

◎热点与综述◎

基于机器学习的医疗决策支持系统综述

梁书彤^{1,2}, 郭茂祖^{1,2}, 赵玲玲³

1. 北京建筑大学 电气与信息工程学院, 北京 100044

2. 建筑大数据智能处理方法研究北京市重点实验室, 北京 100044

3. 哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001

摘要: 利用健康医疗领域的海量临床数据进行辅助医疗决策支持是智慧医疗的核心技术和必然的发展趋势。医疗决策支持主要包括疾病风险预测与疾病智能诊断两方面, 以临床积累和实时获取的多种数据来源为基础, 通过多种机器学习算法实现对患者疾病类型的分类或者对患病风险的预测。从医疗决策支持的概念和方法框架出发, 按照不同疾病种类, 总结了当前采用的机器学习诊断和预测方法, 着重介绍这些方法的特点和区别, 并对存在的挑战和未来发展进行分析。

关键词: 医疗决策支持; 机器学习; 疾病智能诊断; 疾病风险预测

文献标志码: A **中图分类号:** TP181 doi: 10.3778/j.issn.1002-8331.1903-0485

梁书彤, 郭茂祖, 赵玲玲. 基于机器学习的医疗决策支持系统综述. 计算机工程与应用, 2019, 55(19): 1-11.

LIANG Shutong, GUO Maozu, ZHAO Lingling. Survey on medical decision support systems based on machine learning. Computer Engineering and Applications, 2019, 55(19): 1-11.

Survey on Medical Decision Support Systems Based on Machine Learning

LIANG Shutong^{1,2}, GUO Maozu^{1,2}, ZHAO Lingling³

1. College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

2. Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing 100044, China

3. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

Abstract: Using the massive clinical data in the field of health care to support the decision making of auxiliary medicine, is the core technology and an inevitable development trend of intelligent medicine. Medical decision support mainly includes disease risk prediction and intelligent diagnosis. Based on clinical accumulation and real-time acquisition of multiple data sources, it realizes classification of patients' disease types or prediction of disease risks through multiple machine learning algorithms. Based on the concept and method framework of medical decision support, this paper summarizes the current machine learning diagnosis and prediction methods according to different disease types, introduces the characteristics and differences of these methods, and analyzes the existing challenges and future development.

Key words: medical decision support; machine learning; intelligent diagnosis of disease; disease risk prediction

1 引言

近年来,随着ICT技术的迅猛发展,在医疗健康领域产生了大量关于患者的临床检查、治疗报告、医院资源管理记录、电子病历、药品等复杂数据^[1]。许多疾病,基金项目:国家自然科学基金(No.61871020, No.61305013);北京市教委科技计划重点项目(No.KZ201810016019);北京市属高校高水平创新团队建设计划项目(No.IDHT20190506)。

作者简介: 梁书彤(1995—),男,硕士研究生,研究领域为机器学习、智慧城市, E-mail: 446105134@qq.com; 郭茂祖(1966—),男,博士,教授,博士生导师,主要研究方向为机器学习、智慧城市、生物信息学; 赵玲玲(1980—),女,博士,讲师,主要研究方向为城市计算与智能信息处理。

收稿日期: 2019-04-01 **修回日期:** 2019-06-28 **文章编号:** 1002-8331(2019)19-0001-11

CNKI网络出版: 2019-08-21, <http://kns.cnki.net/kcms/detail/11.2127.tp.20190813.1420.006.html>

尤其是慢性病,起病隐匿,病程时间长,病因复杂。传统的医疗决策方法往往难以对这类疾病进行精准分析和诊断,影响了治疗效果。有报告称我国因慢性病死亡的人数已经占到总死亡人数的86.6%,导致的疾病负担已经占据总疾病负担的70%。大量的医疗数据需要更有效的提取和处理方法,通过机器学习技术正确诊断疾病,预测患者患病或病变的概率,将相关领域的人类专家知识与计算机的数据分析处理潜力之间取得平衡,从而获得最佳的疾病诊断结果,改善疾病预防、诊断和研究的现状。作为智慧医疗的核心技术,医疗决策支持将IT、人工智能与医疗相融合,对各类疾病的提前预防、病患护理都有着巨大的益处。

