http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

【作者简介】王岩，助理研究员，发表论文 20 余篇；通信作者：王勇，主任医师。
【基金项目】北京协和医学院中央高校基本科研业务费-青年项目(项目编号：3332025103)。

基于多模型优选的采供血动态预测及系统实现

王岩[1]孙莉[2]巴衣尔策策克[2]陈庆锟[1]万艳丽[1]王勇[2]

（1. 中国医学科学院北京协和医学院，医学信息研究所，北京 100020

2.北京市红十字血液中心，北京 100088）

〔**摘要**〕 **目的/意义** 面向采供血数据的非线性及时序特征，开展多模型比较与优选研究，构建适用于不同时间维度、不同血液制品的采供血预测模型，并将优选模型集成至动态分析与预测系统中，实现预测结果的业务化应用。 **方法/过程** 基于某省血液中心 2015—2025 年脱敏数据，融合气象、节假日等因素，比较 XGBoost、LightGBM、ARIMA、CNN 及两种融合模型，并开发三层 Web 系统。 **结果/结论** 结果表明，采血预测任务，日、月维度最优模型均为 XGBoost+Bi-LSTM（日 sMAPE3.95%，月 9.36%），年维度最佳为 LightGBM。供血预测任务的最优模型：红细胞类日/月/年分别为 XGBoost+Bi-LSTM、LightGBM、LightGBM；血小板类分别为 LightGBM、ARIMA、LightGBM；血浆类分别为 LightGBM、LightGBM、XGBoost+LSTM。依据预测模型结果，完成系统核心功能建设，当前系统已实现数据管理、采供血分析、可视化和趋势预测，可为采血计划、库存管理、血液调配及应急保障提供支持。

〔关键词〕采供血；多模型优选；动态分析；趋势预测；智慧管理

10.12201/bmr.202611.00061V1

Dynamic Forecasting and System Implementation for Blood Collection and Supply Based on Multi-model Optimization

WANG Yan[1], SUN Li[2], BAYIER Ceceke[2], CHEN Qingkun[1], WAN Yanli[1], WANG Yong[2]

(1. Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

2. Beijing Red Cross Blood Center, Beijing 100088, China)

〔**Abstract**〕 **Purpose/Significance** Considering the nonlinear and temporal characteristics of blood collection and supply data, this study compares and selects multiple forecasting models, develops models for different temporal granularities and blood products, and integrates the optimal models into a dynamic analysis and forecasting system to support the operational use of forecasting results. **Method/Process** De-identified data from a provincial blood center covering 2015–2025 were integrated with meteorological and holiday factors. XGBoost, LightGBM, ARIMA, CNN,and two hybrid models were compared, and a three-tier Web system was developed. **Result/Conclusion** The results indicate that for the blood collection prediction task,the XGBoost+Bi-LSTM model achieved optimal performance at both daily and monthly granularities(with daily and monthly sMAPEs of 3.95% and 9.36%,respectively), while LightGBM proved to be the best at the annual granularity. For the blood supply prediction task, the optimal models are as follows: for red blood cells, daily, monthly, and annual predictions are best handled by XGBoost+Bi-LSTM, LightGBM, and LightGBM, respectively; for platelets, by

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

LightGBM, ARIMA, and LightGBM; and for plasma, by LightGBM, LightGBM, and XGBoost+LSTM. Based on these prediction model results, the core functions of the system have been developed. The current system has implemented data management, blood collection and supply analysis, visualization, and trend prediction, providing support for blood collection planning, inventory management, blood allocation, emergency assurance.

〔**Keywords**〕 blood collection and supply; multi-model selection; dynamic analysis; trend forecasting; intelligent management

1 引言

血液作为一种无法人工合成的医疗资源[1]，它不仅在临床急救、外科手术、高危孕产妇、血液病等高风险群体的诊疗中不可替代，也是应对自然灾害或突发公共事件时，应急医疗体系的关键保障[2-3]。近年来，从中央到地方均持续要求强化血液保障体系建设[4-5]，推进血站管理与服务的数字化、智慧化转型[6-7]。

采供血业务数据主要涉及献血方式、采血量、供血机构、供血量、库存水平、血液制品和血型分布等多个指标维度[8]，这些数据沿时间轴会呈现出明显的周期性起伏、非线性波动以及前后时段的依赖关系。影响血液供需平衡的因素不仅包括季节轮转、气象变化等自然条件，还包含节假日安排、人口流动、团体献血组织频率、医院诊疗量等人为因素[9]。传统的描述性统计分析只能反映过去的变化轨迹，难以处理多种因素之间的交叉影响。

国内早期采供血预测多依靠历史同期数据和管理经验[10-11]，难以应对非线性变化及突发事件。W 市[10]、苏州[12]、顺德[13]等血站及重庆某医院基于临床用血数据，建立红细胞、单采血小板或血液制剂 ARIMA 模型[10-13]，为库存管理、献血招募和采集计划提供可量化的依据。但 ARIMA 比较依赖历史规律[14]，难以实时反映政策、气象和突发事件等外部因素，远期预测能力有限[15]。此后，武汉、南宁和哈尔滨某医院的相关研究[16]分别采用多元线性回归[17]、机器学习及 LSTNet 等方法开展预测；模型对比显示，部分改进模型效果在特定数据上优于 ARIMA、GRNN 和 LSTM 等模型，研究表明模型效能与数据特征、预测任务及业务场景密切相关[15-18]。国外研究多聚焦单一血液成分或模型，日本以人口学和统计模型为主[18-19]，美国

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

采用稀疏监督学习和凸优化预测未来 3 日血小板需求[20]，欧洲引入机器学习、深度学习模型开展血液需求、采血量预测[21]，新加坡开展献血者回流预测[22]，但都缺少覆盖采供血、多种血液成分及多时间维度的统一评价[23]。综上，本研究将以采供血动态预测模型的构建与优选为主要研究内容，综合比较 ARIMA、XGBoost、LightGBM、CNN、LSTM 及融合模型的预测效果，筛选适用于采供血任务和血站具体业务场景的最优模型并嵌入构建的 Web 系统[24]中，为献血招募、采集计划、库存调配、医院供应和应急预警提供一体化支持，提升血站体系的智能化水平与应急韧性[25-26]。

2 数据与模型

2.1 数据获取与处理

本研究选取某省域内所有血站 2015 年 1 月至 2025 年 5 月的采血与供血脱敏数据，经数据清洗、标准化及特征构建后，按照时间维度融合外部影响因素，形成用于系统统计分析和模型训练的特征矩阵和数据集。数据具体情况表 1。

