Methods and Applications

Automated Monkeypox Identification from Electronic Medical Records Using Large Language Models — Shenzhen City, Guangdong Province, China, 2023–2025

Diyang Xue[1,&]; Ye Ye[2,&]; Yu Wu[1]; Zhen Zhang[1]; Jianhua Lu[1]; Xuan Zou[1,#]; Qiuying Lv[3,#]

ABSTRACT

Introduction: Monkeypox (mpox) is an emerging infectious disease with diverse clinical manifestations that complicate early case identification. Traditional diagnostic methods relying on laboratory confirmation may delay detection during the early stages of infection. Clinical free-text data in electronic medical records (EMRs), such as chief complaints and histories of present illness, contain valuable information for early identification yet remain underutilized in routine surveillance. This study explored an automated approach to mpox identification from EMRs using large language models (LLMs).

Methods: We conducted a retrospective study of 239 individuals — 126 laboratory-confirmed mpox cases and 113 non-mpox controls — from multiple hospitals in Shenzhen, China. Free-text chief complaints and histories of present illness were processed using the DeepSeek-R1-14B large language model to extract structured clinical variables. Fourteen clinically relevant features were selected for model development. Naïve Bayes, logistic regression, and random forest classifiers were trained on these features, and a Jieba plus term frequency–inverse document frequency (TF-IDF) approach served as a baseline comparison. Qwen3-14B was also applied for feature extraction as a secondary comparison.

Results: mpox cases most frequently presented at fever clinics (55.6%), with considerable variation in care-seeking patterns. Models using DeepSeek-extracted features outperformed those based on traditional TF-IDF representations and Qwen3-14B, with logistic regression achieving the best performance [area under the receiver operating characteristic curve (AUROC), 0.927; accuracy, 87.5%]. DeepSeek demonstrated high accuracy (96.1%) in extracting relevant clinical symptoms, including fever, rash, pustules, and HIV infection.

Conclusion: LLM-based extraction of clinical features from routine text data represents a promising approach for automated mpox case identification and may support future intelligent surveillance and early warning systems.

Monkeypox (mpox) is a zoonotic disease caused by the monkeypox virus, a member of the *Orthopoxvirus* genus that also includes variola, vaccinia, and cowpox viruses (*1*). The disease presents with heterogeneous clinical manifestations, most commonly involving dermatologic and mucocutaneous lesions. Early identification of suspected cases is critical for timely isolation, epidemiologic investigation, and interruption of transmission.

In many clinical settings, mpox remains a relatively low-prevalence disease, and clinicians may have limited experience with its presentation. As a result, mpox may not be considered during the initial diagnostic evaluation, and patients may not be promptly referred for confirmatory testing. This gap in early clinical recognition is particularly concerning given the high transmissibility of mpox, as delayed or missed diagnoses can facilitate ongoing transmission and elevate public health risks.

Current diagnostic practices for mpox rely on epidemiological exposure assessment and clinical evaluation, followed by laboratory confirmation through nucleic acid amplification testing such as real-time or conventional polymerase chain reaction (PCR). Although these methods provide high diagnostic accuracy, laboratory confirmation typically occurs after the initial clinical encounter, limiting their utility for early case detection and real-time surveillance. These limitations underscore the need for complementary approaches that enable earlier identification of suspected mpox cases at the point of care.

Clinical narratives documented in electronic medical record (EMR) systems — including chief complaints, physical examination findings, and histories of present illness — contain rich descriptive information that may facilitate earlier recognition of mpox. Ideally, EMR systems could automatically analyze these narratives and generate alerts for suspected cases.

As an initial investigation, we evaluated the feasibility of automatically identifying suspected mpox cases using unstructured clinical narratives. We developed machine learning models based on clinical features extracted by large language models (LLMs). Through contextual reasoning, LLMs can infer entity boundaries, semantic categories, and attributes — such as symptom presence or negation — from free-text clinical documentation, even without explicit rules. In contrast, traditional natural language processing (NLP) approaches, such as rule-based methods and bag-of-words models, including term frequency–inverse document frequency (TF-IDF) (*2*), require extensive manual curation, are sensitive to linguistic variability, and often fail to capture complex clinical semantics.

