

• 论 著 •

基于数据挖掘的乳腺癌机会性筛查管理系统构建与应用

郭苗苗¹, 许丹丹², 严婷婷¹, 袁玲³

摘要: **目的** 基于数据挖掘构建乳腺癌机会性筛查管理系统并验证其在筛查患者管理中的应用效果。 **方法** 基于数据挖掘构建筛查管理系统,选取2021年1月至2024年5月进行乳腺癌机会性筛查的患者,按照系统应用前后分为对照组和观察组,每组各2 373例。观察组采用系统进行筛查管理,对照组采用常规的筛查管理方法。比较两组明确诊断和入院治疗时间,以及对乳腺癌机会性筛查满意度。 **结果** 观察组明确诊断、入院治疗时间显著短于对照组,其对乳腺癌机会性筛查满意度评分显著高于对照组(均 $P<0.05$)。 **结论** 应用基于数据挖掘的乳腺癌机会性筛查管理系统进行筛查管理,可缩短患者筛查至首次入院治疗的时间,促进患者早诊早治,提高患者满意度。

关键词: 乳腺癌; 机会性筛查; 数据挖掘; 个案管理; 随访管理; 患者满意度; 数智化护理

中图分类号: R473.6 **DOI:** 10.3870/j.issn.1001-4152.2025.05.104

Construction and application of a breast cancer opportunistic screening system based on data mining

Guo Miaomiao, Xu Dandan, Yan Tingting, Yuan Ling. Department of Oncology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing 210008, China

Abstract: **Objective** To construct a breast cancer opportunistic screening system based on data mining and verify its application effect. **Methods** The screening management system was constructed based on data mining. The patients who underwent breast cancer opportunistic screening from January 2021 to May 2024 were selected and divided into a control group and an observation group according to the system before and after application, with 2 373 cases in each group. The observation group used the system for screening management, while the control group adopted the conventional screening management methods. The time of definite diagnosis and hospitalization, and the satisfaction with opportunistic screening for breast cancer were compared between the two groups. **Results** The time for clear diagnosis and hospitalization in the observation group was significantly shorter than those in the control group, and its satisfaction score for breast cancer opportunistic screening was significantly higher than that in the control group (all $P<0.05$). **Conclusion** The application of breast cancer opportunistic screening system based on data mining for screening management can shorten the time from screening to first admission for treatment, promote early diagnosis and treatment of patients, and improve patients' satisfaction.

Keywords: breast cancer; opportunistic screening; data mining; case management; follow-up management; patient satisfaction; digital intelligent nursing

乳腺癌是女性最常见的恶性肿瘤,居女性肿瘤发病率的前三位^[1]。早发现、早诊断、早治疗是我国女性乳腺癌防治的三原则^[2]。乳腺癌筛查分为群体筛查和机会性筛查^[3],我国目前乳腺癌筛查以机会性筛查为主,是医疗保健机构为因各种原因前来就诊的适龄女性进行的乳腺筛查,或女性个体主动或自愿到提供乳腺筛查的医疗保健机构进行检查,具有筛查阳性率高、乳腺癌检出率高、活检阳性率高、筛查费用低的特点^[4-5]。随着乳腺癌发病率的不断上升,乳腺门诊机会性筛查的人数也在不断增加,如何优化就诊流程,使筛查患者尽快得到早诊早治,是医院急需解决的问题^[6]。数据挖掘(Data Mining)又称作数据库知

识发现(Knowledge Discovery from Database, KDD),是从数据中获取价值的一个过程,可以形式化地表示为“数据+工具+方法+目标+行动=价值”^[7],目前已广泛应用到医疗领域,能帮助医护人员从大量不完整、模糊、有噪声、随机的数据中提取出未知的、潜在有用的知识和信息^[8]。黄天海等^[9]采用数据挖掘等技术设计了早期胃癌筛查系统,实现了胃癌的早期筛查。我院利用数据挖掘技术,研发乳腺癌筛查管理系统,优化机会性筛查患者院内管理流程,取得了较好的效果。