表 1 脱敏数据属性情况

数据分类	数据类型	数据属性名称
采血部分	采血点数据	采血点名称、采血点所属区、采血点运行机构、采血工作时间、采血点经纬度、附近地标、采血点类型、启用时间、运行情况
	献血者数据	采血部门、采血地点、组织方式、献血类型、档案 ID、性别、年龄、所属单位组织单位、单位属性、所属系统、单位级别、隶属单位、单位所在行政区、建档日期、登记时间、初筛时间、采血时间、足量、不足量原因、献血码、初检结果、初检不合格项、档案血型、初筛血型、采血量
供血部分	供血机构数据	献血码、供血机构名称、产品码、血液产品名称、ABO、RHD、血量、制备时间、血液有效期、发血时间、发血类型、发血机构、用血机构、用血机构所属行政区
	用血机构数据	供血机构名称、用血机构代码、用血机构名称、用血机构所属行政区、用血机构属性、用血机构类型、用血机构经纬度、用血机构专业类型
外部因素数据	日期数据	日期、星期几、是否周末、是否节假日、是否工作日、年份、月份、日、年内第几天、年内第几周、星期几正弦编码、星期几余弦编码、月份正弦编码、月份余弦编码

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

数据分类	数据类型	数据属性名称
	天气数据	天气状况、平均温度、天气代码。

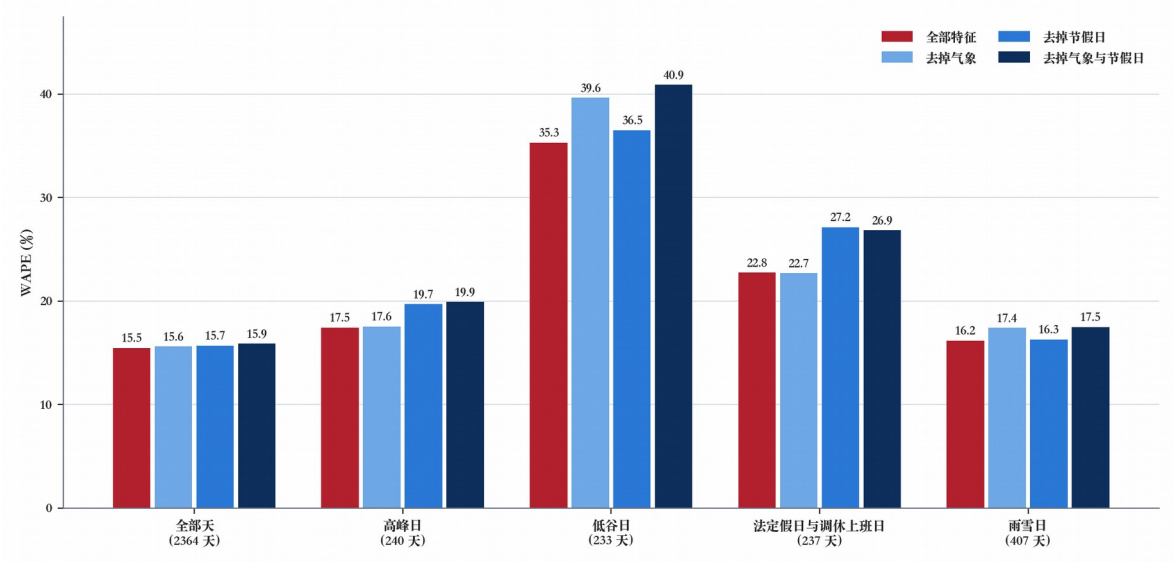

图 1 气象与节假日特征消融：全部特征与去掉某组特征后的 WAPE 结果图

为验证外部指标的有效性，本研究基于 XGBoost 开展了特征消融实验（见图 1）。结果表明，引入气象和节假日特征后，加权绝对百分比误差（WAPE）在全部天（2364 天）从 15.9%降至 15.5%；在法定假日和低谷日场景下效果更加显著，WAPE 分别降低了 4.1%和 5.6%。这证实外部特征能有效纠正节假日突变和低谷期非线性波动，从而提升模型的鲁棒性。

2.2 模型选取与设置

本研究选取 XGBoost、LightGBM、ARIMA 和 CNN 四种单一模型，以及 XGBoost-LSTM 和 XGBoost-BiLSTM 两种融合模型开展对比研究。对六种模型按照统一的数据预处理方法、训练集与测试集划分方式和评价指标开展模型训练与参数调优，并采用平均绝对误差（MAE）、均方根误差（RMSE）、平均绝对百分比误差（MAPE）和加权绝对百分比误差（WAPE）评价模型性能。通过采用时间序列前向验证结合网格搜索进行参数寻优，以验证集的 WAPE 为核心评价指标，遵循验证集划分与信息隔离、相同搜索空间与搜索策略、统一的评判准则，同类模型共用参数的四项原则，以验证误差对各超参数的敏感程度，从而得到模型设置参数。结合预测结果验证各个模型的预测精度、稳定性和泛化效果，选择效果优异的模型与相应趋势预测任务进行适配，模型构成与参数设置见

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

表 2。

表 2 模型构成与参数设置情况

模型名称	模型类型	模型介绍	模型设置参数
XGBoost	传统机器学习模型	适用于包含大量人工特征[27]的采供血量预测，如星期、月份、节假日天气、血型、站点、滞后血量、滚动均值、历史同期值等结构化特征。	日：200 棵、深度 4、学习率 0.05；月：600 棵、深度 3、log 目标、输出与上月值各半融合。
LightGBM	传统机器学习模型	适用于大样本、多特征高维稀疏特征[28]下的采供血量预测，如多站点多血型、多天气、多节假日特征联合建模。	与 XGBoost 同特征、同树数、同学习率；叶子数 15/深度 4（日），叶子数 7/深度 3（月）。
ARIMA(SARIMA)	传统时间序列模型	适用于具有明显时间趋势、周期性和季节性规律[29]的单变量采供血量预测趋势[30]。	log1p 变换，训练段 AIC 小网格选阶（日季节周期 7、月季节周期 12），测试段一步先行预测，无外生特征。
CNN(Conv1D)	深度学习模型	适用于提取采供血量时间序列中的局部变化特征[31]，例如连续几天的波动、短期上升或下降。	两层因果一维卷积+全局平均池化+全连接，回看 30 天（日）/12 个月（月），直接预测下一期。
XGBoost+LSTM	融合模型	适用于同时利用多维结构化特征与时序依赖关系[32]的采供血量预测。先由 XGBoost 生成基线预测，再由 LSTM 学习预测残差的动态变化规律，对基线结果进行残差修正。	XGBoost 基线 + 单向 LSTM（64/32 单元）残差修正头；训练残差取 5 折 OOF 预测；融合 y=y_xgb+clip(残差预测,±50%·|y_xgb|)。
XGBoost+Bi-LSTM	融合模型	适用于挖掘已观测历史窗口内双向时序关联[33]的采供血量预测。以 XGBoost 输出为基线，由 Bi-LSTM 从两个方向提取残差序列特征，增强对周期波动和复杂变化的表征能力。	与 XGBoost+LSTM 完全相同，仅将两层 LSTM 替换为双向 LSTM（Bidirectional），每个方向单元数不变。