医疗过程具有知识专业性强、复杂多样化的特点,医生在诊疗过程中需要诸多信息作为参考依据。医疗行业产生的海量医疗数据,大部分是在医疗机构(如医院、医疗中心)产生的,具有结构复杂、信息量大等特点,不利于医生有效率、高精度地诊断。通过将计算机技术引入医疗行业,建立医疗决策支持系统,能在诊断、治疗、检查、药物成本这一系列诊疗全过程中为医生提供帮助,为医患双方都提供了极大的便利。

基于机器学习的医疗决策支持主要包括疾病风险预测与疾病智能诊断两方面,在多个领域的应用已经广泛开展:谷歌AI在乳腺癌病理识别准确率上已经超过医生^[2];自学习式人工智能已经可协助预测心脏病发作;人工智能已经可以通过分析医疗图像数据进行皮肤癌筛查^[3];在肝炎、登革热等传染性疾病诊断上具有极高的准确率,有效帮助区域卫生安全的维护。

2 医疗决策支持系统概述

医疗决策支持系统(Decision Support System, DSS)通过机器学习、人工智能的方法对结构多样的医疗数据及相关专业知识进行分析推理,从而辅助医生进行疾病的诊断决策或风险预测。医疗决策支持系统可以分为两类:模糊规则系统^[4]和基于机器学习的医疗决策支持系统。区别在于模糊规则系统可以对一些不宜量化的信息,比如某种可由多种疾病引起的症状进行规则提取;另如文献[5]所述,模糊决策支持系统可以从相关领域的人类专家提取规则,这意味着即使没有足够数量的样本,模糊决策支持系统也能较好地完成任务,而机器学习方法需要通过样本对选择的分类模型或预测模型进行训练,再利用训练好的模型进行疾病的诊断或风险预测。本文以基于机器学习的医疗决策支持系统为核心进行介绍。

医疗决策支持主要有疾病智能诊断和疾病风险预测两方面,主要原理是根据已收集到的医疗数据,利用各类机器学习方法构建模型,进行数据挖掘。数据挖掘是识别数据中潜在的、真实的、有价值的知识的重要过

程,本质是挖掘出海量数据中存在但隐藏的关系转化为知识,并提取出来进行有价值的预测和计算,以用于决策支持^[6]。具体流程步骤如下:

第一阶段是医疗数据的采集与获取,医疗数据包括患者的组学数据、表形数据、临床诊断数据、电子病历档案数据(EMR)、医疗影像数据等^[5]。主要来源于患者曾就诊的医院和病例。

第二阶段是对医疗数据进行预处理。医疗病案数据按照类型分有离散型数据和连续型数据两种,离散型数据的属性值的个数是有限的,且不同值之间是间断的、不连续的;连续型数据的值通常为数值型,取值范围为一个完整的区间。因此如若使用分类算法,就需要将连续数据离散化;预处理的另一内容是数据的清理,目的是通过删除冗余数据等方法提高医疗数据准确率,主要处理的对象有空值、噪音数据、不一致数据。

第三个阶段是特征筛选。在案例的特征集合中存在冗余、无效的特征,利用岭回归、桥回归、偏最小二乘、主成分分析、线性判别分析和独立成分分析等方法删除不相关或冗余的特征,从而降低特征维度,提高模型精度,并有效减少运行时间^[7]。

第四、第五阶段就是根据不同的疾病选择恰当的机器学习算法并构建分类器。

第六阶段是根据模型得到的结果进行分析。常见的评估指标有:AUC(Area Under Curve)、C-Index(Concordance Index)等。AUC取值一般在0.5~1.0之间,越接近1.0模型效果越好。

医疗决策支持的整体框架如图1所示。



图1 医疗决策支持框架图

3 机器学习技术概述

机器学习是人工智能的一个领域,涉及到算法的选择、分类器的构建^[1]。算法通过输入的数据进行自动学习获得知识,并基于输入数据建立模型,对新数据进行精确预测。数据的积累有利于分类器性能的提升。机器学习从数据集的数据中识别模式,为诊断和治疗规划的预测和决策过程提供支持^[8],可以分为如图2所示的几类^[9-10]。

3.1 有监督学习(Supervised Learning)

