Adoption of Generative Large Language Models in Pathology: A National Survey of Chinese Pathologists

Peng Xue, Yuting Wang, Victor Yu Cui, Zichen Ye, Qinjing Pan, Xun Zhang, Yao Yang, Youlin Qiao

Submitted to: Journal of Medical Internet Research
on: October 19, 2024

Table of Contents

Adoption of Generative Large Language Models in Pathology: A National Survey of Chinese Pathologists

Peng Xue[1]; Yuting Wang[2]; Victor Yu Cui[3]; Zichen Ye[2]; Qinjing Pan[4]; Xun Zhang[4]; Yao Yang[2]; Youlin Qiao[2]

[1]Chinese Academy of Medical Sciences and Peking Union Medical College Beijing CN
[2]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College Beijing CN
[3]Vanke School of Public Health, Tsinghua University Beijing CN
[4]Department of Cancer Epidemiology, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College Beijing CN

Corresponding Author:
Peng Xue
Chinese Academy of Medical Sciences and Peking Union Medical College
No. 9, Dongdan Sanjie, Dongcheng District
Beijing
CN

Abstract

Background: Pathologists are grappling with high workloads and uneven resource distribution, which can impede professional development and the delivery of quality patient care. The advent of generative large language models (LLMs) has the potential to revolutionize pathological field, where efficiency and resource accessibility are paramount.

Objective: This study aimed to investigate the perceptions and willingness of Chinese pathologists to adopt generative LLMs.

Methods: We conducted a questionnaire survey at the National Pathology Academic Annual Conference in April 2024, involving 339 certified Chinese pathologists. Participant responses were measured with a 5-point Likert scale for the performance of generative LLMs in clinical, research, and educational settings, with statistical analysis using mean and standard deviation (SD). Multivariable logistic regression models were employed to explore factors associated with the adoption of generative LLMs, reporting odds ratios (ORs) and 95% confidence intervals (CIs).

Results: A total of 339 valid questionnaires were returned. The results revealed that pathologists generally supported the performance of generative LLMs in clinical (mean 3.87, SD 0.96), research (mean 3.88, SD 1.09), and educational (mean 4.04, SD 0.82) contexts. Positive attitudes towards the use of generative LLMs were prevalent. Notably, pathologists practicing in less developed urban areas (OR=1.99, 95% CI=1.07 to 3.69, p=0.030), those with higher caseloads (>5000 cases/year; OR=2.12, 95% CI=1.01 to 4.44, p=0.047), and those engaged in research (OR=2.94, 95% CI=1.61 to 5.34, p<0.001) and teaching (OR=2.37, 95% CI=1.42 to 3.96, p=0.001) activities, as well as those with prior experience with generative LLMs (OR=2.45, 95% CI=1.38 to 4.37, p=0.002), showed a greater inclination towards future adoption.

Conclusions: Chinese pathologists are receptive to generative LLMs, showing a positive inclination for their application. The study advocates for fostering the adoption of generative LLMs to improve the efficiency and accuracy of diagnosis, reduce the burden on pathologists, and improve the overall service level in the field of pathology.

(JMIR Preprints 19/10/2024:67720)
DOI: https://doi.org/10.2196/preprints.67720

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Title: Adoption of Generative Large Language Models in Pathology: A National Survey of Chinese Pathologists

Running title: Adoption of AI in Pathology

Authors: Yuting Wang[1†], Victor Yu Cui[2†], Zichen Ye[1], Qinjing Pan[3], Xun Zhang[3], Yao Yang[1*], Peng Xue[1*], Youlin Qiao[1]

†Authors contributed equally to the work

Affiliations:[1]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Vanke School of Public Health, Tsinghua University, Beijing, China, [3]Department of Cancer Epidemiology, National Cancer Center, National Clinical Research Center for Cancer, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

***Correspondence to:**

Peng Xue, PhD

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, 100730

Tel +86-1087788489 Fax +86-1067713648

E-mail: xuepeng_pumc@foxmail.com

Yao Yang, PhD

Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, 100730

Tel +86-1087788489 Fax +86-1067713648

E-mail: yangyao@pumc.edu.cn

Abstract

Background: Pathologists are grappling with high workloads and uneven resource distribution, which can impede professional development and the delivery of quality patient care. The advent of generative large language models (LLMs) has the potential to revolutionize pathological field, where efficiency and resource accessibility are paramount.

Objective: This study aimed to investigate the perceptions and willingness of Chinese pathologists to adopt generative LLMs.

Methods: We conducted a questionnaire survey at the National Pathology Academic Annual Conference in April 2024, involving 339 certified Chinese pathologists. Participant responses were measured with a 5-point Likert scale for the performance of generative LLMs in clinical, research, and educational settings, with statistical analysis using mean and standard deviation (SD). Multivariable logistic regression models were employed to explore factors associated with the adoption of generative LLMs, reporting odds ratios (ORs) and 95% confidence intervals (CIs).

