Biomolecules and Biomedicine

ISSN: 2831-0896 (Print) | ISSN: 2831-090X (Online)
Journal Impact Factor® (2024): 2.2

CiteScore® (2024): 5.2
www.biomolbiomed.com | blog.bjbms.org

RESEARCH ARTICLE

Bu et al: Predicting insurance overspending risk

Enhancing predictions of health insurance overspending risk through hospital departmental performance indicators

Yao Bu[1#], Danqi Wang[2#], Xiaomao Fan[3], Jiongying Li[4], Lei Hua[2], Lin Zhang[5,6], Wenjun Ma[7], Liwen He[8], Hao Zang[9], Haijun Zhang[10], Xingyu Liu[11], Yufeng Gao[12], Li Liu[1,2*]

[1]School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu, China

[2]Big Data Center, Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu, China

[3]College of Big Data and Internet, Shenzhen Technology University, Shenzhen, Guangdong, China

[4]Office of Health Insurance Administration, Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu, China

[5]Suzhou Industrial Park Monash Research Institute of Science and Technology, Monash University, Suzhou, Jiangsu, China

[6]Monash University-Southeast University Joint Research Institute (Suzhou), Southeast University, Suzhou, Jiangsu, China

[7]School of Computer Science, South China Normal University, Guangzhou, Guangdong, China

[8]Wuxi Innovation Center, Shenzhen Research Institute of Big Data, Wuxi, Jiangsu, China

[9]School of Information and Control Engineering China University of Mining and Technology, Xuzhou, Jiangsu, China

[10]Jiangsu Zhisheng Information Technology Co., LTD., Xuzhou, Jiangsu, China

[11]Wuxi Health Statistics and Information Center, Wuxi, Jiangsu, China

[12]Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu, China

[*]Correspondence to **Li Liu**: 9862016027@jiangnan.edu.cn

[#]**Yao Bu** and **Danqi Wang** contributed equally to this work.

DOI: https://doi.org/10.17305/bb.2025.12051

ABSTRACT

The substantial rise in health insurance expenditures, combined with delayed feedback on overspending from administrative departments, highlights the urgent need for timely reporting of such data. This study analyzed a large cohort of 549,910 discharged patients' medical records from the Wuxi Health Commission, covering the period from January 2022 to November 2023. We applied four widely recognized machine learning techniques—Logistic Regression (LR), LightGBM, Random Forest (RF), and Artificial Neural Networks (ANN)—alongside departmental performance indicators (DPIs) to develop Insurance Overspending Risk Prediction (IORP) models at both regional and hospital levels. The dataset was divided into training and testing sets in a 7:3 ratio. Experimental results showed that LightGBM outperformed the other models, achieving an accuracy of 0.82 for both regional and hospital-level predictions. Its weighted F1-score reached 0.78 at the regional level and 0.82 at the hospital level, with corresponding AUC-ROC (Area Under the Receiver Operating Characteristic Curve) values of 0.91 and 0.94, demonstrating strong performance in identifying overspending risks. The model's high recall and precision further ensure reliable predictions and minimize misclassifications. Notably, four key DPIs—Total Amount of Discharged Patients (TADP), Average Inpatient Stay (AIS), Medicine Expenses Percentage (MEP), and Consumable Expenses Percentage (CEP)—were strongly correlated with overspending risks. The integration of IORP models into the Health Insurance Management System (HIMS) at the Affiliated Hospital of Jiangnan University has significantly improved departmental managers' ability to anticipate overspending. By effectively leveraging HIMS in combination with this advanced model, managers can perform timely, accurate assessments, thereby enhancing financial oversight and resource allocation.

Keywords: Health insurance overspending; departmental performance indicators; overspending risk prediction; machine learning; health insurance management system.

INTRODUCTION

The escalating costs of healthcare have become a global concern, with overspending posing significant challenges to the financial sustainability of healthcare systems. In many countries, healthcare expenditures have grown at an unsustainable rate, driven by factors such as aging populations, the increasing prevalence of chronic diseases, and the rising costs of medical technologies and pharmaceuticals[1, 2]. Overspending in healthcare not only strains national budgets but also threatens the equitable allocation of resources, potentially compromising the quality of care and access to essential services[3] . For instance, in China, the rapid expansion of national health insurance coverage has led to increased financial pressures on hospitals, with overspending becoming a critical issue that undermines the efficiency of healthcare delivery [4]. However, we found that delayed feedback on overspending—an issue never before tackled internationally—has significantly hindered hospital departmental managers' ability to make timely, informed adjustments. Therefore, the development of a system for predicting overspending risks is crucial, enabling administrators to make prompt decisions and enhance the management of health insurance expenditures at both regional and hospital levels.

Recently, researchers have investigated various reasons that contribute to high medical insurance costs and proposed a series of corresponding evaluation/prediction approaches[5-15]. These approaches can be categorized into three groups: statistics analysis [5, 10-14], machine learning modeling [6-9, 14], and deep learning methods[6, 15]. As for the statistics analysis-based methods, Z. Mitkova et al. [10]proposed using the Kruskal-Wallis test to analyze cur- rent and extrapolate future trends in the healthcare and pharmaceutical budget based on the National Health Insurance Fund (NHIF). Y. Murakami et al. [11]proposed using Gamma regression to analyze data from 33,213 cardiovascular disease patients, aiming to identify risk factors correlated with medical expenses and reduce overall healthcare costs.

Based on data primarily from 2013 to 2016, I. Papanicolas et al. [12] analyzed information from key international organizations in the Organisation for Economic Co-operation and Development (OECD), and found that services, drug expenses, medical management costs, and employee salaries are critical factors contributing to the high costs incurred by hospitals. As for the machine learning-based methods, N. Ye [14] selected population factors as independent variables and urban basic medical insurance expenditure was selected as dependent variables, establishing a regression model to explore the relationships between these factors. Based on the inpatients of National Health Research Database (NHRD), Y.C. Huang et al. [7] constructed

a predictive model using different machine learning algorithms, including sup- port vector regression (SVR) and extreme gradient boosting (XGBoost), and found that surgical expenses were a major expense factor for patients. K. Kaushik et al. [8] predicted the health insurance cost incurred by individuals on the basis of demographic features and achieved an accuracy of 92.72%. Additionally, several studies have integrated machine learning methods into healthcare information systems to enhance predictive capabilities and decision support [16-20]. These works provide insights into implementation strategies that reinforce the novelty of our HIMS integration. Regarding the deep learning-based methods, G.Z. Zhang and colleagues [15] proposed a framework for detecting fraud in medical insurance using consortium blockchain technology and deep learning, which improved efficiency and effectively identifies fraud. P. Drewe-Boss et al. [6] proposed using a deep neural network and a ridge regression model on a sample of German insurants to predict total one-year healthcare costs. They found that the neural network demonstrated superior performance. These methods achieved competitive prediction performance, but they primarily focused on controlling individual medical expenses. These measures show limited effectiveness in addressing overspending at regional and hospital levels, while imposing heavy administrative burdens on managers overseeing departmental spending.

From a regional and hospital perspective, departmental decision-making plays a more crucial role in managing costs. Strengthening budget management at the departmental level allows hospitals to more effectively control expenditures while maintaining a balance between the quality of medical services and financial stability. Herewith, to address aforementioned issues, we proposed the health insurance overspending risk prediction (IORP) models using departmental performance indicators (DPIs) for regional and hospital administrators. Specifically, we firstly collected 549,910 discharged patient medical records spanning from January 2022 to November 2023 in Wuxi, China. Then, these records were aggregated into regional-level and hospital- level department datasets, containing 8,416 and 44,017 records, respectively. In addition, we utilized the tool of Statistical Process Control (SPC) to categorize the depart- mental overspending into different groups,i.e., high risk, low risk, and no risk. Secondly, we utilized four widely recognized machine learning techniques—Logistic Regression (LR), LightGBM, Random Forest (RF), and Artificial Neural Networks (ANN)—with departmental performance indicators (DPIs) to develop regional-level and hospital- level IORP models. The experimental results show that the LightGBM algorithm exhibited outstanding predictive capabilities with accuracies of 0.82 (regional-level and hospital-level

models). Thirdly, we used the tool of SHapley Additive exPlanations (SHAP) to present the importance of each DPI. Our analysis identified four key indicators that demonstrate a strong correlation with departmental overspending: Total Amount of Discharged Patients (TADP), Average Inpatient Stay (AIS), Medicine Expenses Percentage (MEP), and Consumable Expenses Percentage (CEP). Finally, we integrated the IORP models into the Hospital Information Management System (HIMS) at the Affiliated Hospital of Jiangnan University to enhance the capability of administrators in predicting overspending risks. By effectively utilizing this advanced model within HIMS, the hospital departmental managers can conduct timely and accurate risk assessments, leading to more efficient financial management and optimal resource allocation. To sum up, the primary contributions can be summarized as follows:

•We collected a total of 549,910 discharged patient records from January 2022 to November 2023 in Wuxi, China as well as organizing them into both regional and hospital-level departmental datasets. Meanwhile, we employed SPC (Statistical Process Control) techniques to analyze the data, categorizing departmental overspending into three distinct risk groups: no risk, low risk, and high risk.