METHODS

Study Design and Data Collection

This retrospective, exploratory study used clinical records routinely collected from multiple hospitals in Shenzhen, China. The study included 239 individuals — 126 laboratory-confirmed mpox cases and 113 non-mpox controls — identified between July 18, 2023, and October 18, 2025. All mpox patients met the diagnostic criteria and had complete chief complaint and history of present illness documentation.

Controls were randomly selected from non-mpox patients who visited the same clinical department on the same day as each mpox case, ensuring comparable clinical context and healthcare-seeking patterns. Importantly, controls were not healthy individuals but rather patients presenting with various clinical complaints, including dermatologic conditions, infectious diseases, and other non-mpox diagnoses.

Because the study relied on free-text clinical narratives for feature extraction, records with missing or empty chief complaint and history of present illness fields could not be processed and were excluded. This exclusion resulted in a slight imbalance between cases and controls and may have introduced selection bias.

Given the emerging nature of mpox and the limited number of confirmed cases, this study was designed as an exploratory proof-of-concept investigation using all available eligible cases during the study period. A formal sample size calculation was not performed.

Large Language Model–based Feature Extraction

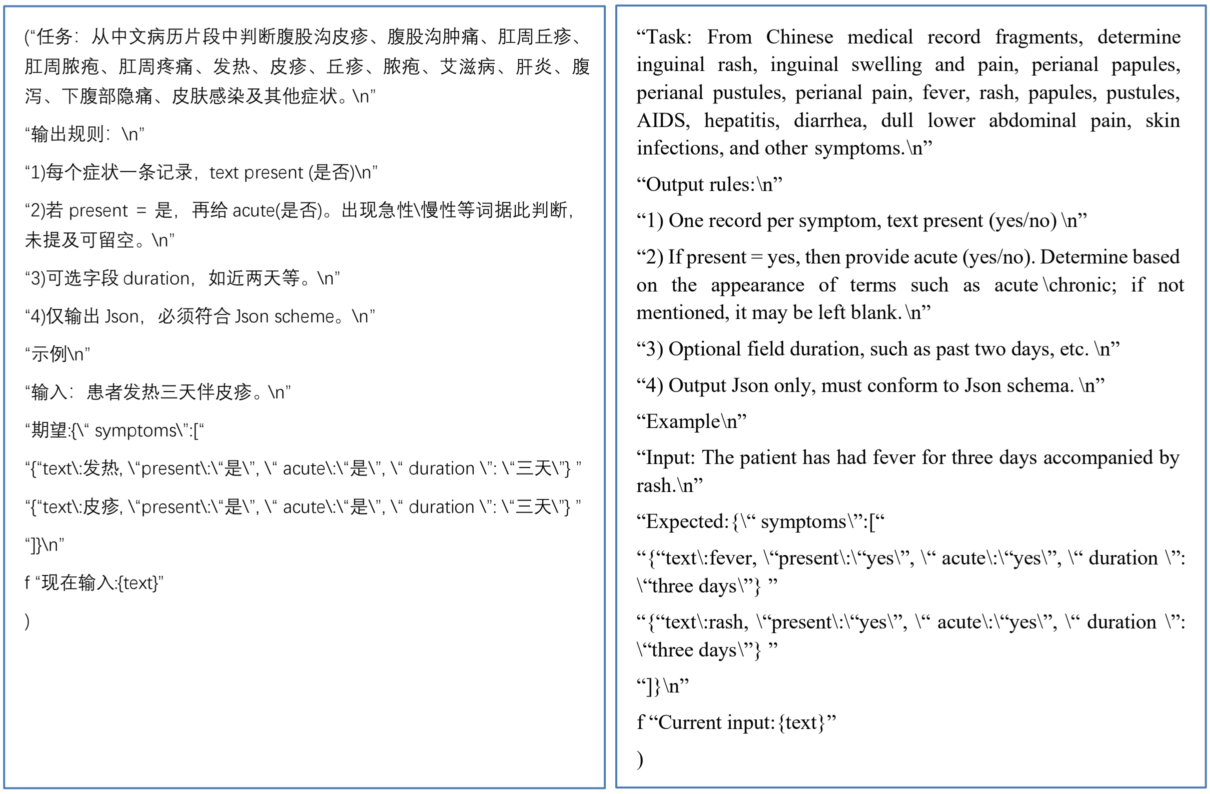
DeepSeek-R1-14B was deployed via an application programming interface (API) for prompt-based extraction of clinical information from free-text chief complaints and histories of present illness. A predefined prompt (Figure 1) instructed the model to perform entity recognition and attribute classification, identify symptoms, clinical signs, and relevant comorbidities, and assign binary presence or absence labels.