Results: A total of 339 valid questionnaires were returned. The results revealed that pathologists generally supported the performance of generative LLMs in clinical (mean 3.87, SD 0.96), research (mean 3.88, SD 1.09), and educational (mean 4.04, SD 0.82) contexts. Positive attitudes towards the use of generative LLMs were prevalent. Notably, pathologists practicing in less developed urban areas (OR=1.99, 95% CI=1.07 to 3.69, p=0.030), those with higher caseloads (>5000 cases/year; OR=2.12, 95% CI=1.01 to 4.44, p=0.047), and those engaged in research (OR=2.94, 95% CI=1.61 to 5.34, p<0.001) and teaching (OR=2.37, 95% CI=1.42 to 3.96, p=0.001) activities, as well as those with prior experience with generative LLMs (OR=2.45, 95% CI=1.38 to 4.37, p=0.002), showed a greater inclination towards future adoption.

Conclusions: Chinese pathologists are receptive to generative LLMs, showing a positive inclination for their application. The study advocates for fostering the adoption of generative LLMs to improve the efficiency and accuracy of diagnosis, reduce the burden on pathologists, and improve the overall service level in the field of pathology.

Keywords: Pathology, Generative Large Language Models, ChatGPT

Introduction:

Pathological diagnosis is essential for accurate disease identification, particularly in cancer, where it dictates personalized treatment plans and patient outcomes. As the aging of the population intensifies and the national awareness of cancer prevention increases, more and more Chinese residents are proactively participating in physical examinations and national screening programs[1]. This not only improves the chances of early detection and treatment of cancer but also leads to a continuous increase in the demand for pathological diagnosis.

Thus, like many other countries, China has certain challenges that predispose pathologists to have a higher-than-expected cancer impact on the population. One of the most significant challenges is the shortage of pathology professionals. According to the 2022 China Health and Health Statistics Yearbook, there are only about 18,700 registered pathologists and assistant physicians in China, far less than the actual demand of over 100,000[2]. This significant talent gap results in an uneven geographical distribution of pathological resources, especially in remote and rural areas, where access to high quality pathological services is even lower. To be specific, the number of pathology departments in Gansu, Qinghai, Ningxia, Inner Mongolia, Tibet and Xinjiang accounted for only 11.7% of the total number of pathology departments in China. Additionally, the China Pathology Department Survey Report indicates that the average annual workload for a Chinese pathologist is approximately 4910 cases[3], a figure that starkly contrasts with the 132 cases handled by American pathologists and the 117 cases by Canadian counterparts[4]. This workload disparity, coupled with a shortage of over 80,000 pathologists and uneven distribution of resources, reveals a critical need for innovative solutions to enhance the overall quality and response speed of pathological services.

In recent years, with artificial intelligence (AI) technologies achieving remarkable breakthroughs in medicine[5-7], digital pathology has made significant strides, enabling precise detection, segmentation, diagnosis, and analysis of digital images, significantly enhancing the efficiency and accuracy of pathological diagnosis[8-10]. Generative large language models (LLMs), a form of AI trained on vast text and image datasets, have demonstrated capabilities in various domains, including question answering and translation, and are now preparing to bring new revolution to the pathology field[11,12]. The development of AI tools like PathChat by Modella AI, in collaboration with Harvard Medical School, exemplifies the potential of these models to provide accurate and rapid diagnostic assistance, promising a significant leap in efficiency and resource accessibility across clinical practice, scientific research, and medical education[13].

This study aims to assess the perceptions and willingness of Chinese pathologists towards adopting generative LLMs, such as ChatGPT, and to identify key factors influencing their acceptance. The insights gathered are intended to provide reference for the health sector to drive the innovative application of generative LLMs in the pathology.

Materials and Methods

Ethics Approval

This study was approved by the Institutional Review Board of the Chinese Academy of Medical Sciences and Peking Union Medical College (No. CAMS and PUMC-IEC-2022-022). Informed consent of the participants was obtained before the study began.

Study design and Participants

A cross-sectional survey was conducted using the Wenjuanxing platform, a professional online tool for designing and distributing questionnaires, to assess perception and attitudes on clinical practice, scientific research, and medical education of generative LLMs usage for pathology. The survey was distributed at the National Pathology Academic Annual Conference held in April 2024 in Xi'an, China, targeting certified pathologists practicing in China Mainland. Eligible participants only included practicing clinical pathologists, excluding medical students and industry representatives. The survey process was designed to be independent, with participants completing the questionnaires without external influence.

Questionnaire

The questionnaire comprises five sections: demographic information, perceptions, assessment, ethical considerations and prospects. In demographic surveys, all participants were required to specify their work location to their city, not just their province. The evaluation and ethical considerations were conducted using a Likert 5-point scale ranging from strongly disagree to strongly agree, with a total of 21 statements (Supplementary file for detailed). Study participants were asked to rate their agreement with each statement. For further interpretation, ratings between 4 and 5 were defined as agreement, and ratings between 1 and 2 were defined as disagreement. A pilot survey was conducted to modify and improve the quality of questionnaire. leading to the selection of the twelve most reliable evaluation items and nine ethical items for the final survey. We obtained the informed consent of the participants before formally began the investigation.

Statistical Analysis

Descriptive statistical analysis was used to calculate the percentage for categorical variable and means with SDs for continuous variables among all participants. The distribution of responses regarding performance assessment of generative LLMs was computed, and mean scores with SD and the proportions of agreed were reported. The Chi-square test was applied to analyze the differences in cognition of generative LLMs among various populations. Multivariable logistic regression models were conducted to identify associated factors of generative LLMs willingness, with reporting Odds ratios (ORs) and 95% confidence intervals (CIs). Statistical analysis is conducted using the SPSS 27.0 software. All analyses in this study use two-tailed tests, with the significance level α set at 0.05.