1 资料与方法

1.1 一般资料 便利选取2021年1月至2024年5月于我院进行乳腺癌机会性筛查的患者,按照系统应用前后分为对照组(2021年1月至2022年5月)和观察组(2023年1月至2024年5月)。纳入标准:①初次到我院行乳腺癌机会性筛查的女性患者,包括乳腺X线摄片、超声、磁共振检查等;②年龄18~80岁;③自愿参加本研究,并签署知情同意书。排除标准:①并存其他恶性肿瘤;②有精神疾病或认知障碍难以配合研究;③有语言沟通障碍;④资料不完整。本研究经医

作者单位:南京大学医学院附属鼓楼医院 1. 肿瘤科 2. 普外科 3. 护理部(江苏 南京,210008)

通信作者:袁玲, yuanling@nju.edu.cn

郭苗苗:女,硕士,副主任护师, gm86@163.com

科研项目:南京鼓楼医院护理部,南京大学护理医学研究所重点课题(2023-A894);徐州医科大学2022年校级劳动教育专项资助项目(Xld202203)

收稿:2024-10-03;修回:2024-11-28

院伦理委员会批准(2023-047-01)。中国临床试验注册中心注册号:ChiCTR2300069635。在患者至我院初诊到确认乳腺癌入院治疗过程中,患者因穿刺或入院等待时间过长、家是外地及医保等原因中途至外院

就诊 76 例,其中观察组 17 例,对照组 59 例,最终两组各 2 373 例患者完成研究。对照组年龄(57.51 ± 15.88)岁,观察组(56.02 ± 16.26)岁,两组比较, $t = 1.357, P = 0.175$ 。两组一般资料比较,见表 1。

表 1 两组一般资料比较 例(%)

组别	例数	文化程度			婚姻状况				医疗付费方式			
		小学及以下	中学	大专及以上	未婚	已婚	离异	丧偶	公费	城镇保险	农村合作医疗	自费
对照组	2 373	241	1 158	974	69	2 101	45	158	18	1 405	925	25
观察组	2 373	238	1 186	949	57	2 135	39	142	25	1 423	912	13
统计量		$Z = -0.577$			$\chi^2 = 2.698$				$\chi^2 = 5.136$			
P		0.564			0.441				0.162			

组别	例数	家庭月收入(元)		BI-RADS 分级			居住地			基础疾病		
		<5 000	≥5 000	<4 类	4 类	5 类	本地	本省外地	外省	高血压	糖尿病	其他
对照组	2 373	127	2 246	2 025	310	38	451	253	1 669	245	105	36
观察组	2 373	119	2 254	2 040	302	31	474	238	1 661	237	118	25
统计量		$\chi^2 = 0.274$		$Z = -0.653$			$\chi^2 = 1.049$			$\chi^2 = 0.148$	$\chi^2 = 0.795$	$\chi^2 = 2.009$
P		0.600		0.514			0.592			0.701	0.373	0.197

1.2 干预方法

对照组患者根据医生建议进行乳腺癌影像检查,包括乳腺 X 线、超声、磁共振等,根据影像筛查结果自行至乳腺专科门诊就诊,根据专科医生建议如需穿刺活检则自行预约并自主查询病理结果,活检结果为恶性肿瘤患者自行至乳腺专科门诊预约住院。观察组构建并应用基于数据挖掘的乳腺癌筛查管理系统进行筛查管理,具体如下。

1.2.1 组建乳腺癌筛查管理团队 团队成员包括:护理部副主任、医务处处长各 1 名,负责项目的决策、协调与推进;乳腺外科主任医师、乳腺内科主任医师、病理科主任医师、门诊护士长、乳腺外科护士长、乳腺内科护士长各 1 名,为患者筛查各环节的业务负责人;信息中心工程师 1 名,负责系统开发及日常维护;个案管理师 1 名,负责协调系统开发过程中的各环节,以及利用系统进行筛查管理。