算法从给定的带标签的训练数据集引入一个映射函数,将新的输入数据映射到它想要的输出。监督学习

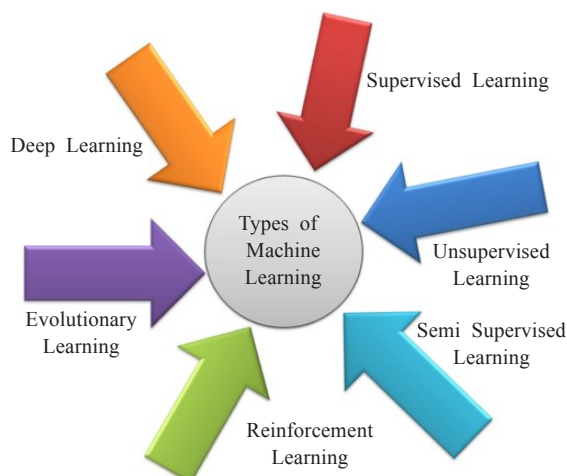


图2 常见的机器学习方法

技术解决的问题基本上分为回归问题和分类问题。在回归问题中,输入变量映射为连续输出函数,而在分类问题中,输入变量映射为离散类别。最广泛使用的监督学习算法有:支持向量机、线性回归、逻辑回归、朴素贝叶斯、决策树、 K -近邻等。

3.2 无监督学习(Unsupervised Learning)

在无监督机器学习的情况下,算法推导出一个映射函数,从无标签的数据集中寻找隐藏的模式和数据之间的相关性,并基于相关性对数据进行分类。输入数据集由示例组成,每个示例是一个没有显式输出值的输入数据。无监督学习任务中研究最多、使用最广泛的方法是聚类^[1]。

3.3 半监督学习(Semi Supervised Learning)

半监督学习是在有监督的分类算法中加入无标签的数据来实现半监督分类,介于有监督学习和无监督学习之间,属于两者结合的一种学习方法^[7]。基本思想是利用数据分布上的模型假设,建立学习器对未标签样例进行标签。

3.4 强化学习(Reinforcement Learning)

强化学习也称再励学习^[12]。如图3所示,在强化学习中,学习器是一个制定决策的智能体,它不会被告知该执行什么动作,而是经过反复尝试运行,来发现能获得最大奖励的行为。因为学习系统的行动会影响到环境,环境又会影响到后续的行动,所以从本质上讲,强化学习是一个闭环控制问题。

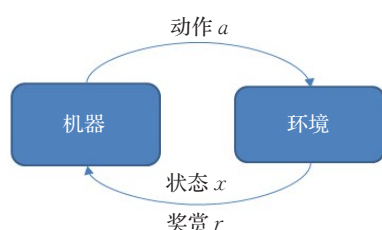


图3 强化学习示意图

3.5 进化学习(Evolutionary Learning)

进化算法的产生的灵感借鉴了大自然中生物的进化操作,是一种成熟的具有高鲁棒性和广泛适用性的全局优化方法,具有自组织、自适应、自学习的特性,能够不受问题性质的限制,有效地处理传统优化算法难以解决的复杂问题。常见的算法有:遗传算法(Genetic Algorithm,GA)、文化基因算法(Memetic Algorithm,MA)、进化多目标优化算法^[13](Multi-Objective Evolution Algorithm,MOEA)。

3.6 深度学习(Deep Learning)

深度学习是机器学习的分支,是一种使用包含复杂结构或由多重非线性变换构成的多个处理层对数据进行高层抽象的算法,即对数据进行表征学习^[14]。至今已有数种深度学习框架,如卷积神经网络^[15-17]、深度置信网络^[18-19]和递归神经网络等已被应用在计算机视觉^[20]、自然语言处理^[21-22]、语音识别、音频识别^[23]与生物信息学^[24-25]等领域,并获取了极好的效果。

4 基于机器学习的医疗决策支持

智能诊断最早源自Ledley^[26]等人在临床学中引入数学模型作为计算机辅助诊断工具,后来陆续出现了各种各样的专家系统。最早的专家系统主要是通过将医生的临床经验与医疗知识汇总,建立一个特定疾病的知识数据库,然后用符号推理的方式进行诊断。这样的系统太过机械简单,面对复杂的医疗数据具有较强的局限性,逐渐被淘汰;随着大数据时代的来临,以及计算机与人工智能技术的突破,众多学者开始利用机器学习的方法来构建智能诊断模型。

目前科学家已经开发了大量的人工智能诊断算法来检测各种疾病,如类风湿关节炎、癌症、肺病、心脏病、糖尿病视网膜病变、肝炎、阿尔茨海默病、肝病、登革热和帕金森病^[1],在某些疾病的诊断和风险预测方面取得了不逊色于人类相关专家的成绩。

