

The BiomolBiomed publishes an "Advanced Online" manuscript format as a free service to authors in order to expedite the dissemination of scientific findings to the research community as soon as possible after acceptance following peer review and corresponding modification (where appropriate). An "Advanced Online" manuscript is published online prior to copyediting, formatting for publication and author proofreading, but is nonetheless fully citable through its Digital Object Identifier (doi®). Nevertheless, this "Advanced Online" version is NOT the final version of the manuscript. When the final version of this paper is published within a definitive issue of the journal with copyediting, full pagination, etc., the new final version will be accessible through the same doi and this "Advanced Online" version of the paper will disappear.

SUPPLEMENTAL DATA

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 Fan³, Jiongying Li⁴, Lei Hua², Lin Zhang^{5,6}, Wenjun Ma⁷, Liwen He⁸, Hao Zang⁹, Haijun Zhang¹⁰, Xingyu Liu¹¹, Yufeng Gao¹², Li Liu^{1,2*}

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

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

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

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

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

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

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

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

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

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

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

¹²Affiliated 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>

Full article is available at the following link:

<https://www.bjbms.org/ojs/index.php/bjbms/article/view/12051>

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	Categorical	208416 (37.9%) 78637 (14.3%) 90185 (16.4%)		0

	3		113281 (20.6%)		
	4		59391 (10.8%)		
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)		<i>P</i> _value	Mean (Std)		<i>P</i> _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.

Feature	Shapiro-Wilk (Levene's test)		
	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.

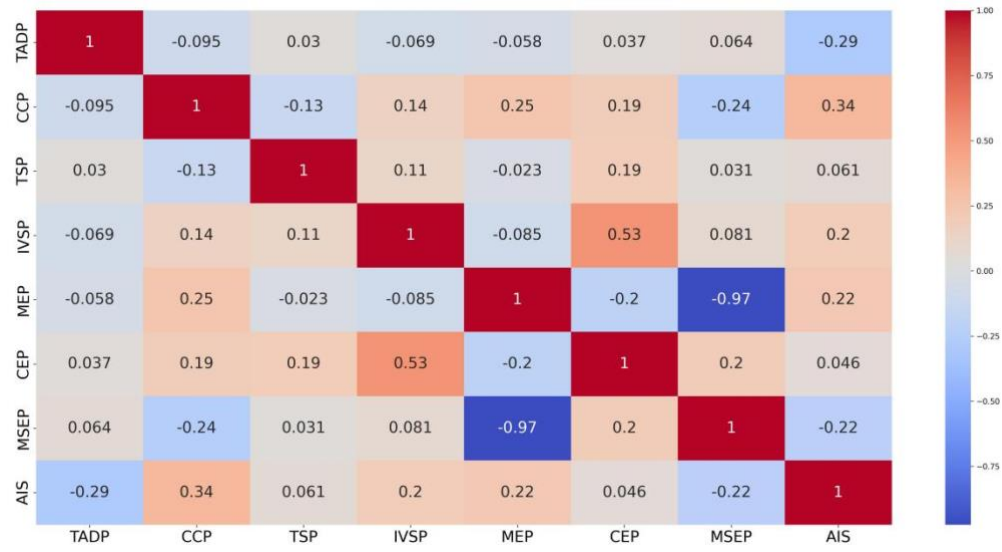
DPI	High-risk vs low-risk	High-risk vs no-risk	Low-risk vs no-risk
TADP	>0.99	0.66	>0.99
CCP	0.84	>0.99	>0.99
TSP	0.98	0.97	0.85
IVSP	0.93	>0.99	>0.99
MEP	>0.99	>0.99	>0.99
CEP	>0.99	0.78	>0.99
AIS	>0.99	>0.99	>0.99
MSEP	>0.99	>0.99	0.98

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)



(B)

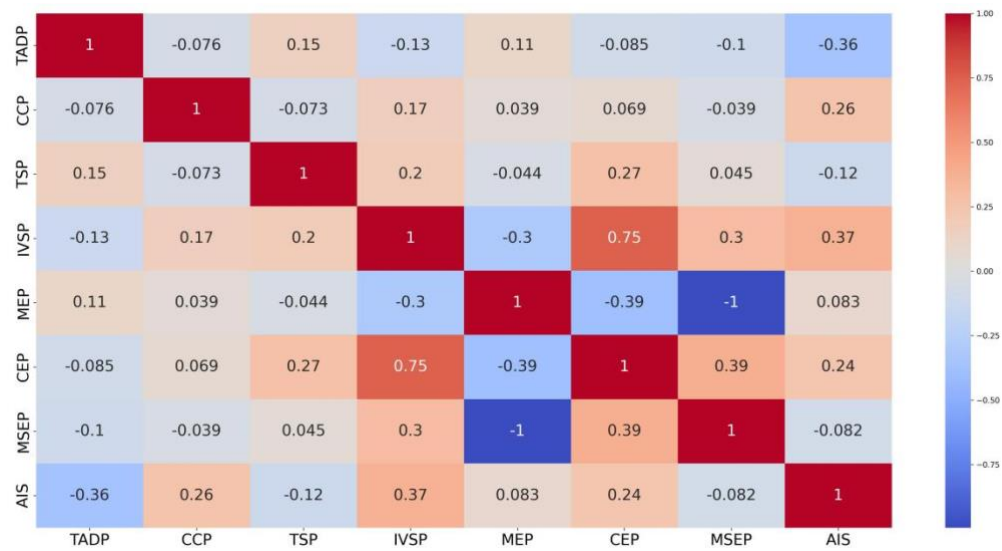


Figure S1. Spearman correlation analysis of the regional-level and hospital-level DPIs.

