

DOI: 10. 12138/j. issn. 1671-9638. 20233547

· 论 著 ·

基于医院临床数据中心的可疑呼吸道传染病发现与预测模型构建

陈廷寅^{1,2}, 冯 嵩^{1,2}

[1. 中南大学湘雅医院网络信息中心, 湖南 长沙 410008; 2. 国家老年疾病临床医学研究中心(湘雅医院), 湖南 长沙 410008]

[摘要] **目的** 通过构建基于临床数据中心的可疑呼吸道传染病发现与预测模型, 实现对可疑传染病的发现与预测。**方法** 选取某三甲医院的临床数据, 基于历史传染病数据进行病历结构化建模, 构建呼吸道传染病知识图谱, 利用 XGboost 算法和知识图谱推理技术形成发现与预测合并决策模型, 并使用医院历史数据做交叉验证, 得到准确度较高的模型。**结果** 发现与预测模型的平均查准率为 92.55%, 查全率为 91.49%, 综合 F1 值为 92.01%, 均优于单独的知识图谱模型或 XGboost 模型, 将模型与医院的电子病历系统和临床辅助决策系统进行集成, 应用于对真实临床病例的预测。**结论** 该方法能够很好地针对新发可疑呼吸道传染病进行预测, 辅助医院及时启动传染病应急预案, 减少传染病发生早期时医务人员的感染概率。

[关键词] 可疑传染病; 人工智能; 知识图谱; 预测模型

[中图分类号] R197.323

Construction of a detection and prediction model for suspected respiratory infectious diseases based on hospital clinical data center

CHEN Ting-yin^{1,2}, FENG Song^{1,2} (1. Network Information Center, Xiangya Hospital, Central South University, Changsha 410008, China; 2. National Clinical Research Center for Geriatric Disorders [Xiangya Hospital], Changsha 410008, China)

[Abstract] **Objective** To construct a detection and prediction model for suspected respiratory infectious diseases (RIDs) based on clinical data center, and achieve the detection and prediction of suspected infectious diseases. **Methods** Clinical data were selected from a tertiary first-class hospital, structural modeling of medical records was constructed based on historical data of infectious diseases, knowledge map of RIDs was formulated, a combined decision model of detection and prediction was formed through XGboost algorithm and knowledge map reasoning technology, and cross validation based on hospital historical data was performed, a model with high accuracy was obtained. **Results** The average precision ratio of the detection and prediction model was 92.55%, with recall ratio of 91.49% and the comprehensive F1 test value of 92.01%, which were superior to the individual knowledge map model or XGboost model. The model was integrated with the hospital's electronic medical record system and clinical decision support system for predicting real clinical cases. **Conclusion** This method can effectively predict emerging suspected RIDs, assist hospitals to initiate emergency plans timely for infectious diseases, and reduce the probability of infection among health care workers at the early stage of infectious diseases.

[Key words] suspected infectious disease; artificial intelligence; knowledge mapping; prediction model

[收稿日期] 2022-11-01

[基金项目] 长株潭一体化医疗大数据应用基础支撑体系研究试点示范(202102-001); 移动医疗教育部中国移动联合实验室课题(2020MHL02015)

[作者简介] 陈廷寅(1986-), 男(汉族), 贵州省贞丰县人, 工程师, 主要从事医疗大数据研究。

[通信作者] 冯嵩 E-mail: fs205@sina.com

我国于 2003 年和 2008 年先后建立了中国传染病疫情和突发公共卫生事件网络直报系统^[1]和传染病自动预警系统^[2],其核心都是围绕已知传染病的上报和监测,采用分级上报和基于规则模型事后集中监测的机制^[3],而医疗机构作为传染病监测哨点机构很难对新发可疑传染病的发生情况和时空聚集性特点进行合理判断与预警^[4]。主要表现在以下几个方面:(1)对于新发可疑传染病提前干预能力不足。目前医院传染病都是通过医生报卡后才知道发生了传染病,而且不知道传染病发生的区域及时间相关特性^[5],无法针对性地进行相应的干预。(2)医院缺少相应的手段对传染病进行监测和预警。医院虽然为非传染病专科医院,但日常诊疗过程中也会收治大量的传染病患者,这些患者会严重地影响到其他患者及医务人员的安全,需要建立起相应的传染病监测与预警机制,减少传染病对日常诊疗工作的影响。(3)无法整合大量临床数据。医院发热门诊、感染病科、放射科、检验科等科室在日常诊疗工作过程中产生了大量的传染病患者全过程临床数据,由于缺少数据的整合基本上都处于无法利用的状态。(4)传染病报告人为影响过大。传染病报告卡以前都是由医生主观决定是否上报,也有可能由于工作较忙而忘记上报,导致传染病患者漏报,调查研究^[6-7]结果显示,河南省综合医院传染病漏报率为 11.43%,四川省法定报告传染病漏报率为 7.69%,需针对这种情况建立监督机制减少漏报。目前,国内外已经开展了相关研究,例如抓取医院信息系统内与传染病相关的诊断名称、关键词^[8],通过对实验室报告数据进行分析处理实现对需要上报传染病病例的识别^[9]。本研究在上述工作的基础上,通过已整合患者全病程数据的临床数据中心,研究建立可疑呼吸道传染病发现与预测的方法和机制。