•We presented four widely recognized machine learning techniques—LR, LightGBM, RF, and ANN—with DPIs to develop regional-level and hospital-level IORP models.

•We identified four key indicators that demonstrate a strong correlation with departmental overspending: TADP, AIS, MEP, and CEP.

•We successfully integrated the IORP models into the Hospital Information Management System (HIMS) at the Affiliated Hospital of Jiangnan University. This integration facilitates timely and accurate risk assessments, significantly improving financial management and resource allocation.

The remainder content of this paper can organized as follows. Section 1 provides a detailed overview of data and methods. Section 2 discusses results, including the models evaluation , explanation and applications. Section 4 explores the principal findings and limitations. Finally, Section 5 concludes the paper with conclusions.

MATERIALS AND METHODS

Study design

The study utilized a local health insurance database containing medical records of discharged patients in China. Our methodology began with the extraction of DPIs from individual patient insurance data to forecast departmental overspending risks. We then constructed two datasets: one for monthly departmental data at the regional level and another for daily departmental data at the hospital level. To classify overspending, i.e., label, across various departments, we employed SPC, a widely recognized method for quality assurance in industrial settings. For the IORP modeling at both regional and hospital levels, we applied four machine learning algorithms: LR, RF, LightGBM, and ANN. The most effective models were selected to predict hospital overspending with much well accuracy. Additionally, we utilized SHAP to interpret and visualize the contribution of each DPI to the target risk status. To support regional and hospital administrators, we developed a HIMS that facilitates the monitoring of health insurance overspending and enables timely adjustments. An overview of our IORP framework is illustrated in Figure 1.

Data collection and preprocessing

We obtained medical records of discharged patients, complete with health insurance information (N=549,910), from the Wuxi Health Commission. These records cover the period from January 2022 to November 2023. We included a total of 12 variables, categorized into three groups: management information (n=2), treatment behavior (n=5), and hospitalization costs (n=5), as detailed in Table S3. The dataset had a 7.82% missing data rate, and among the missing data, a large number of the samples were missing consumable costs. Since consumable costs can vary greatly between departments and the overall impact is small due to the small missing ratio, we decided to exclude 42,993 patient records with missing features to maintain data integrity (Figure 2). After excluding these records, we retained 506,917 samples for analysis. We aggregated the discharged patients' records to create department-level datasets featuring 8 DPIs. The regional datasets were compiled on a monthly basis (N=8,416) to assist regional administrators in tracking the overspending risk status of departments. For hospital-level overspending risk predictions, the dataset was generated cumulatively on a daily basis (N=44,017). Among the 8 DPIs, 5 pertained to treatment behavior and 3 were associated with hospitalization costs (see Table 1). We categorized departments into 7 groups: tumors, burns,

general medicine, integrated sections, surgery, obstetrics and gynecology, and severe illnesses. To assess overspending risk within each group, we utilized SPC to define different risk levels:

•No Risk: An overspending amount less than zero signified that the department was within budget and operating at a surplus, uniformly classified as no risk.

•Low Risk: Overspending below the centerline was classified as low risk, indicating that while budget limits were exceeded, the deviations remained within acceptable limits.

•High Risk: Departments with overspending limits greater than zero had their centerline (mean) calculated. Any overspending above this centerline was classified as high risk, indicating a significant deviation from expected expenditure patterns that necessitates immediate attention and corrective action.

Statistical methods

A descriptive analysis of DPIs was presented in Table 1, where all DPIs were summarized by mean and standard deviation (SD). The Student's t-test was used to compare the groups (no risk vs low risk/ low risk vs high risk/ no risk vs high risk). The statistical methods were handled with the Python scipy package (v1.7.3). Normality was assessed with the Shapiro-Wilk test (all $p > 0.05$), homogeneity of variance with Levene's test (all $p > 0.05$). Full results are provided in Supplementary Table S4.

Univariate analysis was conducted on the training set to evaluate the association between each DPI and the target overspending risks (Python sklearn package (v1.0.2))), considering the DPI with a P-value less than 0.05 to have a significant difference with the label and thus included in modeling. Subsequently, we performed a pairwise Spearman's rank order correlation analysis (Python scipy package (v1.7.3)) on all DPIs. Redundancy was examined for features with coefficients greater than 0.70. Expert opinion and prediction effectiveness were taken into account when selecting DPIs for IORP modeling.

IORP modelling

We randomly divided the total dataset into training and test sets (7:3), followed by scaling the data using MinMaxScaler (Python sklearn package (v1.0.2)). For the overspending modeling we chose four machine learning algorithms, including LR, RF, LightGBM and ANN. Logistic regression maps feature combinations to probabilities using a Sigmoid function for classification[21]. Random Forest enhances generalization by aggregating multiple decision

trees[22]. LightGBM optimizes GBDT with techniques like histogram algorithms and feature bundling for efficiency. ANN, inspired by biological neurons, learns complex patterns through layered computations [23]. We used 5-fold cross-validation in the training set. During each iteration, 4 parts (80% of data) were used for training, and 1 part (20%) for validation. During the cross validation procedure, hyperparameters were optimized using the Optuna framework (tree-structured Parzen estimator (TPE) optimization) (Python Optuna package (v3.0.4)) by maximizing model accuracy. For the RF model, we optimized the number of trees (n_estimators), with the optimal values found to be 181 at the regional modeling and 75 at the hospital modeling. For the LightGBM model, we focused on the max_depth parameter, which controls the maximum depth of the tree, with optimal values of 254 at the regional modeling and 145 at the hospital modeling. Because class imbalance was moderate, no class weighting or resampling was applied. Tree-based ensembles such as LightGBM are robust to moderate imbalance, per-class metrics were used to monitor performance. The hyperparameter tuning ranges for the different algorithms and the corresponding optimal hyperparameter combinations are summarized in Table 2.

Model analysis

Multiple performance metrics were employed to assess the predictive performance of the developed classification models. These metrics included accuracy[24], recall, precision, and F1-score [25] (Python sklearn package (v1.0.2)), the calculation formula was summarized in Table S1. We used SHAP algorithm to determine and visualize the importance of each DPI and its contribution to the prediction [26, 27]. In our study, the best-performing model was explored by examining the importance of each DPI to the high risk overspending, and we found four key indicators that demonstrate a strong correlation with departmental overspending: TADP, AIS, MEP, and CEP. Individualized feature importance plots were created using the test data. Python scikit-learn (v1.0.2) and Shap (v0.41.0) libraries were used for data analysis and visualization.

Ethical statement

The study received approval from The medical ethics committee review board of the Affiliated Hospital of Jiangnan University in 2022 (No. LS2022110), and informed consent was deemed unnecessary.

RESULTS

The characteristics of the study population and departmental datasets

The 12 variables from discharged patients' medical records with respect to the overspending prediction were presented in Table S2. They were mainly comprised of management information (16.7%), treatment behavior (41.7%), and hospitalization costs (41.7%). We removed 42,993 patients (7.82%) with missing data. Comparisons between different risk factor categories were conducted via the Student's t-test (a significant level of 0.05). Numerical variables were presented as mean (SD), and number (percentage) for categorical variables (Table 3).

The selection of DPIs

According to the univariate analysis, all 8 DPIs were statistically significant, with all P-values less than 0.001. As shown in Figure S1B, the hospital-level Spearman correlation analysis showed that the IVSP was highly correlated with the CEP, with a correlation coefficient of 0.75. Since CEP directly reflects the proportion of hospital consumables expenditure, it is a key indicator to measure the risk of overspending in many departments and occupies an important position in the total hospital expenditure. Therefore, we decided to exclude IVSP to help alleviate multicollinearity. As shown in Table S3, the Student's t-test (P- value > 0.05) between training and test sets demonstrated the validity of modeling with the selected DPIs.

To further assess the robustness of the identified differences, we conducted a statistical power analysis for both regional- and hospital-level indicators. The results confirmed adequate power for all pairwise comparisons, supporting the validity of our findings. Detailed results are provided in Supplementary Tables S5 and S6.

Regional-level and hospital-level prediction results

The performance of different algorithms in predicting overspending status were shown in Table 4. At the regional level, LightGBM and RF outweighed other models with the accuracy, weighted precision, recall, and F1-score over 0.70. The F1-scores of LightGBM and RF were 0.78 and 0.72, respectively. The accuracy, weighted precision, recall and AUC-ROC (area under the ROC curve) of LightGBM were 0.82, 0.78, 0.78 and 0.91.