4.1 心脏病

对于心脏病的医疗决策支持系统的研究,主要有两个常用的数据集:一个是印度金奈某糖尿病研究所的数据集;另一个是UCI机器学习知识库提供的心脏病数据集,该数据集由克利夫兰、匈牙利、瑞士和长滩的4个数据集整合而成,共有303个样本和76个特征。

据统计所有基于该数据集的研究都选用了14个特征构成的特征集合的子集,如表1所示,分别是:年龄、性别、胸痛类型、静息血压、血清胆固醇浓度、空腹血糖、静息心电图结果、最大心率、运动诱发心绞痛、相对于休息运动引起的ST抑郁、峰值运动ST段的斜率、主要血管数目、地中海贫血以及心脏病的诊断(血管造影疾病情况)。

表1 UCI心脏病数据集特征信息表

序号	特征名称	特征含义
1	age	年龄
2	sex	性别
3	cp	胸痛类型:具体有典型性心绞痛;非典型性心绞痛;非心绞痛型胸痛;无症状
4	trestbps	静息血压
5	chol	胆固醇浓度
6	fbs	空腹血糖(浓度>120 mg/dl): 1=True;0=False
7	restecg	静息心电图结果: =0,正常; =1,ST-T波形异常; =2,根据Estes标准显示左心室肥厚
8	thalach	最大心率
9	exang	运动诱发心绞痛: =0,无;=1,有
10	oldpeak	相对于休息,运动引起的ST抑郁
11	slope	峰值运动ST段的斜率:upsloping;flat; downsloping
12	ca	主要血管数目
13	thal	地中海贫血: 3=normal; 6=fixed defect; 7=reversible defect
14	num	心脏病诊断(angiographic disease status): Value 0:<50% diameter narrowing Value 1:>50% diameter narrowing

有诸多学者基于UCI的心脏病数据集进行研究。Otoom^[27]等人提出了一个检测系统,用以检测和监测冠状动脉疾病。数据集选择了UCI克利夫兰心脏病数据集,使用了76个特征中的13个。针对检测问题,使用WEKA对3种算法:贝叶斯网络、支持向量机SVM和函数树FT进行了2次测试。结果表明,在交叉验证测试中,SVM和贝叶斯网络的准确率均为83.8%,FT函数树的准确率达到81.5%;而通过对7个使用最佳优先选择算法提取的最优特征的测试,贝叶斯网络的准确率为84.5%,SVM准确率为85.1%,FT分类准确率为84.5%。

Chaurasia和Pal^[28]使用数据挖掘方法进行心脏病检测,基于Weka使用朴素贝叶斯、J48和bagging这3种算法。数据同样选自UCI的心脏病数据集,但仅使用了其中的11个特征进行预测。结果显示,朴素贝叶斯算法的准确率为82.31%。J48的准确率为84.35%,bagging准确率为85.03%,在这个数据集上达到了更好的分类准确率。

Vembandasamy^[29]等人使用朴素贝叶斯算法对心脏病进行了诊断。数据集来自印度金奈的一个糖尿病研究所,由500名患者的病例组成,数据集中病人有142人,358人没有患病。由于朴素贝叶斯使用了贝氏定理,

因此朴素贝叶斯具有很强的独立性假设。通过Weka使用70%的分割来执行分类,最终结果显示Naive Bayes的分类准确率为86.419%。

与文献[29]一样,Parthiban和Srivatsa^[30]也采用了印度金奈研究所收集的500例患者的数据集。使用Weka实现朴素贝叶斯算法和SVM算法,从而诊断糖尿病患者的心脏病。结果显示,朴素贝叶斯算法准确率为74%,SVM的准确率最高可达94.60%。

Tan^[31]等人提出了一种新的思路,相比于大多数研究者使用单一机器学习方法进行分类并评估效果的做法,提出了一种混合技术,利用包装器的方式将遗传算法(GA)和SVM这两种机器学习算法结合起来,并在LIBSVM和WEKA上实现并分析。从加州大学欧文分校的机器学习存储库中提取了5种疾病(虹膜、糖尿病、乳腺癌、心脏病和肝炎)的数据集。实验结果显示,采用遗传算法和支持向量机混合方法,对心脏病数据集的诊断准确率为84.07%,对糖尿病数据集的准确率为78.26%,对乳腺癌的诊断准确率为76.20%,对肝炎数据集的准确率为86.12%。