1 对象与方法

1.1 研究对象 某三级甲等医院在 10 余年近 800 万患者的临床诊疗数据基础上构建了医院临床大数据中心^[10],数据的来源为医院内临床业务数据、区域及联盟机构数据、设备数据。传染病数据库以传染病的数据模型为核心,从临床数据中心采集 5 年内 12 种呼吸道传染病的门诊和住院医疗数据,对传染病所涉及的变量进行全面梳理,形成传染病专病数据库,并建立数据自动增量更新的机制,保证传染病数据库数据的及时、准确和完整。

1.2 研究方法 结合医院历史传染病数据,组织感染科医生、医学信息学专家、技术公司共同对近千份传染病住院病历进行集中多轮标注及审核,构建半监督的机器学习模型,自动学习病历数据,形成电子病历结构化模型。以权威机构出版或发布的传染病医学教材及中英文文献为基础,结合传染病专病临床数据,采用三元组描述医学本体,并采用半自动化的方式抽取医学知识且对知识进行融合,形成以标准化术语为基础的传染病知识图谱。选取典型传染病的全量数据,先进行数据标化、清洗等预处理,采用特征分类和机器学习的方式进行模型训练,同时融合基于医学知识图谱的推理模型,并经过交叉验证后形成可用的传染病预测模型,最终与医院业务系统集成,以工具化形式与业务系统形成一个整体。

1.2.1 病历结构化建模 病历的特征提取采用半监督学习的方式将特征值降维到一个足够建模的小数据集^[11],半监督学习模型首先需要学习数据的内在结构以便合理的组织数据来进行预测,在此学习方式下,输入数据分为两部分,一部分是已经被标注的病历文书,另一部分则是未被标注的病历文书。对已标注的病历文书采用监督学习模型训练数据集,然后对未标记的数据集进行预测,将预测结果与训练数据的实际结果进行比较,不断的调整预测模型,直到模型的预测结果达到一个预期的准确率。对未标注的病历文书采用非监督学习,从无训练的原始数据中找到隐藏的模式或者关系。最后综合两部分得到结构化的病历特征。从病历文书中识别的特征实体类型包括疾病、症状、体征、检验结果、检查结果、体质、生活习惯等。见图 1。

1.2.2 构建传染病知识图谱 本文选取流行性感冒、人感染高致病性禽流感、肺结核、百日咳、白喉等通过呼吸道传染疾病的全量数据。在临床医生及专家的指导下,应用本体技术构建标准的以疾病概念为核心的医学知识图谱模型,并采用人工与机器自动化结合的方式抽取传染疾病医学知识,以三元组描述形式作为基础知识单元。通过人工的方式从诊疗指南、教材、文献等资料中抽取相应的由标准术语描述的医学实体,并基于知识图谱模型构建各类实体之间的语义关系^[12],保障与知识来源中的语义一致性。同时,基于已构建的病历结构化模型、专病库数据模型,分别从传染病专病库中非结构化的电子病历文书及结构化的诊疗过程数据中抽取相应的医学实体及其对应的诊断结果实体,再结合统计分析方法及关联分析方法从中挖掘出各传染病实体

