

从预测到理解：机器学习在住院费用预测中的应用进展

谢昊江¹ 汪芮旭^{2*} (通讯作者)

(1.电子科技大学医学院附属妇女儿童医院·成都市妇女儿童中心医院 四川成都 611731;
2.四川大学华西第二医院 四川成都 610041)

【摘要】DRG/DIP支付方式改革的全面推开对医院精细化管理提出更高要求，住院费用预测将成为医疗成本管控、医保基金科学调控与资源配置优化的关键环节。机器学习方法因其强大的模式识别与泛化能力而受到广泛关注，本文系统梳理机器学习在住院费用预测领域的研究进展，阐述相关主流算法的应用场景与性能表现。研究表明，集成学习算法在预测精度上普遍优于单一模型，SHAP等解释性框架的引入有效提升了模型的临床可信度。未来应着力于多中心数据验证、深度学习方法及预测模型与DRG/DIP分组机制的深度融合，为住院费用精细化管理提供更坚实的理论支撑与决策依据。

【关键词】医保支付方式改革；机器学习；住院费用预测；集成学习；可解释性

From Prediction to Interpretation: Machine Learning Applications and Advances in Forecasting Hospitalization Costs

Xie Haojiang¹ Wang Ruixu² (Corresponding author) *

(1.Chengdu Women's and Children's Central Hospital, School of Medicine, University
of Electronic Science and Technology of China Sichuan Chengdu 611731;
2.,West China Second University Hospital, Sichuan University Sichuan Chengdu 610041)

[Abstract] The extensive implementation of DRG/DIP payment reforms has imposed higher demands on hospital refined management, making inpatient cost prediction a critical component for cost control, scientific regulation of medical insurance funds, and optimization of resource allocation. Machine learning methods have gained considerable attention due to their strong pattern recognition and generalization capabilities. This paper systematically reviews research progress in machine learning for inpatient cost prediction, elaborating on the application scenarios and performance of mainstream algorithms. Studies show that ensemble learning algorithms generally outperform single models in predictive accuracy, and the introduction of interpretability frameworks such as SHAP effectively enhances clinical credibility. Future efforts should focus on multi center data validation, deep learning approaches, and deeper integration of predictive models with DRG/DIP grouping mechanisms, so as to provide more solid theoretical support and decision making basis for refined inpatient cost management.

[Key words] Reform of medical insurance payment, Machine Learning; Hospitalization Cost Prediction; Ensemble Learning; Explainability

2025 年是我国医保支付方式改革的关键之年。国家医疗保障局在《DRG/DIP 支付方式改革三年行动计划》中明确指出，到 2025 年底，疾病诊断相关分组（Diagnosis Related Groups, DRG）和病种分值（Diagnosis-Intervention Packet, DIP）支付方式将覆盖所有符合条件的开展住院服务的医疗机构，基本实现病种及医保基金的全覆盖^[1]。DRG/DIP 支付方式的核心机制是“结余留用、超支分担”，这使药品、耗

材从医院的“收入中心”真正转变为“成本中心”，倒逼医疗机构必须从粗放式扩张转向精细化运营。

在此背景下，构建高精度、可解释的住院费用预测模型，已成为适配支付改革、推进价值医疗的关键举措——从医保部门角度，精准的住院费用预测可为医保基金预算编制、支付标准制定、基金风险预警提供科学依据，保障医保基金可持续运行；从医疗机构角度，提前预测住院费用可优化诊疗

方案、合理调配医疗资源、控制不合理支出,提升运营效率;从患者角度,住院费用的精准预估可减少就医过程中的费用不确定性,减轻就医焦虑与经济负担。然而,住院费用的形成机制极为复杂,呈现出高维、非线性、异质性强及数据分布偏态等特点,传统统计模型如多元线性回归等往往难以充分捕捉其内在规律,制约了预测的准确性与稳健性^[2]。