由表2可以直观看出,在心脏病的诊断上,针对UCI和Chennai两个数据集,SVM都取得了最高的准确率,分别达到85.1%、94.60%。相比其他算法,SVM的优势在于面临有限样本时能正确地实现分类,不容易出现过拟合,同时具有较好的鲁棒性。缺点在于SVM的计算成本较高,因此运行速度较慢,且不适用于大规模训练样本,面对多分类任务时需要组合模式完成。

表2 心脏病诊断汇总表

作者(年份)	疾病种类	使用算法	数据集	准确率/%
Otoom等 (2015)	冠状动脉 疾病	Bayes Net	UCI	84.50
		SVM		85.10
		FT		84.50
Chaurasia、Pal (2013)	心脏病	Naive Bayes	UCI	82.31
		J48		84.35
		Bagging		85.03
Vembandasamy等 (2015)	心脏病	Naive Bayes	金奈的糖尿病研究所	86.42
Parthiban、Srivatsa (2012)	心脏病	SVM	金奈研 究所	94.60
		Navie Bayes		74.00
Tan等 (2009)	心脏病	GA+SVM	UCI	84.07

4.2 糖尿病及其并发症

4.2.1 糖尿病

糖尿病是一组以高血糖为特征的代谢性疾病。长期保持高血糖,会导致各种组织,如:眼、肾、心脏、血管、神经的慢性损害、功能障碍。

Iyer^[32]等人使用WEKA作为数据挖掘工具,选取UCI提供的皮马印第安人(Pima Indian)糖尿病数据集,

利用决策树和朴素贝叶斯进行糖尿病预测。结果表明,在这个数据集中百分比分割(70:30)比交叉验证预测的效果更优异。通过交叉验证J48的准确率为74.869 8%;通过百分比分割,J48的准确率为76.956 5%,朴素贝叶斯算法地准确率为79.565 2%。

元学习(Meta Learning)也称“Learning to Learn”,目的是让模型学会如何学习,即使用以往的知识经验指导新任务的学习,根据新任务调整自己,而非从头开始训练。

Sen和Dash^[33]提出了一个基于元学习算法的糖尿病诊断方案,选取的数据集同样是来自UCI的Pima Indian数据集。分类方面使用了CART、Adaboost、Logiboost和分级学习算法用于预测患者是否患有糖尿病。结果显示,CART的准确率为78.646%,误分类率为21.354%,与其他算法相比取得了最优的效果。其他算法的结果:Adaboost算法的准确率为77.864%;Logiboost的准确率为77.479%;分级算法的准确率为66.406%。

Kumari和Chitra^[34]使用MATLAB 2010a,通过SVM算法进行了一项预测糖尿病的实验,SVM的核函数选用径向基函数(RBF)。数据集同样选取了UCI提供的Pima Indian糖尿病数据集。SVM提供了78%的准确率。

Ephzibah^[35]提出了一个糖尿病诊断模型,使用MATLAB构建模糊逻辑分类器进行分类。模型将遗传算法与模糊逻辑相结合,这样一方面有助于选择特征的最佳子集;另一方面也提高了分类精度。数据集取自UCI,有8个特征,769个样本。在8个特征里通过遗传算法选择了3个最优特征,使用这3个特征模糊逻辑分类器的准确率达到87%,同时将成本降低到原来的一半左右。

如表3所示,在糖尿病的智能诊断研究中,几乎所有学者都选择了UCI的Pima Indian糖尿病数据集。在较早时间使用遗传算法结合模糊逻辑的方法,取得了最高的87%准确率;近年来使用基于机器学习算法的模型,得到的结果相近,准确率浮动在75%~80%的范围。其中朴素贝叶斯取得了79.57%的准确率,因为朴素贝叶斯分类器虽然简单,但是处理高维度输入的任务上性能还要优于复杂的分类方法。其区别于模糊逻辑分类器选择最优特征的方法,朴素贝叶斯算法消除了不相关的特征干扰,提升了分类性能;此外,朴素贝叶斯算法不涉及迭代训练过程,因此大大减少了计算时间。但是相比于在其他疾病上的应用,朴素贝叶斯算法的精度还有待进一步的提升。