Results

Characteristics of Study Participants

A total of 376 participants were involved in the study. The 37 individuals including 11 medical students and 26 staff of companies related to pathology were excluded from statistical analyses. The remaining participants comprised 236 female (69.6%) and 103 male (30.4%), predominantly within the 30–40 years age group, representing 48.4% (n=164) of the study population. Among participants, 63.1% (n=214) held a bachelor's degree, 46.6% (n=158) were attending physicians, 32.7% (n=111) have experienced more than 15 years in clinic, 25.7% (n=87) also occupied in research work and 34.5% (n=117) undertook in teaching work. Participants worked 43.01 hours weekly on average, with 34.2% (n=116) of them read more than 5000 slides annually. The geographic distribution of the participants was broad, with the survey reaching across 29 provinces and autonomous regions in China, as shown in Figure 1. We categorized participants' work locations based on a city-level list published by China Business Network, a financial media outlet. Notably, the 69.6% (n=236) come from third-tier cities and above, and only 30.5% (n=103) are fourth-tier cities and remote areas.Notably, 36% (n=122) of the participants were based in first-tier cities such as Beijing and Xi'an, while 33.6% (n=114) were from second and third-tier cities. The detailed demographic information of participants is shown in Table 1.

Figure 1. Map of participants' work locations

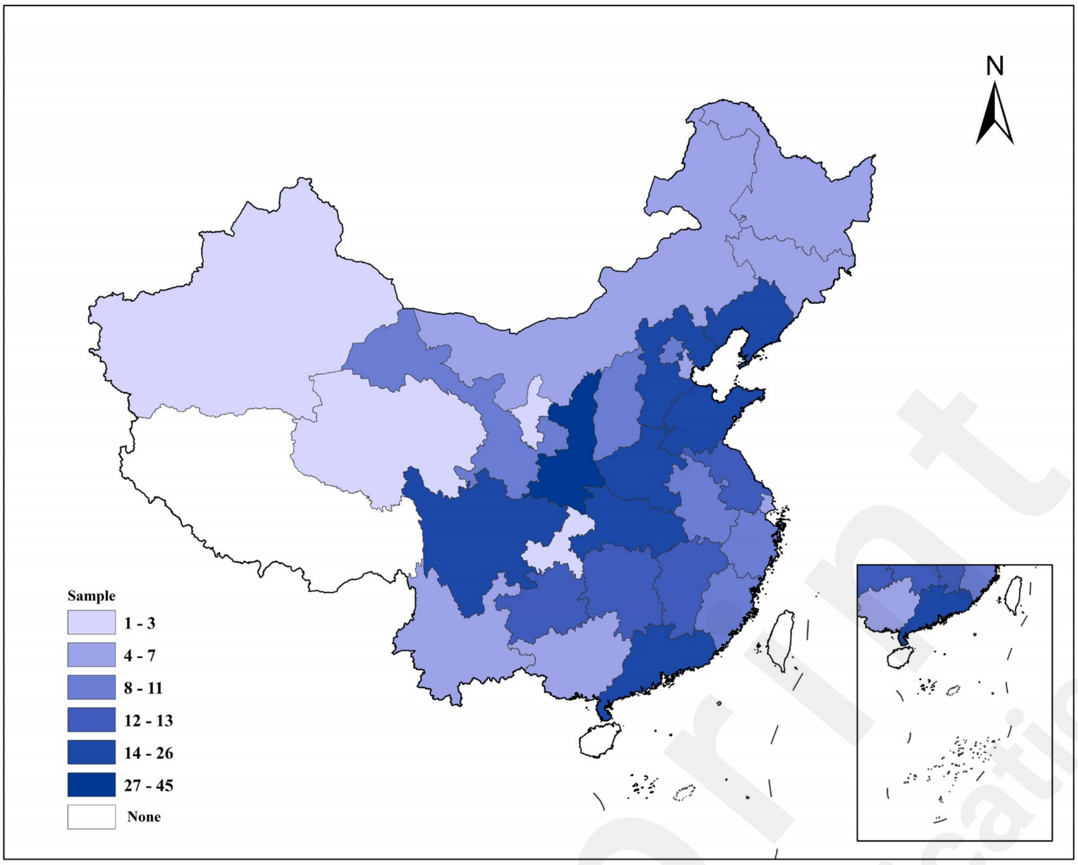


Table 1 Demographic information of participants (n=339)

Variables		N(%)	N_{heard}(%)*	*p*-value**
Gender				**0.021**
	Male	103(30.4)	84(81.6)	
	Female	236(69.6)	164(64.5)	
Age/ year				0.594
	□30	26(7.7)	21(80.8)	
	30-40	164(48.4)	115(70.1)	
	41-50	107(31.6)	81(75.7)	
	□50	42(12.3)	31(73.8)	
City level				**0.010**
	First tier	122(36.0)	101(82.8)	
	Second and third tier	114(33.6)	76(66.7)	
	Fourth tier and below	103(30.4)	71(68.9)	
Education level				0.593
	College degree	13(3.8)	11(84.6)	
	Bachelor degree	214(63.1)	157(73.4)	
	Master degree and above	112(33.1)	80(71.4)	
Title				**0.020**
	Medical intern	70(20.6)	41(58.6)	
	Attending doctor	158(46.6)	123(77.8)	
	Associate chief physician	79(23.3)	59(74.7)	