1.2.2 构建基于数据挖掘的乳腺癌筛查管理系统

1.2.2.1 系统硬件部署架构 ①内外网隔离网络部署:采用先进的网络隔离设备(物理隔离网闸及逻辑隔离防火墙)实现医院内网与外网的严格分离,有效阻断外部网络对内网的直接访问。内网负责存储与处理敏感的医疗数据,外网则提供医护端小程序、医护人员网页工作站及患者端小程序的访问接口。②数据交换机制:为确保内外网之间必要的数据交换,系统设计了安全的数据传输通道及数据交换协议。通过严格的数据加密、身份验证及访问控制机制,实现数据的单向或双向安全传输。③终端访问控制:医护端小程序、医护人员网页工作站及患者端小程序均通过严格的身份认证与访问控制机制接入系统。系统采用多因素认证技术,确保用户身份的真实性与合法性。同时,根据用户的角色与权限,动态调整其访问资源的范围与深度,增强系统的安全性与可控性。

1.2.2.2 乳腺癌筛查平台架构及数据挖掘 系统依

托专业医疗指南,综合利用大数据技术、AI 技术,为乳腺癌筛查提供更高效、准确、全面的技术手段。数据挖掘过程①形成乳腺癌专病数据集:以患者主数据(姓名、身份证等)为索引,从医院信息系统(Hospital Information System, HIS)、实验室(检验科)信息系统(Laboratory Information System, LIS)、放射科信息系统(Radiology Information System, RIS)、影像归档和通信系统(Picture Archiving and Communication System, PACS)等,采集乳腺癌患者就诊记录、检验报告、影像检查结果等信息,形成乳腺癌专病数据集。②制订乳腺癌筛查相关规则:对电子病历系统(Electronic Medical Record, EMR)中的内容进行“后结构化”处理,抽取关键病症实体,如“左乳 OR 右乳 OR 乳腺 OR 双乳”and“肿块 OR 结节 OR 占位”and“BI-RADS 分级:1~5 类”;并基于相关指南^[4,10],制订乳腺癌筛查相关规则,完成“患者临床特征”。同时,对医学数字成像和通信数据(Digital Imaging and Communications in Medicine, DICOM)进行图像分析,形成“影像病灶特征”。最后,综合两方特征,结合神经网络、决策树算法,形成最终的乳腺癌筛查模型。我院信息中心将系统筛查结果与医学影像存档与 PACS 中的结论进行比对,充分证明该系统的有效性后即正式上线。

1.2.2.3 乳腺癌筛查管理系统模块 ①患者档案管理模块:以身份证号作为患者主索引,将当前患者的院内诊疗数据聚合在患者档案详情中,患者档案包括基础档案及管理记录两个部分;可根据筛查管理需要,在患者档案中添加相应的标签和分组。②随访中心模块:系统通过设定随访对象、随访内容、执行方式(短信推送、小程序推送)、随访日期等基本要素,以此来创建随访任务,向患者推送和收集随访数据,动态掌握患者病情。

1.2.3 基于数据挖掘的乳腺癌筛查系统应用 ①每

日凌晨系统自动筛查在我院进行机会性筛查的患者。②个案管理师查看患者档案为每例患者贴上电子标签,进行分类管理;BI-RADS 分级 1~2 类,创建随访任务,通过小程序和短信方式提醒患者定期筛查;BI-RADS 分级 3 类,创建随访任务,通过小程序和短信方式提醒患者 3 个月后复查;BI-RADS 分级 4~5 类,创建随访任务,通过小程序、短信和电话方式提醒患者至乳腺专科门诊明确诊断。③患者通过查看微信小程序提醒、手机短信或电话提醒后到院复诊,个案管理师及时关注系统待随访列表,BI-RADS 分级 4~5 类患者若未按时复诊,再次打电话提醒,询问原因并帮助其解决问题,记录在管理档案中。④优化就诊流程。患者复诊后看诊医生在系统团队界面通知个案管理师,若需要进一步穿刺协助其走绿色通道,若无需穿刺则根据医生意见创建随访任务定期复查。⑤诊断期管理。个案管理师关注患者病理情况,若病理阴性则创建随访任务通知患者专科门诊就诊决定下一步诊疗计划;若病理阳性即时电话通知患者,根据患者意愿通知团队医生为其安排床位。

