of provincial-level macro data. We then use a hybrid method combining manual open coding with a Large Language Model (LLM) to accurately classify the thematic content of each comment. Finally, using a normalized "Topic Index," we apply a suite of statistical models (MLR enhanced by SHAP spatial analysis) to explain the geographical patterns observed. Crucially, to rigorously address RQ3 and bridge the gap between exploratory analysis and theoretical explanation, we formulate three guiding hypotheses (H1, H2a, and H2b) in the subsequent section, which are specifically tested to reveal the mechanisms of supply-side mitigation, economic modernization, and metropolitan aggregation. This study contributes to theory by empirically demonstrating how geographic conditions (e.g., resource density and development levels) amplify and shape public grievance online. Methodologically, it offers a new, robust framework for analyzing large-scale social media data from a geospatial perspective. Practically, its findings provide critical insights for policymakers seeking to craft targeted, regionally aware interventions to help mend China's broken doctor-patient relationship.

2. Theoretical framework and hypotheses

To provide a coherent theoretical basis for this study, we structure our framework into two distinct dimensions: first, the rationale for analyzing social media as a proxy for societal sentiment (Section 2.1); and second, the geographical mechanisms driving the spatial disparity of these sentiments (Section 2.2).

2.1. Digital health geography: Social Amplification and methodological innovation

Why is it academically critical to study online discourse regarding doctor-patient conflict? We draw upon the Social Amplification of Risk Framework (SARF). SARF suggests that risk events interact with psychological and social processes to heighten public perceptions of risk (Kasperson et al., 2003). In the digital age, social media platforms act as powerful "amplification stations," transforming individual medical disputes into collective social narratives that are difficult to capture through traditional official statistics (Wirz et al., 2018).

Building on this framework, recent scholarship in health geography and urban studies has increasingly legitimized the use of social media data as a proxy for understanding complex social phenomena, a concept often termed "social sensing (Liu et al., 2015)." Recent studies have successfully utilized platforms like Twitter and Weibo to map spatiotemporal patterns of public sentiment during crises or urban natural disasters (Ilyas & Sharifi, 2025). Methodologically, these works typically employ traditional Natural Language Processing (NLP) techniques to extract semantic clusters and sentiment polarities (Zhang & Chen, 2025). For instance, research has demonstrated that online discourse regarding public services often correlates strongly with regional socioeconomic indicators, providing a real-time, bottom-up supplement to traditional top-down survey data (Zhang, Wei, et al., 2026). These findings confirm that digital footprints serve as a valid mirror of physical societal structures.

However, while traditional quantitative NLP methods have laid the groundwork, capturing these nuanced narratives at a large scale presents a challenge for traditional quantitative methods. Recent urban and social scholarship advocates for the use of Large Language Models (LLMs) as a "sociological instrument" capable of decoding complex, context-dependent social sentiments that keyword counting or traditional topic modeling (like LDA) often miss (Fatemi et al., 2025; Lin et al., 2025; Wang et al., 2024; Zhang et al., 2024). By integrating LLMs, this study aims to bridge the gap between qualitative depth and quantitative breadth (Zhang, Wei, & Kong, 2026), validating the use of unstructured social data to monitor the pulse of public sentiment (Zhang, Jiang, et al., 2026; Ziems et al., 2024).

2.2. Mechanisms of regional disparity: supply vs. expectations

While Section 2.1 establishes the "how" and "where" of the discourse, this section addresses the "why" of its spatial variation. We integrate concepts from Health Geography and Spatial Justice, conceptualizing "regional disparities" not merely as statistical differences, but

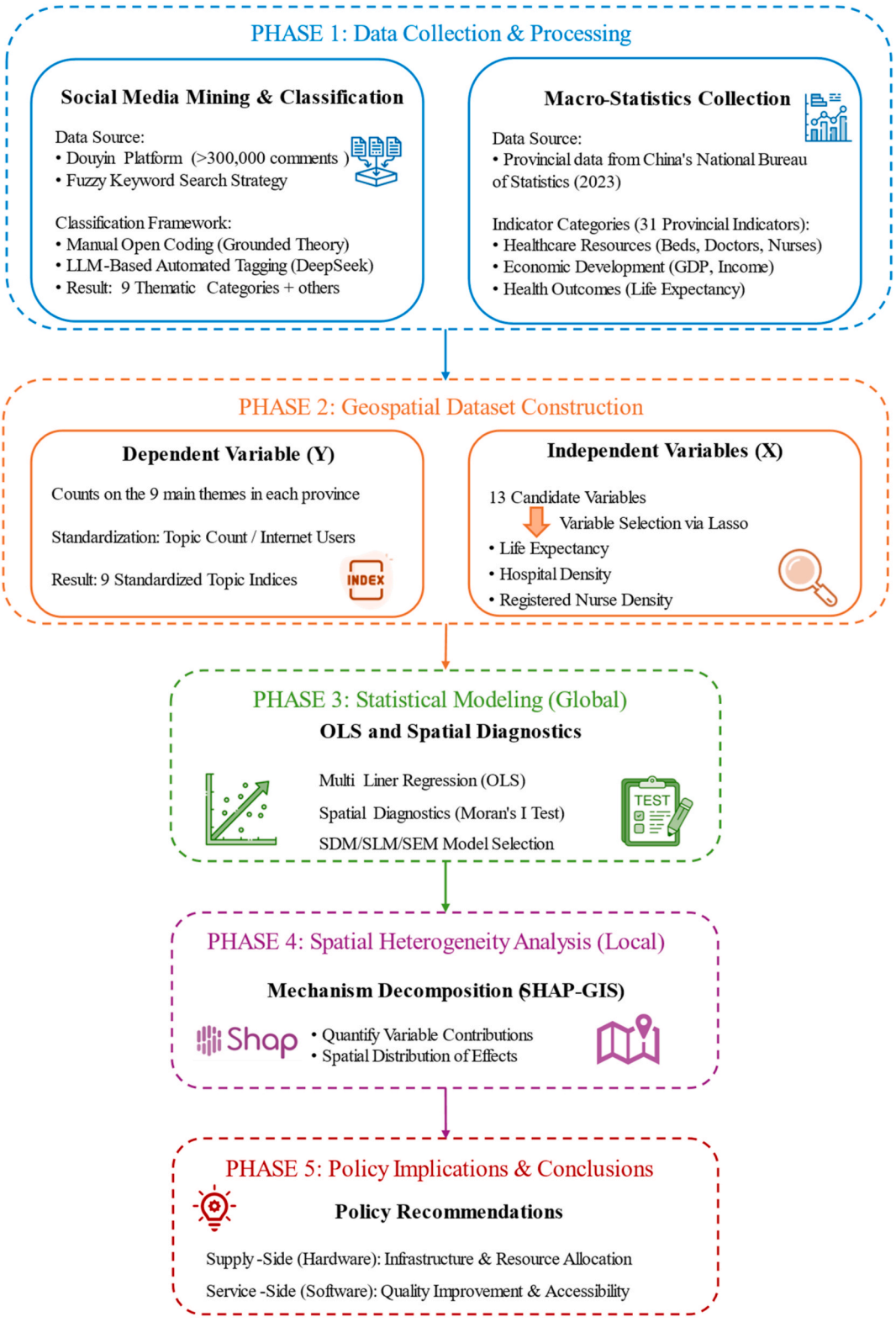


Fig. 1. The overall research design roadmap.

through the lens of Health Equity (Whitehead et al., 2019). We posit that online grievances are the digital manifestation of physical spatial injustices (Parr, 2002).

Existing literature validates this spatial stratification. Research indicates that the manifestation of doctor-patient tension varies significantly across China's diverse development contexts (He & Qin, 2023). In resource-concentrated eastern metropolises, conflicts are frequently triggered by high expectations and systemic efficiency issues (e.g., long waiting times despite high-quality care), leading to a "crisis of complexity" (Zhang & Sleeboom-Faulkner, 2011). In contrast, in less developed western and central regions, grievances are often rooted in fundamental accessibility barriers and the affordability of basic medical services (Pan et al., 2015). This observed East-West dichotomy suggests that the physical geography of resource allocation fundamentally shapes

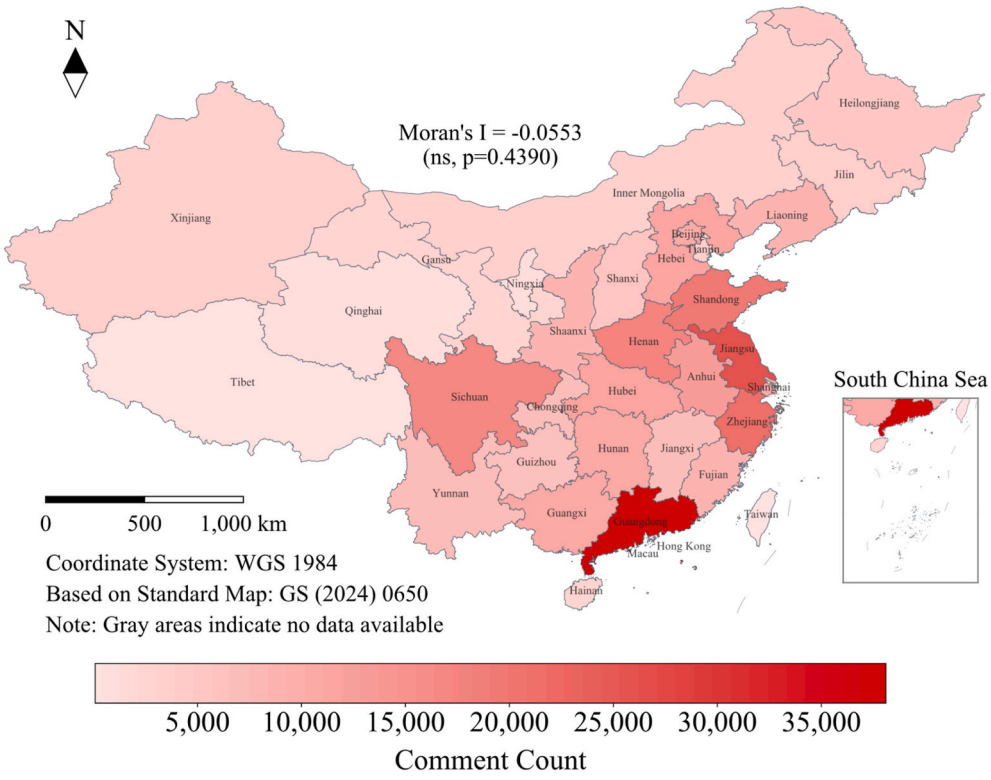


Fig. 2. Provincial distribution of valid comments.

the nature of public complaint (Sun et al., 2017).

Specifically, these empirical observations point to two competing mechanisms governing the tension between healthcare supply and public demand: Classical economic theory suggests that increasing supply should alleviate scarcity-induced friction. Therefore, regions with a higher density of physical resources (e.g., Hospitals and Nurses) should theoretically exhibit reduced grievance levels due to improved access (Lakin & Kane, 2022; MacKinnon et al., 2023). Conversely, sociological theories of "Rising Expectations" argue that in more modernized regions (characterized by higher Life Expectancy), public standards for dignity, efficiency, and service quality rise disproportionately. This "modernization penalty" may offset the benefits of resource expansion, as the public becomes more critical of the system's "software" rather than just its "hardware" (Chandra & Foster, 2005; Pan & Chen, 2022).

2.3. Hypotheses

Based on these theoretical underpinnings, we propose the following hypotheses to be tested:

H1. (The Supply-Side Hypothesis): Regions with a higher density of healthcare resources (specifically Hospital Density) will exhibit a lower intensity of grievance-related discourse, reflecting the mitigating effect of improved physical access ("The Supply Buffer").

H2. (The Modernization & Complexity Hypothesis): Socioeconomic modernization will qualitatively and quantitatively reshape public discourse in ways that counteract supply-side benefits.

H2a. (The Discursive Shift): In regions with higher development levels (proxied by Life Expectancy), public attention will shift from basic cost concerns to higher-order systemic reflections (e.g., institutional ecology), driven by a "hierarchy of needs."

H2b. (The Modernization Tension): Highly urbanized regions will face a "Developmental Paradox." The stabilizing effect of resource abundance (H1) will be offset or overwhelmed by the amplifying effect of modernization pressures (H2a), creating a high-intensity discourse environment driven by conflicting structural forces rather than simple scarcity.