的医学知识。最后,将前两步构建的知识图谱进行 的传染病知识图谱。相互校验、相互融合,最终形成符合理论和实践经验

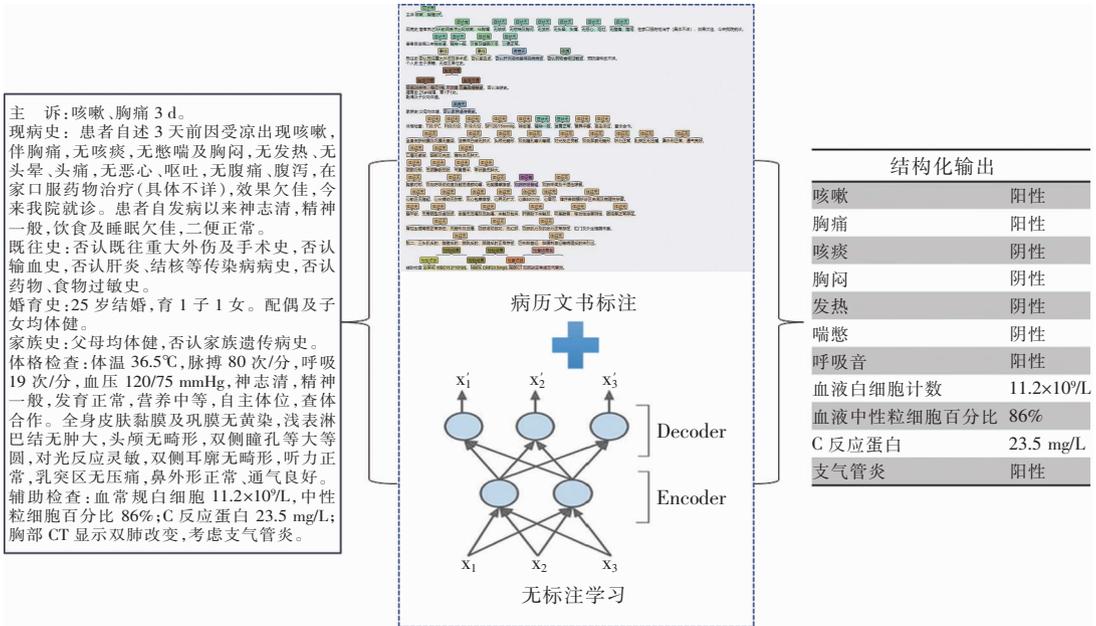


图 1 病历半监督学习模式

Figure 1 Semi-supervised learning mode of medical records

1.2.3 发现与预测模型 基于传染病专病库、电子病历结构化模型以及传染病知识图谱的构建,本文

提出了一种可疑传染病发现与预测模型。模型的整体执行流程如图 2 所示。

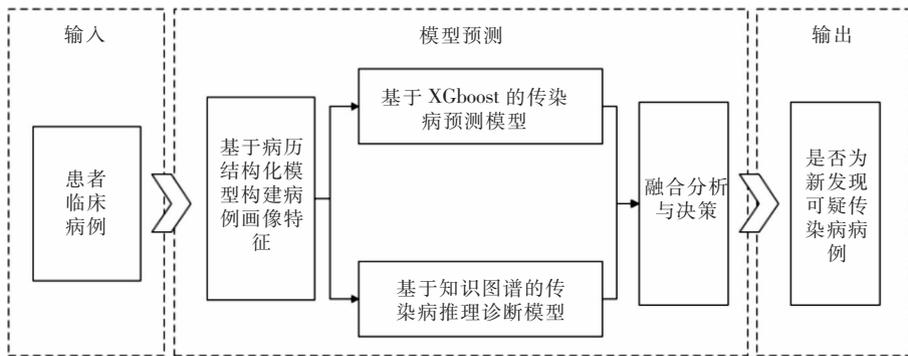


图 2 可疑传染病发现与预测模型整体流程

Figure 2 Overall process of the detection and prediction model for suspected infectious diseases

模型的输入为患者临床病例数据信息,该信息包括患者的基本信息以及患者临床诊疗过程数据。可疑传染病发现与预测模型是由多个子模型构建而形成的流程化发现与预测模型,其中包括:(1)病例画像特征构建模型。结合患者临床病例数据,提取患者基本信息,包括年龄、性别、体质等特征,同时利用病历结构化模型从病历文书中提取患者病例特征信息包括症状、体征、检验结果、检查结果、体质、生

活习惯等实体,构建病例特征画像。(2)基于 XG-Boost 的传染病预测模型。基于病例画像特征的构建,利用传染病专病库历史病例数据信息,构建分析样本,其中因变量为病例画像特征信息,自变量为是否传染病的标签,属于二分类的建模研究。本文采用 XGboost 算法^[13]对构建的病例画像特征及标签样本进行学习建模,目标函数采用 binary:logistic 构建二分类模型。基于该模型可以对病例画像特征