3 系统设计

3.1 架构设计

本系统面向某省级血液中心的智慧化管理需求，按照数据管理、标化处理、模型调用、预测结果展示和业务应用的业务流程进行设计，其中模型调用和预测结果应用是系统设计的核心环节，数据管理、动态分析和可视化功能主要用于支撑预测模型运行及其结果解释。本系统采用用户层、业务应用层和数据层的三层系统架构。数

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

据层包括用户数据库、血站信息数据库、采血点信息数据库、采血明细数据库、供血明细数据库和外部特征数据库（天气、温度、节假日、工作日及开学日等）。通过完成数据标化，并由业务应用层通过接口调用数据层资源，构建数据上传、采血分析和供血分析三个主要功能模块，实现采供血业务的动态统计与趋势预测。用户层采用 Web 方式提供系统服务，并通过智慧大屏、桌面计算机等终端，面向血液中心用户、血站用户、血库用户和系统管理员展示最终结果。系统总体架构如图 2 所示。

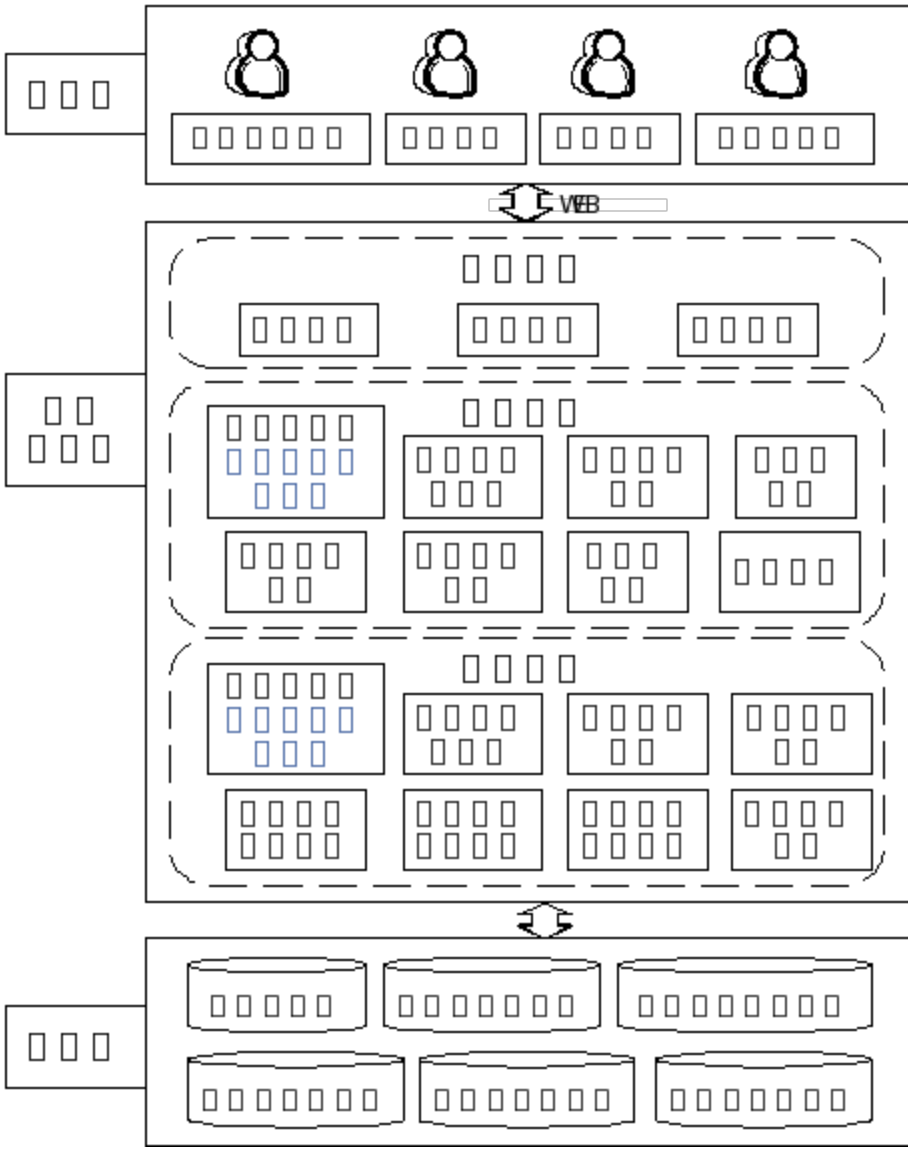

图 2 采供血动态分析与预测系统框架设计图

3.2 功能设计

3.2.1 数据上传 数据上传模块由文件上传、文件管理和数据管理三个子功能构成。该模块支持省级血液中心 Excel 业务数据的批量导

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

入，系统会按照既定规则进行解析，并完成格式转换和标准化处理。导入的文件并非直接进入后续流程，上传人员还需对文件数据进行人工审核，包括查询、校验、修订、补录和删除等操作，确保数据质量，通过“自动处理+人工审核”的双重保障机制，提高数据的完整性与规范性，为后续采供血分析提供可靠的数据基础。

3.2.2 采血分析 采血分析模块包含采血量情况、各行政区采血量、采血机构排行、采血点排行、献血区域排行、献血类型排行、献血者分析和血型占比等八个子功能。本模块分别从采血趋势、献血分布、绩效排行和献血结构四个层面展开分析与预测。首先，通过“采血量情况”汇总指定机构在日、月、年维度的历史数据并预测其未来趋势。其次，深入挖掘空间分布特征，既包括“各行政区采血量”中全血、血小板及人次的细分统计，同时涵盖“献血区域排行”对区域规模的比较。在绩效评估方面，系统支持对全省 11 家“采血机构”及具体“采血点”进行排序，并按独立设施、交通枢纽、商业中心、旅游景区等七类场所展示“献血类型排行”。最后，采血分析模块利用可视化图表呈现血液与人群结构：“献血者分析”以环形图刻画年龄、性别等人群特征，“血型占比”则以饼状图展示 A 型、B 型、O 型和 AB 型四种血型的具体构成情况。