1.3 评价方法 由经过统一培训的 1 名调查员进行数据收集,收集方法包括面对面、电话、全病程管理系统线上沟通等,所有资料及时录入 Excel 表,保证数据完整无漏项。①明确诊断的时间:初次就诊到明确是否为乳腺恶性肿瘤的天数。②入院治疗时间:患者穿刺病理确诊乳腺恶性肿瘤到首次收治入院的预约

表 3 两组乳腺癌机会性筛查满意度比较

组别	例数	就诊效率	就诊流程	服务质量	服务态度	问题解决能力	响应速度	医患沟通流畅性	总分
对照组	2 373	5.91±1.56	5.91±1.56	5.47±2.12	5.87±1.75	5.86±1.77	5.70±1.83	5.97±1.60	40.68±5.55
观察组	2 373	9.04±1.06	9.06±0.99	8.75±0.62	8.75±0.64	8.77±0.73	9.03±0.92	8.89±0.91	62.30±1.81
<i>t</i>		32.671	32.217	27.758	29.103	28.836	31.110	30.739	69.457
<i>P</i>		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

3 讨论

3.1 基于数据挖掘的乳腺癌筛查管理系统可快速准确识别筛查阳性患者 本研究结果表明,观察组和对照组的筛查阳性率分别为 14.03%和 14.66%,这与国内其他研究结果类似^[13]。目前我国乳腺癌筛查最主要的方式,还需依靠主动机会性筛查^[14]。北欧国家 2017—2020 年乳腺癌相关数据表明,早筛早诊可提高乳腺癌患者的 10 年生存率^[15]。目前各级医疗机构的乳腺癌机会性筛查干预机制不明确、可操作性不强,如我院两院区每个月 1 000 多人进行机会性筛查,若单靠人工进行随访追踪,其工作量过于庞大。医护人员只有依靠信息化并提高患者筛查期间的管理水平,才能满足筛查管理需求。赵秀芬等^[16]通过数据挖掘找出鼻肠管非计划性拔管的高风险指标并进行风险管理,非计划性拔管率从 2.22%降至 1.45%。本研究通过院内数据挖掘分析,并根据 BI-RADS 分级等数据建立有价值的强关联规则,将临床中难以察觉的规律进行可视化呈现,形成个人的电子

档案。③患者满意度调查:根据相关文献^[11-12]自制乳腺癌机会性筛查满意度调查表,包括就诊效率、就诊流程、服务质量、服务态度、问题解决能力,响应速度,医患沟通流畅性 7 个条目,每个条目 0~10 分,总分 0~70 分,得分越高表示满意度越好。

1.4 统计学方法 数据录入统计软件时由双人核对,采用 SPSS19.0 软件进行统计分析,分类变量采用频数、构成比进行描述,采用 χ^2 检验进行统计;等级资料采用秩和检验;连续变量采用 $(\bar{x} \pm s)$ 进行描述,采用独立样本 *t* 检验进行统计。检验水准 $\alpha=0.05$ 。

2 结果

2.1 两组明确诊断和入院治疗时间比较 本研究观察组 333 例(14.03%)筛查阳性(BI-RADS 分级 ≥ 4 类),对照组 348 例(14.66%)筛查阳性,两组进一步确诊及入院治疗时间比较,见表 2。