	Chief physician	32(9.5)	25(78.1)	
Experience working/ year				0.714
	≤5	60(17.7)	42(70.0)	
	6-10	91(26.8)	64(70.3)	
	11-15	77(22.7)	57(74.0)	
	□15	111(32.8)	85(76.6)	
Weekly working hours		43.01 (9.66)		0.483
	≤43	217(64.0)	156(71.9)	
	>43	122(36.0)	92(75.4)	
Slide reading cases / year				**0.004**
	≤1000	67(19.8)	46(68.7)	
	1001-3000	80(23.6)	48(60.0)	
	3001-5000	76(22.4)	58(76.3)	
	>5000	116(34.2)	96(82.8)	
Research work				0.458
	Yes	87(25.7)	61(70.1)	
	No	252(74.3)	187(74.2)	
Teaching work				**0.007**
	Yes	117(34.5)	96(82.1)	
	No	222(65.5)	152(68.5)	

*This section only analyzes participants who have heard the generative LLMs.

**The Chi-square test was used to analyze demographic differences among heard participants.

Perceptions and Assessment of Generative LLMs

Table 2 shows the participants' perceptions of the generative LLMs. A total of 248 (73.2%) participants demonstrated an understanding of generative LLMs. Notably, male participants (*p*=0.021), residing in first-tier cities (*p*=0.010), holding the title of chief physician (*p*=0.020), with an average annual slide reading caseload exceeding 5,000 (*p*=0.004), and those engaged in teaching activities (*p*=0.007), were more likely to be aware of these models (Table 1). ChatGPT by OpenAI was the most recognized model, mentioned by 74.6% of the participants, followed by the domestic model ERNIE Bot by Baidu (35.9%) and various mirror websites (29.8%). The primary source of information about generative LLMs was social media like Weibo and WeChat public accounts, mentioned by 71% (n=176) of the participants. Among the participants, 64 (25.8%) reported having utilized generative LLMs, with 11 having subscribed to paid services. Of the subscribers, nine individuals noted that the paid version met their initial expectations and requirements. ChatGPT was also the most used model, with 53.1% of users selecting it, again followed by ERNIE Bot at 23.4%. When inquiring about the reasons for non-use, the majority of non-users (45.8%) stated they were unaware of how to access these models, while 34.9% indicated they had no need for them. Table 3 presents the descriptive results of pathologists' assessments of the clinical (mean 3.87, SD 0.96), research (mean 3.88, SD 1.09), and educational (mean 4.04, SD 0.82) performance of generative LLMs using a Likert 5-point scale. Factors influencing the willingness to use generative LLMs included working in less developed cities (fourth-tier and below, OR=1.99, 95% CI=1.07 to 3.69, *p*=0.030), annual reading more than 5,000 slide cases (OR=2.12, 95% CI=1.01 to 4.44, *p*=0.047), being engaged in both research (OR=2.94, 95% CI=1.61 to 5.34, *p*<0.001) or teaching activities (OR=2.37, 95% CI=1.42 to 3.96, *p*=0.001), and having prior experience with generative LLMs at work (OR=2.45, 95% CI=1.38 to 4.37, *p*=0.002). Willingness to use generative LLMs was not significantly associated with gender, age, educational level, professional title, years in practice, weekly working hours, or attitudes toward LLMs potentially replacing pathologists (Table 4).

Table 2 Perception of generative LLMs

Variables		N(%)
Have you heard about generative LLMs before today? (n=339)		
	Yes	248(73.2)
	No	91(26.8)
Know the way (n=248)		
	Friends or colleagues	31(12.5)
	Social media	176(71.0)
	Academic paper	15(6.0)
	Academic conference	26(10.5)
Have you used about generative LLMs before today? (n=248)		
	Yes	64(25.8)
	No	184(74.2)
Reason for not using (n=184)		
	I have no need to use it	56(30.4)
	I do not have an account that I can access.	18(9.8)
	I do not know where to access	90(48.9)
	I'm concerned that the answers provided may not be accurate.	11(6.0)
	I am concerned about privacy or data breaches.	8(4.3)
	Other	3(1.6)
Language (n=64)		
	English	8(12.5)
	Chinese	56(87.5)
Whether you have subscribed to paid services? (n=64)		

Yes	11(17.2)
No	53(82.8)

Table 3 Performance evaluation of generative LLMs (n=248)

Statements	**Mean±SD**	**Agree, *N* (%)**
Clinical practice(n=248)		
It can help me analyze patient data quickly and correctly, and improve the accuracy and efficiency of diagnosis.	3.83±0.95	167 (67.3%)
It can help me organize and analyze medical history more efficiently and optimize the medical record process.	3.89±0.97	182 (73.4%)
It can help me better understand the mechanism of disease, diagnosis, treatment and other questions.	3.88±0.96	177 (71.4%)
Total-clinic practice	3.87±0.96	
Scientific research (n=75)		
It can help me systematically sort out the experimental steps and methods to ensure that the experimental process is clear and clear.	3.92±1.08	55 (73.3%)
It can help me quickly complete the data visualization work such as scientific research drawing.	3.91±1.09	54 (72.0%)
It can help me write code efficiently.	3.83±1.12	52 (69.3%)
Total-scientific research	3.88±1.09	
Medical education (n=96)		
It can change the way we teach in the future.	4.03±0.83	75 (78.1%)
It can clearly explain complex concepts and help students build a systematic and complete knowledge system.	4.03±0.83	77 (80.2%)
It can improve teaching efficiency, and students can quickly acquire accurate knowledge through interaction with natural language.	4.05±0.83	76 (79.2%)
Total-medical education	4.04±0.82	

LLMs performance on clinic, research and teaching performance were measured by items, respectively, with a 5-point Likert scale, and response options higher than “neutral” would be classified into “agree”.