3. Research design

This study investigates the spatial heterogeneity and socioeconomic drivers of public discourse on doctor-patient conflict in China. As outlined in the research roadmap (Fig. 1), our analytical framework follows a structured workflow: We constructed a dataset of over 300,000 user comments, classified via a hybrid method combining manual open coding and LLM automated tagging. We then quantified public attention into a normalized "Topic Index" and employed Lasso regression to rigorously screen candidate variables for the most robust predictors. Finally, we applied a multi-step modeling approach combining Multiple Linear Regression (MLR), SHAP analysis, and spatial mapping to identify global drivers and decompose localized mechanisms.

3.1. Data preparation

This study's empirical analysis relies on two primary types of data: a large-scale collection of user comments from a leading social media platform and a set of provincial-level statistics on socioeconomic and healthcare attributes.

3.1.1. Social media data

The social media data was sourced from Douyin, China's most popular short-video platform. The data acquisition and cleaning process was conducted as follows: First, on May 15, 2025, we performed a comprehensive video search using the keyword '医患矛盾' ("doctor-patient conflict"). The platform's fuzzy search algorithm returned to a broad set of potentially relevant videos. To ensure the relevance and significance of our data, we manually filtered this initial pool through a two-step process: (1) we removed videos whose content was not directly

related to doctor-patient issues, and (2) we excluded videos with fewer than 1000 user comments to focus on discussions with substantial public engagement. This refinement process yielded a final sample of 96 highly relevant videos. Next, on May 16, 2025, we collected the entire corpus of user comments from these 96 videos. This initial raw dataset contained 523,901 comments, each including its province-level IP address and the number of "likes" it had received.

To create a high-quality "core database" for analysis, we implemented a rigorous cleaning protocol. We began by removing emojis and other non-textual, irrelevant characters. While we acknowledge that emojis often carry emotional sentiment, this removal was a deliberate methodological choice driven by our research objective. Our study focuses on thematic topic classification (identifying the specific substance of the grievance, e.g., "high costs" vs. "poor attitude") rather than sentiment analysis (assessing emotional polarity). In this context, emojis were treated as noise that could obscure the semantic clarity required for accurate topic identification by the LLM. Subsequently, to filter out low-information content and ensure that each comment had substantive value for analysis, we deleted any comment containing fewer than six Chinese characters. This cleaning procedure resulted in a final, validated dataset of 326,323 comments. The provincial-level spatial distribution of these comments was visualized using GIS, as shown in Fig. 2.

To assess potential selection bias introduced by the 1000-comment threshold, we conducted a sensitivity analysis using a supplementary dataset of 11,056 comments from 'medium-engagement' videos (100-999 comments). The results (see Table 4 in **Appendix**) indicate that while medium-engagement videos contain more general noise (higher 'Others' category) and slightly more focus on daily hassles like 'Medical Costs' and 'Efficiency', the structural composition of the discourse remains robust. Crucially, systemic themes such as 'Medical Resources' and 'Perception and Empathy for Professionals' remain consistently prominent across both datasets. This confirms that our main dataset captures the core structural dimensions of public sentiment, while understandably amplifying the most acute friction points (Conflicts and Attitude) that drive broad social attention.

3.1.2. Socioeconomic and healthcare data

To investigate the underlying causes of regional variations in social media discourse, we must situate these online conversations within the real-world contexts from which they emerge. We posit that the climate of opinion surrounding the doctor-patient relationship is shaped by two tiers of influence: the macro-level socioeconomic conditions that define residents' lives, and the more immediate characteristics of the local healthcare system they directly experience. Accordingly, we constructed our explanatory framework by collecting two distinct categories of provincial-level data from the official website of the National Bureau of Statistics (NBS) of China.[1]

First, we examine what we term socioeconomic and demographic indicators. These variables collectively sketch the "initial conditions" and "social mindset" of the populace when they engage with the healthcare system. Economic Indicators, including GDP per capita (2023), Disposable Income per capita (2023), and Consumption Expenditure per capita (2023), measure the financial capacity and pressures faced by residents. The economic standing of a region directly relates to its citizens' ability to afford medical care and their sensitivity to costs, forming a critical foundation for their healthcare experience and subsequent evaluations. Social and Demographic Indicators, such as Life Expectancy (2020), Natural Population Growth Rate (2023), Gender Ratio (2020), and the Average Years of Education (Age 15+, 2020), offer a broader portrait of a province's social maturity, demographic dynamics, and the overall literacy of its residents. A higher level of education, for instance, may correlate with greater health literacy and rights consciousness, profoundly influencing how individuals interact with and what they expect from the medical system. Together, these socioeconomic factors form the "background canvas" against which public sentiment is painted.

Second, we shift our focus to the healthcare system indicators, which capture the tangible environment that the public most directly perceives and experiences. If socioeconomic context is a distal cause, then a region's healthcare resource allocation, service capacity, and operational pressures are the proximate drivers of discourse surrounding doctor-patient conflict (Potter & McKinlay, 2005). We assessed this through two sets of metrics: Healthcare Resource Supply, including Hospitals per 10,000 people, Healthcare Workers per 10,000 people, Practicing Physicians per 10,000 people, and Registered Nurses per 10,000 people. These metrics quantify the "hardware" of the healthcare system. The availability of adequate facilities and professional staff is fundamental to ensuring quality medical care and directly impacts patient convenience and experience. Healthcare System Load, measured by Healthcare Visits per capita people and Hospital Admissions per capita. These two indicators reflect the "operational pressure" on the system. Even in a region with a substantial supply of resources, excessive demands seen in high visit and admission rates—can lead to staff burnout, long wait times, and other frictions that are common catalysts for doctor-patient conflict.

3.2. Text classification

To systematically analyze the content of the 326,323 comments, we implemented a two-stage classification strategy. This process began with a qualitative, human-driven approach to develop a valid thematic framework, followed by a quantitative, AI-driven method to apply this framework at scale and evaluate its performance.

3.2.1. Thematic framework development via open coding

The foundation of our classification schema was built through a rigorous qualitative analysis of a representative data sample, following a systematic "Open-Axial-Selective" coding protocol to mitigate subjectivity.

First, we drew a random sample of 5000 comments from the core database. This sample size was deemed sufficient to reach thematic saturation while remaining manageable for in-depth manual review. Next, to ensure reliability, the coding process was conducted by two independent researchers who were blind to each other's annotations. The procedure involved three distinct stages:

Open Coding (Deconstruction) and Axial Coding (Reconstruction): Researchers performed line-by-line analysis to identify initial concepts. For instance, raw comments like "The doctor didn't look at me" or "The tone was very impatient" were labeled with initial codes such as "lack of eye contact" and "verbal impatience." These initial codes were then grouped into clusters based on semantic similarity. Continuing the example, "lack of eye contact" and "verbal impatience" were aggregated into a broader category of "Service Attitude Deficiencies."

Selective Coding (Theming): Finally, these categories were synthesized into core themes. "Service Attitude Deficiencies" was formalized into the final topic category: "Healthcare Staff Attitude and Communication".

Throughout this process, inter-coder reliability was rigorously monitored. Upon completion of the independent coding, the initial agreement rate was calculated at 94.5% **(**McHugh, 2012**)**. Discrepancies (2.5%) were resolved through a "negotiated agreement" process involving a third senior researcher to reach a consensus. This collaborative and iterative process resulted in the finalization of a robust thematic framework consisting of nine distinct topic categories and one "Others" category (Yan et al., 2024), as visually summarized in Fig. 3 which maps the trajectory from raw text to final themes.

Table 1 are the ten final topic categories used in our analysis, each with its definition and an illustrative example.

[1] https://www.stats.gov.cn/sj/.

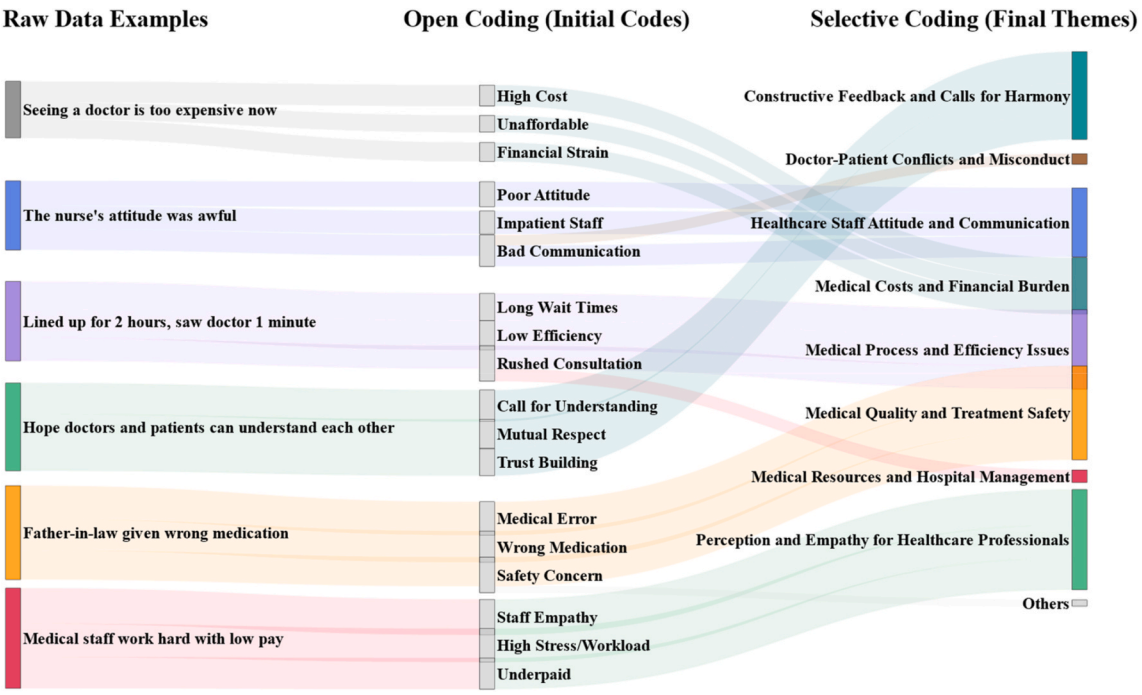


Fig. 3. Thematic framework development trajectory.

3.2.2. LLM-based classification and performance evaluation

With the human-validated thematic framework established, we aimed to classify the entire dataset. While traditional probabilistic topic modeling methods, such as Latent Dirichlet Allocation (LDA) or Structural Topic Models (STM), have been widely used in social science, they face significant limitations when applied to social media comments (Chen & Mankad, 2024). These methods rely on "bag-of-words" co-occurrence statistics, often struggling with the "short-text sparsity problem", where short, informal comments lack sufficient word counts to form coherent topic clusters. Furthermore, traditional models often fail to capture complex semantic nuances, sarcasm, and implicit grievances which are prevalent in Chinese internet slang.

In contrast, Large Language Models (LLMs) represent a paradigm shift in automated text analysis. They leverage vast pre-trained semantic knowledge to understand context beyond mere keyword frequency. Crucially, our research required a supervised classification task (assigning comments to our validated 9 categories) rather than an unsupervised clustering task (which is what LDA performs). LLMs demonstrate superior performance in such zero/few-shot classification scenarios, ensuring that the analysis adheres strictly to our theoretical framework rather than generating random or interpretable topics.

For this task, we employed *DeepSeek R1 671B*, an open-source LLM renowned for its excellent reasoning performance. We utilized a few-shot learning approach, which allows the model to perform a complex classification task based on a small number of high-quality examples (Song et al., 2023; Wang et al., 2020). We constructed a detailed prompt that provided the model with the name and precise definition of each of the ten categories, supplemented by several clear examples for each. The model was instructed to act as a "social media comment classification assistant" and to assign only the single most relevant topic to each comment, focusing on its core meaning.

To validate the reliability of this automated approach, we conducted a rigorous performance evaluation. We created a gold-standard test set by randomly sampling 1000 comments from the core database (separate from the initial 5000 used for theme development). We then compared the model's classifications on this test set against the ground truth established by our human coders. The model demonstrated high fidelity to our framework, achieving an impressive overall accuracy of 92.1%. A detailed breakdown of the model's performance for each topic is presented in Table 2. For context, the human-coded ground truth itself was established with a very high degree of consistency; the two independent researchers achieved an inter-coder agreement rate of 97.5% on the same test set (McHugh, 2012). The model's strong performance, approaching the level of human agreement, confirmed its suitability for reliably classifying the entire dataset of 326,323 comments and provided a solid foundation for the subsequent quantitative analysis. The detailed Prompt and core Python Code could be found in Supporting Materials.

3.3. Analysis of driving factors

Having identified and quantified the key themes of public discourse, the final stage of our methodology aims to statistically explain their geographical variations. Our analytical strategy involves a rigorous process of variable construction, model selection, and advanced interpretation.