Feature extraction was conducted under deterministic inference by fixing the temperature parameter at 0 to ensure reproducibility and minimize stochastic variation. The model was applied in a zero-shot setting without task-specific fine-tuning. Raw model outputs were post-processed into standardized binary variables, with consistency checks applied to ensure conformity with the predefined schema. This process generated 98 candidate binary clinical variables per individual (Supplementary Material, available at https://weekly.chinacdc.cn/), spanning dermatologic manifestations, systemic symptoms, gastrointestinal symptoms, and relevant medical history.

Feature Selection Based on Clinical Knowledge

From the 98 candidate variables extracted by the LLM, feature selection was guided by existing clinical guidelines and expert consultation with infectious disease and public health specialists (*3*–*4*). Given the exploratory nature of the study, the relatively small sample size, and the need for model interpretability, features were selected manually rather than through automated methods. Additional analyses using all 98 variables and least absolute shrinkage and selection operator (LASSO)-based feature selection were conducted to evaluate the impact of manual feature selection.

Selection prioritized variables associated with characteristic mpox lesions, high-risk populations, and sufficient prevalence in the study population. Fourteen variables were retained for model development: inguinal rash, inguinal swelling and pain, perianal papules, perianal pustules, perianal pain, fever, rash, papules, pustules, AIDS, hepatitis, diarrhea, dull lower abdominal pain, and skin infections.



```
("任务：从中文病历片段中判断腹股沟皮疹、腹股沟肿痛、肛周丘疹、肛周脓疱、肛周疼痛、发热、皮疹、丘疹、脓疱、艾滋病、肝炎、腹泻、下腹部隐痛、皮肤感染及其他症状。\n"

"输出规则：\n"

"1)每个症状一条记录，text present (是否)\n"

"2)若 present = 是，再给 acute(是否)。出现急性\慢性等词据此判断，未提及可留空。\n"

"3)可选字段 duration，如近两天等。\n"

"4)仅输出 Json，必须符合 Json scheme。\n"

"示例\n"

"输入：患者发热三天伴皮疹。\n"

"期望:{\" symptoms\":["

"{"text\:发热, \"present\:\"是\", \" acute\:\"是\", \" duration \": \"三天\"} "

"{"text\:皮疹, \"present\:\"是\", \" acute\:\"是\", \" duration \": \"三天\"} "

"]}\n"

f "现在输入:{text}"

)
```

```
"Task: From Chinese medical record fragments, determine inguinal rash, inguinal swelling and pain, perianal papules, perianal pustules, perianal pain, fever, rash, papules, pustules, AIDS, hepatitis, diarrhea, dull lower abdominal pain, skin infections, and other symptoms.\n"

"Output rules:\n"

"1) One record per symptom, text present (yes/no) \n"

"2) If present = yes, then provide acute (yes/no). Determine based on the appearance of terms such as acute \chronic; if not mentioned, it may be left blank. \n"

"3) Optional field duration, such as past two days, etc. \n"

"4) Output Json only, must conform to Json schema. \n"

"Example\n"

"Input: The patient has had fever for three days accompanied by rash.\n"

"Expected: {\" symptoms\":["

"{"text\:fever, \"present\:\"yes\", \" acute\:\"yes\", \" duration \": \"three days\"} "

"{"text\:rash, \"present\:\"yes\", \" acute\:\"yes\", \" duration \": \"three days\"} "

"]}\n"

f "Current input: {text}"

)
```

FIGURE 1. Prompt for symptom extraction from Chinese clinical text for monkeypox cases and controls.
Note: left panel: original Chinese; right panel: English translation.