表 2 两组明确诊断和入院治疗时间比较

组别	明确诊断		入院治疗	
	例数	时间	确诊	时间
对照组	2 373	10.72±2.56	101	6.16±1.57
观察组	2 373	6.13±1.42	92	4.06±0.99
<i>t</i>		-30.249		-22.06
<i>P</i>		<0.001		<0.001

2.2 两组乳腺癌机会性筛查满意度比较 见表 3。

分, $\bar{x} \pm s$

档案,为更为高效地筛查干预提供可能性^[17]。

3.2 应用乳腺癌机会性筛查管理系统可缩短患者诊断及入院治疗时间 本研究结果显示,观察组明确诊断和入院治疗时间显著短于对照组(均 $P<0.05$),与国内其他研究结果相似^[6]。我院原有的乳腺癌日常诊疗以治病为中心,筛诊流程不连贯,导致患者拿到影像报告后缺乏专业诊疗指导,往往会贻误最佳治疗时机。本研究团队通过系统及时识别高危人群并进行流程再造,个案管理师快速提炼出与决策相关的信息,通过系统创建随访任务,以线上、短信、电话等形式对患者的下一步诊疗作出提醒,督促筛查阳性的患者及时就医,提升工作效率的同时也提高患者自我管理能力^[18],在穿刺活检和入院时必要时走绿色通道,大大缩短了患者从筛查到首次入院治疗的时间,有助于提高患者的生存期^[19]。

3.3 应用乳腺癌机会性筛查管理系统可提高乳腺癌患者满意度 本研究结果表明,观察组患者对于筛诊期管理的就诊效率、就诊流程、服务质量、服务态度、

问题解决能力、响应速度、医患沟通流畅性以及总满意度显著优于对照组(均 $P<0.05$),这与国内一项基于智能平台的门诊管理研究结果类似^[20]。从脱落的病例来看,观察组筛查期间至外院就诊病例数少于对照组。随访管理不完善可能是影响患者筛查行为的障碍因素,我院原有的筛查患者在门诊就诊后缺乏连续性的随访,本研究通过系统的应用帮助医护人员识别需要随访的人群,个案管理师为患者建立档案并分类、主动管理,及时解决患者在门诊筛查诊断过程中遇到的问题,解答其困惑,使就医效率提高,沟通更加顺畅,提升患者就医体验感,从而提高患者满意度。

4 结论

本研究利用数据挖掘技术将患者影像报告、检验报告、就诊信息等数据进行有效分析和管理的,促进了乳腺癌患者的早诊早治并提高了患者的满意度,但本研究随访时间较短,对筛查阴性患者的远期影响尚未探究。利用信息技术助力乳腺癌筛查已经成为一种趋势,未来有望通过数据挖掘整合各级医院、社区卫生中心及三甲医院资源,以实现社区和医院的上下联动,从而提高乳腺癌筛查率和筛查水平,构建一体化肿瘤随访及健康管理平台。

参考文献:

[1] 尹周一,王梦圆,游伟程,等. 2022 美国癌症统计报告解读及中美癌症流行情况对比[J]. 肿瘤综合治疗电子杂志,2022,8(2):54-62.

[2] 成娟. 天津市某社区女性乳腺癌机会性筛查开展情况及相关因素分析[D]. 天津:天津医科大学,2018.

[3] Quintin C, Chatignoux E, Plaine J, et al. Coverage rate of opportunistic and organised breast cancer screening in France: department-level estimation [J]. *Cancer Epidemiol*,2022,81:102270.

[4] 中华医学会外科学分会乳腺外科学组,刘森,王殊,等. 中国浸润性乳腺癌诊治临床实践指南(2022 版)[J]. 中国实用外科杂志,2022,42(2):121-127.

[5] 沈松杰,徐雅莉,周易冬,等. 中国女性乳腺癌群体筛查与机会性筛查的比较研究[J]. 中华外科杂志,2021,59(2):109-114.