3.2.3 供血分析 供血分析模块包括供血量情况、机构供血量排行、各区供血情况、供血产品占比、医疗机构供血排行、医疗机构供血详情、各区医院供血排行和医院供血分布等八个子功能。本模块分别从供血趋势、供血流向、绩效排行和用血结构四个层面展开分析与预测。首先，通过“供血量情况”汇总指定机构在日、月、年维度的红细胞、血小板及血浆供应数据情况与未来预测趋势，并结合“供血产品占比”以饼图和数据表展示 20 种具体制品的构成。其次，在供血流向分析上，“各区供血情况”创新引入横向占比（机构向区域供血量/机构总供血量）与纵向占比（机构向区域供血量/区域总接收量）双重指标，精准刻画血液的空间调配关系。在绩效排行方面，系统支持对全省 11 家“供血机构”进行供应规模排序，对 193 家“医疗机构”按属性和类型分类统计，并通过“各区医院供血排行”展示不同区域内各医院的血液供应水平。从用血结构上，“医

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

疗机构供血详情”支持模糊检索定位目标医院并查看其供应来源构成，“医院供血分布”则以环形图从机构属性和医院类型双维度分析供血量构成详情。

4 系统构建

4.1 采血模型优选

采血模型的日维度预测以 2015 年 1 月 1 日至 2025 年 6 月 22 日的自愿献血日采血量为研究对象，共获得 3826 个日观测值。按照时间先后顺序，将末 383 个观测值设置为独立测试集，测试期为 2024 年 6 月 5 日至 2025 年 6 月 22 日；其余 3443 个观测值构成训练集，时间范围为 2015 年 1 月 1 日至 2024 年 6 月 4 日。月维度预测首先将纯街头无偿献血日采血量按自然月汇总，并剔除数据不完整的 2025 年 6 月。2015 年 1 月至 2025 年 5 月共形成 125 个月度观测值。由于模型需要构造 12 个月滞后及滚动统计特征，前 12 个月不纳入有效样本，最终得到 113 个月度观测值。其中，2024 年 6 月至 2025 年 5 月的 12 个月作为测试集，其余 101 个月作为训练集。年维度预测采用滚动年度回测，分别以 2023 年、2024 年、2025 年 1 月至 6 月为 3 个测试年，每个测试年仅使用该年之前的全部数据训练日模型，对整年逐日预测后求和得到年总量，并同时统计年内 12 个月的月聚合误差。在同一预测维度下，六种模型均采用完全相同的训练集与测试集。

日维度和月维度预测均以加权绝对百分比误差（WAPE）作为主要评价指标，并采用平均绝对误差（MAE）、均方根误差（RMSE）和对称平均绝对百分比误差（sMAPE）进行辅助评价。

WAPE 的计算公式为：$WAPE = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{\sum_{t=1}^{n} |y_t|} \times 100\%$，年维度使用年总量绝对百分比误差 APE 与年内月聚合 WAPE。各项误差指标数值越小，

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

表示模型预测结果越接近实际观测值。三维度六模型的测试 WAPE 对比结果与预测趋势见图 3 和图 4。

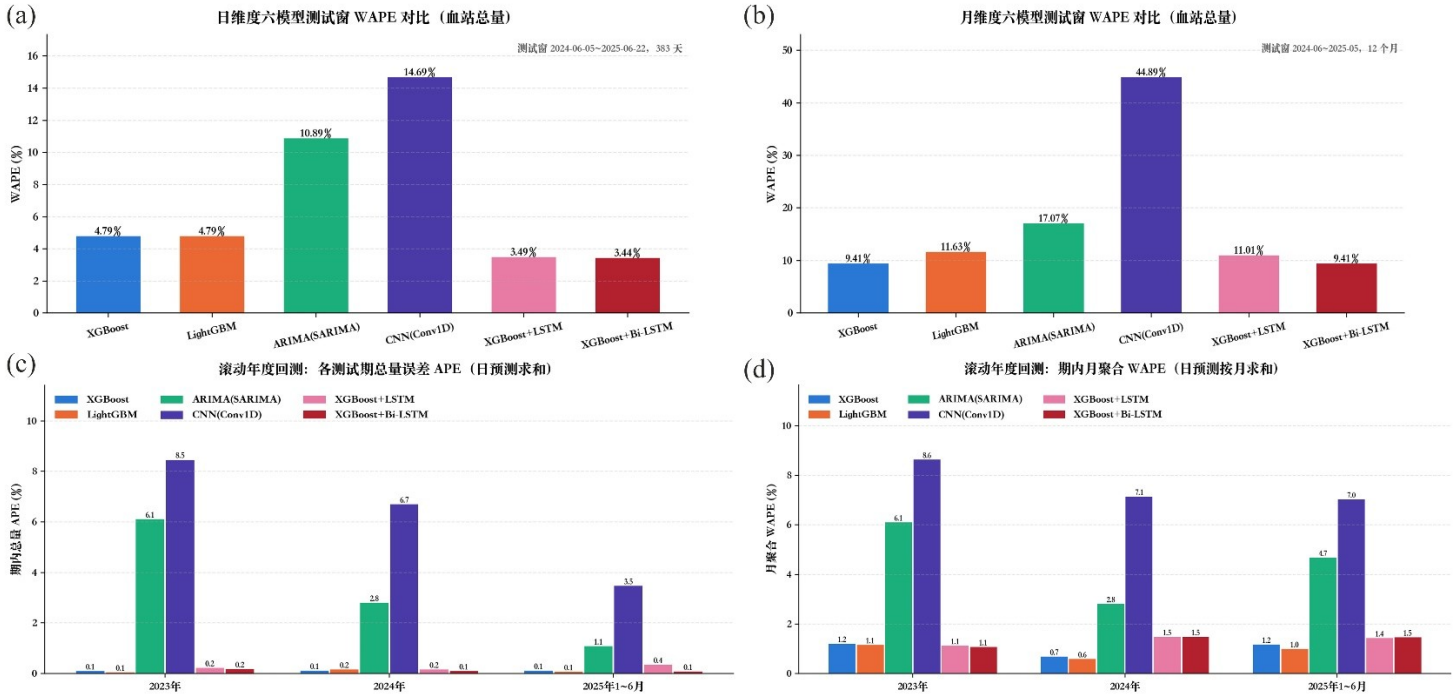

图 3 日月年三维度-六模型测试 WAPE 对比图

（a）日维度六模型测试窗 WAPE 对比图 （b）月维度六模型测试窗 WAPE 对比图
（c）年维度（日预测求和）APE 对比图 （d）年维度（日预测按月求和）WAPE 对比图