4.2.2 糖尿病并发症

糖尿病在危害人体的同时还会引发一系列并发症状,如糖尿病视网膜病变,病因是糖尿病对视网膜的损伤,有如图4所示的几个阶段。据统计在患糖尿病20年

表3 糖尿病诊断汇总表

作者(年份)	疾病种类	使用算法	数据集	准确率/%
Iyer等 (2015)	糖尿病	J48	UCI	76.96
		Naive Bayes		79.57
		CART		78.65
Sen、Dash (2014)	糖尿病	Adaboost	UCI	77.86
		Logiboost		77.48
		分级学习		66.41
Kumari、Chitra (2013)	糖尿病	SVM	UCI	78.00
Ephzibah (2011)	糖尿病	GA+Fuzzy Logic	UCI	87.00

以上的患者中有80%都受到了不同程度的影响,最严重的将致盲。但研究表明,若糖尿病患者对视力变化保持关注和监测,可以降低90%的新增病例^[36]。因此,早期检测、控制病变方面的研究,引起了越来越多学者的关注。

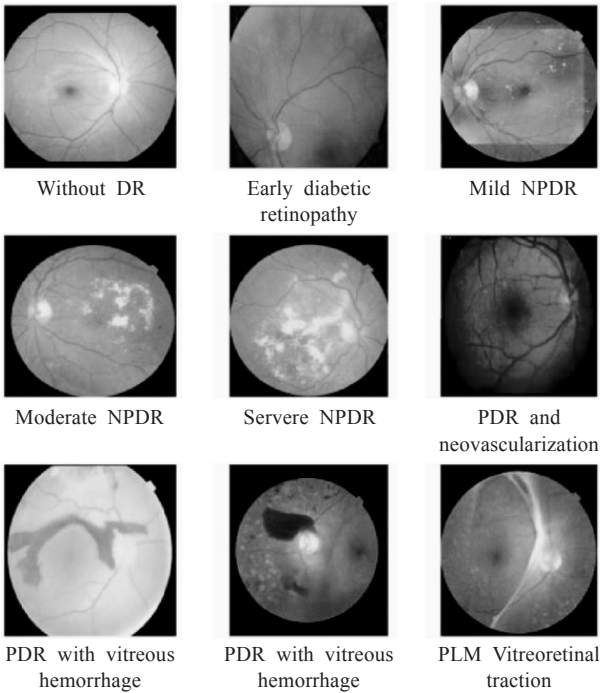


图4 糖尿病性视网膜病变阶段图

Wang^[37]等人使用CNN作为特征提取器,随机森林作为分类器来解决视网膜血管分割问题。这种集合方法的优势有:避免了人工提取特征的误差,降低了成本;模型的泛化能力较为优异。

曹文哲^[38]等人应用随机森林算法和Logistic回归算法,对2型糖尿病并发视网膜病变进行风险预测,并通过两者的ROC曲线面积比较预测效果。结果显示,在随机森林模型的39个变量重要性评分中,糖化血红蛋白、空腹血糖、尿素、肌酐、尿酸、年龄、冠心病和慢性肾病得分较高且具有临床意义;Logistic回归模型最终纳入性别、血糖控制情况(糖化血红蛋白浓度)、慢性肾病、冠心病、心梗和癌症6个因素。在检验水平 $\alpha=0.05$ 下,

对两种模型作差异性检验,得 $P=0.001\ 9$ 。结合 ROC 曲线,随机森林模型预测效果优于 Logistic 回归模型。

Chandrakumar 和 Kathirvel^[39]提出如图 5 所示的深度学习架构进行糖尿病性视网膜病变分类。该架构适用于病变的全阶段,且通过过拟合可以得到精准的分类;但是由于架构的复杂度和计算密集度,对图像处理单元的性能有较高的要求。

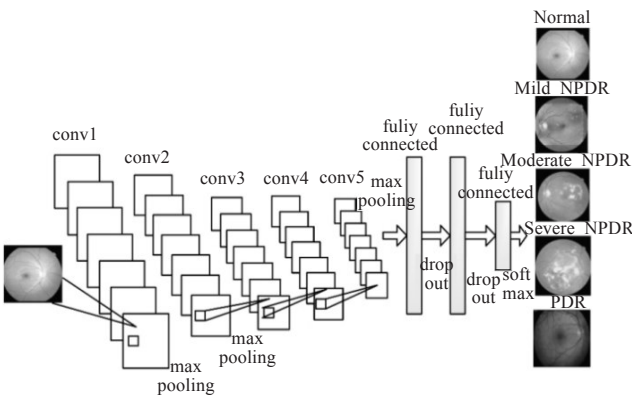


图5 深度卷积神经网络架构

对于糖尿病性视网膜病变,由于相关的医疗数据类型通常包括图像,因此利用深度学习构建模型的研究较多,或像文献[37]一样使用深度学习模型进行特征提取,再输入传统机器学习分类器(RF),集两者优点于一身,得到良好的效果。局限性在于若输入随机森林分类器的数据具有较大的噪声时,容易出现过拟合的情况。