3.3.1. Dependent variable: construction of the normalized Topic Index

To ensure comparability across provinces with vastly different populations and levels of digital infrastructure, we constructed a normalized metric: the **Topic Index**. This index is designed to measure the *relative intensity* of public attention toward a specific topic, rather than the absolute volume of discussion, which is naturally biased by the size of the local user base. The Topic Index T_{ij} for topic i in province j is calculated as follows:

$$T_{ij} = \frac{C_{ij}}{U_j}$$

where: C_{ij} is the **raw frequency** count of comments related to topic i in province j. U_j is the number of mobile internet users in province j (also ger from National Bureau of Statistics of China). While province-level user data specific to the Douyin platform is proprietary and unavailable, we utilize "mobile internet users" as the most robust proxy. Given the exceptionally high penetration rate of short-video applications in China, Douyin functions as a near-universal platform. Therefore, normalizing by the mobile internet population effectively controls for

Table 1
Topic categories, definitions and illustrative examples.

Category	Definition	Illustrative Example
Constructive Feedback and Calls for Harmony	Comments that offer specific suggestions for improvement, express gratitude for good medical service, share positive experiences, or call for mutual understanding, trust, and cooperation between doctors and patients to build a more harmonious relationship.	I hope both doctors and patients can have more understanding and respect for each other. This would reduce a lot of conflicts.
Doctor-Patient Conflicts and Misconduct	Comments recounting direct grievances, including allegations of misconduct, verbal and physical altercations, and condemnations of disruptive behaviors (yinao) from patients or perceived mistreatment by staff.	Regardless of who was right or wrong, resorting to physical violence is already a mistake.
Healthcare Ecology of Specific Institutions or Regions	Critiques that attribute systemic problems to the perceived poor medical environment or chronic mismanagement at a specific, named hospital or within a particular region.	Shanghai Sixth Hospital again? How does this hospital always have problems?
Healthcare Staff Attitude and Communication	Expressions of frustration and anger directed at the perceived indifference, impatience, or poor communication of healthcare staff. These comments often detail feelings of being disrespected or dehumanized during clinical encounters.	The nurse's attitude was awful. She was so impatient when I asked a question, as if I owed her money.
Medical Costs and Financial Burden	Expressions of outrage regarding the high cost of care, with users often criticizing the perceived lack of transparency in billing, alleging profit-driven motives behind prescriptions and tests, and detailing the severe financial strain placed on families.	Seeing a doctor is too expensive now. A simple cold cost me several hundred yuan.
Medical Process and Efficiency Issues	Critiques of systemic inefficiencies within the medical process. These comments voice frustration with excessive wait times, convoluted procedures for appointments and tests, and the prevalent "3-h wait, 3-min consultation" phenomenon.	Lined up for 2 h to see the doctor for 1 min. The efficiency is just too low.
Medical Quality and Treatment Safety	Expressions of deep-seated mistrust in the quality of care and treatment safety. These comments often detail fears or allegations of misdiagnosis, medical errors, negligence, and unsafe practices that put patient safety at risk.	My father-in-law was given the wrong medication right after being transferred from the ICU to a general ward. He nearly died.
Medical Resources and Hospital Management	Criticisms aimed at hospital management and resource allocation. Users in this category often blame systemic problems on a perceived shortage of staff or beds, poor institutional policies, and unaccountable or incompetent hospital administration.	With so many patients, why doesn't the hospital add more emergency doctors? What is the management thinking?
Perception and Empathy for Healthcare Professionals	Comments that explore the unique nature of the medical profession and express understanding, sympathy, respect, or defense for the hard work and pressure faced by healthcare staff.	Doctors and nurses have it really tough. They face so many patients every day for low pay and under great stress.
Others	Comments that do not fall into any of the above categories, such as irrelevant topics, pure emotional venting without specific context, or content whose core intent is difficult to determine.	This comment section is really something else. Setting aside the incident, it's all just personal stories.

Table 2
Performance of the LLM classifier on the validation set.

Topic Category	Frequency	True Count	Accuracy Rate (%)
Constructive Feedback and Calls for Harmony	46	44	95.65
Doctor-Patient Conflicts and Misconduct	221	206	93.21
Healthcare Ecology of Specific Institutions or Regions	25	24	96
Healthcare Staff Attitude and Communication	163	156	95.71
Medical Costs and Financial Burde	51	45	88.24
Medical Process and Efficiency Issues	23	22	95.65
Medical Quality and Treatment Safety	75	67	89.33
Medical Resources and Hospital Management	128	113	88.28
Perception and Empathy for Healthcare Professionals	122	111	90.98
Others	146	133	91.1
Total	1000	921	92.1

the regional "digital divide," ensuring that the index reflects genuine variations in public sentiment rather than artifacts of internet accessibility.

3.3.2. *Independent variables: selection and collinearity diagnostics*

To comprehensively capture the socioeconomic drivers of public discourse, we initially constructed a pool of 13 candidate independent variables spanning economic development, demographic structure, healthcare infrastructure, and service capacity (see Fig. 10 for the distribution and descriptive statistics of all standardized candidate variables). Given the limited sample size (N = 31 provinces) and potential multicollinearity among the 13 candidate variables, traditional manual selection risks subjectivity and overfitting. To address this, we employed Least Absolute Shrinkage and Selection Operator (Lasso) regression for objective feature selection. We ran Lasso regression models for each of the nine topic indices. Since Lasso automatically shrinks coefficients of less relevant features to zero, we counted the frequency with which each variable was selected (i.e., had a non-zero coefficient) across the nine models. Fig. 4 illustrates this selection frequency.

To adhere to the principle of parsimony and ensure sufficient degrees of freedom (EPV ratio ≈ 10:1), we restricted our final OLS and spatial models to these Top 3 variables (Austin & Steyerberg, 2015). As shown in Fig. 4, a distinct "natural break" appears after the top three variables. Life Expectancy, Hospital Density, and Registered Nurse Density emerged as the dominant predictors. Following the Lasso selection, we

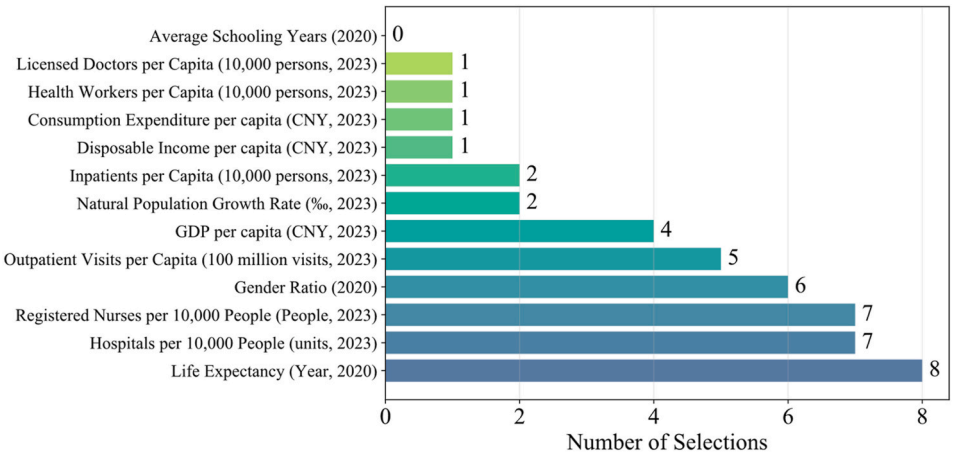


Fig. 4. Variable selection frequency via Lasso regression.

conducted a collinearity diagnostic on the final three-variable model. The Variance Inflation Factor (VIF) analysis confirms that multicollinearity is negligible (O'brien, 2007), with all values remaining well below the strict threshold of 5 (Mean VIF <2, see Appendix Fig. 11), ensuring the statistical stability of our regression estimates.

To finalize our variable selection, we visualized the spatial distribution of the three chosen predictors and calculated their Global Moran's I to detect inherent spatial structures (see Appendix Fig. 12). The analysis reveals distinct spatial profiles: Both Life Expectancy ($I = 0.534$, $p < 0.01$) and Hospital Density ($I = 0.484$, $p < 0.01$) exhibit strong positive spatial autocorrelation. This confirms that socioeconomic development and physical infrastructure are highly clustered, likely driving the regional concentration of public discourse. In contrast, Registered Nurse Density shows no significant spatial clustering ($I = -0.106$, $p = 0.228$). This suggests that unlike physical infrastructure, the distribution of the nursing workforce is more spatially randomized across provinces, strictly following administrative allocations rather than regional economic clusters.

3.3.3. *Baseline modeling and spatial specification*

We constructed a separate Multiple Linear Regression (MLR) model (Uyanık & Güler, 2013) for each of our nine core topic indices (C_k) to estimate the average effect of various provincial indicators (collectively referred to as socioeconomic development explanations, D_i). This allowed us to identify which provincial characteristics have a statistically significant relationship with public concerns on a national level. The general form of our MLR model is as follows:

$$C_{k,j} = \beta_0 + \sum_{i=1}^{13} \beta_i D_{i,j} + \epsilon_j \quad (1)$$

where: $C_{k,j}$ is the Topic Index for topic k (e.g., Medical Costs and Financial Burden) in province j. $D_{i,j}$ is the value of the i-th socioeconomic development explanation (e.g., Life Expectancy) in province j. β_i is the regression coefficient representing the average change in the Topic Index for a one-unit change in D_i. β_0 is the model intercept. ϵ_j is the error term for province j. We constructed a separate MLR model for each of our nine core topic indices (C_k) to estimate the average effect of various provincial indicators.

Given the potential for spatial dependence among provinces and the limitations of Lagrange Multiplier (LM) tests in small samples ($N = 31$), we adopted a robust specification strategy. Instead of relying solely on post-OLS diagnostics, we proactively estimated the Spatial Durbin Model (SDM), a general specification that nests both the Spatial Lag Model (SLM) and Spatial Error Model (SEM) to capture potential spatial spillovers in both dependent and independent variables (Rüttenauer, 2019). Given the small sample size, we adopted a conservative model selection strategy based on comprehensive assessment of spatial effects. A spatial model was retained only when strong spatial dependence was detected, as indicated by statistically significant LR tests supporting the specific spatial structure (i.e., spatial lag or spatial error effects). In cases where spatial effects were weak or statistically insignificant, the parsimonious OLS model was preferred to preserve degrees of freedom and ensure robust inference.

To ensure the statistical validity of our inferences, we implemented two critical robustness measures: For models where the Breusch-Pagan test indicated heteroskedasticity (e.g., "Healthcare Ecology"), we employed HC3 robust standard errors (Hayes & Cai, 2007). Recognizing the risk of Type I errors associated with testing nine concurrent models, we applied the Benjamini-Hochberg procedure to control the False Discovery Rate (FDR) (Thissen et al., 2002). We set the significance threshold (α) at 0.05, ensuring that the reported significant relationships are robust and not artifacts of multiple hypothesis testing.

3.3.4. *Model interpretation: SHAP value analysis*

While the MLR coefficients indicate the direction of effects, SHAP analysis allows us to quantify the magnitude of impact for each predictor. We employed SHAP (SHapley Additive exPlanations) to gain a deeper, more nuanced understanding of each factor's impact. By integrating MLR for robust statistical inference with SHAP for rich, detailed interpretation, our analysis can precisely identify not only *which* factors are significant drivers of public discourse but also *how* their influence varies across different provincial contexts (Ponce-Bobadilla et al., 2024; Zhang, Li, et al., 2026). Given our dataset is limited to 31 provincial units, applying GWR poses significant risks of overfitting and parameter instability due to insufficient degrees of freedom. Consequently, we adopted a robust alternative: using SHAP values to visualize the spatially heterogeneous contributions of features derived from the stable global MLR model.

4. Results

This section presents the empirical findings of our study. We begin by outlining the overall thematic landscape of public discourse on doctor-patient conflict, followed by an analysis of the geographical distribution of these themes and the results of our regression models exploring their underlying drivers.

4.1. *The landscape of public discourse: volume vs. resonance*

Our initial analysis examines the thematic distribution of the 306,323 classified comments. We distinguish between the volume of discussion (raw frequency) and the public resonance (average likes per

Table 3
Thematic Distribution of Public Discourse: Volume vs. Resonance.

