A fine-tuned large language model for predicting provincial AIDS case increases

0883

Xinsheng Wu[1,2], Jinyuan Wu[1,2], Huachun Zou [1,2,3,4]

[1]Shanghai Institute of Infectious Disease and Biosecurity, School of Public Health, Fudan University, Shanghai, China, [2]Center for AIDS Research (CFAR), School of Public Health, Key Laboratory of Public Health Safety (Fudan University), Ministry of Education, Fudan University, Shanghai, China, [3]School of Public Health, Southwest Medical University, Luzhou, China, [4]Melbourne Sexual Health Centre, Faculty of Medicine Nursing and Health Sciences, Monash University, Melbourne, Australia; Corresponding authors: Prof. Huachun Zou

BACKGROUND

Short-horizon, province-level signals of AIDS case increases can help anticipate testing capacity and targeted outreach. Conventional time-series and machine-learning methods typically demand long look-back windows and numerous predictors for both training and inference, and they underperform when only a few inputs are available, thereby heightening data and implementation burdens. We examined whether a low-rank adaptation (LoRA)-adapted pretrained large language model (LLM) can forecast short-term increases using limited recent observations.

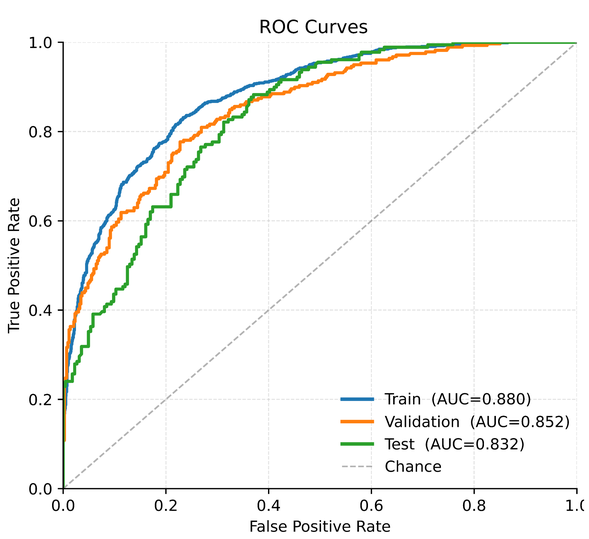
METHODS

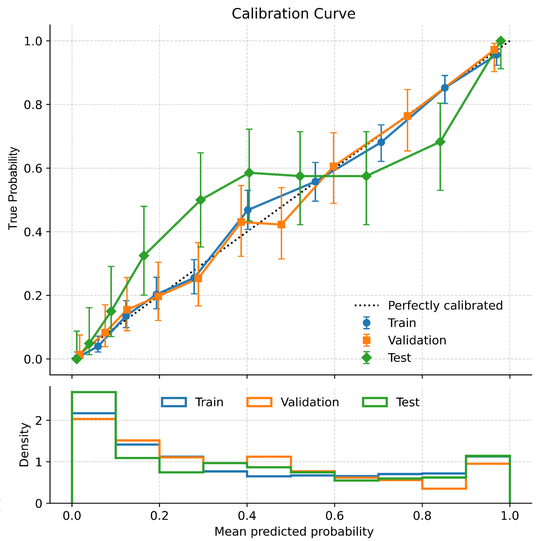
We used provincial monthly counts of newly reported AIDS cases from the National Notifiable Diseases Surveillance System (NNDSS) and the provincial Baidu Search Index (BSI) for “AIDS” from January 2011 to December 2020 to train and evaluate an LLM-based model forecasting 10% AIDS cases increases in the next month for all 31 provinces/municipalities in China. Data were split in temporal order using a 7:2:1 ratio for training, validation, and testing. Leveraging pretrained LLMs (Qwen3-0.6B), we trained a single, multi-province model that learns jointly from all 31 provinces/municipalities and forecasted 10% AIDS cases increase in the next month from a rolling three-month input window using LoRA. Primary performance summaries included accuracy, F1 score, balanced accuracy, recall, and ROC-AUC. Calibration was assessed with quintile-based reliability plots (Wilson 95% CIs) plus a flexible LOWESS/monotone calibration curve (isotonic fallback). Clinical utility was evaluated by decision-curve analysis.

A LoRA-adapted pretrained large language model can deliver accurate, multi-provincial AIDS increase forecasts from minimal inputs

RESULTS

In the validation window, the model achieved accuracy 0.770, F1 score 0.722, precision 0.683, recall 0.766, specificity 0.772, and ROC-AUC 0.852. In the test window, performance was comparable: accuracy 0.739, F1 score 0.709, precision 0.703, recall 0.715, specificity 0.759, and ROC-AUC 0.832. On the test set, Brier=0.175; the flexible curve and quintile means lay close to the 45° line. DCA showed positive net benefit for thresholds 0.02–0.80. The policy-optimal threshold was 0.440, at which the model’s net benefit was 0.203 per province-month. Expressed per 100 province-months, this corresponds to about 20.3 additional correctly identified growth months relative to the best baseline.





CONCLUSIONS

Our study showed that a LoRA-adapted pretrained LLM can deliver accurate, multi-provincial AIDS increase forecasts from minimal inputs without long look-back windows or high-dimensional covariates. This data-light design markedly lowers collection and implementation burdens, simplifies deployment across all provinces with a single model, and preserves strong predictive performance even where surveillance inputs are sparse or delayed. The approach is readily integrable into routine workflows and is extensible to other notifiable diseases and subnational settings.

Author Contact Information



Huachun Zou, MD, PhD, Professor
Department of Epidemiology
School of Public Health, Fudan University
Email: zouhuachun@fudan.edu.cn

PLAIN LANGUAGE SUMMARY

Using national monthly AIDS surveillance reports and Baidu search activity from 2011–2020, we built one AI early-warning model that predicts—for each of China’s 31 provinces/municipalities—whether newly reported AIDS cases will jump by ≥10% next month, using only the most recent three months of data. In out-of-sample testing, the model was consistently accurate (about 74% correct; AUC 0.83) and, in a decision-focused analysis, it would identify more true surge months than standard methods while requiring far less information to run.