进行预测,预测结果为“是”(表示该病例疑似为传染病)或“否”(表示该病例不是传染病)。在此分类模型中,还可以得出各特征的重要性分析,特征重要性是通过数据集的每个特征进行计算,并进行排序得到,特征重要性越高,其对结果的影响越大。(3)基于知识图谱的传染病诊断模型。同样基于构建的病例画像特征,结合传染病知识图谱,利用知识图谱推理技术,来分析当前病例特征可诊断的传染疾病。基于临床诊断思维,分别从疾病的可解释性以及可诊断性两方面进行分析,采用综合评分方法来为候选疾病 d 进行得分计算,其中可解释性 explanatory 计算如下:

$$\text{explanatory}_{s,d} = \frac{|S \cap G_d|}{|G_d|}$$

其中 S 表示当前病例中的特征集合, G_d 表示知识图谱 G 中疾病实体 d 关联的实体特征集合。若病例特征集合包含了疾病 d 在知识图谱中关联的实体特征集合,那么可解释性(explanatory)为 1,即表示当前病例特征可以完全解释当前疾病 d 。同时,疾病的可诊断性(diagnostic)计算如下:

$$\text{diagnostic}_{s,d} = \frac{\sum_x^S |G_x \rightarrow d|}{\sum_x^S |G_x \rightarrow ?|}$$

其中 $x \in S$, $G_x \rightarrow ?$ 表示特征实体 x 与其他任意疾病存在诊断关系的知识集合, $G_x \rightarrow d$ 表示特征实体 x 与疾病 d 存在诊断关系的知识集合。可诊断性越高,表示当前特征集合 S 指向疾病 d 的特异性越强。

基于病例画像特征集合 S 与疾病 d 的可解释性与可诊断性的计算,进一步提出疾病诊断推荐综合评分计算方法如下:

$$\text{score}_{s,d} = \frac{2 \times \text{explanatory}_{s,d} \times \text{diagnostic}_{s,d}}{\text{explanatory}_{s,d} + \text{diagnostic}_{s,d}}$$

即给定病例画像特征集合 S ,针对候选疾病 d 可计算出其诊断推荐评分 $\text{score}_{s,d}$ 。综合评分 score 的取值范围为 0~1,值越大即表示该病的诊断可靠性越高,基于该评分可实现对疑似传染病诊断结果进行排序。(4)合并分析与决策。综合传染病预测模型以及传染病诊断模型的分析结果,若传染病预测模型预测当前病例为“是”,即该病例为疑似传染病,同时传染病诊断模型未诊断出结果,即传染病诊断模型中最高得分的诊断结果低于一个给定的阈值 α (默认为 0.5),那么此时,即认为当前病例为新发现的可疑传染病并输出,否则认为未发现新的可疑传染病。

1.3 评估指标 在本文中,将采用查准率(precision)、查全率(recall)以及 $F1$ 值(fscore)来进行评估,评估值是通过预测结果计算得出的,true 与 false 代表着样本是否预测正确,positive 和 negative 表示为样本的预测结果——被预测为正样本或者负样本。根据两两搭配的四组组合方式,分别为:真正(true positive, TP)、假正(false positive, FP)、真负(true negative, TN)和假负(false negative, FN)。

查准率是指预测的正样本被预测正确多少。

$$\text{precision} = \frac{TP}{TP + FP}$$

查全率是指实际的正样本被预测正确多少。

$$\text{Recall} = \frac{TP}{TP + FN}$$

综合查准率与查全率,采用指标 $F1$ 值进行评估。

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

1.4 统计学方法 应用 SPSS 软件(版本 22.0)对电子病历的特征进行统计分析,采用卡方检验对不同症状、不同部位、不同检查结果和不同疾病的主体进行比较评价。