10.12201/bmr.202611.00061V1

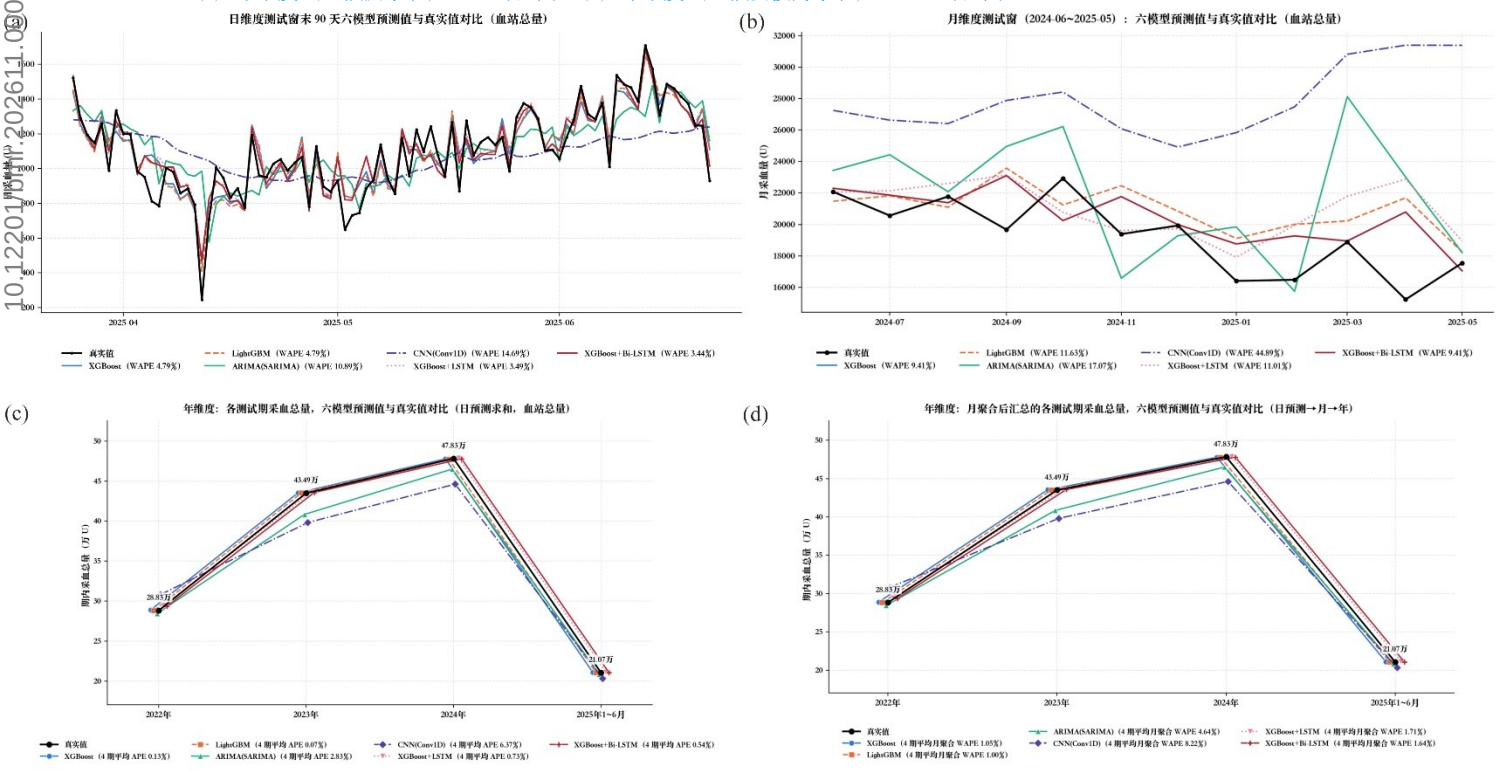

图 4 日月年三维度-六模型预测趋势图

（a）日维度测试窗 90 天六模型预测真实对比图（b）月维度测试窗 12 月六模型预测真实对比图
（c）年维度（日预测求和）六模型预测真实对比图 （d）年维度（日预测按月求和）六模型预测真实对比图

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

4.1.1 采血日维度预测 XGBoost+Bi-LSTM 模型的 WAPE、MAE、RMSE 和 sMAPE 分别为 3.44%、44.3U、65.6U 和 3.95%，均为六种模型中的最低值。与单一 XGBoost 模型相比，该模型的 WAPE 降低 1.35 个百分点，相对降幅为 28.2%；与 XGBoost+LSTM 模型相比，其 WAPE 进一步降低 0.05 个百分点，相对降幅为 1.4%。此外，XGBoost+LSTM 模型相较 XGBoost 模型的 WAPE 相对下降 27.1%。上述结果表明，在本测试窗口内，引入时序残差修正后，融合模型在主要误差指标上取得了更好的预测效果；采用双向 LSTM 后，模型性能得到小幅提升。具体结果见表 3。

表 3 日维度测试窗六种模型指标

模型	MAE(U)	RMSE(U)	WAPE(%)	sMAPE(%)	测试窗总量偏差(%)
XGBoost	61.60	82.20	4.79	5.37	0.01
LightGBM	61.70	79.80	4.79	5.33	0.28
ARIMA(SARIMA)	140.20	190.50	10.89	11.50	2.15
CNN(Conv1D)	189.20	239.00	14.69	15.47	5.75
XGBoost+LSTM	44.90	67.00	3.49	3.99	0.49
XGBoost+Bi-LSTM	44.30	65.60	3.44	3.95	0.18

4.1.2 采血月维度预测 XGBoost 基线模型与 XGBoost+Bi-LSTM 模型取得相同的测试结果，MAE、WAPE、sMAPE 和测试窗总量偏差分别为 1811U、9.41%、9.36%和 6.32%，在六种模型中并列最优。与 XGBoost+LSTM 模型相比，二者的 WAPE 降低 1.60 个百分点，相对降幅为 14.5%。在当前月度样本规模和参数配置下，Bi-LSTM 残差修正未带来额外的精度提升，而 LSTM 残差修正后的预测误差有所增加。考虑到 XGBoost 基线模型结构相对简单，根据模型简约性原则将 XGBoost 模型确定为月维度预测模型，具体结果见表 4。

表 4 月维度测试窗六种模型指标

模型	MAE(U)	WAPE(%)	sMAPE(%)	测试窗总量偏差(%)
XGBoost	1811.00	9.41	9.36	6.32
LightGBM	2238.00	11.63	11.42	9.06
ARIMA(SARIM	3285.0	17.07	15.65	13.45

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

模型	MAE(U)	WAPE(%)	sMAPE(%)	测试窗总量偏差(%)
A)	0			
CNN(Conv1D)	8636.00	44.89	36.85	44.89
XGBoost+LSTM	2118.00	11.01	10.79	8.94
XGBoost+Bi-LSTM	1811.00	9.41	9.36	6.32