机器学习(Machine Learning, ML)作为人工智能的重要分支,能够从数据中自动学习规律和模式,为复杂的预测问题提供了可行的解决路径。与传统统计方法不同,机器学习不依赖严格的数据分布假设,能较好处理高维数据、非线性关系和变量间的交互作用^[3]。然而,机器学习在医疗费用预测领域也面临模型“黑箱”属性与临床决策对可解释性要求之间的冲突。此外,医保支付政策不断变化,如何让模型适应这种动态环境也是一项挑战。近年来,从单纯追求预测精度到兼顾模型可解释性的研究范式转变尤为值得关注——以SHAP(SHapley Additive exPlanations)为代表的解释技术使研究者能够剖析“黑箱”模型的内在决策逻辑,揭示各因素对住院费用的贡献方向与强度。

本文旨在系统综述机器学习在住院费用预测领域的研究进展。梳理主流算法的应用现状与性能比较,探讨特征工程与模型解释等关键技术问题,分析当前研究面临的关键挑战与存在的局限性,并在此基础上展望未来具有潜力的研究方向,以期期为医疗管理研究者、政策制定者及临床实践者提供理论参考与实践启示。

1 预测算法的演进:从线性回归到集成学习

住院费用预测的方法演进大致可分为三个阶段:以传统统计模型(线性回归、岭回归、套索回归)为主的探索期、以集成学习类(随机森林、梯度提升、XGBoost、LightGBM)模型为主的发展期、以深度学习(多层感知机、卷积神经网络)模型为主的深化期。

1.1 算法类型与应用现状

从现有研究看,集成学习方法的优势在多病种比较研究中得到系统验证,但其预测表现因疾病类型而异。在心血管

与代谢性疾病领域,机器学习模型已取得较为成熟的预测效果。Arnold等^[4]基于美国国家住院样本数据库,针对慢性心力衰竭(CHF)、慢性阻塞性肺疾病(COPD)和糖尿病酮症酸中毒(DKA)三类高负担疾病住院费用构建了六种机器学习模型,结果显示XGBoost和随机森林的预测精度普遍优于传统线性回归和岭回归,其中对于CHF费用的预测效果最为稳定。针对恶性肿瘤患者,Ozair等^[5]基于美国MarketScan商业保险数据库开展脑转移瘤患者住院费用预测,采用随机森林算法获得了稳健的预测性能(RMSE=0.85),验证了集成学习在肿瘤特殊人群中的适用性。此外,针对慢性肾脏病患者,Yabana Kiremit等^[6]发现随机森林在费用金额估算中表现最佳,支持向量机与M5P树则在住院天数预测上更为突出。

但在部分疾病中,机器学习模型的预测性能仍有提升空间。Malhotra等^[7]分析重度抑郁症患者的住院费用后发现,XGBoost虽比传统负二项回归的RMSE降低19%,模型整体解释力却依然有限,且不同算法识别出的关键驱动因素存在差异。这提示精神类疾病的费用预测可能面临比器质性疾病更复杂的社会心理因素交织,仅靠临床与人口学特征难以完整刻画其费用形成机制。另外,深度学习在住院费用预测中的表现并不总是优于传统集成学习。Dhulipala等^[8]指出,当面对高维、小样本或非线性关系复杂的费用数据时,深度神经网络(DNN)容易出现过度拟合或训练不稳定,预测精度反而不及XGBoost或随机森林。原因可能在于,医疗费用数据多为高维稀疏表格结构,而深度学习的强项在于处理图像、序列等具有局部相关性的数据;样本量不足时,其庞大的参数空间反而容易导致过拟合。

1.2 算法性能比较的方法学启示

综合多项研究的证据,可初步形成以下认识:第一,集成学习算法(尤其是XGBoost、随机森林和LightGBM)在住院费用预测精度上普遍优于传统单一模型,其优势源于通过Boosting/Bagging策略降低方差、自动捕捉非线性关系与特征交互、稳健处理缺失值的能力。第二,不同病种、不同数据源下模型性能存在差异,现有证据表明,同一算法在不同疾