According to the evaluation results of the regional models, LR and ANN perform poorly with accuracy, weighted precision and recall metrics below 0.7; especially in the key metrics LR's

accuracy and weighted F1-score are 0.63 and 0.56, and ANN's accuracy and weighted F1-score are 0.69 and 0.67. RF and LightGBM, on the other hand, show higher accuracy and weighted F1-score. Given these results, we decided to exclude LR and ANN were excluded from the final hospital-level experiment because they did not meet the performance thresholds required for reliable hospital-level overspending prediction. As shown in Table 4, LightGBM achieved accuracy and weighted F1-score of 0.82, and weighted AUC-ROC of 0.94. Meanwhile, RF achieved accuracy, weighted precision, recall, F1-score of 0.74, and weighted AUC-ROC of 0.88.

For the LightGBM model, which performed the best, we report accuracy, precision, recall, F1-score, as well as AUC-ROC scores and curves (Figures 3A and 3B) for each classification (high risk, low risk, and no risk). As shown in Table 5, at the regional level, LightGBM achieved a high-risk classification accuracy of 0.85 and AUC-ROC of 0.91. At the hospital level, the model maintained excellent performance for the high-risk class, with an accuracy of 0.82 and AUC-ROC of 0.97. In addition, we present the PR-AUC curves for each classification of the LightGBM model at the regional and hospital levels in Figures 3C and 3D. At the regional and hospital levels, the high-risk class achieved PR-AUC values of 0.93 and 0.90, respectively, indicating a good precision-recall trade-off and excellent identification capability for high-risk departments. To further assess classification performance, the confusion matrices are provided in Supplementary Figure S2. Moreover, Calibration analysis (Supplementary Fig. S3) showed Brier scores were 0.06 and 0.05 for the regional and hospital models, respectively. In both cases, the calibration curves closely followed the 45° diagonal, suggesting acceptable probability calibration.

These results demonstrate that the LightGBM model can provide reliable predictions for high-risk cases, which is crucial for practical deployment in HIMS.

Discoveries from IORP modeling

As shown in SHAP summary plots (Figure 4), the vertical axis represented DPIs, while the right side of the horizontal axis indicated a positive correlation with high risk overspending, and the left side indicated a negative correlation with high risk overspending. The values of DPIs were presented in colors: red indicated larger values, while blue indicated smaller values. Figures 4A and 4B showed that the top ranking DPIs of LightGBM for high risk overspending were TADP, AIS, MEP, CEP. In these cases, higher values were positively associated with higher risk of overspending. In addition, the similar ranking of DPIs for the hospital was shown

in Figure 4. Combining the SHAP summary plots (Figures 4C and 4D) of RF, high risk overspending was also positively correlated with TADP and AIS for the region and hospital. However, MSEP had a greater impact for the region and hospital.

HIMS

We have devised HIMS based on IORP designed specifically. Due to privacy concerns, we provide a demonstration of the system's functionalities using partial test data in this context (http://prediction.overspending.risk.zxstech.com/). As shown in Figure 5A, regional administrators have the access to select specific hospitals and their departments, accessing the latest monthly prediction of overspending risk. HIMS presented the individual interpretative analysis, the relevant explanations of DPIs (left bottom) and historical overspending amount (right bottom). The daily overspending risk was also provided for hospital monitoring, allowing administrators to adjust the budget and prevent overspending in time (Figure 5B).

DISCUSSION

Principal findings

In the field of medical cost control, people were extremely concerned about patient-level interventions for high expenditure [28-36]. By leveraging patient and departmental performance data, traditional statistical methods (linear regression and significance analysis) identified factors related to high-expenditure department, primarily attributing to complex patient cases, escalated drug expenditures, increased patient numbers, inpatient services and prolonged hospital stays [5, 9, 13, 37]. However, there was a lack of study on discovering overspending risk factors for specific department within regional and hospital contexts. Our study validated some factors for high expenditures of health insurance, while also identifying other factors such as consumables expenses percentage and total surgery percentage that impact the balance of medical insurance expenditure. In light of these findings, our IORP modeling tested whether the 8 DPIs presented from patient medical records could facilitate the prediction of departmental overspending risks. For facilitating the classification of risk factors, the tool of SPC have been applied in the modeling process. To the best of our knowledge, machine learning approaches have not been used for that purpose. Here, we utilized four machine-learning algorithms (i.e., LR, RF, LightGBM, and ANN) and constructed the models for regional and hospital- level overspending prediction. The LightGBM achieved the F1-score of 0.78 and 0.82 for both regional and hospital-level overspending prediction, which illustrated

the medical records data contain information that can be used to better predict departmental overspending status. In addition, we developed the overspending risk system that provided risk predictions for regional and hospital administrators, with the function of monitoring departmental health insurance overspending risk. Considering the concerns of hospital departmental management, overspending on health insurance can impact the quality of healthcare services. According to the IORP, drug and consumable expenses were identified as prominent predictors and therefore require stringent control within hospital management practices. AIS emerged as another significant predictor of high-cost overspending, which indicated that reducing patients' hospital stays could potentially alleviate the risk of excessive spending. Several strategies can be implemented for hospital departmental managers. For instance, efforts to reduce postoperative infections and advance medical technology to accelerate patient recovery, along with enhancements to hospital management protocols regarding patient waiting times. In addition, we found TADP was a key determinant of high-risk overspending and proposed actionable interventions. These included optimizing bed scheduling, introducing night-time procedures, and improving diagnostic appointment systems to alleviate resource constraints. By implementing these measures, hospital administrators could proactively manage departmental expenditures, enhance financial oversight, and optimize resource allocation.

This work explores an enhancing prediction method of health insurance overspending risk through designed hospital departmental performance indicators. Our proposed insurance overspending risk prediction models and integrated health information medical system demonstrate good adaptability. The data type and format input into the integrated system are broadly applicable. However, our study has several limitations. When applied to different hospital or cities, the model's hyperparameters may differ with the change of population samples. Therefore, the specific application in each region may require appropriate adjustments and optimizations based on the data characteristics and task requirements. Moreover, the presented DPIs used in our study reflected primary two aspects of healthcare administration, namely the quality of care (TADP, CCP, TSP, IVSP, AIS) and operational efficiency (MEP, CEP, MSEP). In future, our studies could further enhance generalizability by incorporating socioeconomic variables (e.g., insurance coverage rates, rural/urban disparities) and automating DPI adjustments via federated learning techniques.

CONCLUSION

In this paper, we developed IORP models utilizing DPIs tailored for regional and hospital administrators. Our process began with the collection of 549,910 discharged patient medical records from January 2022 to November 2023 in Wuxi, China. These records were organized into regional and hospital-level departmental datasets, comprising 8,416 and 44,017 records, respectively. To analyze departmental overspending, we employed SPC to categorize the data into three risk groups of high risk, low risk, and no risk. Subsequently, we built regional and hospital-level IORP models with machine learning methods of LR, LightGBM, RF, and ANN. Our experimental results indicated that the LightGBM algorithm demonstrated exceptional predictive capabilities, achieving accuracies of 0.82 for both regional and hospital-level models. To further enhance our analysis, we utilized SHAP to assess the importance of each DPI. This analysis highlighted four critical indicators strongly associated with departmental overspending: TADP, AIS, MEP, and CEP.

Finally, we integrated the IORP models into the HIMS at the Affiliated Hospital of Jiangnan University. Using the steps outlined in Figure 6, administrators can monitor health insurance overspending. This integration significantly enhances the departmental administrators' ability to predict overspending risks, facilitating timely and accurate risk assessments. By optimizing departmental performance, this model supports the sustainable management of healthcare expenditures, ultimately contributing to better financial health within healthcare institutions. As a result, this system proved instrumental in significantly reducing overall hospital expenses in just one year of 2023 and 2024, based on the same department conditions: per capita medical costs decreased by 6.28%, per capita drug expenditures dropped by 12.18%, and per capita consumables costs were reduced by 14.1%. By the application of the system, it has enabled regional and hospital departmental managers to optimize fiscal resources, resulting in enhanced financial management capabilities and more sustainable budgetary control across hospital departments.

Funding: The study has been funded by the National Key R&D Program of China (NO. 2021YFC0122701), the Scientific Research Program of Wuxi Health Commission (NO. Z202309), and the Taihu Light Technology Research (Medical and Health; NO. Y20232001).

Conflicts of interest: Authors declare no conflicts of interest.

Data availability: The datasets utilized and/or analyzed in this study can be obtained from the primary corresponding author upon reasonable request.