Table 4 Predictors for willingness of generative LLMs in a multivariable logistic regression model (n= 248)

Variable	Willingness		
	OR	95%CI	*p*-value
Gender (ref. Male)			
Female	1.17	0.71-1.92	.545
Age (ref. <30)			
30-40	1.15	0.42-3.20	.786
41-50	0.84	0.27-2.64	.764
≥51	1.61	0.43-6.11	.483
Educational level (ref. College degree)			
Bachelor degree	1.16	0.37-3.62	.802
Master degree and above	1.53	0.47-5.02	.482
Urban level (ref. First tier)			
Second and third tier	1.14	0.63-2.04	.670
Fourth tier and below	**1.99**	**1.07-3.69**	**.030**
Title (ref. Medical intern)			
Attending doctor	0.77	0.36-1.64	.499
Associate chief physician	0.51	0.20-1.29	.154
Chief physician	0.82	0.27-2.50	.730
Experience working (ref. ≤5)			
6-10	1.23	0.51-3.00	.643
11-15	1.00	0.39-2.53	.999
□15	1.51	0.54-4.21	.435
Weekly working hours (ref. ≤ 43)	1.18	0.72-1.93	.514
Slide reading cases/ year (ref. ≤1000□			
1001-3000	1.18	0.53-2.59	.686
3001-5000	1.18	0.54-2.56	.682
>5000	**2.12**	**1.01-4.44**	**.047**
Attitude to LLMs replacing pathologist (ref. No)			
Natural	1.28	0.68-2.44	.447
Yes	0.86	0.38-1.84	.722
Research status (ref. No)	**2.94**	**1.61-5.34**	**.000**
Teaching status (ref. No)	**2.37**	**1.42-3.96**	**.001**
Using experience (ref. No)	**2.45**	**1.38-4.37**	**.002**

Ethical Considerations and Future Suggestions

Nearly 80% of participants acknowledged potential ethical issues with generative LLMs in pathology. The most pressing concerns were the exacerbation of health inequalities due to limited model availability in underprivileged areas (S8, 34.34%), the risk of plagiarism and intellectual property infringement in academic writing (S4, 33.84%), and data leaks involving sensitive information (S6, 33.33%). Detailed ethical issues see Figure 2a. For future suggestions, participants prioritized the integration of generative LLMs with AI cytopathology to improve diagnostic accuracy (I2, 84.68%), the improvement of training datasets for better recognition of pathological features (I6, 71.37%), and regular updates to the knowledge base for more precise diagnostic support (I4, 70.79%). Figure 2b displays the desirabilities of the six provided directions for future improvements.

Figure 2a. Percentage bar-chart of assessment for the severity of ethical consideration