2 结果

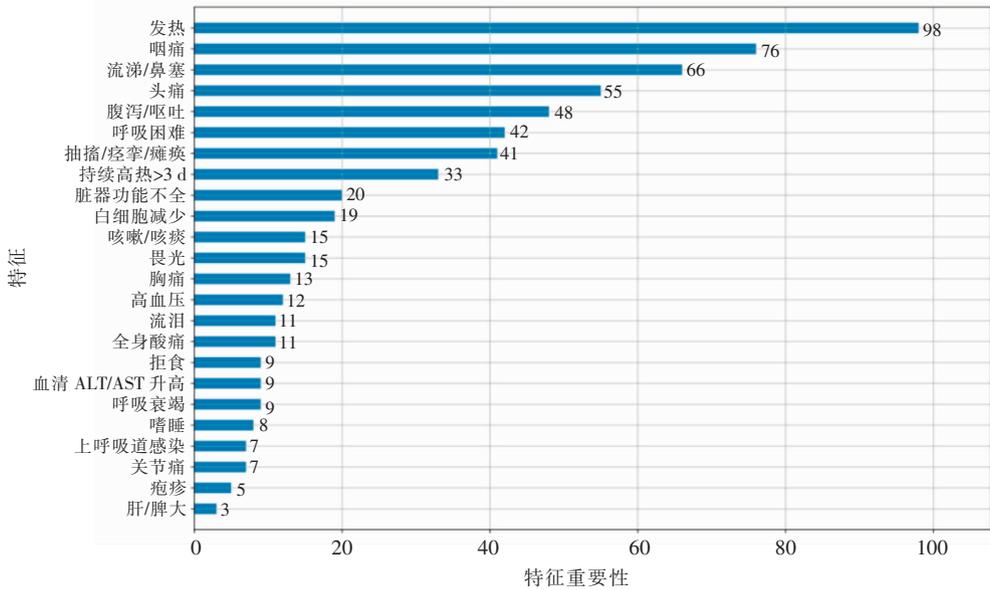
2.1 基本数据 本文数据来自真实医疗机构电子病历中产生的诊疗数据,纳入 2016—2021 年的流行性感、人感染高致病性禽流感、肺结核、百日咳、白喉等 12 个病种共计 2 035 例病例,同时获取 2 300 份非传染病病例,共计 4 335 份病例数据。为了验证模型的有效性,基于交叉验证方法构建了 10 套数据集,每套数据集中对所有病例数据按照 7:3 进行随机拆分^[14],分别用于模型训练和测试。

2.2 基于 XGboost 模型的预测结果 本试验采用 python 语言开发 XGboost 模型,并针对 10 个数据集中的训练集和测试集分别进行训练和验证,试验预测结果如表 1 所示。采用 XGboost 模型对数据集进行十次交叉验证,模型平均查准率为 91.88%,查全率为 88.06%,综合 $F1$ 值达到 0.899 1,模型整体表现效果良好。此外进一步基于 XGboost 模型还可以实现特征的重要性分析,如图 3 所示,其中 TOP10 重要特征分别为发热、咽痛、流涕/鼻塞、头痛、腹泻/呕吐、呼吸困难、抽搐/痉挛/瘫痪、持续高热 > 3 d、脏器功能不全、白细胞异常。

表 1 基于 XGboost 的模型预测结果

Table 1 Model prediction results based on XGboost

序号	真实情况		预测情况		TP	FN	FP	TN	Precision	Recall	F1
	正例	反例	正例	反例							
1	620	680	588	712	540	80	48	632	0.918 4	0.871 0	0.894 0
2	605	695	585	715	542	63	43	652	0.926 5	0.895 9	0.910 9
3	610	690	593	707	547	63	46	644	0.922 4	0.896 7	0.909 4
4	598	702	591	709	535	63	56	646	0.905 2	0.894 6	0.899 9
5	600	700	587	713	532	68	55	645	0.906 3	0.886 7	0.896 4
6	615	685	576	724	533	82	43	642	0.925 3	0.866 7	0.895 0
7	630	670	569	731	529	101	40	630	0.929 7	0.839 7	0.882 4
8	601	699	587	713	539	62	48	651	0.918 2	0.896 8	0.907 4
9	595	705	588	712	533	62	55	650	0.906 5	0.895 8	0.901 1
10	625	675	580	720	539	86	41	634	0.929 3	0.862 4	0.894 6



注:ALT 为丙氨酸转氨酶;AST 为天冬氨酸转氨酶。

图 3 特征重要性分析

Figure 3 Analysis on characteristic importance

2.3 基于知识图谱的模型预测结果 采用 python 语言实现基于知识图谱的诊断模型,对 10 次随机数据集中的测试集进行验证,其试验结果详情如表 2 所示。采用基于医学知识图谱的诊断推理模型对测试集进行验证,模型平均查准率为 90.9%,查全率为 86.74%,综合 F1 值为 0.887 6。模型整体识别效果良好,但是略低于 XGboost 模型。

2.4 融合模型预测结果 综合机器学习及知识图谱技术,将 XGboost 预测模型及知识图谱诊断模型

进行融合形成合并分析决策模型,结果如表 3 所示。在融合了基于 XGboost 的预测模型及基于知识图谱的推理模型后,模型的平均查准率为 92.55%,查全率为 91.49%,综合 F1 值为 0.920 1。融合后的模型预测效果均优于前两者。同时,通过比较 XGboost 模型、知识图谱诊断模型、合并分析决策模型的受试者工作特征曲线(ROC 曲线)可知,合并分析决策模型的性能也优于其他两种模型,见图 4。