Categories	Count	%	Average Likes
Constructive Feedback and Calls for Harmony	10,312	3.37	14.61
Doctor-Patient Conflicts and Misconduct	65,627	21.42	12.87
Healthcare Ecology of Specific Institutions or Regions	11,573	3.78	14.27
Healthcare Staff Attitude and Communication	54,071	17.65	10.58
Medical Costs and Financial Burde	14,936	4.88	6.73
Medical Process and Efficiency Issues	6141	2.00	18.02
Medical Quality and Treatment Safety	23,722	7.74	16.64
Medical Resources and Hospital Management	39,230	12.81	18.85
Perception and Empathy for Healthcare Professionals	38,527	12.58	21.75
Others	42,184	13.77	12.96

comment). The results are presented in Table 3. Based purely on frequency, the discourse is dominated by negative personal narratives. "Doctor-Patient Conflicts and Misconduct" is the most frequent topic (21.4%), followed closely by "Healthcare Staff Attitude" (17.7%). Together, these grievance-based categories comprise nearly 40% of the dataset, depicting a highly contentious online environment.

However, examining the average likes per comment reveals a divergence between what users *say* and what the community *endorses*. While direct complaints about costs and conflicts are voluminous, they garner relatively lower community validation (e.g., only 6.7 likes/comment for "Medical Costs"). In contrast, constructive and systemic topics exhibit significantly higher resonance. Notably, comments related to "Perception and Empathy for Healthcare Professionals" received the highest level of endorsement, averaging 21.75 likes per comment significantly higher than the conflict-related topics. Similarly, discussions on "Medical Resources" (18.85 likes) and "Process Efficiency" (18.02 likes) also garnered strong support.

This "Volume-Resonance Gap" suggests a nuanced public sentiment: while individual users frequently vent about immediate pain points, the broader silent majority is more likely to endorse reflective content that acknowledges the difficulty of medical work or critiques systemic resource constraints.

4.2. Geographical distribution of public discourse

To explore the regional variations in public discourse, we mapped the normalized Topic Index for each of the nine core themes. Figs. 5 and 6 presents the spatial distribution of these indices using choropleth maps (see Appendix Fig. 13 for detailed numerical data). The visualization reveals distinct geographical patterns: The Coastal-Inland Dichotomy in Grievances The most intense conversations regarding friction points occur in China's economically developed, urbanized regions. As shown

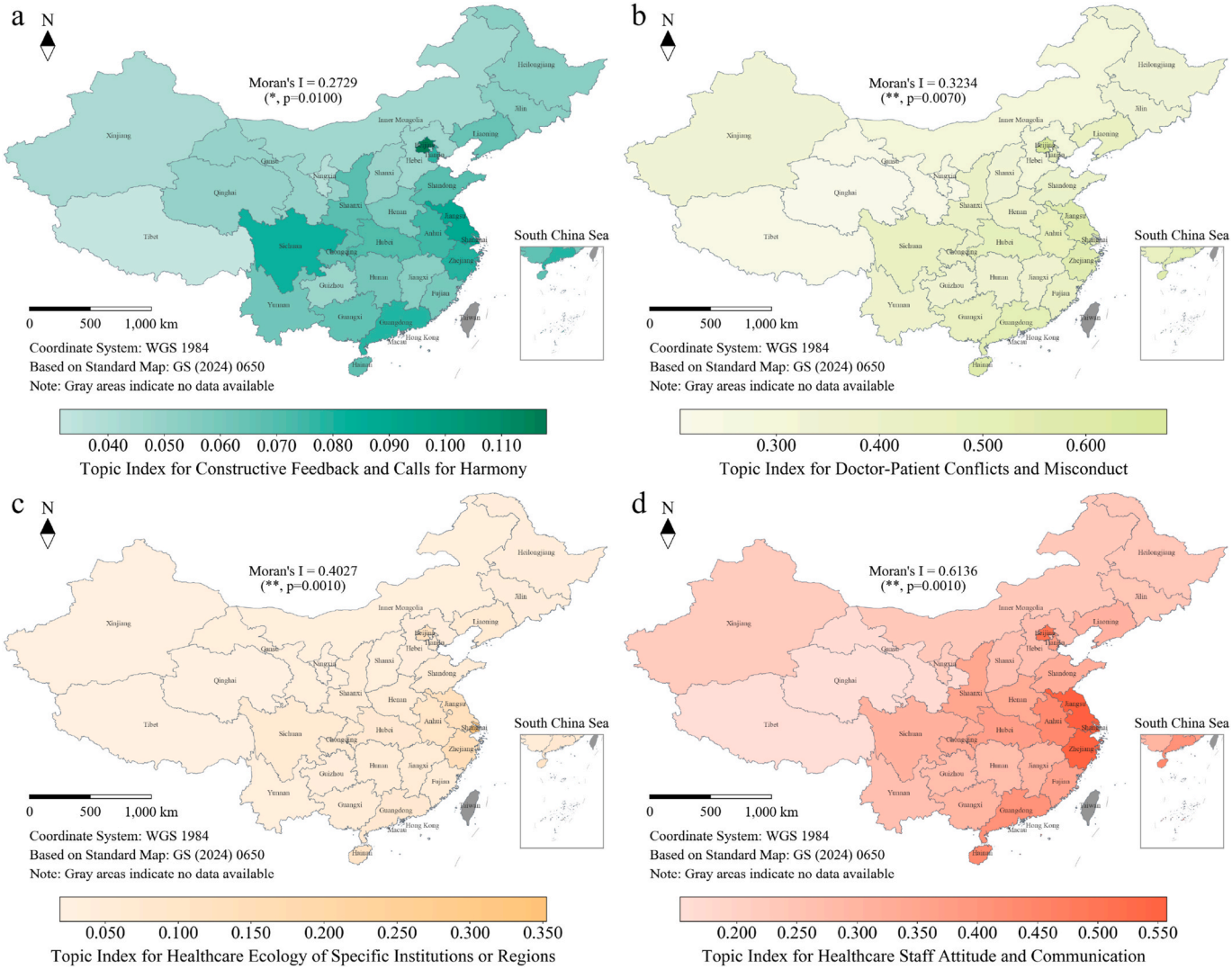


Fig. 5. Spatial Distribution of the first four Topic Indices across 31 Provinces in Mainland China.

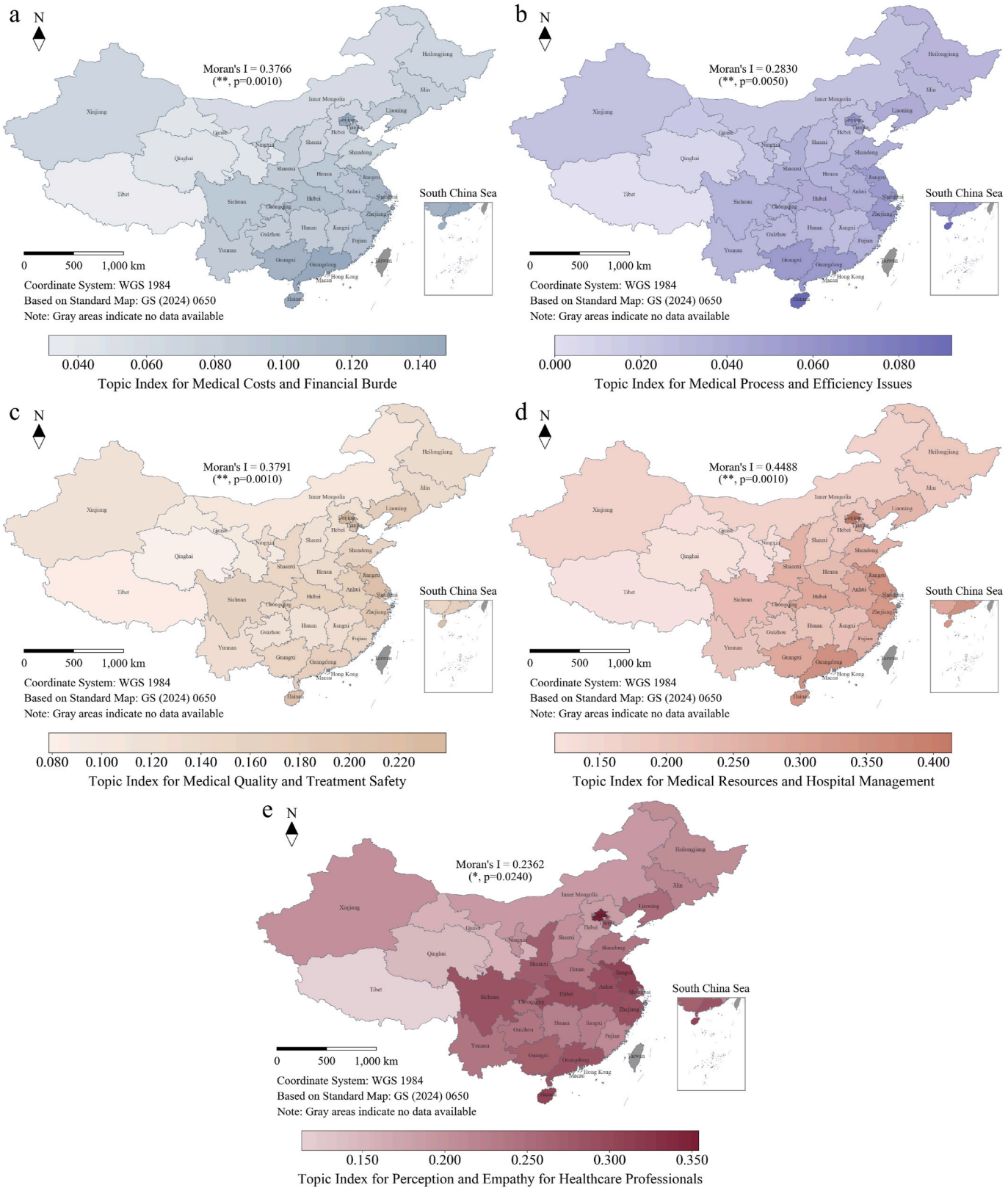
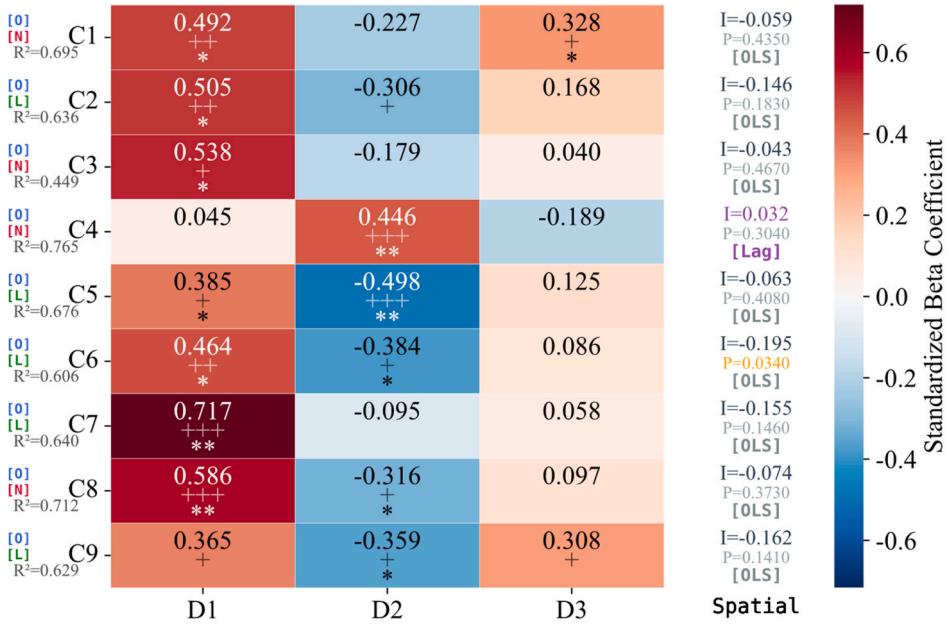


Fig. 6. Spatial Distribution of the else Topic Indices across 31 Provinces in Mainland China.

in Fig. 5 b ("Conflicts") and Fig. 5 d ("Staff Attitude"), high-intensity clusters (indicated by darker shades) are consistently found in Beijing, Shanghai, and the coastal provinces of Jiangsu and Zhejiang. This spatial agglomeration indicates that the epicenters of modernization are also the epicenters of online criticism regarding service quality and interpersonal conflict.