Submitted: 13 January 2025

Accepted: 24 May 2025

Published online: 28 June 2025

REFERENCES

[1] G. Yevgeniy, T.S. P, C. Sébastien, L.M. Aliénor, C. Michele, F.A. B, V. Bruno, Assessing the future medical cost burden for the European health systems under alternative exposure-to-risks scenarios, PloS one 15(9) (2020) e0238565-e0238565.
[2] J.F. Lu, W.C. Hsiao, Does universal health insurance make health care unaffordable? Lessons from Taiwan, Health Aff (Millwood) 22(3) (2003) 77-88.
[3] J.H. Hung, L. Chang, Has cost containment after the National Health Insurance system been successful? Determinants of Taiwan hospital costs, Health Policy 85(3) (2008) 321-35.
[4] Q. Meng, H. Fang, X. Liu, B. Yuan, J. Xu, Consolidating the social health insurance schemes in China: towards an equitable and efficient health system, Lancet 386(10002) (2015) 1484-92.
[5] G.F. Anderson, J. Hurst, P.S. Hussey, M. Jee-Hughes, Health spending and outcomes: trends in OECD countries, 1960-1998, Health Aff (Millwood) 19(3) (2000) 150-7.
[6] P. Drewe-Boss, D. Enders, J. Walker, U. Ohler, Deep learning for prediction of population health costs, BMC Med Inform Decis Mak 22(1) (2022) 32.
[7] Y.C. Huang, S.J. Li, M. Chen, T.S. Lee, The Prediction Model of Medical Expenditure Appling Machine Learning Algorithm in CABG Patients, Healthcare (Basel) 9(6) (2021).
[8] K. Kaushik, A. Bhardwaj, A.D. Dwivedi, R. Singh, Machine Learning-Based Regression Framework to Predict Health Insurance Premiums, Int J Environ Res Public Health 19(13) (2022).
[9] M. Laudicella, K.R. Olsen, A. Street, Examining cost variation across hospital departments--a two-stage multi-level approach using patient-level data, Soc Sci Med 71(10) (2010) 1872-81.
[10] Z. Mitkova, M. Dimitrova, M. Doneva, K. Tachkov, M. Kamusheva, L. Marinov, N. Gerasimov, D. Tcharaktchiev, G. Petrova, Budget cap and pay-back model to control spending on medicines: A case study of Bulgaria, Front Public Health 10 (2022) 1011928.
[11] Y. Murakami, T. Okamura, K. Nakamura, K. Miura, H. Ueshima, The clustering of cardiovascular disease risk factors and their impacts on annual medical expenditure in Japan: community-based cost analysis using Gamma regression models, BMJ Open 3(3) (2013).
[12] I. Papanicolas, L.R. Woskie, A.K. Jha, Health Care Spending in the United States and Other High-Income Countries, JAMA 319(10) (2018) 1024-1039.
[13] P. Raeissi, F.E. Fard Azar, A. Rezapour, M. Afrouzi, S.S. Gholami, N. Niknam, Cost analysis based on performance indicators during Healthcare Reform Plan in selected educational hospitals, J Educ Health Promot 8 (2019) 206.
[14] N. Ye, The Influence of Demographic Factors on Urban Medical Insurance Expenditure under Big Data, Mathematical Problems in Engineering 2022 (2022).
[15] G.Z. Zhang, XY ; Bilal, M ; Dou, WC ; Xu, XL ; Rodrigues, JJPC, Identifying fraud in medical insurance based on blockchain and deep learning, Future Generation Computer Systems-The international journal of escience 130 (2022) 140-154.
[16] D. Atkins, C.A. Makridis, G. Alterovitz, R. Ramoni, C. Clancy, Developing and Implementing Predictive Models in a Learning Healthcare System: Traditional and Artificial Intelligence Approaches in the Veterans Health Administration, Annual Review of Biomedical Data Science 5 (2022) 393-413.
[17] H. Ghaith, B. Sigurd, A. Christopher, P. Thomas, M. Thomas, Real-World Implementation of Artificial Intelligence/Machine Learning for Managing Surgical Spine Patients at 2 Academic Health Care Systems, International journal of spine surgery 17(S1) (2023).
[18] J. Stuart, Y. Maha, L.C. Xi, The Agile Deployment of Machine Learning Models in Healthcare, Frontiers in Big Data 1 (2019) 7.

[19] P.Y. Wang, J. Li, Implementation of Real-Time Medical and Health Data Mining System Based on Machine Learning, Journal of Healthcare Engineering 2021 (2021).
[20] N. Yuvaraj, K.R. SriPreethaa, Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster, Cluster Computing 22(1s) (2019) 1-9.
[21] C.J. Biesheuvel, Y. Vergouwe, E.W. Steyerberg, D.E. Grobbee, K.G.M. Moons, Polytomous logistic regression analysis could be applied more often in diagnostic research, Journal of Clinical Epidemiology 61(2) (2008) 125-134.
[22] J.L. Speiser, M.E. Miller, J. Tooze, E. Ip, A comparison of random forest variable selection methods for classification prediction modeling, Expert Systems with Applications 134 (2019) 93-101.
[23] C. Zhang, Y. Guo, M. Li, Review of Development and Application of Artificial Neural Network Models, Computer Engineering and Application 57(11) (2021) 57-69.
[24] P. Eusebi, Diagnostic Accuracy Measures, Cerebrovascular diseases 36(4) (2013) 267-272.
[25] C. Meaney, T.A. Stukel, P.C. Austin, R. Moineddin, M. Greiver, M. Escobar, Quality indices for topic model selection and evaluation: a literature review and case study, BMC Med Inform Decis Mak 23(1) (2023) 132.
[26] Y. Huang, X. Wang, Y. Cao, M. Li, L. Li, H. Chen, S. Tang, X. Lan, F. Jiang, J. Zhang, Multiparametric MRI model to predict molecular subtypes of breast cancer using Shapley additive explanations interpretability analysis, Diagn Interv Imaging (2024).
[27] Y. Song, D. Zhang, Q. Wang, Y. Liu, K. Chen, J. Sun, L. Shi, B. Li, X. Yang, W. Mi, J. Cao, Prediction models for postoperative delirium in elderly patients with machine-learning algorithms and SHapley Additive exPlanations, Transl Psychiatry 14(1) (2024) 57.
[28] A.R. Barker, K.E. Joynt Maddox, E. Peters, K. Huang, M.C. Politi, Predicting Future Utilization Using Self-Reported Health and Health Conditions in a Longitudinal Cohort Study: Implications for Health Insurance Decision Support, Inquiry 58 (2021) 469580211064118.
[29] A.M. Jödicke, U. Zellweger, I.T. Tomka, T. Neuer, I. Curkovic, M. Roos, G.A. Kullak-Ublick, H. Sargsyan, M. Egbring, Prediction of health care expenditure increase: how does pharmacotherapy contribute?, BMC Health Serv Res 19(1) (2019) 953.
[30] Y.J. Kim, H. Park, Improving Prediction of High-Cost Health Care Users with Medical Check-Up Data, Big Data 7(3) (2019) 163-175.
[31] B. Langenberger, T. Schulte, O. Groene, The application of machine learning to predict high-cost patients: A performance-comparison of different models using healthcare claims data, PLoS One 18(1) (2023) e0279540.
[32] J.C. Lauffenburger, M. Mahesri, N.K. Choudhry, Not there yet: using data-driven methods to predict who becomes costly among low-cost patients with type 2 diabetes, BMC Endocr Disord 20(1) (2020) 125.
[33] J.L. Li, QL; Zhu, EY; Xu, Y; Zhu, D, A Study of Health Insurance Fraud in China and Recommendations for Fraud Detection and Prevention, Journal of Organizational and End User Computing 34(4) (2022).
[34] Y. Nomura, Y. Ishii, Y. Chiba, S. Suzuki, A. Suzuki, S. Suzuki, K. Morita, J. Tanabe, K. Yamakawa, Y. Ishiwata, M. Ishikawa, K. Sogabe, E. Kakuta, A. Okada, R. Otsuka, N. Hanada, Does Last Year's Cost Predict the Present Cost? An Application of Machine Leaning for the Japanese Area-Basis Public Health Insurance Database, Int J Environ Res Public Health 18(2) (2021).
[35] J. Sun, Y. Wang, Y. Zhang, L. Li, H. Li, T. Liu, L. Zhang, Research on the risk governance of fraudulent reimbursement of patient consultation fees, Front Public Health 12 (2024) 1339177.

[36] C.I. ul Hassan, J ; Hussain, S ; AlSalman, H ; Mosleh, MAA ; Ullah, SS, A Computational Intelligence Approach for Predicting Medical Insurance Cost, Mathematical Problems in Engineering 2021 (2021).
[37] G.N. Stock, C. McDermott, Operational and contextual drivers of hospital costs, J Health Organ Manag 25(2) (2011) 142-58.

TABLES AND FIGURES WITH LEGENDS

Table 1. The description of departmental performance indicators (DPIs) for departmental datasets.

Category	DPIs	Description
Treatment behavior	Total amount of discharged patients (TADP)	Total amount of discharged patients
	Critical cases percentage (CCP)	Proportion of critical patients to the total amount of discharged patients
	Total surgery percentage (TSP)	Proportion of discharged patients undergoing surgeries to the total amount of discharged patients
	IV-surgery percentage (IVSP)	Proportion of discharged patients undergoing IV-surgeries to discharged patients undergoing surgeries
	Average inpatient stay (AIS)	Average length of inpatient stay
Hospitalization costs	Medicine expenses percentage (MEP)	Proportion of medicine expenses to total expenses
	Consumables expenses percentage (CEP)	Proportion of consumables expenses to total expenses
	Medical service expenses percentage (MSEP)	The deduction of total cumulative expenses to medicine and consumables expenses

Table 2. The hyperparameter tuning range of different algorithms and the optimal hyperparameter combination for each algorithm.