表 2 基于知识图谱的诊断推理结果

Table 2 Diagnostic reasoning results based on knowledge mapping

序号	真实情况		预测情况		TP	FN	FP	TN	Precision	Recall	F1
	正例	反例	正例	反例							
1	620	680	591	709	534	86	57	623	0.903 6	0.861 3	0.881 9
2	605	695	582	718	527	78	55	640	0.905 5	0.871 1	0.888 0
3	610	690	599	701	547	63	52	638	0.913 2	0.896 7	0.904 9
4	598	702	589	711	526	72	63	639	0.893 0	0.879 6	0.886 3
5	600	700	583	717	524	76	59	641	0.898 8	0.873 3	0.885 9
6	615	685	576	724	528	87	48	637	0.916 7	0.858 5	0.886 6
7	630	670	572	728	525	105	47	623	0.917 8	0.833 3	0.873 5
8	601	699	581	719	531	70	50	649	0.913 9	0.883 5	0.898 5
9	595	705	569	731	517	78	52	653	0.908 6	0.868 9	0.888 3
10	625	675	577	723	530	95	47	628	0.918 5	0.848 0	0.881 9

表 3 合并分析决策模型预测结果

Table 3 Prediction results of combined analysis decision model

序号	真实情况		预测情况		TP	FN	FP	TN	Precision	Recall	F1
	正例	反例	正例	反例							
1	620	680	606	694	563	57	43	637	0.929 0	0.908 1	0.918 4
2	605	695	593	707	549	56	44	651	0.925 8	0.907 4	0.916 5
3	610	690	599	701	551	59	48	642	0.919 9	0.903 3	0.911 5
4	598	702	589	711	544	54	45	657	0.923 6	0.909 7	0.916 6
5	600	700	603	697	558	42	45	655	0.925 4	0.930 0	0.927 7
6	615	685	605	695	568	47	37	648	0.938 8	0.923 6	0.931 1
7	630	670	627	673	585	45	42	628	0.933 0	0.928 6	0.930 8
8	601	699	591	709	541	60	50	649	0.915 4	0.900 2	0.907 7
9	595	705	597	703	553	42	44	661	0.926 3	0.929 4	0.927 9
10	625	675	619	681	568	57	51	624	0.917 6	0.908 8	0.913 2

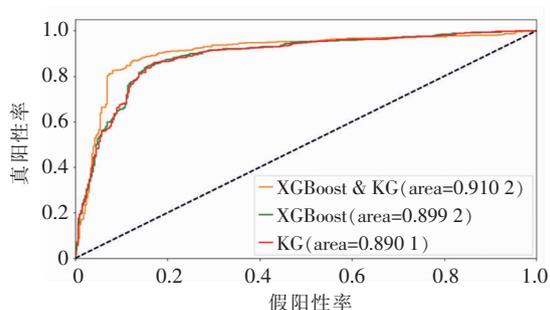


图 4 三类模型的 ROC 曲线对比

Figure 4 Comparison of the ROC curves of three types of models

2.5 信息系统集成 将训练好的模型与医院的电子病历系统和临床辅助决策系统进行集成,合理设置触发节点,当患者关键信息形成后,电子病历系统将患者基本信息传递给临床辅助决策系统,系统从数据汇聚库和临床数据中心自动抽取检验、检查等相关指标并输入发现与预测模型,如果模型计算结果符合设定的阈值,则弹出提示警告,并引导相关管理部门进行上报,见图 5。



图 5 与医疗信息系统集成

Figure 5 Integration with medical information system

3 讨论

本文的创新性表现在将可疑呼吸道传染病的发现通过信息化手段前移至临床一线,克服了既往模式下基于国家或地区传染病预警系统内置规则判断是否传染病的时限性和准确性问题,有利于医院快速启动医院感染防控应急预案,提前筹备防控物资;其先进性在于利用医院临床数据中心的基础数据支撑和计算能力支撑,不依赖特定诊断,而是通过分析多维度的临床特征数据,构建了符合呼吸道传染病特点的发现预警模型,并提出合并传染病预测模型和传染病诊断模型的分析与决策模型,可以通过调节阈值形成符合数据规律的最佳决策方案。目前基于多种呼吸道传染病症状、体征、临床特征建立的呼吸道可疑传染病发现与预测模型,以医院历史的发热门诊患者最终确诊为传染病的患者进行模型验证,通过对可疑传染病的预测,辅助医院管理者及时启动传染病应急预案,很大程度上减少在传染病发生早期医务人员的感染概率。然而,本模型也存在一定的局限性,其平均查准率、查全率均超过 90%,综合 F1 值超过 0.92,这得益于模型中的临床数据均来自同一所医院,并且经过了严格的数据标准化和结构化,病历的书写习惯和名词做了有针对性的预训练,使得模型的预测效果较好。如果推广至其他医疗机构,需要利用其既往的临床数据再次进行多轮模型训练,并调整分析与决策模型的阈值,同时,还要与电子病历等临床应用系统对接,才能发挥模型的发现与预测作用。