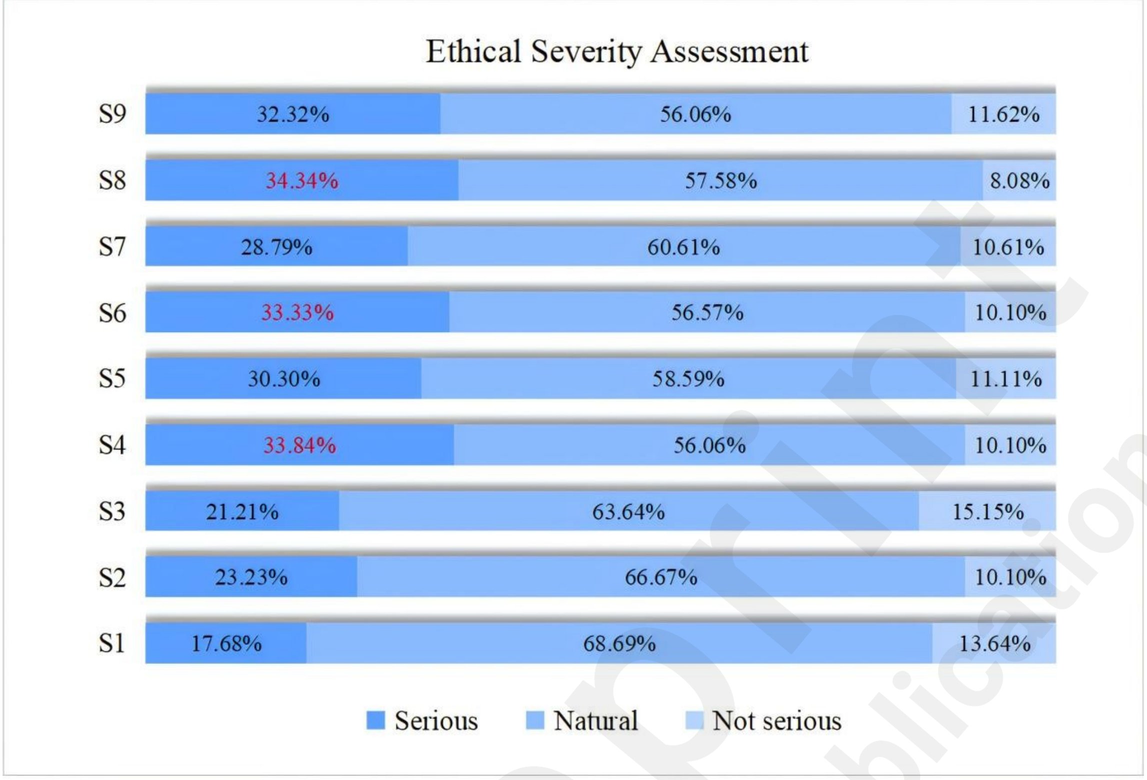
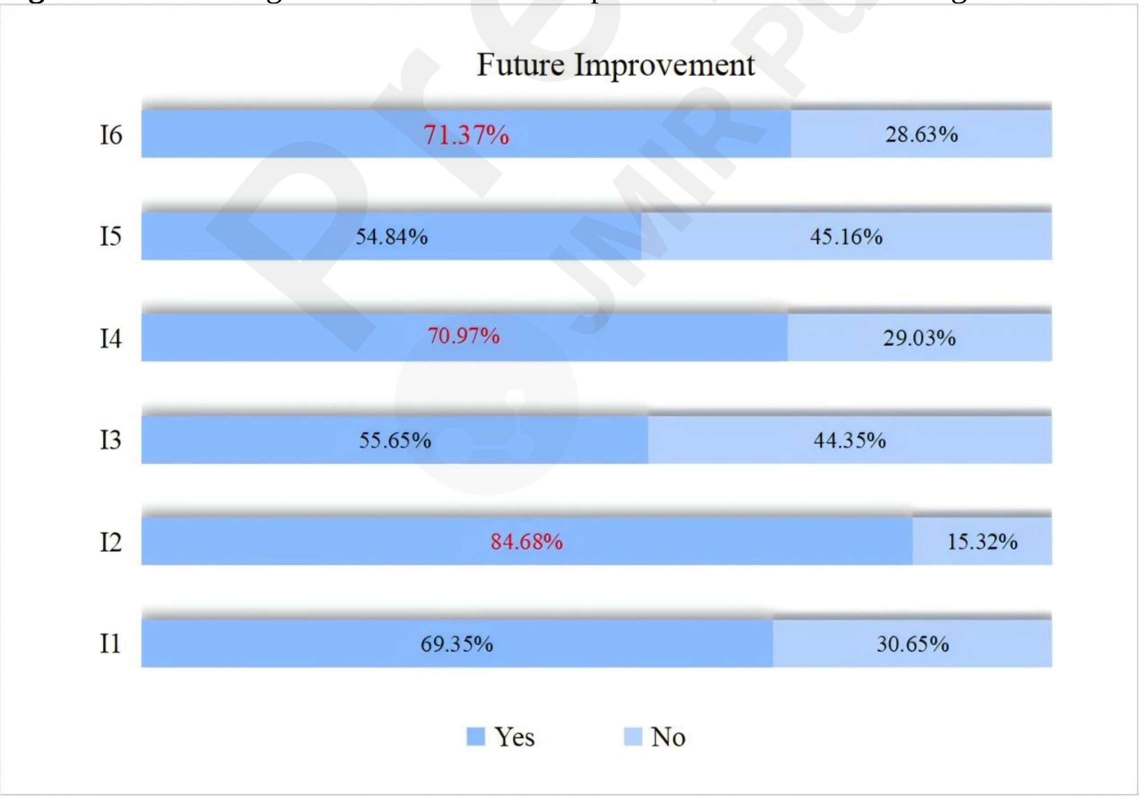


Figure 2b. Percentage bar-chart of future improvement direction of the generative LLMs



Discussion

To our knowledge, this is the first systematic survey conducted on the perception and willingness to use generative LLMs among Chinese pathologists. It helps them better understand and assess the potential application of generative LLMs in the medical field and their impact on the workflow of physicians.

The findings indicated majority of Chinese pathologists (n=248, 73.2%) have heard of generative LLMs and generally hold a positive assessment for those models in clinical practice, scientific research, and teaching. However, widespread adoption is not prevalent, with only 25.8% (n=64) utilizing the models, with not knowing where to access the model being the leading constraint. Moreover, the higher acceptance of generative LLMs among pathologists in China's less developed areas may reflect the critical need for enhanced diagnostic capabilities. These regions, particularly fourth-tier cities and below, face challenges in attracting and retaining pathology professionals due to lower salaries and limited career advancement opportunities. The concentration of advanced degrees in higher-tier cities exacerbates this disparity, with all doctoral degree holders and the majority of specialized degree holders from first-tier cities. This imbalance results in an uneven distribution of pathology resources, potentially compromising healthcare quality and accessibility. The integration of LLMs could improve diagnostic services in these regions, offering a solution to the limitations posed by resource constraints.

Pathologists reading over 5,000 slides annually are more open to adopting generative LLMs, likely due to the potential for these technologies to alleviate the workload and eye strain associated with extensive screen time during digital image analysis. LLMs designed for pathology can perform preliminary automated analyses of slide images, identifying common pathological features, thus allowing pathologists to focus on complex cases requiring expert judgment. Moreover, LLMs can streamline the process of issuing pathology reports, which are pivotal for clinical decision-making. By extracting and structuring information from unformatted text in electronic medical records, LLMs can assist in generating comprehensive reports that include diagnostic insights and molecular markers. Truhn et al.[14] have demonstrated that LLMs, such as GPT-4, can accurately parse unstructured pathology reports and convert them into structured data, integrating medical reasoning to enhance clinical interpretation. This capability not only reduces the writing burden on pathologists but also supports the formulation of optimal patient care strategies.