4.1.3 采血年维度预测 三个 XGBoost 系模型的平均累计量 APE 均低于 1%。其中，XGBoost、XGBoost+LSTM 和 XGBoost+Bi-LSTM 的三期平均 APE 分别为 0.11%、0.25%和 0.12%，均低于 ARIMA 的 3.33%和 CNN 的 6.21%。LightGBM 在三个回测窗口中的 APE 分别为 0.06%、0.16%和 0.07%，三期平均值为 0.09%，为六种模型中的最低值；其三期平均月聚合 WAPE 为 0.90%，同样取得最优结果。虽然 LightGBM 的日 WAPE 为 4.98%，高于 XGBoost+LSTM 和 XGBoost+Bi-LSTM，但其在累计总量和月度聚合层面的预测误差更低，故采血年维度预测选用 LightGBM 模型，具体数据见表 5。

表 5 年维度两种方法测试窗六种模型指标

年维度	模型	2023 年 APE(%)	2024 年 APE(%)	2025 年 1~6 月 APE(%)	3 期平均(%)	日 WAPE3 期平均(%)
滚动年度回测各测试期总量误差 APE	XGBoost	0.10	0.11	0.12	0.11	—
	LightGBM	0.06	0.16	0.07	0.09	—
	ARIMA(SARIMA)	6.11	2.81	1.08	3.33	—
	CNN(Conv1D)	8.46	6.70	3.47	6.21	—
	XGBoost+LSTM	0.23	0.16	0.35	0.25	—
	XGBoost+Bi-LSTM	0.18	0.10	0.09	0.12	—
滚动年度回测期内月聚合 WAPE 与日 WAPE	XGBoost	1.20	0.68	1.16	1.02	5.03
	LightGBM	1.15	0.58	0.99	0.90	4.98
	ARIMA(SARIMA)	6.11	2.81	4.67	4.53	12.96
	CNN(Conv1D)	8.63	7.14	7.04	7.60	13.22
	XGBoost+LSTM	1.12	1.47	1.43	1.34	4.13
	XGBoost+Bi-LSTM	1.08	1.47	1.46	1.34	3.97

4.1.4 采血预测模型结果 采血预测任务嵌入系统的模型情况：日维度和月维度的最优模型均为 XGBoost+Bi-LSTM（日 sMAPE 为 3.95%，月 sMAPE 为 9.36%），年维度最佳模型为 LightGBM，APE 三期平均为 0.09%。

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

4.2 供血模型优选

供血模型以按日期、发血机构、血液制品类别和 ABO 血型汇总的净供血量为研究对象，净供血量定义为发血量与回退量之差。分别汇总 A 型、B 型、O 型和 AB 型血液的净供血量，形成红细胞类、血小板类和血浆类的供血总量，并依据各品类的计量单位分别建模与评价，其中红细胞类以 U 计量，血小板类以治疗量计量，血浆类以 mL 计量。

日维度预测按照时间顺序划分训练集和测试集，将末 384 天设置为独立测试窗口，测试期为 2024 年 6 月 12 日至 2025 年 6 月 30 日。月维度预测采用直接月模型和日预测聚合两种方法。直接月模型将日供血数据按自然月汇总，并在剔除因 12 个月滞后特征构造所产生的前置无效样本后，将 2024 年 7 月至 2025 年 6 月的 12 个月设置为测试集。日预测聚合方法则将日维度模型在 2024 年 7 月至 2025 年 6 月的逐日预测结果按自然月求和，并与相同月份的实际供血总量进行比较。年维度预测采用扩展窗口滚动回测方法，分别将 2023 年、2024 年和 2025 年 1 至 6 月设置为测试窗口。每次回测仅使用测试期开始前的历史数据训练日维度模型，并将测试期内的逐日预测结果汇总为测试期累计供血量。2023 年和 2024 年的预测结果作为完整年度结果进行评价，2025 年 1 至 6 月则作为部分年度累计结果处理。

在相同血液制品类别和预测任务下，六种模型采用一致的目标变量、有效样本范围和数据划分方式。日维度与月维度预测以 WAPE 为主要评价指标，并结合 MAE、RMSE 和 sMAPE 评价预测误差；滚动年度回测采用测试期累计量 APE 评价总体偏差，并以期内月聚合 WAPE 评价月度预测表现。具体核心指标结果见表 6。

表 6 三类制品与六种模型核心指标结果

品类	指标	XGBoost	LightGBM	ARIMA (SARIMA)	CNN (Conv1D)	XGBoost +LSTM	XGBoost +Bi-LSTM
红细胞类	日 WAPE(%)	2.79	2.55	10.19	3.89	2.32	2.18
	月 WAPE(%)·直接月模型	6.93	7.19	5.88	10.01	8.33	8.90
	月 WAPE(%)·日预测聚合	0.42	0.39	2.33	2.57	0.49	0.41

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

品类	指标	XGBoost	LightGBM	ARIMA (SARIMA)	CNN (Conv1D)	XGBoost +LSTM	XGBoost +Bi-LSTM
血小板类	年总量 APE(%)·3 期均值	0.41	0.37	1.66	1.57	0.56	0.57
	日 WAPE(%)	4.81	4.51	9.67	8.86	5.26	5.24
	月 WAPE(%)·直接月模型	11.40	11.47	6.65	29.81	11.40	15.21
	月 WAPE(%)·日预测聚合	3.70	3.39	2.10	5.76	4.36	4.34
血浆类	年总量 APE(%)·3 期均值	1.10	0.88	1.78	4.44	1.47	1.62
	日 WAPE(%)	5.11	4.77	30.61	10.36	5.06	4.78
	月 WAPE(%)·直接月模型	12.59	13.15	14.53	17.76	12.59	12.59
	月 WAPE(%)·日预测聚合	1.62	1.22	12.75	6.82	1.53	1.56
	年总量 APE(%)·3 期均值	0.66	0.74	14.23	5.45	0.44	0.52

4.2.1 供血三类制品日维度预测 红细胞类供血量预测中，XGBoost+Bi-LSTM 的 WAPE 最低，为 2.18%；相较 XGBoost 和 XGBoost+LSTM，其 WAPE 分别降低 22.00%和 6.10%。血小板类预测中，LightGBM 的 WAPE 最低，为 4.51%；XGBoost+Bi-LSTM 的 WAPE 为 5.24%，相较 XGBoost 升高 9.10%，相较 XGBoost+LSTM 降低 0.20%，说明双向结构仅带来较小改善。血浆类预测中，LightGBM 取得最低 WAPE，为 4.77%，XGBoost+Bi-LSTM 以 4.78%次之；相较 XGBoost 和 XGBoost+LSTM，XGBoost+Bi-LSTM 的 WAPE 分别降低 6.5%和 5.5%。