Model Construction and Baseline Comparison

The dataset was randomly split into training and validation sets at an 8:2 ratio. Using the same 14 LLM-extracted variables, three classification models were developed: naïve Bayes, logistic regression, and random forest. No additional internal feature selection was applied, ensuring comparability across models.

For the naïve Bayes model, binary variables were modeled under a Bernoulli distribution assumption. Logistic regression was implemented with L2 regularization, and random forest models consisted of ensembles of decision trees.

To further assess model robustness and reduce the risk of overfitting, five-fold cross-validation was additionally performed on the training dataset for the primary model (logistic regression using DeepSeek-extracted features).

To evaluate the added value of LLM-based feature extraction, a traditional NLP baseline was implemented using the Jieba word segmentation tool (*5*) combined with TF-IDF feature representation. Naïve Bayes, logistic regression, and random forest models were constructed using the same workflow for comparison.

In addition to DeepSeek-R1-14B, the Qwen3-14B large language model (*6*) was applied using the same prompt, inference settings, and feature selection procedures to provide a comparative assessment of different LLM-based feature extraction approaches. Model performance was evaluated using area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and Brier score for calibration assessment.

RESULTS

Clinical Care-Seeking Patterns of Monkeypox Cases

Mpox cases presented across multiple clinical departments, most commonly fever clinics (55.6%), followed by infectious disease outpatient clinics (16.7%) and HIV/AIDS clinics (9.5%) (Table 1). Dermatology-related clinics accounted for 11.1% of initial visits, with the remaining cases distributed across emergency departments and other outpatient services.

This distribution reflects heterogeneous care-seeking

TABLE 1. Distribution of mpox cases across clinical departments.

Clinical Department	Number of Cases (*n*)	Proportion (%)
Fever clinic	70	55.6
Infectious disease outpatient clinic	21	16.7
HIV/AIDS clinic	12	9.5
Dermatology outpatient clinic	10	7.9
Research follow-up clinic	6	4.8
Dermatology and venereology clinic	3	2.4
Emergency department outpatient clinic	1	0.8
Dermatologic aesthetic and laser clinic	1	0.8
Wound, ostomy, and continence nursing specialty clinic	1	0.8
Proctology night clinic	1	0.8
Total	126	100.0

Abbreviation: mpox=monkeypox; HIV=human immunodeficiency virus; AIDS=acquired immunodeficiency syndrome.

patterns and highlights the limitations of department-specific surveillance strategies for early mpox case identification.

Accuracy of DeepSeek LLM in Clinical Symptom Extraction

As shown in Table 2, DeepSeek-R1-14B demonstrated strong performance in extracting 14 mpox-related clinical symptoms from unstructured clinical text, achieving a precision of 0.771, recall of 0.919, F1 score of 0.839, and accuracy of 0.961. The model showed particularly high precision and recall for commonly documented symptoms, including fever, rash, pustules, and HIV infection. Overall, these results suggest that DeepSeek can reliably extract clinically relevant symptom information from routine clinical narratives.

Comparative Classification Performance Across Feature Extraction Methods

As shown in Table 3, models using LLM-extracted features consistently outperformed those based on traditional Jieba plus TF-IDF representations across multiple evaluation metrics, including AUROC, accuracy, sensitivity, specificity, PPV, NPV, and calibration (Brier score).