Algorithm	Range of Hyperparameters	Regional modeling hyperparameters	Hospital modeling hyperparameters
LR	C:(1e-5, 100); max_iter:(100,1000); solver:{'liblinear', 'lbfgs', 'newton-cg', 'sag', 'saga'}	'C': 33.37; 'max_iter': 872; 'solver': 'liblinear'	/
RF	n_estimators:(20,200); max_depth:(2,256); min_samples_leaf:(1,64); max_samples:(0.5,1.0); criterion:{'gini', 'entropy'}; random_state:(1,100)	'n_estimators': 181; 'max_depth': 92; 'min_samples_leaf': 1; 'max_samples': 0.9; 'criterion': 'entropy'; 'random_state': 70	'n_estimators': 75; 'max_depth': 150; 'min_samples_leaf': 1; 'max_samples': 0.85; 'criterion': 'gini'; 'random_state': 14
LightGBM	n_estimators:(20,200); max_depth:(2,256); learning_rate(0.01,0.2); min_child_samples(5,100)	'n_estimators': 30; 'max_depth': 254; 'learning_rate': 0.1; 'min_child_samples': 10	'n_estimators': 200; 'max_depth': 145; 'learning_rate': 0.19; 'min_child_samples': 45
ANN	layers:(1,3); units_per_layer:(32,512);	'layers': 2; 'units_per_layer': 436;	/

	activation{'relu', 'tanh', 'sigmoid'}	'activation': 'relu'	

Table 3. Baseline characterization of departmental data.

DPIs	**Mean (SD)**				***P*-value**		
	Overall	High risk	Low risk	No risk	High vs Low	High vs No	Low vs No
Region							
TADP	81.87	92.68	60.01	86.14	< 0.001	0.02	< 0.001
CCP	0.38 (0.33)	0.35 (0.31)	0.4 (0.36)	0.44 (0.35)	< 0.001	< 0.001	< 0.001
TSP	0.78 (0.29)	0.77 (0.31)	0.77 (0.27)	0.81 (0.21)	< 0.001	< 0.001	< 0.001
IVSP	0.14 (0.22)	0.13 (0.21)	0.16 (0.24)	0.17 (0.25)	< 0.001	< 0.001	0.15
MEP	0.23 (0.12)	0.22 (0.11)	0.23 (0.11)	0.25 (0.14)	< 0.001	< 0.001	< 0.001
CEP	0.15 (0.14)	0.16 (0.14)	0.13 (0.13)	0.15 (0.16)	< 0.001	0.01	< 0.001
MSEP	0.77 (0.12)	0.78 (0.11)	0.77 (0.11)	0.75 (0.14)	< 0.001	< 0.001	< 0.001
AIS	7.87 (7.5)	5.87 (3.3)	9.34 (6.81)	12.69	< 0.001	< 0.001	< 0.001
Hospital							
TADP	54.38	62.4	33.66	61.52	< 0.001	0.29	< 0.001
CCP	0.29 (0.27)	0.37 (0.28)	0.35 (0.3)	0.25 (0.25)	< 0.001	< 0.001	< 0.001
TSP	0.76 (0.26)	0.75 (0.22)	0.74 (0.26)	0.77 (0.27)	< 0.001	< 0.001	< 0.001
MEP	0.26 (0.1)	0.28 (0.1)	0.27 (0.1)	0.26 (0.1)	< 0.001	< 0.001	< 0.001
CEP	0.16 (0.13)	0.17 (0.13)	0.17 (0.13)	0.16 (0.13)	< 0.001	< 0.001	< 0.001
MSEP	0.77 (0.12)	0.78 (0.11)	0.77 (0.11)	0.75 (0.14)	< 0.001	< 0.001	< 0.001
AIS	7.87 (7.5)	5.87 (3.3)	9.34 (6.81)	12.69(14.8)	< 0.001	< 0.001	< 0.001

Table 4. Regional and hospital performance across models.

Algorithm	Accuracy	Weighted avg.			
		Precision	Recall	F1-score	AUC-ROC
Region					
LR	0.63	0.63	0.63	0.56	0.71
RF	0.72	0.72	0.72	0.72	0.87
LightGBM	0.82	0.78	0.78	0.78	0.91
ANN	0.69	0.68	0.69	0.67	0.79
Hospital					
RF	0.74	0.74	0.74	0.74	0.88
LightGBM	0.82	0.82	0.82	0.82	0.94

Table 5. Prediction performance of the LightGBM model across region and hospital for each risk category.

	Accuracy	Precision	Recall	F1-score	AUC-ROC
Region					
No Risk	0.86	0.87	0.86	0.86	0.95
Low Risk	0.73	0.73	0.73	0.73	0.87
High Risk	0.85	0.81	0.85	0.83	0.91
Hospital					

No Risk	0.90	0.87	0.90	0.89	0.94
Low Risk	0.65	0.72	0.65	0.68	0.90
High Risk	0.82	0.80	0.82	0.81	0.97

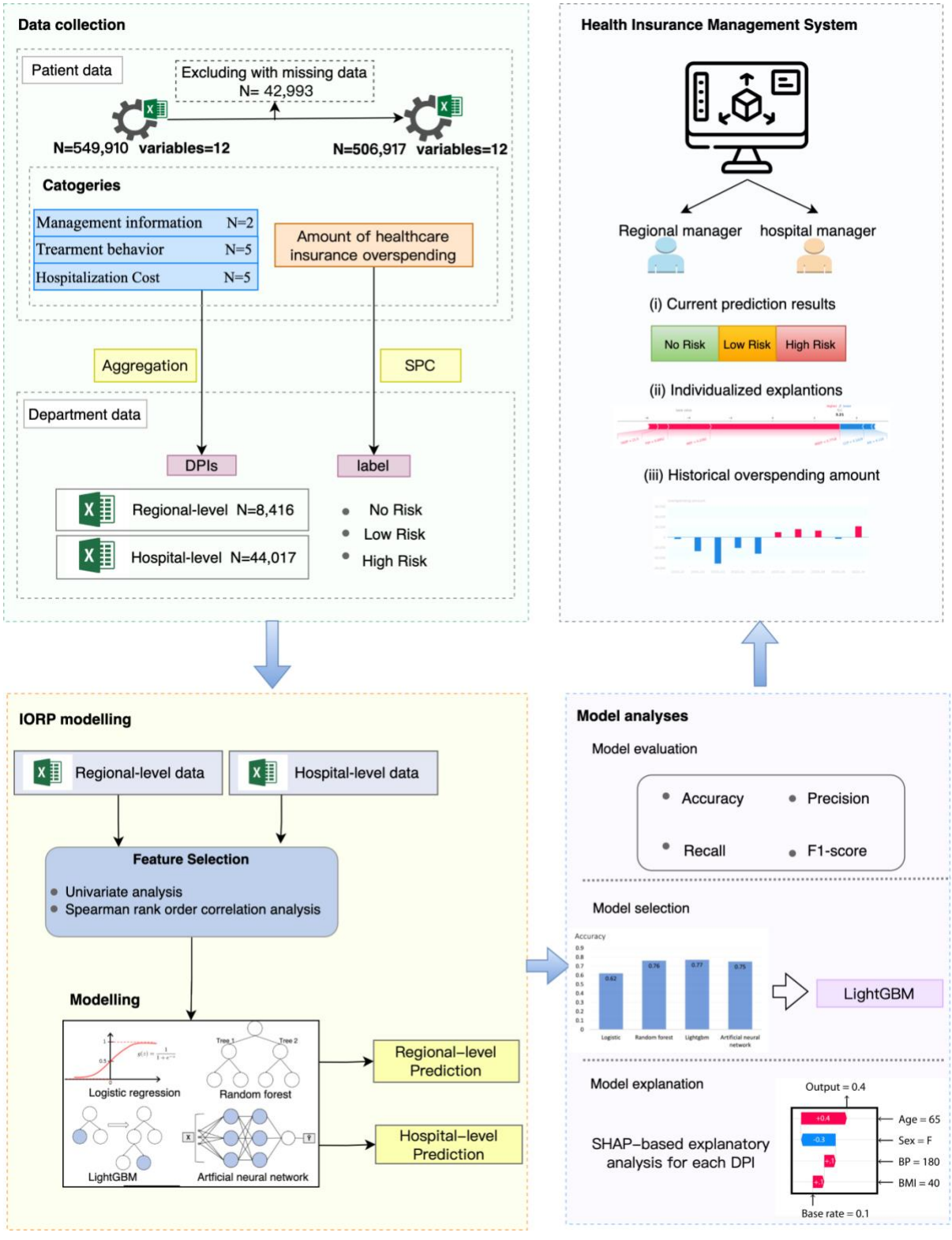


Figure 1. The overview of IORP model and analyses.