Some participants engaged in research or teaching activities also showed a heightened interest in generative LLMs, recognizing their utility in these domains. In research, these models excel at automating tasks such as classifying and analyzing extensive textual data, thereby accelerating the academic process. They facilitate rapid compilation of research reports by extracting and summarizing information from diverse sources[15]. In education, LLMs offer customized resources that cater to the complex medical curriculum, enhancing students' learning efficiency through vivid teaching methods and focused content[16]. They also serve as interactive study companions, capable of creating mock exams and providing immediate feedback, enriching the learning experience[17,18]. For physicians, particularly those with limited access to quality training materials, generative LLMs are a vital source for medical knowledge acquisition. By applicating these models, pathologists can access a wider array of information and reduce training expenses, ultimately improving their professional competencies and the level of pathological diagnosis in underdeveloped regions.

However, the adoption of a generative LLMs does present ethical challenges. Participants expressed concerns that the integration of generative LLMs could widen health inequalities, which refers to the higher systemic health risks faced by certain populations compared to others due to social factors such as poverty or race. These risk disparities are traced back to Social Determinants of Health (SDoH) that influence health outcomes throughout an individual's life[19]. SDoH are pivotal for patient prognosis, enabling pathologists to tailor diagnostic and treatment plans by understanding underlying health factors and connecting patients to necessary social services. However, capturing SDoH in electronic health records (EHRs) remains a challenge due to incomplete structured data[20].

Promisingly, research at Harvard Medical School has demonstrated that finely-tuned models like

ChatGPT can effectively identify SDoH indicators in medical records, outperforming traditional ICD coding[19]. This capability suggests that with proper training, generative LLMs could be instrumental in addressing health inequities. But it has to be mentioned, the high maintenance costs of these models pose a threat to their widespread application, particularly in low- and middle-income countries[21]. Future efforts should focus on reducing these costs to ensure equitable access to generative LLMs in healthcare, thereby enhancing medical standards and patient outcomes.

Our survey identified key areas for future enhancement of generative LLMs in pathology. The foremost desire was the integration of generative LLMs with automated diagnostic tools to boost the efficiency and accuracy (84.7%). Such integrations have made progress in the field of dentistry, which could aid in the identification of heteromorphic cells or tissues, thereby facilitating image-based diagnoses[22]. The second priority was improving the quality and diversity of training data to enhance the models' ability to recognize pathological features (71.4%). Studies, such as one by Haemmerli et al.[23] on diagnosing central nervous system tumors, have shown that while LLMs can propose treatment plans, they sometimes struggle with precise tumor classifications. This suggests the need for more multi-center datasets to be trained or validated to improve the generalization ability of models, or even to train more models specific to the diagnosis of pathology like PathChat[24]. Lastly, there was a strong call for continuous learning and knowledge base updates to ensure pathologists receive up-to-date diagnostic support (71.0%). Addressing the models' limitations in accessing the latest research and data is crucial to avoid providing outdated information[25].

Strengths and Limitations

This study presents several strengths. The study's approach to categorizing regional economic levels by city tiers, rather than traditional geographical divisions, offers a nuanced understanding of regional disparities. This method considers political status, economic strength, development level, and urban influence, providing a scientific foundation for more effective regional development strategies and medical resource allocation. Moreover, our study covered 29 provinces, municipalities or autonomous regions across the country, ensuring a national representation of the study object. However, the study did not account for subjective perceptions of eye fatigue, which could influence the willingness to adopt new technologies. Additionally, the cross-sectional design limits our ability to track actual usage patterns over time.

Conclusion

The survey reveals a predominantly open and positive stance towards generative LLMs among Chinese pathologists, with minimal concern about these technologies replacing their professional expertise. While actual usage remains limited, the attitudes indicate a readiness to embrace LLMs. A key takeaway is the importance of fostering the adoption of generative LLMs to alleviate the workload of pathologists and to break the barrier of the imbalance of pathological resources. This has critical implications for advancing the quality and equity of medical services across regions.

Acknowledgments

This work was supported by the CAMS Innovation Fund for Medical Sciences [CIFMS 2021-I2M-1-004]; China Postdoctoral Science Foundation [2024T170072]; and Postdoctoral Fellowship Program of CPSF [GZB20230076].