病类型间的预测效果存在非均衡性,这提示算法选择不仅取决于技术性能,更需充分考虑疾病本身的临床特征、变量构成及数据质量,算法选择需结合具体情境进行优化。第三,算法的选择应与研究目的相匹配——在以结构化数据为主的费用预测场景中,集成学习仍占主流,而深度学习在针对融合影像、文本等多模态数据时更具潜力。

2 数据基础与预测因子

2.1 特征变量的筛选与构建

在住院费用的预测建模中,特征工程是将原始医疗数据转化为最能代表预测问题特征的过程,直接影响模型性能的上限。从现有研究来看,住院费用预测的特征变量主要涵盖以下维度:①患者人口学特征—年龄、性别是绝大多数研究纳入的基础变量。夏泳等的研究^[9]显示,慢性病患者中年龄与住院费用呈现非线性关联,高龄患者费用更高。Zafar 等人^[10]的相关研究也证实年龄是驱动费用的重要特征。②临床特征—包括疾病诊断(主诊断、合并症、疾病严重程度)、手术操作(手术次数、手术级别)、住院天数等。其中住院天数几乎在所有研究中均被证明是最重要的预测因子之一。③机构特征:医院等级、医院类型(综合/专科)、地区分布等。陈默等^[11]发现三级医院、专科医院的住院费用中位数显著高于其他机构。④治疗与支付特征:是否手术、手术级别、是否实施 DRG 支付等。

2.2 特征选择方法

针对高维数据,合理的特征选择既能提升模型性能,也能增强可解释性。在特征筛选方法上,Boruta 算法、递归特征消除及 LASSO 回归等是常用的降维技术,旨在识别真正具有预测价值的变量,降低过拟合风险并提升模型可解释性。

2.3 特征-算法的匹配性思考

值得关注的是,不同算法对特征工程的依赖程度存在差异。基于树模型的集成学习(如随机森林、XGBoost)天然具备非线性交互捕捉能力,对特征缩放和标准化不敏感,但若原始特征中混杂大量噪声变量,仍可能影响模型稳定性。相比之下,深度学习虽能自动学习特征表示,但在小样本场

景下易过拟合。因此,特征选择不仅是预处理步骤,更应与算法选择协同优化——在样本量有限时,优先采用集成学习并结合稳健的特征筛选;在多模态数据丰富的场景下,深度学习可能释放更大潜力。

3 从“预测”到“理解”:可解释性分析的兴起

3.1 可解释性需求的背景

机器学习模型常因“黑箱”特性而受诟病——即使预测精度很高,研究者也难以理解模型为何作出特定判断。在医疗领域,这一局限尤为突出:临床决策和政策制定不仅需要“知道”费用高低,更需要“理解”费用背后的驱动机制。传统的特征重要性排序(如 Gini 重要性)虽能告知哪些变量重要,却无法揭示变量对预测结果的正负向影响及其交互作用。

3.2 SHAP理论及其应用

SHAP 理论源于博弈论,通过计算每个特征对预测结果的边际贡献来统一解释模型输出。SHAP 值具有可加性:所有特征 SHAP 值之和等于预测值与基线值之差,因此每个预测可分解为各特征贡献的线性组合^[12]。在住院费用预测中,SHAP 分析提供两个层次的解释。个体层面:针对单次住院,可直观看出哪些特征推高或拉低费用。例如有研究^[13]展示了两例——一例因“住院时间短、未手术、一级医院”被预测为低费用(4,502元),另一例因“男性、三级医院、伴合并症”被预测为高费用(36,540元)。这种个案解释能力让模型从统计工具变为临床决策的可信助手。全局层面:汇总所有样本的 SHAP 值可判断各特征的整体影响方向。比如 SHAP 总结图显示“住院天数”的 SHAP 值与费用呈单调正相关——住院越久,正向贡献越大;而“DRG 实施”呈负向贡献,反映支付改革对费用的抑制效果。