4.2.2 供血三类制品月维度预测 采用直接月模型时，红细胞类和血小板类均由 ARIMA（SARIMA）取得最低 WAPE，分别为 5.88%和 6.65%，对应的 XGBoost+Bi-LSTM 结果分别为 8.90%和 15.21%；血浆类由 XGBoost 取得最低 WAPE，为 12.59%，XGBoost+Bi-LSTM 同为 12.59%。采用日预测聚合方式时，红细胞类和血浆类均以 LightGBM 表现最优，WAPE 分别为 0.39%和 1.22%，对应的 XGBoost+Bi-LSTM 结果分别为 0.41%和 1.56%；血小板类以 ARIMA（SARIMA）最优，WAPE 为 2.10%，XGBoost+Bi-LSTM 为 4.34%。在本测试窗口内，日预测聚合方式的月度误差整体低于直接月模型，因此系统采用日预测结果聚合生成月维度预测值。

4.2.3 供血三类制品年维度预测 采用滚动回测方式对 2023 年、2024 年和 2025 年 1 至 6 月三个测试期进行评价，并以日预测结果的期内累计值计算三期平均 APE。红细胞类和血小板类均以 LightGBM 表现最优，平均 APE 分别为 0.37%和 0.88%；其中，红细胞类 XGBoost、XGBoost+LSTM 和 XGBoost+Bi-LSTM 的平均 APE 分别为 0.41%、0.56%和 0.57%，血小板类对应结果分别为

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

1.10%、1.47%和 1.62%。血浆类以 XGBoost+LSTM 表现最优，平均 APE 为 0.44%， XGBoost 和 XGBoost+Bi-LSTM 分别为 0.66%和 0.52%。

4.2.4 供血三类制品预测模型结果 供血预测任务嵌入系统的模型情况：日维度预测红细胞类选取的是 XGBoost+Bi-LSTM，WAPE 为 2.18%，血小板类和血浆类选用的模型是 LightGBM，WAPE 分别是 4.51%和 4.77%。月维度红细胞雷与血浆类的选取模型为 LightGBM，AWPE 分别是 0.39%和 1.22%，血小板类月最佳模型为 ARIMA，WAPE 为 2.10%。年维度预测红细胞类和血小板类的模型为 LightGBM，WAPE 分别为 0.37%和 0.88%，血浆类年维度模型是 XGBoost+LSTM，WAPE 为 0.44%。

4.3 系统建设效果



图 5 采供血数据平台主页面加采血预测模块截图

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。



图 6 供血模块-医疗机构供血排行页面截图

本研究已完成采供血数据平台的设计、开发、部署与功能联调，数据上传、采血分析、供血分析等功能模块均已上线运行。系统采用统一的 Web 交互界面，对采供血业务数据进行集中管理，并通过统计图表、趋势曲线、结构占比和排行列表等形式呈现业务运行状态。系统功能验证结果表明，各模块衔接顺畅，分析结果能够按照业务条件动态更新，实现了从数据管理、业务统计到趋势预测和可视化展示的一体化应用，可为采血计划制定、供血结构分析、库存风险研判及血液资源调配提供数据支持，基本满足省级血液中心采供血动态分析与趋势预测的智慧化管理需求。具体页面截图见图 5 和图 6。

5 总结

本研究基于血站连续十年脱敏业务数据，面向采供血任务需求和业务场景开展多模型比较与优选，选取 XGBoost、LightGBM、ARIMA、CNN、XGBoost-LSTM 和 XGBoost-BiLSTM 六种模型，在统一的数据处理、样本划分和评价指标下开展对比，并针对不同时间、机构、血液制品及预测任务选用性能较优的模型，在此基础上，以预测结果展示为核心功能完成采供血动态分析与预测系统的开发，能为采供血应急保障提供数据支持。目前研究主要基于单一血液中心，预测模型的外部验证推广与跨区域泛化能力有限；模型库只涉及六种模型，尚未充分覆盖更

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

多的先进模型；预测颗粒度和业务场景仍可进一步细化，系统实际应用效果与业务协同能力仍需进一步评估。未来可开展多地区、多机构的外部验证，增强模型泛化能力；引入更多模型及融合策略；将预测细化到市级血站、科室、病种和献血者群体等维度；完善数据治理与隐私安全保护；推进系统落地评估与业务协同，并按计划拓展献血者画像、应急血液保障大语言模型对话、医院用血订单分析及趋势预测等模块。

作者贡献：王岩负责模型调试、功能设计、论文撰写；孙莉负责研究设计；巴衣尔策策克、陈庆锟负责数据整理、页面设计；万艳丽负责系统开发、模型验证；王勇负责业务指导，论文审核。
利益冲突：所有作者均声明不存在利益冲突。