Authors' Contributions

YTW, VYC and PX performed study concept and design; YTW, VYC, YY and PX performed questionnaire collection; YTW, VYC and ZCY performed data analysis, YTW and VYC performed manuscript drafting; YTW, VYC, ZCY, QJP, XZ, YY, PX and YLQ performed manuscript revision. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Reference

1. Zheng, R. S. *et al.* Cancer incidence and mortality in China, 2022. *Zhonghua Zhong Liu Za Zhi* **46**, 221-231 (2024). https://doi.org:10.3760/cma.j.cn112152-20240119-00035
2. China Health Statistical Yearbook 2022. *National Health Commission of the People's Republic of China* (2023).
3. Liu, H. Investigation and consideration on the status of pathology departments in 3 831 hospitals of 31 provinces, municipalities and autonomous regions. *Zhonghua Bing Li Xue Za Zhi* **49**, 1217-1220 (2020). https://doi.org:10.3760/cma.j.cn112151-20200904-00695
4. Metter, D. M., Colgan, T. J., Leung, S. T., Timmons, C. F. & Park, J. Y. Trends in the US and Canadian Pathologist Workforces From 2007 to 2017. *JAMA Netw Open* **2**, e194337 (2019). https://doi.org:10.1001/jamanetworkopen.2019.4337
5. Xue, P. *et al.* Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis. *NPJ Digit Med* **5**, 19 (2022). https://doi.org:10.1038/s41746-022-00559-z
6. Xue, P. *et al.* Development and validation of an artificial intelligence system for grading colposcopic impressions and guiding biopsies. *BMC Med* **18**, 406 (2020). https://doi.org:10.1186/s12916-020-01860-y
7. Xue, P., Ng, M. T. A. & Qiao, Y. The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence. *BMC Med* **18**, 169 (2020). https://doi.org:10.1186/s12916-020-01613-x
8. Niazi, M. K. K., Parwani, A. V. & Gurcan, M. N. Digital pathology and artificial intelligence. *Lancet Oncol* **20**, e253-e261 (2019). https://doi.org:10.1016/s1470-2045(19)30154-8
9. Xue, P. *et al.* Improving the Accuracy and Efficiency of Abnormal Cervical Squamous Cell Detection With Cytologist-in-the-Loop Artificial Intelligence. *Mod Pathol* **36**, 100186 (2023). https://doi.org:10.1016/j.modpat.2023.100186
10. Xue, P. *et al.* Assessing artificial intelligence enabled liquid-based cytology for triaging HPV-positive women: a population-based cross-sectional study. *Acta Obstet Gynecol Scand* **102**, 1026-1033 (2023). https://doi.org:10.1111/aogs.14611
11. OpenAI GPT-4 technical report. *arXiv* (Preprint posted online on March 27, 2023).
12. Schukow, C. *et al.* Application of ChatGPT in Routine Diagnostic Pathology: Promises, Pitfalls, and Potential Future Directions. *Adv Anat Pathol* **31**, 15-21 (2024). https://doi.org:10.1097/pap.0000000000000406
13. Lu, M. Y. *et al.* A Multimodal Generative AI Copilot for Human Pathology. *Nature* (2024). https://doi.org:10.1038/s41586-024-07618-3
14. Truhn, D. *et al.* Extracting structured information from unstructured histopathology reports using generative pre-trained transformer 4 (GPT-4). *J Pathol* **262**, 310-319 (2024). https://doi.org:10.1002/path.6232
15. Dergaa, I., Chamari, K., Zmijewski, P. & Ben Saad, H. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biol Sport* **40**, 615-622 (2023). https://doi.org:10.5114/biolsport.2023.125623
16. Holderried, F. *et al.* A Language Model-Powered Simulated Patient With Automated Feedback for History Taking: Prospective Study. *JMIR Med Educ* **10**, e59213 (2024). https://doi.org:10.2196/59213

17 Frommeyer, T. C., Fursmidt, R. M., Gilbert, M. M. & Bett, E. S. The Desire of Medical Students to Integrate Artificial Intelligence Into Medical Education: An Opinion Article. *Front Digit Health* **4**, 831123 (2022). https://doi.org:10.3389/fdgth.2022.831123

18 Fang, Q. *et al*. Artificial Intelligence (AI)-driven dental education: Exploring the role of chatbots in a clinical learning environment. *J Prosthet Dent* (2024). https://doi.org:10.1016/j.prosdent.2024.03.038

19 Guevara, M. *et al*. Large language models to identify social determinants of health in electronic health records. *NPJ Digit Med* **7**, 6 (2024). https://doi.org:10.1038/s41746-023-00970-0

20 Truong, H. P. *et al*. Utilization of Social Determinants of Health ICD-10 Z-Codes Among Hospitalized Patients in the United States, 2016-2017. *Med Care* **58**, 1037-1043 (2020). https://doi.org:10.1097/mlr.0000000000001418

21 Denecke, K., May, R. & Rivera Romero, O. Potential of Large Language Models in Health Care: Delphi Study. *J Med Internet Res* **26**, e52399 (2024). https://doi.org:10.2196/52399

22 Huang, H. *et al*. ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. *Int J Oral Sci* **15**, 29 (2023). https://doi.org:10.1038/s41368-023-00239-y

23 Haemmerli, J. *et al*. ChatGPT in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board? *BMJ Health Care Inform* **30** (2023). https://doi.org:10.1136/bmjhci-2023-100775

24 The Lancet Digital, H. Pathology in the era of generative AI. *Lancet Digit Health* **6**, e536 (2024). https://doi.org:10.1016/s2589-7500(24)00157-2

25 Kwong, J. C. C., Wang, S. C. Y., Nickel, G. C., Cacciamani, G. E. & Kvedar, J. C. The long but necessary road to responsible use of large language models in healthcare research. *NPJ Digit Med* **7**, 177 (2024). https://doi.org:10.1038/s41746-024-01180-y

Supplementary Files

Multimedia Appendixes

Questionnaire.
URL: http://asset.jmir.pub/assets/3d010f654001e48cd0f1ce58efc1177a.docx