Intriguingly, the spatial distribution of positive sentiment mirrors that of grievances. Fig. 6 e ("Perception and Empathy") shows that developed regions like Beijing and Jiangsu also exhibit the highest indices for empathy towards professionals. This suggests a polarized



FDR-Corrected Significance (α=0.05): * q<0.05, ** q<0.01, *** q<0.001
Original Significance: + p<0.05, ++ p<0.01, +++ p<0.001
Ramsey RESET Test: [L]=Linearity supported (p≥0.05), [N]=Potential nonlinearity (p<0.05)
Breusch-Pagan Test: [O]=Homoskedasticity (p≥0.05)
Spatial Analysis: I= Residual Moran's I, [Lag]=SLM, [Err]=SEM, [SDM]=SDM, [OLS]=OLS
Independent Variables:
D1: Life Expectancy (Year, 2020) D2: Hospitals per 10,000 People (Units, 2023)
D3: Registered Nurses per 10,000 People (People, 2023)
Topic Category:
C1: Constructive Feedback and Calls for Harmony
C2: Doctor-Patient Conflicts and Misconduct
C3: Healthcare Ecology of Specific Institutions or Regions
C4: Healthcare Staff Attitude and Communication
C5: Medical Costs and Financial Burde C6: Medical Process and Efficiency Issues
C7: Medical Quality and Treatment Safety
C8: Medical Resources and Hospital Management
C9: Perception and Empathy for Healthcare Professionals

Fig. 7. Heatmap of standardized beta coefficients from regression models.

public discourse in these hubs: they are not merely centers of complaint but also foster a sophisticated counter-narrative acknowledging the systemic pressures faced by medical staff. In sharp contrast, the intensity of discourse is significantly lower in western and inland provinces such as Gansu, Qinghai, and Tibet (lighter shades across most maps). This "silence" in less developed regions may reflect different local priorities, where issues of basic access might outweigh service-level critiques, or differences in digital engagement with these specific topics.

Finally, certain topics defy a simple East-West divide. "Medical Costs" (Fig. 6 a) displays a more dispersed pattern, with high indices appearing in both affluent areas (Guangdong) and regions with lower average incomes (Guangxi). This demonstrates that financial burden is a universal pressure point that transcends regional economic boundaries, whereas "institutional ecology" (Fig. 5 c) is strictly clustered in developed regions, marking it as a "modernization-specific" concern.

4.3. Socioeconomic drivers of public discourse

We employed the parsimonious Top 3 variable model (selected via Lasso) to identify the drivers of regional discourse variations. Fig. 7 summarizes the regression results, and the detailed regression coefficients, standard errors, and diagnostic statistics for all nine models are provided in the **Supplementary Excel File**. The model demonstrates robust explanatory power, with an average R^2 of 0.645 across the nine topics. Based on the adaptive model selection, the MLR specification was optimal for eight topics, while the Spatial Lag Model (SLM) was adopted for *"Healthcare Staff Attitude"* to account for spatial contagion (McFadden's Pseudo $R^2 = 0.765$). Regression results reveal two distinct, and at times opposing, mechanisms driving public sentiment, directly addressing our theoretical hypotheses. While for *"Medical Process and Efficiency Issues"*, Lagrange Multiplier tests indicated no significant spatial dependence (LM-Lag $p = 0.827$; LM-Error $p = 0.666$), OLS regression was adopted as the final model ($R^2 = 0.606$).

Consistent with the Modernization Hypothesis (H2a), our proxy for regional development and societal expectations (Life Expectancy) emerges as a powerful positive predictor for grievance intensity. Systemic & Quality Concerns: Life Expectancy is significantly positively associated with "Healthcare Ecology" (beta = 0.538, $p < 0.05$) and "Medical Quality" (beta = 0.717, $p < 0.01$). This confirms H2a (Discursive Shift): as regions modernize, public focus shifts from basic access to higher-order critiques of institutional governance and technical safety. Notably, Life Expectancy also positively predicts "Doctor-Patient Conflicts" (beta = 0.505, $p < 0.05$). This suggests that modernization does not automatically bring harmony; rather, the

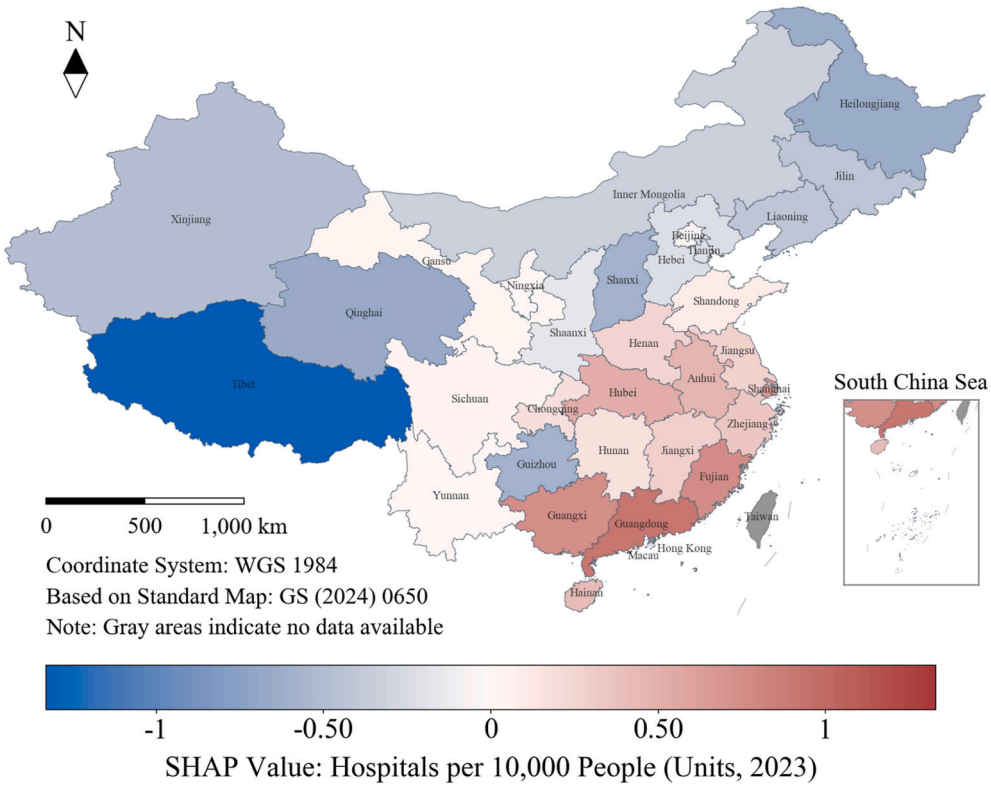


Fig. 8. Spatial distribution of SHAP values: The contribution of Hospital per 10,000 people to the "medical cost and financial burde" topic index.

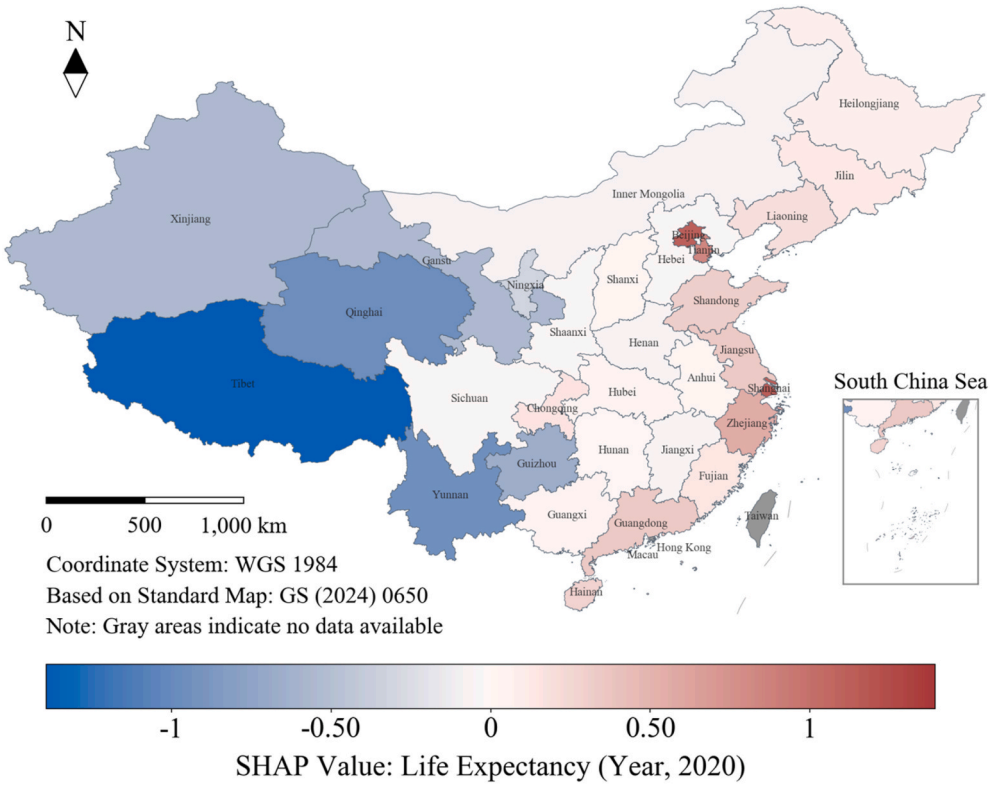


Fig. 9. Spatial distribution of SHAP values: The contribution of life expectancy to the "healthcare ecology of specific institutions or regions" topic index.

heightened expectations in developed regions may lower the threshold for conflict perception.

In contrast, Hospital Density functions as a critical buffer, validating the Supply-Side Hypothesis (H1). Higher hospital density is significantly associated with *lower* grievance levels regarding "Medical Costs" (beta $= -0.498$, $p < 0.01$) and "Medical Process Efficiency" (beta $= -0.384$, p

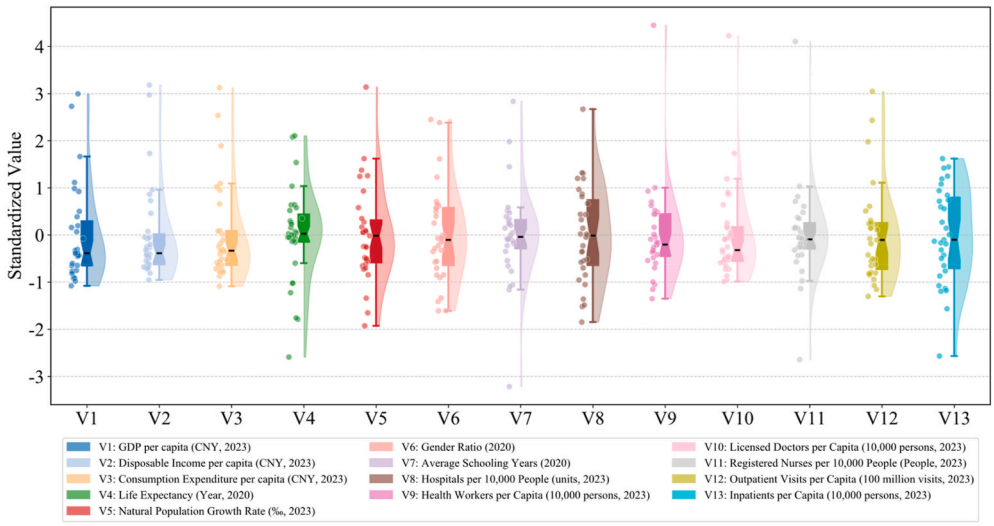


Fig. 10. Distribution of 13 standardized candidate variables.

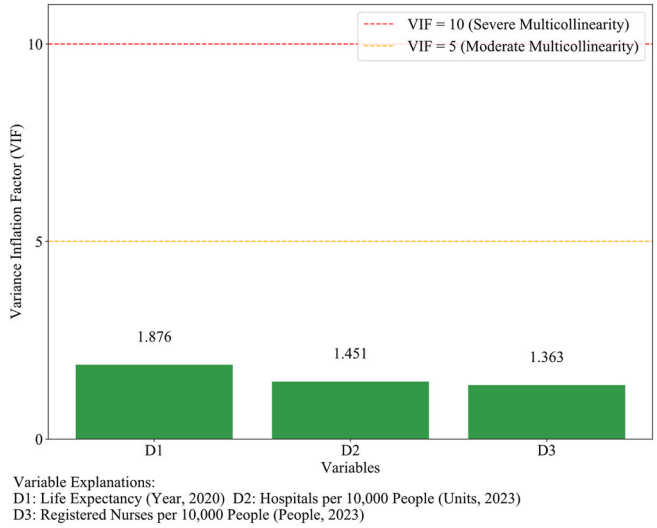


Fig. 11. Variance Inflation factor (VIF) diagnostics for the final model.

< 0.05). This indicates that expanding physical infrastructure effectively mitigates friction points related to affordability and accessibility. However, for "Healthcare Staff Attitude" (modeled via SLM), hospital density shows a significant positive association (beta = 0.446, $p < 0.01$). This suggests that while infrastructure solves efficiency problems, the high patient volume associated with dense medical hubs may strain patient-provider interactions, negatively impacting service attitude—a precursor to the "Modernization Paradox" discussed in the next section.