4.3 癌症

医学上,癌是指起源于上皮组织的恶性肿瘤。一般人们所说的“癌症”泛指所有恶性肿瘤。据统计全球每年新增约1 410万癌症患者,致死人数达到820万^[40]。男性发病率最高的是肺癌,女性发病率最高的是乳腺癌,除此之外还有肝癌、胃癌、食管癌等。癌细胞一经查出,多数已经转移或者扩散,因此早期诊断非常重要。许多学者也就此问题展开相关研究,并取得良好的结果。癌症诊断的部分研究成果见表4。

4.3.1 乳腺癌

肿瘤肿块是乳腺癌的主要类型之一,当癌变肿块被不同密度的实质组织结构所嵌入和掩盖时,乳房X线照片很难直观地检测到它们。针对此问题许多学者引入机器学习来解决。

Zhen和Chan^[41]提出了一种结合多种人工智能技术和离散小波变换(DWT)检测乳腺肿块的算法。涉及的人工智能技术包括分形维数分析、多分辨率马尔可夫随机场和犬兔算法。分形维数分析作为一种预处理器,用于确定乳房X线照片中可疑肿瘤区域的大致位置。采用犬兔聚类算法对乳腺X线片三级DWT分解的LL子带进行分割,最后通过构建树型分类器来确定给定区域

表4 癌症诊断汇总表

作者(年份)	疾病种类	使用算法	数据集	准确率/%
Zhen、Chan (2001)	乳腺癌	分形维数分析、 多分辨率马尔 可夫随机场、 犬兔算法	乳腺X 线图片	97.30
Jhajharia等 (2016)	乳腺癌	PCA+ANN	—	96.00
Spanhol等 (2016)	乳腺癌	CNN(AlexNet)	BreaKHis	—
Sun等 (2016)	肺癌	CNN DBN SDAE	LIDC	79.76 81.79 79.29
Pan等 (2015)	肺癌	CNN	数百张弱 注释图像	—
Xu等 (2016)	子宫颈癌	Deep Learning	—	88.92

癌变的可能性。实验结果表明,该算法的准确率为97.3%,每幅图像的误报数为3.92。

Jhajharia等^[42]提出了一种基于神经网络的乳腺癌预测模型,并对其进行了主成分分析(PCA)处理。实验采用多元统计方法结合人工智能学习技术实现预测模型。主成分分析对数据进行预处理,以最相关的形式提取特征,用于人工神经网络的训练。最终的实验结果准确率达到96%。

Spanhol等^[43]使用 AlexNet,通过将图像块进行多种组合进行乳腺癌分类,取得了良好的结果。实验所用图像选自 BreaKHis 的乳腺癌组织病理学图集。

4.3.2 肺癌

肺癌是发病率和死亡率最高的恶性肿瘤之一^[44],在男性中尤甚。随着早期诊断在美国的开展,已经可以将早期肺癌5年的生存率提升到60%~80%。

Sun等^[45]采用3种深度学习算法:卷积神经网络(CNN)、深度信念网络(DBN)、堆叠去噪自编码(SDAE),对肺癌图像数据库(Lung Image Database Consortium, LIDC)的肺癌诊断案件进行实验。结果显示:CNN、DBNs和SDAE的准确率分别是79.76%、81.79%和79.29%。

Pan等^[46]将数百张弱注释的图像作为训练集,通过训练CNN进行特征提取,实现对肺癌细胞的自动检测,在准确率等方面取得了不错的结果。

4.3.3 子宫颈癌

子宫颈癌的发病率在女性恶性肿瘤中仅次于乳腺癌。在2017年,世界范围内有20万人死于子宫颈癌,其中中国死亡5.3万人。子宫颈癌晚期和复发性宫颈癌的治疗效果很差,但在早期5年生存率可达90%以上,这引起了许多学者的关注。2017年Kaggle发布挑战赛,寻找能最准确筛查女性子宫颈癌的医疗图像的算法。