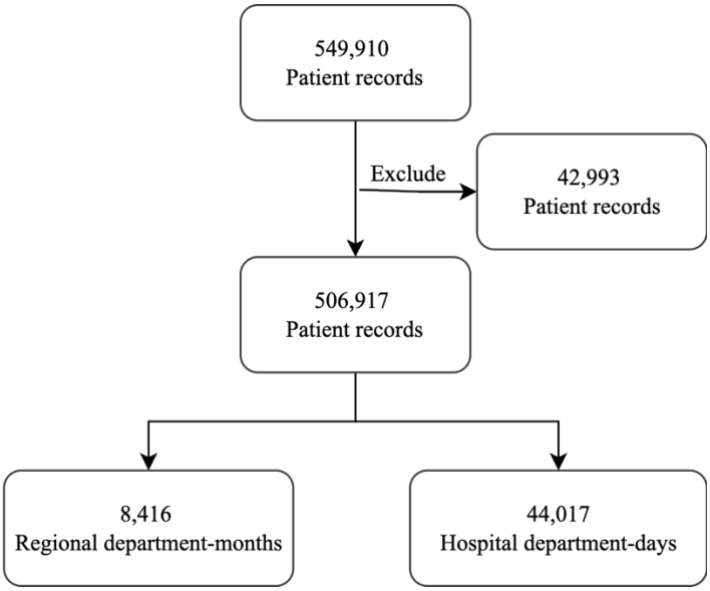


Figure 2. Data processing flowchart for patient record inclusion and department-level aggregation.

(A) (B)

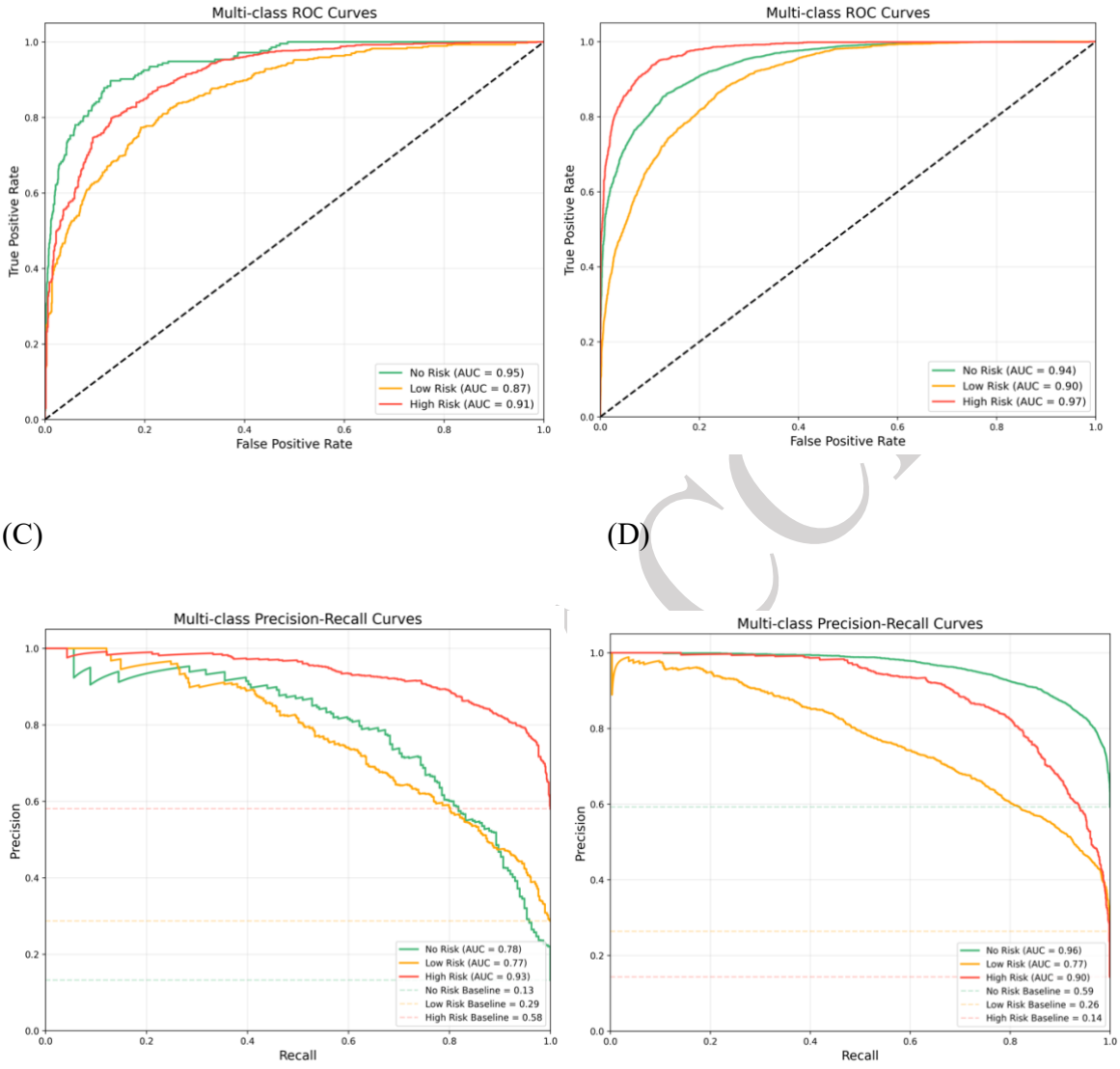


Figure 3. ROC and PR curves of the LightGBM model for regional- and hospital-level predictions. (A) Regional-level ROC curve. (B) Hospital-level ROC curve. (C) Regional-level PR curve. (D) Hospital-level PR curve.

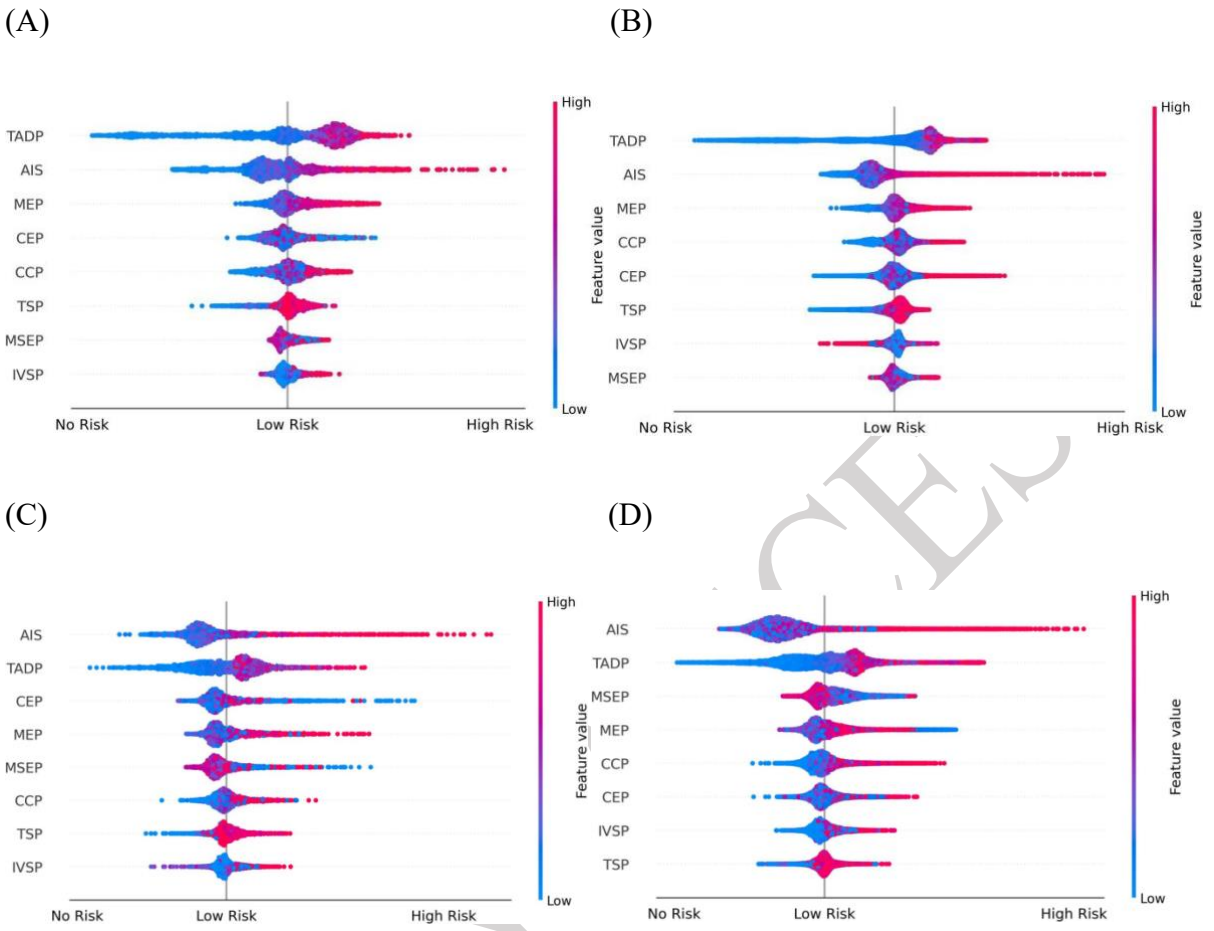
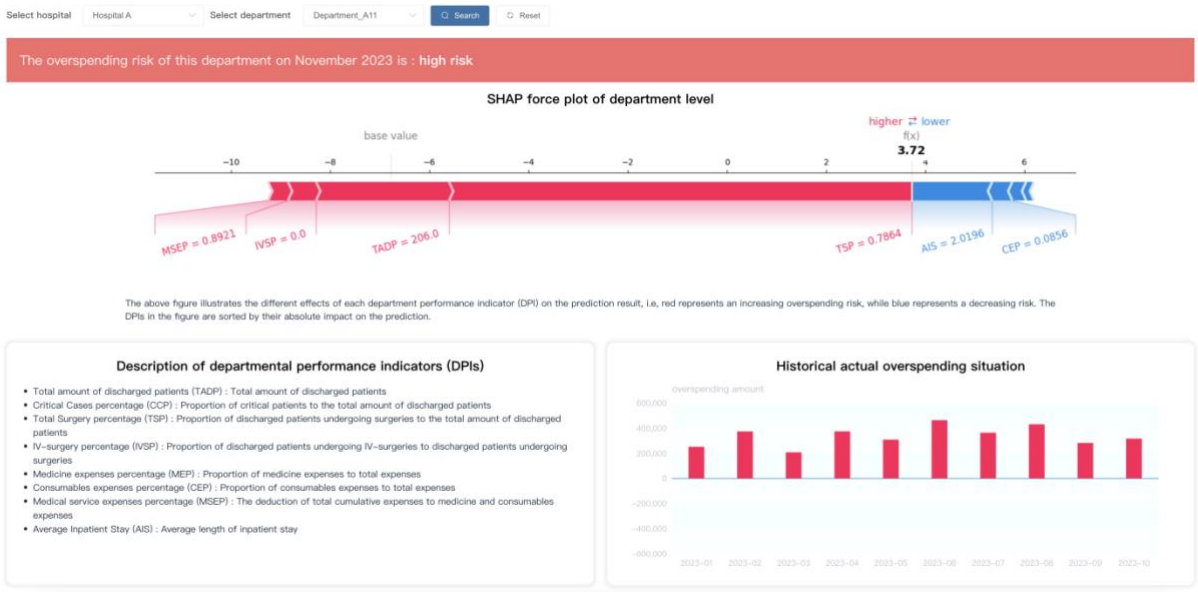