本研究进一步计划首先将知识图谱扩充至 20 种传染病,并分为呼吸道传染病、消化道传染病、血液传染病和体表传染病四大类分别构建模型^[15],不断扩展可疑传染病预测的知识图谱知识,将检查、检

验等知识融合,从疾病预防控制中心、区域人口健康信息平台等渠道获取更多的数据,将更多传染病预测模型进行融合后应用到临床及管理决策部门。其次,与地理信息系统相结合进行可疑传染病时空预警模型算法研究^[16-17],提供传染病时空聚集性的智能分析,探索多模型的综合预警,通过模拟手段评估医院通行限制、病区防控和医疗资源调配等措施的效果。此外,还可以与医院互联网平台结合,开设传染病信息快讯类模块,提升公众主动查询传染病预警信息的便捷性^[18],同时也作为收集社会信息的渠道,丰富模型数据维度,提升模型的发现和预警效率。

利益冲突:所有作者均声明不存在利益冲突。

[参考文献]

[1] 任南,文细毛,吴安华. 全国医院感染监测与数据直报系统的研制及使用[J]. 中国感染控制杂志, 2008, 7(3): 170-172.
Ren N, Wen XM, Wu AH. Development and application of national nosocomial infection surveillance and data reporting system[J]. Chinese Journal of Infection Control, 2008, 7(3): 170-172.

[2] 杨维中,兰亚佳,李中杰,等. 国家传染病自动预警系统的设计与应用[J]. 中华流行病学杂志, 2010, 31(11): 1240-1244.
Yang WZ, Lan YJ, Li ZJ, et al. The application of national outbreak automatic detection and response system, China[J]. Chinese Journal of Epidemiology, 2010, 31(11): 1240-1244.

[3] 余斐,张洪龙,赖圣杰,等. 国家传染病自动预警系统在基层应用效果初步分析[J]. 中华流行病学杂志, 2011, 32(5): 446-449.
Yu F, Zhang HL, Lai SJ, et al. The effectiveness of China infectious disease automated-alert and response system (CIDARS) in the local regions[J]. Chinese Journal of Epidemiology, 2011, 32(5): 446-449.