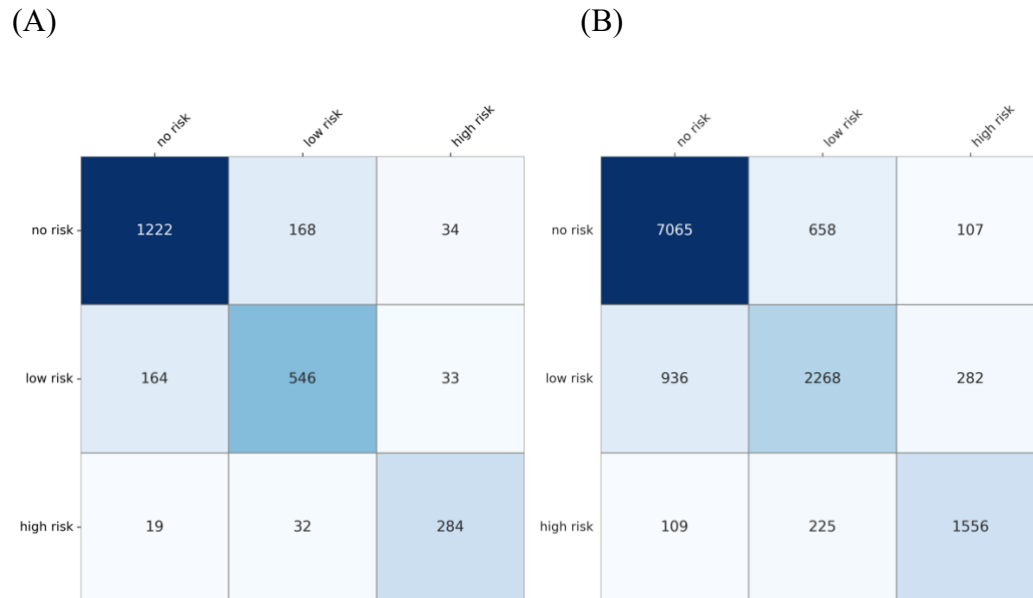


Figure S2. Confusion matrices for LightGBM models at regional (A) and hospital (B) levels.

EARLY ACCESS

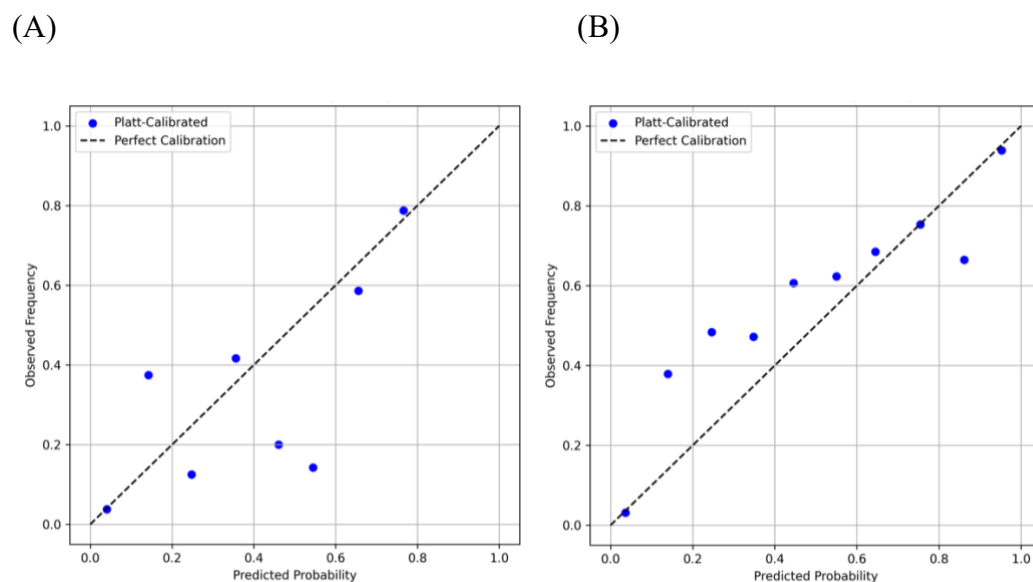


Figure S3. Calibration curves for high-risk class predictions in regional and hospital LightGBM models.