Figure 4. The SHAP summary plots for the overspending forecasting (LightGBM (A and B), RF (C and D)).

(A)



(B)

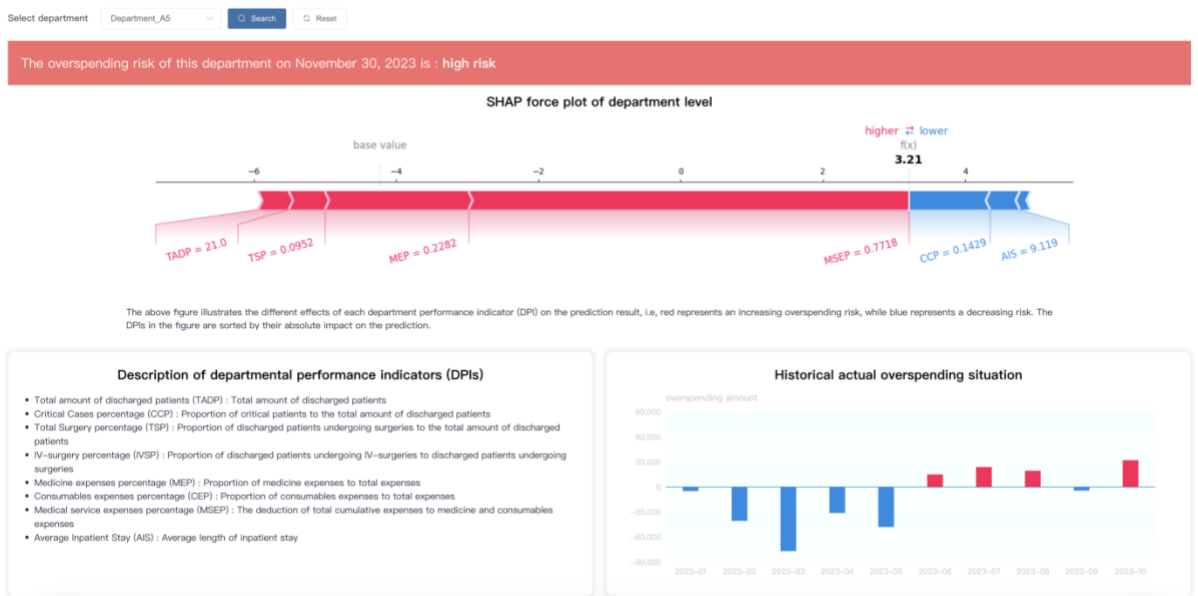


Figure 5. System User Interface for regional (A) and hospital (B) administrators.

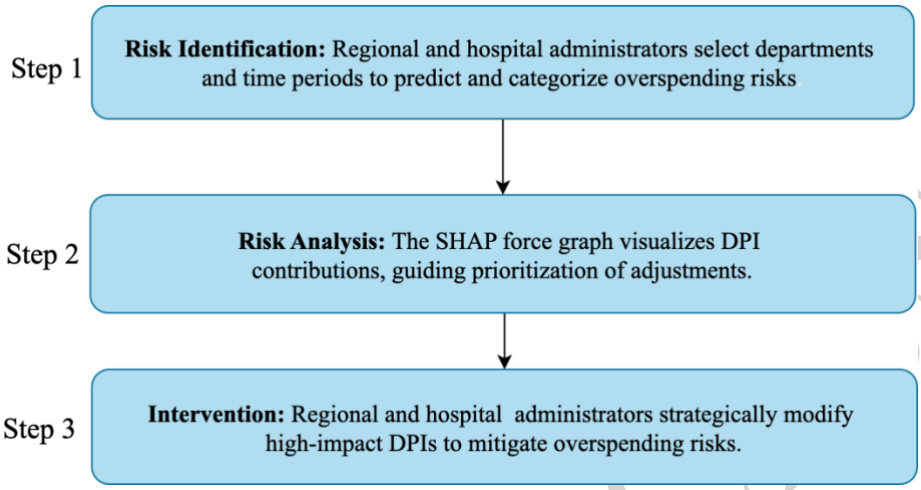


Figure 6. Practical steps actionable by regional and hospital administrators.

SUPPLEMENTAL DATA

Table S1. Calculation formula for each evaluation method.

Evaluation method	**Formula**
Accuracy	(TP+TN) / (TP+TN+FP+FN)
Precision	TP / (TP+FP)
Recall	TP / (TP+FN)
F1-score	2*Precision*Recall / (Precision+Recall)

Table S2. The characteristics of patient population in the study.

Name	**Category**	**Data type**	**Mean (std) /N (%)**	**(P25, P75)**	**Missing (%)**
Management information					
Hospital name		Categorical			0
Department name		Categorical			0
Treatment behavior					
Discharge time		Numerical			0
Critical condition	Yes No	Categorical	193259 (35.2%) 356651 (64.8%)		0

Surgical procedure	Yes No	Categorical	430147 (78.3%) 119763 (21.7%)		0
Surgical category	0 1 2 3 4	Categorical	208416 (37.9%) 78637 (14.3%) 90185 (16.4%) 113281 (20.6%) 59391 (10.8%)		0
Inpatient stay		Numerical	9.51 (12.11)	(4.00, 10.00)	0
Hospitalization costs					
Medicine expenses (western medicine, traditional Chinese medicine)		Numerical	3752.37 (10989.56)	(614.58, 3937.84)	0.04
Consumables expenses (examination, surgery)		Numerical	3696.46 (14511.98)	(53.64, 1619.01)	0.07
Medical services expenses		Numerical	8171.33 (1463.95)	(5718.96, 11621.34)	0
Health insurance overspending amount		Numerical	-1473.30 (16863.80)	(-3609.64, 3172.37)	0

Total expenses		Numerical	15620.16 (24037.58)	(6387.18, 17178.18)	0.01

Table S3. Comparison between the training and test sets in the regional and hospital datasets.