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

参考文献

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

1[] 国家卫生健康委员会等.国卫医急函[2026]25 号.关于印发无偿献血倡议书的通知[EB/OL].(2026-02-02)[2026-09-11].https://www.nhc.gov.cn/ylyjs/gzdt/202602/4b7fbba3d6324f84b83d41b378e35192.shtml
2[] 国家卫生健康委员会法规司.中华人民共和国献血法[EB/OL].(1997-12-29)[2026-09-10].https://www.nhc.gov.cn/fzs/c100048/201808/ea9fb73f399d4cedb1b3438825646505.shtml
3[] 国家卫生健康委员会.卫生部令第 85 号.医疗机构临床用血管理办法[EB/OL].(2012-06-07)[2026-09-11].https://www.nhc.gov.cn/wjw/c100221/202201/7b7da6f69f0746e1acba7a4562232511.shtml
4[] 中共中央.国民经济和社会发展第十五个五年规划[EB/OL].(2025-10-23)[2026-09-10].https://www.gov.cn/gongbao/2025/issue_12386/202511/content_7047415.html
5[] 国务院.国发[2026]23 号.国民健康“十五五”规划[EB/OL].(2026-07-07)[2026-09-10].https://www.gov.cn/zhengce/content/202607/content_7075213.htm
6[] 国家卫生健康委员会.国卫医发[2021]42 号.全国血站服务体系建设发展规划(2021-2025 年)[EB/OL].(2021-12-30)
[2026-09-10].http://www.nhc.gov.cn/yzygj/s3589/202112/35bf5b087fd74ea08b0ee509978e4841.shtml
7[] 北京市人民政府.京政发[2021]38 号.“十四五”时期健康北京建设规划[EB/OL].(2021-12-21)[2026-09-11].https://www.beijing.gov.cn/zhengce/zhengcefagui/202112/t20211229_2575955.html
8[] 裘君娜,孔长虹,胡伟. WS/T 865-2025《血站采供血基本数据集》的构建与研究[J]. 中国输血杂志,2026,39(8):999-1003. DOI:10.13303/j.cjbt.issn.1004-549x.2026.08.001.
9[] 谢桂芸,陈锦艳,欧阳剑,等. 基于多源时空数据的采血点效能诊断与网络优化研究——以广州为例[J].中国输血杂志,2026,39(6):776-783.DOI:10.13303/j.cjbt.issn.1004-
10[] 刘芸男. 基于 W 市血液中心临床用血情况分析的血液需求预测研究[D]. 重庆:重庆医科大学,2020.
11[] 谢丽娟,郑敏,陈剑羽. 基于 ARIMA 模型的血液制剂使用量时间预测模型的建立与应用[J]. 现代医药卫生,2020,36(24):4047-4051. DOI:10.3969/j.issn.1009-5519.2020.24.054.
12[] 谢淑红,张思静,严伟斌,等. 基于 ARIMA 模型的临床红细胞类血液需求预测研究[J]. 蚌埠医学院学报,2023,48(5):633-636. DOI:10.13898/j.cnki.issn.1000-2200.2023.05.019.
13[] 吕艺通,刘志泉,莫巧频,等. ARIMA 模型在基层血站单采血小板临床需求量预测中的可行性分析[J]. 中国实用医药,2023,18(23):144-148. DOI:10.14163/j.cnki.11-5547/r.2023.23.036.
14[] 苏武锦,陈楠. 弱季节性视角下的采供血需求预测方法论——从竞争模型到归因范式的探讨[J]. 中国输血杂志,2026,39(8):1067-1073. DOI:10.13303/j.cjbt.issn.1004-549x.2026.08.011.
15[] 刘奇. 基于数据驱动的血液供应量预测方法研究[D]. 成都:西南交通大学,2021.
16[] 于新雨. 基于优化 LSTNet 和列线图的临床用血量预测与影响因素研究[D]. 哈尔滨:哈尔滨理工大学,2025.
17[] 洪语. 武汉市血液采供能力预测研究[D]. 武汉:华中科技大学,2021.
18[] Akita T, Tanaka J, Ohisa M, Sugiyama A, Nishida K, Inoue S, Shirasaka T. Predicting future blood supply and demand in Japan with a Markov model: application to the sex- and age-specific probability of blood donation. Transfusion. 2016 Nov;56(11):2750-2759. doi: 10.1111/trf.13780. Epub 2016 Sep 5. PMID: 27595917.
19[] Imada H, Akita T, Sugiyama A, Tanaka J. Trend of blood donors entering the coronavirus pandemic era and challenges: Age-period-cohort analysis using 75.5 million all blood donations data during 2006-2020 in Japan. Transfusion. 2023 Jun;63(6):1184-1194. doi: 10.1111/trf.17387. Epub 2023 May 9. PMID: 37158239.
20[] Guan L, Tian X, Gombar S, Zemek AJ, Krishnan G, Scott R, Narasimhan B, Tibshirani RJ, Pham TD. Big data modeling to predict platelet usage and minimize wastage in a tertiary care system. Proc Natl Acad Sci U S A. 2017 Oct 24;114(43):11368-11373. doi: 10.1073/pnas.1714097114. Epub 2017 Oct 9. PMID: 29073058; PMCID: PMC5664553.
21[] Dauphinet M, Henriques R, Monteiro P T. Context-Aware Deep Forecasting: Principles for the Nation-Wide Management of Blood Products[C]//EPIA Conference on Artificial Intelligence. Cham: Springer Nature Switzerland, 2024: 359-372.
22[] Gammon RR, Hindawi S, Al-Riyami AZ, Ang AL, Bazin R, Bloch EM, Counts K, de Angelis V, Goel R, Grubovic Rastvorceva RM, Pati I, Lee CK, La Raja M, Mengoli C, Oreh A, Patidar GK, Rahimi-Levene N, Ravula U, Rexer K, So-Osman C, Thachil J, Nevessignsky MT, Vermeulen M. The use of predictive modelling to determine the likelihood of donor return during the COVID-19 pandemic. Transfus Med. 2024 Oct;34(5):333-343. doi: 10.1111/tme.13071. Epub 2024 Aug 8. PMID: 39113629.
23[] Engelke M, Çinar-Koraş OA, Kleesiek J, Horn PA. AI and Blood Banking: Predicting Transfusion Demand - A Systematic Review of Forecasting Approaches. Transfus Med Hemother. 2026 Jun 12. doi: 10.1159/000552969. Epub ahead of print. PMID: 42541243; PMCID: PMC13427359.
24[] 刘磊. 基于大数据挖掘的中心血站采供血结构智能统计分析研究[J].消费电子,2025(22):29-31. DOI:10.3969/j.issn.1674-7712.2025.22.010.
25[] 王自强,王凯敏,阮天任,等. 中山市血液保障管理体系实践研究[J]. 临床医学研究与实践,2025,10(27):18-22. DOI:10.19347/j.cnki.2096-1413.202527005.
26[] 中共北京市委,北京市人民政府.关于加强首都公共卫生应急管理体系建设的若干意见[EB/OL].

10.12201/bmr.202611.00061V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00061
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

(2020-05-19)[2026-09-11]. https://www.beijing.gov.cn/zhengce/zhengcefagui/202005/t20200520_1902580.html
27[] CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2016: 785-794. DOI: 10.1145/2939672.2939785.
28[] KE G, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C/OL]//Advances in Neural Information Processing Systems 30. Red Hook: Curran Associates, 2017: 3146-3154[2026-09-16].
https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.
29[] BOX G E P, JENKINS G M, REINSEL G C, et al. Time series analysis: forecasting and control[M]. 5th ed. Hoboken: John Wiley & Sons, 2015.
30[] 谢淑红,张思静,严伟斌,等. 基于 ARIMA 模型的临床红细胞类血液需求预测研究[J]. 蚌埠医学院学报,2023,48(5):633-636. DOI:10.13898/j.cnki.issn.1000-2200.2023.05.019.
31[] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324. DOI: 10.1109/5.726791.
32[] SAJID M, MALIK K R, ALMOGREN A, et al. Enhancing intrusion detection: a hybrid machine and deep learning approach[J]. Journal of Cloud Computing, 2024, 13: 123. DOI: 10.1186/s13677-024-00685-x.
33[] EID M H, KOVÁCS A, SZŰCS P. Enhancing karst spring discharge simulation through a hybrid XGBoost–BiLSTM machine learning framework[J]. Water, 2026, 18(9): 1038. DOI: 10.3390/w18091038.

10.12201/bmr.202611.00061V1