Among all models, logistic regression using DeepSeek-R1-14B–extracted features achieved the best overall performance, with the highest AUROC (0.927), accuracy (87.5%), balanced sensitivity (88.0%) and specificity (87.0%), and the lowest Brier score (0.106), indicating superior calibration.

Models based on Qwen3-14B–extracted features demonstrated moderate performance, generally achieving higher sensitivity but lower specificity and less favorable calibration compared with DeepSeek-based models. In contrast, Jieba plus TF-IDF–based models showed the weakest overall performance, with reduced specificity and PPV.

A five-fold cross-validation on the training dataset yielded a mean AUROC of 0.904, consistent with the performance observed on the independent validation set.

Feature Selection Analysis

To evaluate the impact of manual feature selection, additional analyses were conducted using all 98 LLM-extracted variables and a LASSO logistic regression for data-driven feature selection. Logistic regression using all 98 variables achieved an AUROC of 0.910 and an accuracy of 85.4%, while the LASSO-based model achieved an AUROC of 0.902 and an accuracy of 87.5%. These results were comparable to the 14-feature clinically selected model and did not demonstrate clear improvement.

LASSO-based feature selection identified several clinically relevant dermatologic and systemic symptoms associated with mpox, suggesting that meaningful signals can be captured directly from high-dimensional feature space. However, models based on clinically selected features demonstrated more stable and interpretable performance, particularly given the relatively small sample size.

Error Case Analysis

Error case analysis revealed several patterns underlying model misclassification. False-positive cases were commonly associated with nonspecific

TABLE 2. Performance evaluation of DeepSeek-R1-14B for extracting mpox-related clinical symptoms.

Symptom	TP	FP	TN	FN	Precision	Recall	F1 Score	Accuracy
Inguinal rash	10	0	228	0	1.000	1.000	1.000	1.000
Inguinal swelling and pain	2	0	236	0	1.000	1.000	1.000	1.000
Perianal papules	9	7	222	0	0.563	1.000	0.720	0.970
Perianal pustules	7	2	227	2	0.778	0.778	0.778	0.983
Perianal pain	6	1	231	0	0.857	1.000	0.923	0.996
Fever	86	11	141	0	0.887	1.000	0.940	0.954
Rash	97	12	128	1	0.890	0.990	0.940	0.945
Papules	31	63	143	1	0.330	0.969	0.492	0.731
Pustules	34	3	182	19	0.919	0.624	0.756	0.908
AIDS	37	1	199	1	0.974	0.974	0.974	0.992
Hepatitis	7	1	229	1	0.875	0.875	0.875	0.992
Diarrhea	10	0	225	3	1.000	0.769	0.870	0.987
Dull lower abdominal pain	1	0	237	0	1.000	1.000	1.000	1.000
Skin infections	3	0	233	2	1.000	0.600	0.750	0.992
Total	340	101	2,861	30	0.771	0.919	0.839	0.961

Note: Precision=TP / (TP + FP); Recall=TP / (TP + FN); F1=2 × Precision × Recall / (Precision + Recall); Accuracy=(TP + TN) / (TP + FP + TN + FN).

Abbreviation: mpox=monkeypox; AIDS=acquired immunodeficiency syndrome; TP=true positive; FP=false positive; TN=true negative; FN=false negative.

TABLE 3. Comparison of classification performance using DeepSeek-R1-14B, Qwen3-14B, and Jieba + TF-IDF features.

Feature extraction	Model	AUROC	Accuracy, %	Sensitivity, %	Specificity, %	PPV, %	NPV, %	Brier score
DeepSeek-R1-14B	Naive Bayes	0.910	85.4	88.0	82.6	84.6	86.4	0.141
DeepSeek-R1-14B	Logistic Regression	0.927	87.5	88.0	87.0	88.0	87.0	0.106
DeepSeek-R1-14B	Random Forest	0.890	85.4	88.0	82.6	84.6	86.4	0.118
Qwen3-14B	Naive Bayes	0.770	72.9	76.0	69.6	73.1	72.7	0.271
Qwen3-14B	Logistic Regression	0.838	81.2	92.0	69.6	76.7	88.9	0.147
Qwen3-14B	Random Forest	0.861	81.2	92.0	69.6	76.7	88.9	0.141
Jieba + TF-IDF	Naive Bayes	0.790	68.8	80.0	56.5	66.7	72.2	0.189
Jieba + TF-IDF	Logistic Regression	0.819	75.0	80.0	69.6	74.1	76.2	0.188
Jieba + TF-IDF	Random Forest	0.877	79.2	92.0	65.2	74.2	88.2	0.152