[4] 张洪龙,曾令佳,赖圣杰,等. 2016 年国家传染病自动预警信

- 息系统运行情况分析[J]. 疾病监测, 2018, 33(2): 159 - 167.
- Zhang HL, Zeng LJ, Lai SJ, et al. Performance of China infectious disease automated-alert and response system in 2016[J]. *Disease Surveillance*, 2018, 33(2): 159 - 167.
- [5] 夏祎祺, 陈非非, 马露, 等. 我国公立综合医院传染病预警系统建设现状与发展建议[J]. 中华医院管理杂志, 2022, 38(7): 544 - 547.
- Xia YQ, Chen FF, Ma L, et al. Status quo and development recommendations for infectious disease early warning system in public general hospitals in China[J]. *Chinese Journal of Hospital Administration*, 2022, 38(7): 544 - 547.
- [6] 马莉珍, 罗湘蜀, 杨长虹, 等. 2015 年四川省医疗机构传染病漏报情况调查[J]. 预防医学情报杂志, 2017, 33(3): 247 - 251.
- Ma LZ, Luo XS, Yang CH, et al. Assessment on reporting rates of notifiable infectious disease in medical institutions in Sichuan province in 2015[J]. *Journal of Preventive Medicine Information*, 2017, 33(3): 247 - 251.
- [7] 李桂芹. 医院传染病报告的质量管理[J]. 中国病案, 2014, 15(12): 57 - 58, 68.
- Li GQ. Quality management of hospital infectious disease report[J]. *Chinese Medical Record*, 2014, 15(12): 57 - 58, 68.
- [8] 张丽, 韩月红. 北京安贞医院传染病网络实时监测效果评价[J]. 传染病信息, 2018, 31(5): 463 - 465.
- Zhang L, Han YH. Effect evaluation of real-time surveillance network for infectious diseases in Beijing Anzhen Hospital[J]. *Infectious Disease Information*, 2018, 31(5): 463 - 465.
- [9] 冷姝芳, 王键, 杨晓青, 等. 基于医院信息系统的传染病监测报告系统的应用效果评价[J]. 上海交通大学学报(医学版), 2015, 35(12): 1888 - 1891.
- Leng SF, Wang J, Yang XQ, et al. Evaluation of application effect of infectious disease reporting system based on hospital information system[J]. *Journal of Shanghai Jiaotong University (Medical Science)*, 2015, 35(12): 1888 - 1891.
- [10] 段鹤李, 陈廷寅. 基于信息技术的临床科研数据平台建设[J]. 中国现代医学杂志, 2020, 30(9): 124 - 128.
- Duan JL, Chen TY. Construction of clinical scientific research data platform based on information technology[J]. *China Journal of Modern Medicine*, 2020, 30(9): 124 - 128.
- [11] 李村合, 张振凯, 朱洪波. 基于半监督学习的多示例多标签改进算法[J]. 电子技术应用, 2019, 45(7): 32 - 35, 39.
- Li CH, Zhang ZK, Zhu HB. A multi-instance multi-label improved algorithm based on semi-supervised learning[J]. *Application of Electronic Technique*, 2019, 45(7): 32 - 35, 39.
- [12] Chen TY, Hu YM. Entity relation extraction from electronic medical records based on improved annotation rules and BiLSTM-CRF[J]. *Ann Transl Med*, 2021, 9(18): 1415.
- [13] 周彤彤, 俞凯, 袁贞明, 等. 基于 LSTM 与 XGBOOST 混合模型的孕妇产后出血预测[J]. 计算机系统应用, 2020, 29(3): 148 - 154.
- Zhou TT, Yu K, Yuan ZM, et al. Predictive analysis of postpartum hemorrhage based on LSTM and XGBoost hybrid model[J]. *Computer Systems & Applications*, 2020, 29(3): 148 - 154.
- [14] 艾正琳, 洪珊, 胡居龙, 等. 食管胃静脉曲张治疗后再出血预测模型的建立与评价[J]. 临床肝胆病杂志, 2019, 35(9): 1954 - 1957.
- Ai ZL, Hong S, Hu JL, et al. Establishment and evaluation of a predictive model for rebleeding after endoscopic treatment of esophageal and gastric varices[J]. *Journal of Clinical Hepatology*, 2019, 35(9): 1954 - 1957.
- [15] Burkom H, Loschen W, Wojcik R, et al. Electronic surveillance system for the early notification of community-based epidemics (ESSENCE): overview, components, and public health applications[J]. *JMIR Public Health Surveill*, 2021, 7(6): e26303.
- [16] 裴韬, 王席, 宋辞, 等. COVID-19 疫情时空分析与建模研究进展[J]. 地球信息科学学报, 2021, 23(2): 188 - 210.
- Pei T, Wang X, Song C, et al. Review on spatiotemporal analysis and modeling of COVID-19 pandemic[J]. *Journal of Geo-information Science*, 2021, 23(2): 188 - 210.
- [17] 劳晓洁, 徐贤丽, 林琨, 等. 基于地理信息系统的一起腺病毒感染暴发疫情的调查处置[J]. 中国感染控制杂志, 2020, 19(10): 894 - 899.
- Lao XJ, Xu XL, Lin K, et al. Investigation and disposal of an outbreak of adenovirus infection based on geographic information system[J]. *Chinese Journal of Infection Control*, 2020, 19(10): 894 - 899.
- [18] Sharma D, Panigrahi A. Health alert network[J]. *Arch Community Med Public Health*, 2020, 6(1): 26 - 27.

(本文编辑:陈玉华)

本文引用格式:陈廷寅, 冯嵩. 基于医院临床数据中心的可疑呼吸道传染病发现与预测模型构建[J]. 中国感染控制杂志, 2023, 22(8): 964 - 971. DOI:10.12138/j.issn.1671-9638.20233547.

Cite this article as: CHEN Ting-yin, FENG Song. Construction of a detection and prediction model for suspected respiratory infectious diseases based on hospital clinical data center[J]. *Chin J Infect Control*, 2023, 22(8): 964 - 971. DOI: 10.12138/j.issn.1671-9638.20233547.