Registered Nurse Density exhibits a unique role in fostering positive sentiment. It is a significant positive predictor for "Constructive Feedback and Calls for Harmony" (beta = 0.328, $p < 0.05$). This implies that a sufficient nursing workforce, often the primary interface for emotional labor, is key to cultivating a more empathetic and constructive online discourse.

4.4. *Spatial non-stationarity of driving mechanisms*

While the global regression models identify the average direction of effects, they mask the significant spatial heterogeneity in *how* these drivers influence public sentiment locally. To deconstruct these mechanisms, we mapped the SHAP values for two paradigmatic relationships that validate our core hypotheses (Figs. 8 and 9).

Fig. 8 maps the localized impact of Hospital Density on "Medical Costs." The Stabilizer Effect (Validating H1): Across most healthcare resource-abundant provinces, such as Heilongjiang (SHAP = −0.65), Guizhou (SHAP = −0.60), and Xinjiang (SHAP = −0.48), the SHAP values are strongly negative (blue). This confirms that in these regions, higher per capita infrastructure effectively suppresses grievances regarding financial burden, validating the supply-side buffer hypothesis. The Modernization Tension (Validating H2b): Crucially, this buffering effect is reversed in southeast coastal areas. Guangdong (SHAP = 0.92), Fujian (SHAP = 0.76), and Hainan (SHAP = 0.42) exhibit significant positive contributions (red). This indicates that in these economically vibrant or high-inflow areas, high hospital density fails to mitigate cost concerns and may even correlate with higher spending pressures. This reversal visually captures the "Developmental Paradox" (H2b), where structural complexity in hubs overwhelms the benefits of simple resource expansion.

Fig. 9 visualizes the contribution of Life Expectancy to the discourse on "Healthcare Ecology." A striking "East-West" dichotomy emerges. In China's most modernized regions, higher life expectancy acts as a powerful driver for systemic critique. In eastern and coastal developed regions, such as Shanghai (SHAP = 1.13), Beijing (SHAP = 1.12), and Tianjin (SHAP = 0.83), high positive SHAP values (red) are observed, indicating that in these longevity-rich societies, development actively pushes public attention towards higher-order institutional reflections. Conversely, in less developed western provinces like Tibet (SHAP = −1.39) and Yunnan (SHAP = −0.95), this effect is strongly negative (blue). This spatial polarity robustly validates Hypothesis H2a, confirming a "hierarchy of needs" where systemic reflection is a distinct feature of modernization that has not yet permeated the hinterland.

5. Discussion

This study set out to investigate the spatial heterogeneity of online doctor-patient discourse in China, testing three core hypotheses derived from Health Geography and Risk Amplification theories. We began by mapping the thematic landscape and geographical distribution of over 300,000 user comments, classified via a validated LLM framework. By subsequently employing a parsimonious Top 3 variable model and distinguishing between global regression effects and SHAP-GIS, our findings paint a nuanced picture of a healthcare system in transition. We reveal a complex landscape where the stabilizing effects of resource expansion collide with the rising expectations of a modernizing society and the unique pressures of metropolitan aggregation.

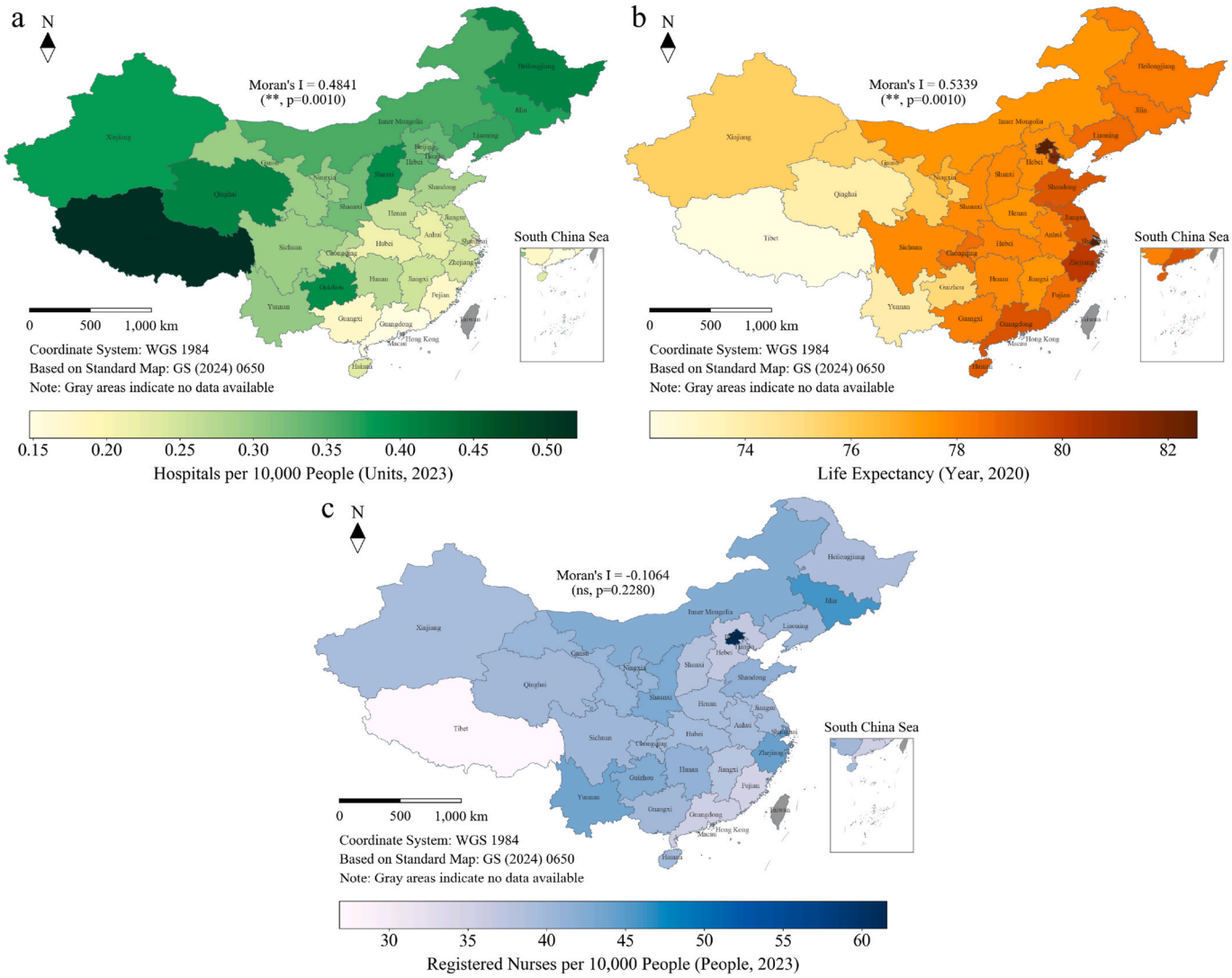


Fig. 12. Spatial distribution and Moran's I of the final three predictors.

5.1. *The "buffer effect" of infrastructure*

Our first hypothesis (H1) posited that increasing physical healthcare resources would mitigate public grievance. The regression results provide robust empirical support for this "supply-side" mechanism. The consistent negative coefficients for Hospital Density across critical grievance topics, most notably "Medical Costs" and "Process Efficiency" indicate on a national scale, expanding physical infrastructure serves as an effective stabilizer. Regions with higher per capita hospital availability consistently exhibit a lower baseline of online complaints regarding affordability and access.

This finding challenges the pessimistic view that resource investment is futile. Instead, it confirms that the Chinese government's massive investment in healthcare infrastructure over the past decade has yielded tangible benefits in dampening fundamental livelihood grievances. The "Supply Buffer" map visually reinforces this, showing that in resource-abundant provinces like Heilongjiang and Guizhou, infrastructure density effectively suppresses potential discontent.

5.2. *The Modernization of discourse: validating the rising expectations*

However, resource expansion is not the only force at play. Validating Hypothesis H2a, we found that socioeconomic modernization (proxied by Life Expectancy) qualitatively reshapes the nature of public sentiment. In highly developed coastal regions, higher life expectancy acts as a powerful positive driver for "Healthcare Ecology" and "Medical Quality", as visualized in the SHAP maps. This spatial divergence reveals a "hierarchy of needs" in public discourse. In regions where basic survival and access needs are secured, public attention migrates from immediate pain points to higher-order reflections on institutional governance, medical ethics, and technical safety. This confirms that online grievances in modernized China are not merely complaints about service failure but are increasingly sophisticated critiques of the system's structural integrity, driven by a rights-conscious public with elevated standards for "satisfactory care."

5.3. *The Modernization paradox: structural tension and saturation*

Perhaps the most critical theoretical contribution of this study is the empirical validation of the "Modernization Paradox" (H2b) through the lens of spatial non-stationarity. By decomposing the contribution of Hospital Density using SHAP analysis, we identified a fundamental shift in how physical infrastructure shapes public sentiment across different geographies. While Hypothesis H1 (the supply-side buffer) holds true for most inland provinces, Fig. 8 reveals a striking structural reversal in China's most developed economic hubs. In regions like Guangdong, Fujian, and Hainan, the SHAP values for Hospital Density flip from negative (mitigating) to strongly positive (amplifying). This indicates

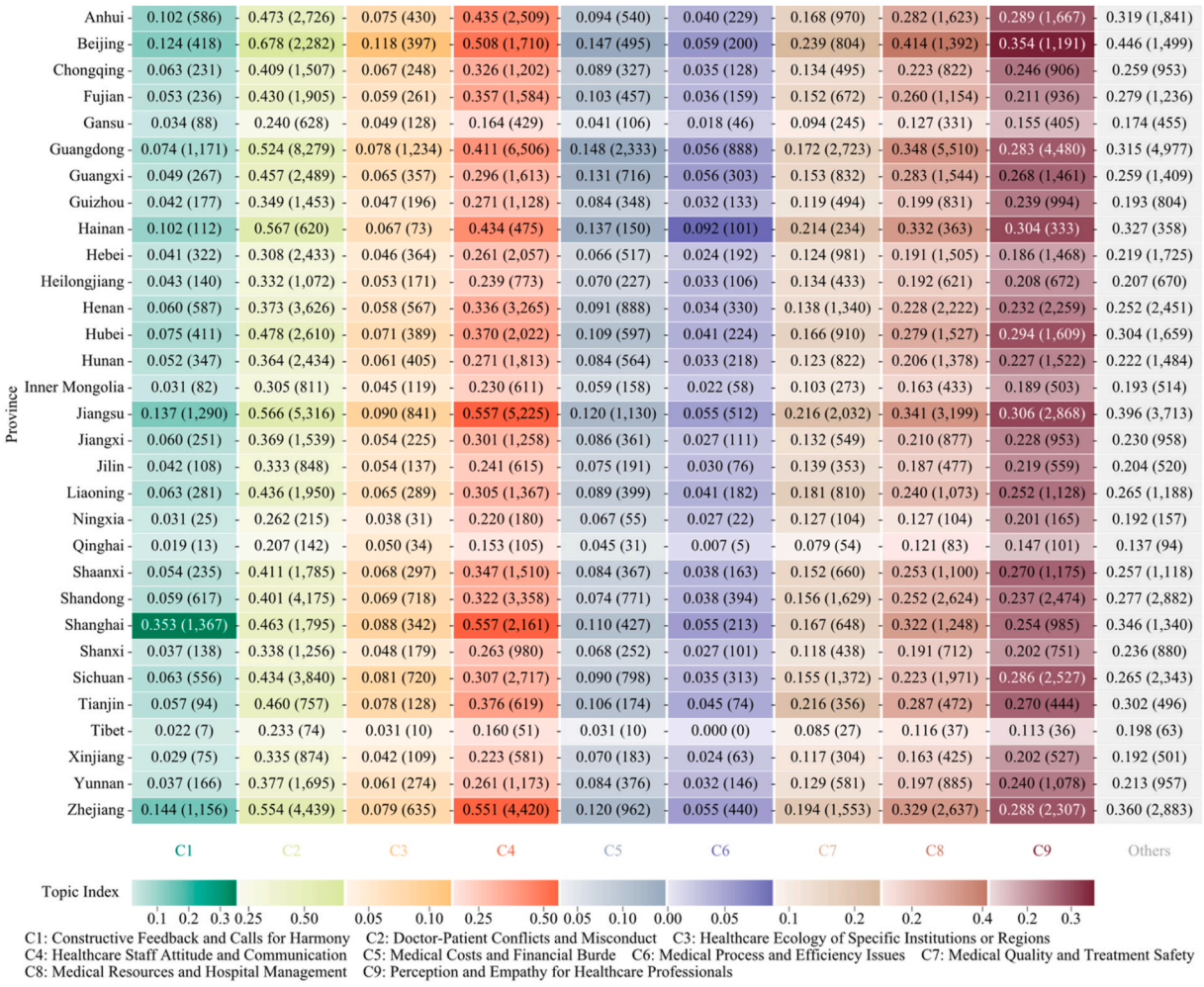


Fig. 13. Heatmap of provincial topic indices.

that in these specific high-pressure environments, the sheer density of hospitals is no longer a "stabilizer" but correlates with intensified grievances regarding medical costs. This counter-intuitive finding confirms H2b: in highly urbanized settings, the stabilizing effect of resource abundance is overwhelmed by countervailing structural forces.