Abbreviation: AUROC=area under the receiver operating characteristic curve; PPV=positive predictive value; NPV=negative predictive value; TF-IDF=term frequency–inverse document frequency.

dermatologic symptoms, such as erythema or papules, that were over-interpreted as characteristic mpox lesions in the absence of systemic indicators. For example, a patient presenting with perianal erythema and itching was clinically diagnosed with localized eczema, yet the model assigned a high probability of mpox based on the anatomical location of the symptoms.

In contrast, false-negative cases primarily involved patients presenting with systemic symptoms such as fever and myalgia but lacking documented characteristic rashes or pustules at the initial encounter. These findings suggest that further refinement is needed to improve differentiation between mpox and common dermatologic conditions with overlapping presentations.

Comparison Between DeepSeek and Qwen3 Feature Extraction

To further investigate the performance differences between DeepSeek-R1-14B and Qwen3-14B, we conducted a qualitative comparison of their feature extraction outputs using identical prompts and inference settings.

Although both models were capable of extracting structured clinical variables from free-text records, DeepSeek demonstrated greater robustness in handling ambiguous or indirectly expressed symptoms. For example, when clinical narratives described "perianal discomfort" or "localized irritation," DeepSeek more consistently inferred dermatologic involvement, whereas Qwen3 frequently failed to capture these signals when explicit terms such as "rash" or "pustules" were absent.

Additionally, Qwen3 exhibited greater variability in feature extraction consistency across similar clinical descriptions, occasionally producing incomplete or less reliable binary representations. These findings suggest that stronger contextual understanding and instruction-following capability may contribute to the superior downstream classification performance observed with DeepSeek.

DISCUSSION

Main Findings

This study demonstrated that LLM-extracted clinical features can improve mpox case identification from routine clinical text, with DeepSeek-R1-14B yielding the strongest performance among the evaluated models. The logistic regression model built on DeepSeek-derived features achieved an AUROC of 0.927 and an accuracy of 87.5%, indicating that the model can capture clinically relevant signals from unstructured medical narratives.

Beyond overall model performance, the favorable results may also reflect the clinical relevance of the symptom features used for classification. The selected features represent common and emerging mpox presentations observed in recent clinical practice, particularly dermatologic manifestations involving the inguinal and perianal regions, which are frequently described in routine medical records (*7*). Local symptoms such as inguinal swelling and perianal pain capture inflammatory responses — including lymphadenopathy and rectal involvement — that may occur even when typical skin lesions are limited or absent. Systemic and gastrointestinal symptoms, including fever and diarrhea, further account for heterogeneous clinical presentations and early-stage disease. Additionally, hepatitis reflects systemic involvement or coexisting conditions observed in some patients, particularly in the context of immune dysfunction or concurrent infections (*8*). Moreover, incorporating immune status–related information such as HIV/AIDS enables the model to capture clinically meaningful variation in symptom burden and disease severity (*9*). Together, these features provide a clinically grounded representation of mpox manifestations and explain the model's ability to extract discriminative signals from unstructured clinical text.

Clinical Translation

The observed performance of DeepSeek highlights the potential utility of LLMs in supporting infectious disease surveillance and clinical decision-making. Because the proposed framework relies solely on routinely collected EMR data — including chief complaints and histories of present illness — it may be feasible to integrate into existing hospital information systems without requiring additional data collection or workflow disruption.