3.3 从黑箱预测到机制探索:特征交互的挖掘

SHAP 分析还能探索特征间的交互作用。通过 SHAP 交互值,可以观察某一特征对住院费用的贡献如何随另一特征变化,从而揭示变量间的协同或拮抗关系。但是 SHAP 值并不等同于因果推断,而是反映模型中的边际贡献,本质上是

对相关关系的分解。即便如此,该分析框架仍可为研究者提供数据驱动的假设生成工具,指导后续的因果研究或临床验证,推动研究从“黑箱预测”走向“机制探索”。

4 结语

机器学习正在改变住院费用预测的研究方式。集成学习凭借优异的预测性能成为主流;可解释性技术的进步让研究

者能够打开“黑箱”,理解费用形成的内在逻辑;预测模型也开始融入医保支付方式改革与医院精细化管理实践。当然,数据质量、模型泛化、算法偏见等挑战依然存在。未来研究在追求更高精度的同时,需更重视可解释性、公平性及临床落地价值,使机器学习真正成为理解和控制住院费用的有效工具。

参考文献:

- [1]国家医疗保障局.关于印发 DRG/DIP 支付方式改革三年行动计划的通知.中国政府网,2026.3.19.
- [2]石菊,王小倩,王熙.微观医疗费用预测模型:从线性回归到机器学习[J].经济学(季刊),2023,23(6):2247-2263.
- [3]孙冲.多种机器学习算法预测呼吸系统疾病住院费用效果比较分析[D].吉林大学,2023.
- [4]ARNOLD M, LIOU L, BOLAND M R. Development, evaluation and comparison of machine learning algorithms for predicting in-hospital patient charges for congestive heart failure exacerbations, chronic obstructive pulmonary disease exacerbations and diabetic ketoacidosis[J/OL]. BioData Mining, 2024, 17(1): 35. DOI:10.1186/s13040-024-00387-9.
- [5]JOZAI A, MUKHERJEE D. P10.19.A MACHINE LEARNING-BASED PREDICTION OF INPATIENT CARE COSTS FOR PATIENTS WITH BRAIN METASTASES: A REAL-WORLD ANALYSIS[J]. Neuro-Oncology, 2024, 26(Sup5): 2.
- [6]KIREMIT B Y, ŞAHİN D Ö. Comparison of machine learning algorithms for predicting length of stay in chronic kidney disease patients[J/OL]. Computers in Biology and Medicine, 2025, 196: 110825. DOI:10.1016/j.combiomed.2025.110825.
- [7]MALHOTRA P, JHA R, MALIK K. MSR58 Comparing Predictive Models: Traditional Versus Modern Statistical Approaches for Economic Outcomes in Major Depressive Disorder[J]. Value in Health, 2024, 27(6-Sup): 1.
- [8]DHULIPALA S, MARAM B, RAO R, et al. A Comparative Analysis of Deep Learning and Ensemble Methods for Time Series Forecasting[C]//2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI). IEEE, 2025: 1713-1716.
- [9]夏泳,陈任.基于 BP 神经网络模型分析的老年人慢性病住院费用及影响因素[J].山东第一医科大学(山东省医学科学院)学报,2022,44(5):43.
- [10]ZAFAR S, HAJIANI M A, AFFAN M, et al. Enhancing Accuracy in Pediatric Inpatient Hospital Cost Estimation: A Machine Learning-Based Approach[J]. 2023.
- [11]陈默,蔡苗,黄阿红,等.基于 K-means 聚类与支持向量机的大病患者住院费用影响因素与控制策略研究[J].中国医院管理,2019,39(5):4.
- [12]LUNDBERG S, LEE S I. A Unified Approach to Interpreting Model Predictions[J]. 2017.
- [13]涂朗,李树蕾,樊若曦,等.缺血性心脏病的住院费用预测:基于 SHAP 理论的机器学习[J].2025,(8):881-887.