In developed coastal regions, high hospital density often signifies a concentration of tertiary "super-hospitals" equipped with advanced but expensive technologies. Unlike the basic care facilities in the hinterland, these institutions handle high-stakes, capital-intensive medicine. Consequently, the "presence of hospitals" in these regions becomes a proxy for high medical expenditures, fueling the specific discourse on financial burden identified in our analysis. The positive SHAP values also suggest a saturation point. Once basic physical access is saturated (as in Beijing or Shanghai), the marginal utility of adding more beds diminishes, while the associated frictions—such as the "siphon effect" of attracting non-local patients and the higher living costs inherent to these clusters—begin to dominate the public narrative.

5.4. Policy implications: toward a spatially differentiated intervention framework

The empirical findings presented herein underscore the inherent limitations of universalist policy approaches in addressing healthcare system deficiencies. Rather than adopting homogeneous interventions, policymakers must calibrate strategies according to regional developmental trajectories and institutional configurations.

For Underdeveloped and Peripheral Regions: The evidence indicates that sustained capital investment in tangible healthcare infrastructure—encompassing medical facilities and nursing personnel—constitutes the most efficacious mechanism for ameliorating fundamental grievances pertaining to service affordability and accessibility. In these contexts, resource scarcity remains the predominant structural constraint, rendering supply-side expansion the primary policy imperative.

For Metropolitan and Economically Advanced Centers: In urbanized areas characterized by the aforementioned "Magnet Effect," continued expansion of physical infrastructure demonstrates diminishing marginal returns. Consequently, policy orientation must undergo a paradigmatic shift from the "hardware" to the "software" dimensions of healthcare delivery. This transition necessitates the institutionalization of hierarchical referral mechanisms to rationalize patient distribution across care tiers, alongside the systematic integration of medical humanities principles into clinical performance evaluation frameworks. Such measures are essential for mitigating the inherent tensions embedded within patient-provider dynamics that persist despite adequate resource provision.

5.5. Limitations and future research directions

It is important to acknowledge the limitations of this study, which position our findings as exploratory and macro-level in nature. First and foremost is the constraint of spatial aggregation. While our raw dataset is substantial (over 300,000 comments), aggregating these data to the provincial level results in a limited effective sample size (N = 31). This

constrained the degrees of freedom for complex spatial econometric modeling, necessitating the use of parsimonious variable selection (Top 3) rather than comprehensive multivariate controls. Consequently, our spatial findings reflect broad regional patterns rather than definitive causal mechanisms at the local scale. Future research is urgently needed to downscale this analysis to the city or county level. Finer granularity would not only increase statistical power but also allow for the application of local modeling techniques (e.g., Geographically Weighted Regression) to capture intra-provincial heterogeneity.

Second, platform and demographic biases exist. Relying on a single platform (Douyin) may skew the sample towards a younger demographic, although the substantial scale of our dataset helps mitigate the idiosyncratic influence of individual video tones on aggregate sentiment (Liang & Ye, 2025). Third, the "location-identity mismatch" remains a challenge. The IP address displayed at the time of commenting does not necessarily indicate the user's actual location of medical treatment. This limits our ability to fully disentangle the views of residents from those of medical tourists. Furthermore, due to the platform's algorithmic distribution mechanism, our analysis focuses on "visible" public discourse (videos with significant engagement). Sentiments expressed in low-engagement content (<100 comments) are inherently underrepresented, as they do not reach the threshold of collective social attention. Incorporating patient mobility flow data in future studies would allow for a direct test of the "magnet effect" hypothesis.

Finally, the limitations of our cross-sectional design point toward avenues for longitudinal and mixed-methods research. A temporal analysis tracking discourse over several years could reveal how public sentiment shifts in response to major health policy reforms. Furthermore, combining large-scale quantitative findings with in-depth qualitative methods, such as interviews with patients and doctors in identified "grievance hotspots", which would provide a deeper understanding of the lived experiences behind the digital data.

6. Conclusion

This study represents one of the first attempts to systematically map and explain the spatial heterogeneity of public grievance regarding the Chinese healthcare system in the digital era. By integrating large-scale social media data with provincial macro-statistics through a rigorous geospatial analytical framework, we move the conversation beyond anecdotal evidence to a structural understanding of the forces shaping online doctor-patient discourse.

Our findings reveal a dual reality characterizing China's healthcare landscape. On a national scale, the "supply-side" logic holds true: increasing the density of physical infrastructure, specifically hospitals, serves as an effective stabilizer that significantly mitigates baseline levels of public complaint regarding costs and efficiency. This empirically validates the achievements of recent healthcare reforms in improving physical access. However, a critical "Modernization Paradox" defines the country's most developed hubs. In these modernized regions, the buffering effect of resource abundance is increasingly offset by the "Modernization Penalty", where higher development levels (proxied by Life Expectancy) actively drive a shift towards critical systemic discourse. The tension observed in Beijing, Shanghai, and coastal provinces underscores that in high-pressure urban environments, the challenge has evolved from a crisis of scarcity to a crisis of rising expectations and structural complexity.

Methodologically, this research demonstrates the power of combining Large Language Models (LLMs) with Lasso-selected regression and SHAP-GIS visualization to decode the "black box" of digital sentiment. We show that online discourse is not merely a chaotic reflection of individual emotions but is deeply rooted in the physical realities of place—governed by the interplay between regional economic stages and resource allocation.

Ultimately, tackling the doctor-patient crisis requires a paradigm shift in governance. For policymakers, the message is clear: while building more hospitals remains necessary for under-resourced inland regions, it is a diminishing-return strategy for developed modernization centers. Future reforms must prioritize the "software" of healthcare optimizing patient flow, humanizing service interactions, and reforming institutional management, to bridge the widening gap between the advancing physical reality of Chinese medicine and the escalating standards of its people. In this context, social media serves not just as a channel for venting, but as a crucial, real-time diagnostic tool for aligning healthcare delivery with the evolving needs of a modernizing society.

CRediT authorship contribution statement

Pu Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yiliang Li:** Writing – original draft, Methodology, Formal analysis, Data curation. **Zheng Wei:** Validation, Formal analysis, Data curation. **Pan Hui:** Supervision, Project administration, Funding acquisition. **Na Jiang:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Conceptualization.

Ethical approval and informed consent statements

This article does not contain any studies with human or animal participants.

Data availability statement

The datasets generated and analyzed during the current study are not publicly available due to privacy concerns related to social media user data. The core code and the LLM prompts used for the analysis can be found online at https://doi.org/10.1016/j.apgeog.2026.103954.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2024YFC3307602) and the Guangdong Provincial Talent Program (2023JC10X009).

Appendix

A.1 Sensitivity Analysis: Robustness Check on Sampling Strategy

To address potential concerns regarding selection bias introduced by the primary sampling criterion (videos with >1000 comments), we conducted a sensitivity analysis. A supplementary dataset was collected from 30 ″medium-engagement" videos (100–999 comments), yielding 11,056 valid comments. These comments were classified using the same LLM-based methodology as the main dataset.

Table 4 compares the thematic distribution of this medium-engagement sample with the primary high-engagement dataset. The results demonstrate a high degree of consistency across key structural topics. While medium-engagement videos contained a higher proportion of general chatter (categorized as "Others") and daily operational complaints (e.g., "Medical Costs"), critical systemic themes such as "Perception and Empathy for Healthcare Professionals" and "Medical Resources and Hospital Management" remained top priorities in both datasets. This consistency confirms that the core patterns of public discourse identified in this study are robust and not merely artifacts of high-profile viral incidents.

Table 4
Sensitivity Analysis: Comparison of Topic Distribution between High-Engagement and Medium-Engagement Videos

Topic Categories	Main Dataset N = 306,323	Robustness Check N = 11,056	Observation
Perception and Empathy for Professionals	12.58%	8.99%	Consistent. Remains a top-tier topic in both datasets, confirming prevalent public empathy.
Medical Resources and Hospital Management	12.81%	10.19%	Consistent. Systemic issues remain a core focus regardless of video virality.
Medical Quality and Treatment Safety	7.74%	6.34%	Consistent. Stable proportion of discourse.
Medical Process and Efficiency Issues	2.00%	4.52%	Slightly higher in medium videos, likely reflecting daily operational complaints.
Medical Costs and Financial Burden	4.88%	6.50%	Higher in medium videos; daily financial burdens are a constant, non-viral grievance.
Doctor-Patient Conflicts and Misconduct	21.42%	12.64%	Difference noted. High-profile videos naturally attract more conflict-focused comments.
Healthcare Staff Attitude and Communication	17.65%	4.00%	Difference noted. Viral events often trigger specific critiques of attitude.
Others	13.77%	42.25%	Medium engagement videos contain significantly more "noise" and general chatter.

A.2 Distribution of 13 Standardized Candidate Variables

Raincloud plots (Fig. 10) visualize the probability density, box plots (interquartile range), and raw data points for each standardized variable. This visualization confirms the data quality and variability prior to variable selection.

A.3 Variance Inflation Factor (VIF) Diagnostics for the Final Model

To verify the stability of the final model specification derived from Lasso regression, we calculated the Variance Inflation Factor (VIF) for the three selected predictors: Life Expectancy, Hospital Density, and Registered Nurse Density. As illustrated in the figure, the VIF values for all variables are exceptionally low (ranging from approximately 1.5 to 2.5), significantly below the conventional cutoff of 5 or 10 (see Fig. 11). This confirms that the final parsimonious model is free from problematic multicollinearity, ensuring that the regression coefficients reflect the independent contribution of each factor rather than redundant variance.

A.4 Spatial Distribution and Moran's I of the Final Three Predictors

A.5 Topic Index Heatmap

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeog.2026.103954.

References

Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology, 68*(6), 627–636. https://doi.org/10.1016/j.jclinepi.2014.12.014

Chandra, S., & Foster, A. W. (2005). The "Revolution of Rising Expectations," relative deprivation, and the urban social disorders of the 1960s: Evidence from state-level data. *Social Science History, 29*(2), 299–332. https://doi.org/10.1017/S0145553200012967

Chen, X., Giles, J., Yao, Y., Yip, W., Meng, Q., Berkman, L., Chen, H., Chen, X., Feng, J., Feng, Z., Glinskaya, E., Gong, J., Hu, P., Kan, H., Lei, X., Liu, X., Steptoe, A., Wang, G., Wang, H., & Zhao, Y. (2022). The path to healthy ageing in China: A Peking University–Lancet Commission. *The Lancet, 400*(10367), 1967–2006. https://doi.org/10.1016/S0140-6736(22)01546-X

Chen, S., Guo, X., Wu, T., & Ju, X. (2021). Exploring the influence of doctor–patient social ties and knowledge ties on patient selection. *Internet Research, 32*(1), 219–240. https://doi.org/10.1108/INTR-07-2020-0403

Chen, L., & Mankad, S. (2024). A structural topic and sentiment-discourse model for text analysis. *Management Science, 71*(7), 5767–5787. https://doi.org/10.1287/mnsc.2022.00261

Dong, Z., & Xie, T. (2024). Why do people love short-form videos? The motivations for using Chinese TikTok (Douyin) and implications for well-being. *Current Psychology, 43*(26), 22283–22296. https://doi.org/10.1007/s12144-024-05927-4

Du, J., Wang, Y., Wu, A., Jiang, Y., Duan, Y., Geng, W., Wan, L., Li, J., Hu, J., Jiang, J., Shi, L., & Wei, J. (2023). The validity and IRT psychometric analysis of Chinese version of Difficult Doctor-Patient Relationship Questionnaire (DDPRQ-10). *BMC Psychiatry, 23*(1). https://doi.org/10.1186/s12888-023-05385-5

Fatemi, S., Hu, Y., & Mousavi, M. (2025). A comparative analysis of instruction fine-tuning large language models for financial text classification. *ACM Transactions on Management Information Systems, 16*(1), 1–30. https://doi.org/10.1145/3706119