Automated screening based on clinical narratives may facilitate earlier identification of suspected mpox cases before laboratory confirmation, particularly in settings where patients present across multiple clinical departments. In addition, the LLM-based approach demonstrated some ability to capture semantically related symptom descriptions beyond explicit keywords, suggesting potential advantages in handling heterogeneous clinical documentation.

Ethical and privacy considerations are also important for practical implementation. This study used retrospective, de-identified clinical data, and any future deployment would require strict adherence to data protection regulations and institutional governance. Importantly, the proposed system is intended to support — rather than replace — clinical decision-making, and human oversight remains essential.

Finally, this study was based on retrospective data and did not include temporal stratified analysis or prospective validation. Future work should focus on evaluating model performance across different time periods and conducting prospective validation in real-world clinical settings to assess generalizability and robustness.

Limitations

Several limitations should be considered when interpreting these findings. First, this study should be regarded as a proof-of-concept investigation based on a relatively small retrospective dataset, which may limit generalizability and increase the risk of overfitting despite consistent five-fold cross-validation performance. External validation across different healthcare systems and prospective evaluation in real-world settings were not performed.

Second, the approach relies heavily on free-text clinical documentation, which may be incomplete, nonstandardized, or ambiguous. Although DeepSeek demonstrated some capability in handling indirectly expressed symptoms, performance may decrease when clinical descriptions are highly vague or atypical. In addition, the exclusion of records with incomplete narratives may have introduced selection bias.

Third, clinically guided manual feature selection may introduce verification bias by constraining the model to predefined symptom patterns. Although additional analyses using all 98 variables and LASSO-based feature selection demonstrated comparable performance, clinically selected features provided more stable and interpretable results.

Finally, the baseline comparison in this study was limited to Jieba plus TF-IDF feature extraction. More advanced traditional clinical NLP approaches, such as rule-based named entity recognition systems or neural sequence labeling models, may provide stronger baselines for future evaluation.

Future Directions

Despite these limitations, the proposed framework demonstrates the feasibility of using LLMs for automated extraction of clinically relevant information from routine medical narratives. Because the approach operates in a zero-shot manner without task-specific annotation or model fine-tuning, it may offer practical advantages for rapid deployment in emerging infectious disease surveillance settings.

Future studies should evaluate the framework using larger multicenter datasets, external validation cohorts, and prospective real-world implementation. Incorporating additional data modalities — such as medical imaging or population-level surveillance data — may further improve early identification of atypical presentations. More broadly, similar approaches may be adaptable to other emerging or reemerging infectious diseases that rely heavily on narrative clinical documentation.

The DeepSeek-R1-14B large language model, combined with clinically informed feature selection, demonstrated strong performance in automated mpox case identification using routine clinical text. These findings suggest that LLM-based processing of clinical narratives represents a promising and scalable approach for supporting infectious disease surveillance. Despite the limitations of this proof-of-concept study, the proposed framework may be adaptable to other infectious diseases and integrated into existing surveillance systems following further validation.

Conflicts of interest: No conflicts of interest.

Funding: Supported by the Shenzhen Medical Research Fund (Grant No. E24010011, E24010010) and the Shenzhen Science and Technology Program (Grant No. SYSPG20241211173921049). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

doi: 10.46234/ccdcw2026.102

[#] Corresponding authors: Qiuying Lv, lvqiuying@wjw.sz.gov.cn; Xuan Zou, zouxuan@wjw.sz.gov.cn.

---

[1] Shenzhen Center for Disease Control and Prevention, Shenzhen City, Guangdong Province, China; [2] School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen City, Guangdong Province, China; [3] School of Public Health, Southern Medical University, Guangzhou City, Guangdong Province, China.
[&] Joint first authors.