DPIs	**Region**			**Hospital**		
	Mean (Std)		***P*_value**	**Mean (Std)**		***P*_value**
	Training	**Test**		**Training**	**Test**	
TADP	0.1 (0.09)	0.1 (0.1)	0.29	0.13 (0.13)	0.14 (0.14)	0.24
CCP	0.38 (0.33)	0.38 (0.33)	0.57	0.3 (0.27)	0.29 (0.27)	0.50
TSP	0.78 (0.29)	0.77 (0.29)	0.41	0.76 (0.26)	0.76 (0.26)	0.22
IVSP	0.14 (0.22)	0.15 (0.23)	0.10	/	/	/
MEP	0.26 (0.14)	0.3 (0.15)	0.23	0.3 (0.12)	0.31 (0.12)	0.65
CEP	0.18 (0.17)	0.18 (0.17)	0.38	0.18 (0.15)	0.19 (0.16)	0.53
MSEP	0.77 (0.12)	0.77 (0.12)	0.23	0.74 (0.1)	0.69 (0.12)	0.17
AIS	0.05 (0.05)	0.04 (0.04)	0.35	0.06 (0.04)	0.04 (0.03)	0.31

Table S4. Shapiro-Wilk and Levene's test P-values for assessing normality and homogeneity of variance across feature groups.

	Shapiro-Wilk (Levene's test)

Feature	High vs Low	High vs No	Low vs No
			Region
TADP	0.24 (0.13)	0.22 (0.30)	0.3 (0.25)
CCP	0.18 (0.25)	0.27 (0.41)	0.2 (0.37)
TSP	0.33 (0.08)	0.19 (0.15)	0.26 (0.17)
IVSP	0.08 (0.19)	0.48 (0.38)	0.34 (0.09)
MEP	0.21 (0.23)	0.23 (0.63)	0.29 (0.10)
CEP	0.49 (0.44)	0.07 (0.18)	0.67 (0.55)
AIS	0.25 (0.07)	0.19 (0.22)	0.42 (0.57)
MSEP	0.31 (0.67)	0.26 (0.12)	0.32 (0.23)
			Hospital
TADP	0.12 (0.15)	0.17 (0.37)	0.26 (0.54)
CCP	0.21 (0.28)	0.25 (0.51)	0.19 (0.14)
TSP	0.38 (0.36)	0.08 (0.16)	0.42 (0.25)
MEP	0.24 (0.09)	0.29 (0.07)	0.37 (0.31)
CEP	0.59 (0.41)	0.38 (0.43)	0.3 (0.27)
AIS	0.31 (0.35)	0.3 (0.33)	0.32 (0.12)
MSEP	0.27 (0.26)	0.25 (0.29)	0.28 (0.17)

Table S5. Statistical Power Analysis of Regional DPIs.

DPI	High-risk vs low-risk	High-risk vs no-risk	Low-risk vs no-risk
TADP	>0.99	0.66	>0.99
CCP	0.84	>0.99	>0.99
TSP	0.98	0.97	0.85
IVSP	0.93	>0.99	>0.99
MEP	>0.99	>0.99	>0.99
CEP	>0.99	0.78	>0.99
AIS	>0.99	>0.99	>0.99
MSEP	>0.99	>0.99	0.98

Table S6. Statistical Power Analysis of Hospital DPIs.

DPI	High-risk vs low-risk	High-risk vs no-risk	Low-risk vs no-risk

TADP	>0.99	0.78	>0.99
CCP	>0.99	>0.99	>0.99
TSP	0.98	0.90	>0.99
MEP	>0.99	>0.99	>0.99
CEP	0.86	>0.99	>0.99
AIS	>0.99	>0.99	>0.99
MSEP	>0.99	>0.99	>0.99

(A)


TADP 1 -0.095 0.03 -0.069 -0.058 0.037 0.064 -0.29
CCP -0.095 1 -0.13 0.14 0.25 0.19 -0.24 0.34
TSP 0.03 -0.13 1 0.11 -0.023 0.19 0.031 0.061
IVSP -0.069 0.14 0.11 1 -0.085 0.53 0.081 0.2
MEP -0.058 0.25 -0.023 -0.085 1 -0.2 -0.97 0.22
CEP 0.037 0.19 0.19 0.53 -0.2 1 0.2 0.046
MSEP 0.064 -0.24 0.031 0.081 -0.97 0.2 1 -0.22
AIS -0.29 0.34 0.061 0.2 0.22 0.046 -0.22 1
TADP CCP TSP IVSP MEP CEP MSEP AIS
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

(B)


TADP 1 -0.076 0.15 -0.13 0.11 -0.085 -0.1 -0.36
CCP -0.076 1 -0.073 0.17 0.039 0.069 -0.039 0.26
TSP 0.15 -0.073 1 0.2 -0.044 0.27 0.045 -0.12
IVSP -0.13 0.17 0.2 1 -0.3 0.75 0.3 0.37
MEP 0.11 0.039 -0.044 -0.3 1 -0.39 -1 0.083
CEP -0.085 0.069 0.27 0.75 -0.39 1 0.39 0.24
MSEP -0.1 -0.039 0.045 0.3 -1 0.39 1 -0.082
AIS -0.36 0.26 -0.12 0.37 0.083 0.24 -0.082 1
TADP CCP TSP IVSP MEP CEP MSEP AIS
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

Figure S1. Spearman correlation analysis of the regional-level and hospital-level DPIs.

(A) (B)

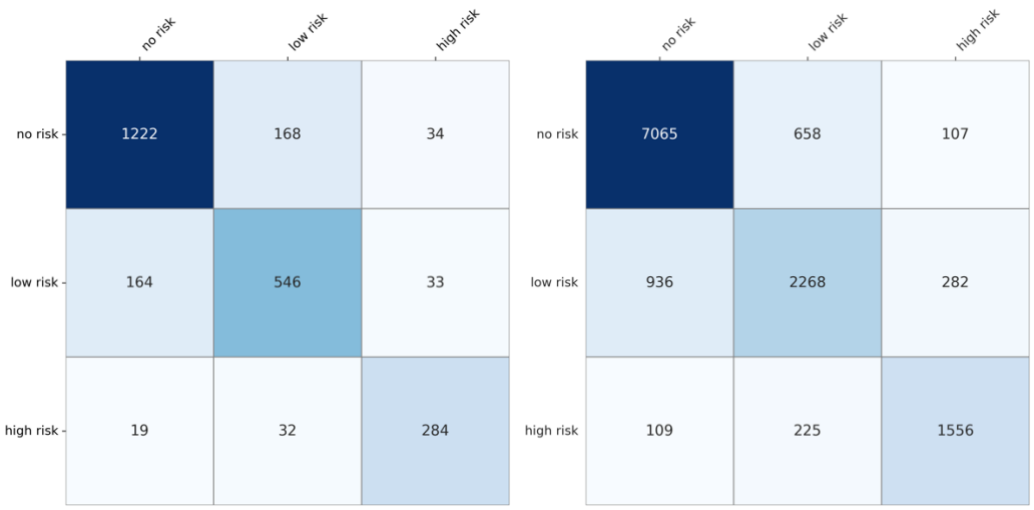


Figure S2. Confusion matrices for LightGBM models at regional (A) and hospital (B) levels.

(A) (B)

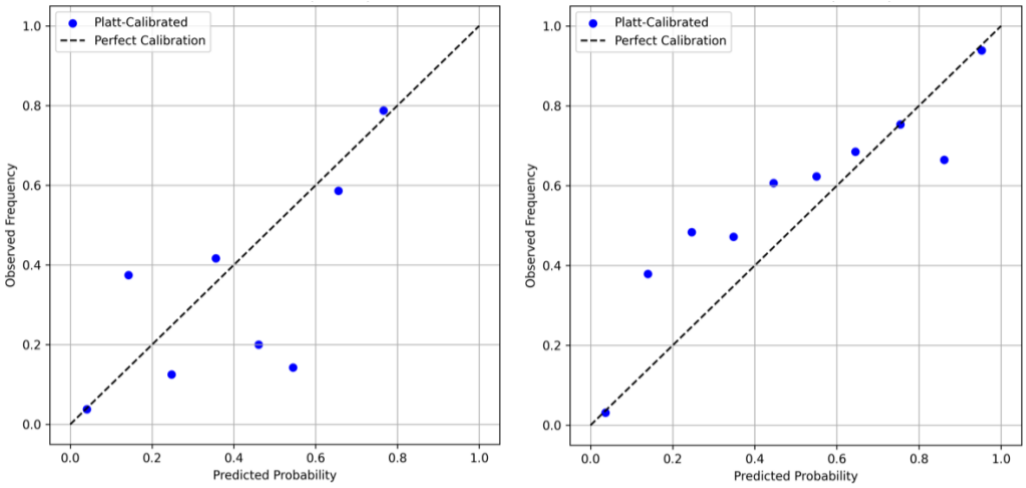


Figure S3. Calibration curvs for high-risk class predictions in regional and hospital LightGBM models.