Fu, L., Fang, Y.n., & Dong, Y. (2022). The healthcare inequality among middle-aged and older adults in China: A comparative analysis between the full samples and the homogeneous population. *Health Economics Review, 12*(1). https://doi.org/10.1186/s13561-022-00383-x

Fung, A., & Hu, Y. (2022). Douyin, storytelling, and national discourse. *International Communication of Chinese Culture, 9*(3), 139–147. https://doi.org/10.1007/s40636-022-00259-z

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*(4), 709–722. https://doi.org/10.3758/BF03192961

He, A. J., & Qian, J. (2016). Explaining medical disputes in Chinese public hospitals: The doctor–patient relationship and its implications for health policy reforms. *Health Economics, Policy and Law, 11*(4), 359–378. https://doi.org/10.1017/S1744133116000128

He, L., & Qin, Y. (2023). Developing a systematic-dynamic approach to critical discourse analysis: A study protocol in the context of the doctor–patient relationship in Western China. *International Journal of Qualitative Methods, 22*, Article 16094069221149510. https://doi.org/10.1177/16094069221149510

Hu, Y., Huang, J., Zhao, D., Zhang, C., Xia, J., & Lu, X.-m. (2024). Effect of safety and security equipment on patient and visitor violence towards nurses in multiple public hospitals of China during the COVID-19 pandemic: A retrospective, difference-in-

difference analysis. *BMJ Open, 14*(1), Article e078598. https://doi.org/10.1136/bmjopen-2023-078598

Ilyas, B., & Sharifi, A. (2025). A systematic review of social media-based sentiment analysis in disaster risk management. *International Journal of Disaster Risk Reduction, 123*, Article 105487. https://doi.org/10.1016/j.ijdrr.2025.105487

Jia, P., Wang, Y., Yang, M., Wang, L., Yang, X., Shi, X., Yang, L., Wen, J., Liu, Y., Yang, M., Xin, J., Zhang, F., Jiang, L., Chi, C., Zhang, L., Ma, X., Ma, X., Zhao, L., & Li, W. (2022). Inequalities of spatial primary healthcare accessibility in China. *Social Science & Medicine, 314*, Article 115458. https://doi.org/10.1016/j.socscimed.2022.115458

Kasperson, J. X., Kasperson, R. E., Pidgeon, N., & Slovic, P. (2003). The social amplification of risk: Assessing fifteen years of research and theory. In N. Pidgeon, R. E. Kasperson, & P. Slovic (Eds.), *The social amplification of risk* (pp. 13–46). Cambridge University Press. https://doi.org/10.1017/CBO9780511550461.002.

Lakin, K., & Kane, S. (2022). Peoples' expectations of healthcare: A conceptual review and proposed analytical framework. *Social Science & Medicine, 292*, Article 114636. https://doi.org/10.1016/j.socscimed.2021.114636

Liang, M., & Ye, L. (2025). Algorithmic pedagogy: How Douyin constructs algorithmic imaginaries for content creators. *Platforms & Society, 2*. https://doi.org/10.1177/29768624251365615

Lin, X., Yu, L., Wang, B., & Li, H. (2025). Decoding nested entities from classical Chinese with LLMs. *International Journal of Geographical Information Science*, 1–22. https://doi.org/10.1080/13658816.2025.2583431

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers, 105*(3), 512–530. https://doi.org/10.1080/00045608.2015.1018773

Liu, J., Zhou, H., Liu, L., & Wang, C. (2020). The weakness of the strong: Examining the squeaky-wheel effect of hospital violence in China. *Social Science & Medicine, 245*, Article 112717. https://doi.org/10.1016/j.socscimed.2019.112717

Luo, A., Qin, L., Yuan, Y., Yang, Z., Liu, F., Huang, P., & Xie, W. (2022). The effect of online health information seeking on physician-patient relationships: Systematic review. *Journal of Medical Internet Research, 24*(2), Article e23354. https://doi.org/10.2196/23354 [Review].

MacKinnon, N. J., Emery, V., Waller, J., Ange, B., Ambade, P., Gunja, M., & Watson, E. (2023). Mapping health disparities in 11 high-income nations. *JAMA Network Open, 6*(7), Article e2322310. https://doi.org/10.1001/jamanetworkopen.2023.22310

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/BM.2012.031

Meng, F., Zhang, X., Liu, L., & Ren, C. (2021). Converting readers to patients? From free to paid knowledge-sharing in online health communities. *Information Processing & Management, 58*(3), Article 102490. https://doi.org/10.1016/j.ipm.2021.102490

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity, 41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Pan, J., & Chen, C. (2022). Reducing universal health coverage regional disparities in China. *The Lancet Public Health, 7*(12), e985–e986. https://doi.org/10.1016/S2468-2667(22)00256-0

Pan, J., Liu, D., & Ali, S. (2015). Patient dissatisfaction in China: What matters. *Social Science & Medicine, 143*, 145–153. https://doi.org/10.1016/j.socscimed.2015.08.051

Pan, J., Zhao, H., Wang, X., & Shi, X. (2016). Assessing spatial access to public and private hospitals in Sichuan, China: The influence of the private sector on the healthcare geography in China. *Social Science & Medicine, 170*, 35–45. https://doi.org/10.1016/j.socscimed.2016.09.042

Parr, H. (2002). New Body-Geographies: The embodied spaces of health and medical information on the internet. *Environment and Planning D: Society and Space, 20*(1), 73–95. https://doi.org/10.1068/d41j

Peng, W., Lin, S., Chen, B., Bai, X., Wu, C., Zhang, X., Yang, Y., Cui, J., Xu, W., Song, L., Yang, H., He, W., Zhang, Y., Li, X., & Lu, J. (2024). Area-level socioeconomic inequalities in mortality in China: A nationwide cohort study based on the ChinaHEART project. *The Lancet Public Health, 9*(12), e1014–e1024. https://doi.org/10.1016/S2468-2667(24)00154-3

Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science, 17*(11). https://doi.org/10.1111/cts.70056

Potter, S. J., & McKinlay, J. B. (2005). From a relationship to encounter: An examination of longitudinal and lateral dimensions in the doctor–patient relationship. *Social Science & Medicine, 61*(2), 465–479. https://doi.org/10.1016/j.socscimed.2004.11.067

Rüttenauer, T. (2019). Spatial regression models: A systematic comparison of different model specifications using monte carlo experiments. *Sociological Methods & Research, 51*(2), 728–759. https://doi.org/10.1177/0049124119882467

Song, Y., Wang, T., Cai, P., Mondal, S. K., & Sahoo, J. P. (2023). A Comprehensive Survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys, 55*(13s). https://doi.org/10.1145/3582688. Article 271.

Sun, Y., Mu, J., Wang, D. W., Ouyang, N., Xing, L., Guo, X., Zhao, C., Ren, G., Ye, N., Zhou, Y., Wang, J., Li, Z., Sun, G., Yang, R., Chen, C.-S., He, J., Wang, C., Qiao, L., Shi, C., & He, H. (2022). A village doctor-led multifaceted intervention for blood pressure control in rural China: An open, cluster randomised trial. *The Lancet, 399* (10339), 1964–1975. https://doi.org/10.1016/S0140-6736(22)00325-7

Sun, J., Wang, J., Liu, S., Liu, Q., Wang, Z., Chen, Y., Zhao, P., Huang, Z., Wu, S., Huang, H., Ma, J., & Liu, Y. (2017). The impact of adverse media reporting on doctor–patient relationships in China: An analysis with propensity-score matching. *The Lancet, 390*, Article S100. https://doi.org/10.1016/S0140-6736(17)33238-5

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the benjamini-hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*(1), 77–83. https://doi.org/10.3102/10769986027001077

Uyanık, G. K., & Güler, N. (2013). A Study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences, 106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

Wang, Y., & Du, S. (2023). Time to rebuild the doctor-patient relationship in China. *Hepatobiliary Surgery and Nutrition, 12*(2), 235–238. https://doi.org/10.21037/hbsn-23-104

Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a Few Examples: A Survey on few-shot learning. *ACM Computing Surveys, 53*(3). https://doi.org/10.1145/3386252. Article 63.

Wang, W., Zhu, X., Lu, P., Zhao, Y., Chen, Y., & Zhang, S. (2024). Spatio-temporal evolution of public opinion on urban flooding: Case study of the 7.20 Henan extreme flood event. *International Journal of Disaster Risk Reduction, 100*, Article 104175. https://doi.org/10.1016/j.ijdrr.2023.104175

Wei, S., & Mao, Y. (2023). Small talk is a big deal: A discursive analysis of online off-topic doctor-patient interaction in Traditional Chinese Medicine. *Social Science & Medicine, 317*, Article 115632. https://doi.org/10.1016/j.socscimed.2022.115632

Whitehead, J., Pearson, A. L., Lawrenson, R., & Atatoa-Carr, P. (2019). How can the spatial equity of health services be defined and measured? A systematic review of spatial equity definitions and methods. *Journal of Health Services Research & Policy, 24*(4), 270–278. https://doi.org/10.1177/1355819619837292

Wirz, C. D., Xenos, M. A., Brossard, D., Scheufele, D., Chung, J. H., & Massarani, L. (2018). Rethinking social amplification of risk: Social media and zika in three languages. *Risk Analysis, 38*(12), 2599–2624. https://doi.org/10.1111/risa.13228

Wu, Y., Jiang, F., Ma, J., Tang, Y.-L., Wang, M., & Liu, Y. (2021a). Experience of Medical Disputes, Medical Disturbances, Verbal and Physical Violence, and Burnout Among Physicians in China [Original Research]. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.556517, 2020.

Wu, Q., Jin, Z., & Wang, P. (2021b). The relationship between the physician-patient relationship, physician empathy, and patient trust. *Journal of General Internal Medicine, 37*(6), 1388–1393. https://doi.org/10.1007/s11606-021-07008-9

Yan, L., Echeverria, V., Fernandez-Nieto, G. M., Jin, Y., Swiecki, Z., Zhao, L., Gašević, D., & Martinez-Maldonado, R. (2024). *Human-AI Collaboration in thematic analysis using ChatGPT: A user Study and design recommendations* extended abstracts of the 2024 CHI conference on. *Human Factors in Computing Systems*. https://doi.org/10.1145/3613905.3650732

Yan, X., He, S., Webster, C., & Yu, M. (2022). Divergent distributions of physicians and healthcare beds in China: Changing patterns, driving forces, and policy implications. *Applied Geography, 138*, Article 102626. https://doi.org/10.1016/j.apgeog.2021.102626

Zhang, R., & Chen, A. (2025). Public risk perception in extreme weather events: Topic distribution, spatiotemporal analysis, and sentiment comparison on social media. *International Journal of Disaster Risk Reduction, 127*, Article 105692. https://doi.org/10.1016/j.ijdrr.2025.105692

Zhang, P., Jiang, N., Li, Y., Wei, Z., & Hui, P. (2026a). Leveraging large language models to explain provincial heterogeneity in online discussion of water pollution in China: A socioeconomic and environmental perspective. *Journal of Cleaner Production, 543*, Article 147517. https://doi.org/10.1016/j.jclepro.2026.147517

Zhang, P., Li, Y., Wei, Z., & Hui, P. (2026b). The geography of climate concern: A large-scale analysis of public discourse on extreme heat in China using social media and explainable AI. *Environmental Impact Assessment Review, 117*, Article 108227. https://doi.org/10.1016/j.eiar.2025.108227

Zhang, X., & Sleeboom-Faulkner, M. (2011). Tensions between medical professionals and patients in mainland China. *Cambridge Quarterly of Healthcare Ethics, 20*(3), 458–465. https://doi.org/10.1017/S0963180111000144

Zhang, P., Wei, Z., & Kong, F. (2026c). Reconfiguring responsibility: An empirical analysis of crisis discourse and situational crisis communication on Douyin. *International Journal of Disaster Risk Science*. https://doi.org/10.1007/s13753-026-00693-2

Zhang, P., Wei, Z., Zhou, M., Evans, J., & Hui, P. (2026d). Beyond aggregate feedback: A spatially-aware computational framework for understanding citizen-state interaction on social media. *ACM Transactions on Social Computing*. https://doi.org/10.1145/3790251

Zhang, P., Zhang, H., & Kong, F. (2024). Research on online public opinion in the investigation of the "7–20" extraordinary rainstorm and flooding disaster in Zhengzhou, China. *International Journal of Disaster Risk Reduction, 105*, Article 104422. https://doi.org/10.1016/j.ijdrr.2024.104422

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large Language models transform computational social science? *Computational Linguistics, 50*(1), 237–291. https://doi.org/10.1162/coli_a_00502