Submitted: January 23, 2026
Accepted: May 08, 2026
Issued: May 15, 2026

REFERENCES

1. Mitjà O, Ogoina D, Titanji BK, Galvan C, Muyembe JJ, Marks M, et al. Monkeypox. Lancet 2023;401(10370):60 – 74. https://doi.org/10.1016/S0140-6736(22)02075-X.
2. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Inf Process Manage 1988;24(5):513 – 23. https://doi.org/10.1016/0306-4573(88)90021-0.
3. National Health Commission of the People's Republic of China. Guidelines for the diagnosis and treatment of monkeypox (2022 version). Chin J Viral Dis 2022;12(4):255 – 6. https://doi.org/10.16505/j.2095-0136.2022.0040.
4. World Health Organization. Clinical management and infection prevention and control for mpox: living guideline, May 2025. Geneva: World Health Organization; 2025. http://dx.doi.org/10.2471/B09434.
5. Sun J. Jieba: Chinese text segmentation. 2012. https://github.com/fxsjy/jieba. [2025-11-15].
6. Yang A, Li AF, Yang BS, Zhang BC, Hui BY, Zheng B, et al. Qwen3 technical report. 2025. https://arxiv.org/abs/2505.09388. [2025-12-03].
7. Pittman PR, Martin JW, Kingebeni PM, Tamfum JJM, Wan QW, Reynolds MG, et al. Clinical characterization of human monkeypox infections in the Democratic Republic of the Congo. medRxiv 2022. http://dx.doi.org/10.1101/2022.05.26.22273379. [2025-12-03].
8. Calin R, Périllaud-Dubois C, Marot S, Kerrou K, Peytavin G, Bachir M, et al. Mpox hepatic and pulmonary lesions in HIV/hepatitis B virus co-infected patient, France. Emerg Infect Dis 2024;30(11):2445 – 7. https://doi.org/10.3201/eid3011.241331.
9. Saldana CS, Kelley CF, Aldred BM, Cantos VD. Mpox and HIV: a narrative review. Curr HIV/AIDS Rep 2023;20(4):261 – 9. https://doi.org/10.1007/s11904-023-00661-1.

SUPPLEMENTARY MATERIAL

A total of 98 clinical features were extracted from electronic medical records using DeepSeek-R1-14B. The 98 clinical features extracted were: inguinal rash, inguinal swelling and pain, perianal papules, perianal pustules, perianal pain, fever, rash, papules, pustules, acquired immunodeficiency syndrome (AIDS), hepatitis, diarrhea, dull lower abdominal pain, skin infections, dizziness, fatigue, hyperlipidemia, perianal herpes, cough, sore throat, nasal congestion, vomiting, abdominal pain, dysuria, perianal rash, perianal abscess, syphilis, headache, myalgia, chills, rigors, human immunodeficiency virus positive (HIV-positive), diabetes mellitus, genital ulcer, muscle and joint pain, rhinorrhea, cough with tinnitus, soft tissue infection, penile swelling and pain, pruritus, pain, productive sputum, itching sensation, perianal inflammation, herpesvirus infection of the genital and genitourinary tract, herpes, generalized muscle and joint pain, cough with white sputum, facial swelling, difficulty in defecation, mild myalgia, recurrent condyloma acuminatum, painful rash, bronchitis, respiratory tract infection, cough-variant asthma, renal insufficiency, nausea and vomiting, generalized muscle and joint discomfort, dry throat, dry cough with scant sputum, gingival pain, nasal congestion and rhinorrhea, pulmonary tuberculosis, night sweats, dyspnea, chest pain, hemoptysis, parasitic infection, cough with yellow sputum, neurosyphilis, poor appetite, poor general condition, chest mass, multiple lipomas, poor mental status and appetite, poor sleep, parasitic disease, cough with yellow mucoid sputum, venous thrombosis, Grade 3 hypertension (very high risk), abdominal pain and diarrhea, bacterial infection, viral warts, human papillomavirus (HPV) infection, abdominal distension, bilateral breast tenderness, mild headache, chest tightness, productive cough, nausea, throat itching, rhinorrhea and nasal congestion, chronic hepatitis B, poor postoperative wound healing, androgenetic alopecia, and kidney stones.