[6] 张雪,陈华英,吴绍勇. 多学科团队全程管理模式建立门诊乳腺癌绿色通道 的实践[J]. 西南国防医药,2020,30(5):479-482.

[7] 邢海龙. 大数据联盟数据挖掘服务模式研究[D]. 哈尔

滨:哈尔滨理工大学,2020.

[8] Kaur I, Doja M N, Ahmad T. Data mining and machine learning in cancer survival research:an overview and future recommendations[J]. *J Biomed Inform*,2022,128:104026.

[9] 黄天海,褚永华,张桦. 基于计算机辅助决策的早期胃癌筛查系统的设计[J]. 中国医院建筑与装备,2021,22(11):96-99.

[10] 中国研究型医院学会乳腺专业委员会中国女性乳腺癌筛查指南制定专家组. 中国女性乳腺癌筛查指南(2022 年版)[J]. 中国研究型医院,2022,9(2):6-13.

[11] 陈婕君,李金花,李旭英,等. 乳腺癌日间化疗患者基于信息化平台的预住院管理[J]. 护理学杂志,2020,35(20):6-9.

[12] 杨林宁,杨艳,胡嘉乐,等. 决策辅助工具用于乳腺癌患者的研究进展[J]. 护理学杂志,2020,35(2):110-113.

[13] 刘潇霞,乔良,钟志刚,等. 2019—2023 年度四川省城市癌症早诊早治项目人群风险评估及筛查分析[J]. 中国肿瘤,2024,33(7):542-549.

[14] 郭香娣,赵丹,陈佳娣. 围绝经期女性乳腺癌机会性筛查影响因素分析及对策[J]. 中国妇幼保健,2022,37(20):3867-3870.

[15] Zitricky F, Försti A, Hemminki A, et al. Conditional survival in breast cancer up to 10 years in the Nordic countries[J]. *Cancer Med*,2023,12(17):17945-17951.

[16] 赵秀芬,梁世耀,李影,等. 数据挖掘在降低住院患者鼻肠管非计划拔管率的应用与效果分析[J]. 医院管理论坛,2021,38(4):51-55,94.

[17] Subrahmanya S V G, Shetty D K, Patil V, et al. The role of data science in healthcare advancements: applications, benefits, and future prospects[J]. *Ir J Med Sci*,2022,191(4):1473-1483.

[18] 张希,王峥嵘,胡梦云,等. 个案管理对甲状腺相关眼病患者自我管理效能及并发症的影响[J]. 护理学杂志,2024,39(13):40-43.

[19] Muratov S, Canelo-Aybar C, Tarride J E, et al. Monitoring and evaluation of breast cancer screening programmes:selecting candidate performance indicators[J]. *BMC Cancer*,2020,20(1):2-10.

[20] 洪学敏. 基于智能平台的管理方案在妇科门诊护理管理中的价值分析[J]. 医学理论与实践,2024,37(11):1973-1975.

(本文编辑 钱媛)

(上接第 103 页)

[18] 夏丽霞,林征,顾则娟,等. 基于真实世界数据的住院患者跌倒成本效益研究[J]. 中国实用护理杂志,2022,38(22):1728-1734.

[19] Cuttler S J, Barr-Walker J, Cuttler L. Reducing medical-surgical inpatient falls and injuries with videos, icons and alarms[J]. *BMJ Open Qual*,2017,6(2):e000119.

[20] Wright K, Singh S. Reducing falls in dementia inpatients using vision-based technology[J]. *J Patient Saf*,2022,18(3):177-181.

[21] Collins C E, Schultz K, Mathew P, et al. A personalized approach empowering successful aging: patient perspective on fall prevention education[J]. *PM R*,2022,14(7):786-792.

[22] Dykes P C, Burns Z, Adelman J, et al. Evaluation of a patient-centered fall-prevention tool kit to reduce falls and injuries:a nonrandomized controlled trial[J]. *JAMA Netw Open*,2020,3(11):e2025889.

(本文编辑 钱媛)