Xu等^[47]提出了一种利用多模态信息的深度学习框架用以检测子宫癌发育的程度。该方法在数据集上取得了87.83%的灵敏度和90%的特异性。

现基于人工智能的子宫颈癌诊断已产品化,2018年问世的宫颈癌诊断机器人“兰丁”,能以极高的工作效率从数字化图像里找到癌细胞的踪迹,实现完全自动化的诊断。

4.4 肝脏疾病

肝脏疾病按照发病机理可以分为病毒性肝病和非病毒性肝病^[48]。病毒性肝病是由多种不同的肝炎病毒引起的传染病^[49]。非病毒性肝病多由不健康的生活方式所引起,主要有:酒精性肝病、脂肪性肝病^[50]、新陈代谢异常性肝病等。这两类肝脏疾病发病机理不同,提取特征存在差异,因此不能统一归为肝脏疾病构建模型,需要分别展开讨论分析。

4.4.1 非病毒性肝病

针对患者的非病毒性肝脏病变,有学者进行了以下研究:

Vijayarani和Dhayanand^[51]基于MATLAB使用SVM和朴素贝叶斯算法对样本是否患有肝脏疾病进行预测分析。实验所用的ILPD数据集来自UCI,由560个样本和10个特征组成。实验在考虑执行时间的基础上对两种算法进行了比较。结果显示,朴素贝叶斯算法在1 670 ms下的准确率为61.28%,而SVM在3 210 ms下的准确率为79.66%;因此准确率方面SVM高于朴素贝叶斯,但是在时间执行方面,朴素贝叶斯算法花费的时间更少。实验结果与两种算法的原理分析基本吻合。

Gulia等^[52]开展了一项基于多种机器学习算法对肝脏患者进行分类并评估研究。采用WEKA数据挖掘工具与J48、MLP、随机森林、SVM和贝叶斯网络5种算法,进行了两次实验。第一次将5种算法应用到原始的完整数据集上进行分类,并计算每种算法的准确率;第二次实验使用特征选择的方法进行降维,得到一个整个数据集的子集,然后使用5种算法对子集进行分类。特征选择前后的结果比较显示,在特征选择之后,5种算法的准确率均有了不同程度的提高。其中J48准确率为70.669 0%;MLP算法准确率为70.840 5%;SVM的准确率为71.355 1%;随机森林的准确率为71.869 6%;贝叶斯网络的准确率为69.125 2%。

4.4.2 病毒性肝病

病毒性肝病也就是肝炎。许多病原微生物如病毒、细菌、真菌、立克次氏体、螺旋体及某些原虫和寄生虫的感染都可能引起肝脏发炎。机器学习技术在肝炎诊断上的应用由来已久,大部分研究人员使用的数据集都来自UCI的肝炎数据集,包括155个样本和20个特征,特征信息如表5所示。

表5 UCI肝炎数据集特征信息表

序号	特征名称	特征含义
1	CLASS	类别:DIE,LIVE
2	AGE	年龄:10,20,30,40,50,60,70,80
3	SEX	性别: male, female
4	STEROID	类固醇: no, yes
5	ANTIVIRALS	抗病毒药: no, yes
6	FATIGUE	疲劳: no, yes
7	MALAISE	萎靡不适: no, yes
8	ANOREXIA	厌食症: no, yes
9	LIVER BIG	肝脏较正常值偏大: no, yes
10	LIVER FIRM	肝硬化: no, yes
11	SPLEEN PALPABLE	明显易怒: no, yes
12	SPIDERS	血管蛛化: no, yes
13	ASCITES	腹腔有积液: no, yes
14	VARICES	静脉曲张: no, yes
15	BILIRUBIN	胆红素: 0.39, 0.80, 1.20, 2.00, 3.00, 4.00
16	ALK PHOSPHATE	碱性磷酸: 33, 80, 120, 160, 200, 250
17	SGOT	谷草转氨酶: 13, 100, 200, 300, 400, 500
18	ALBUMIN	白蛋白: 2.1, 3.0, 3.8, 4.5, 5.0, 6.0
19	PROTIME	凝血酶原时间: 10, 20, 30, 40, 50, 60, 70, 80, 90
20	HISTOLOGY	组织结构: no, yes

2013年, Ba-Alwi和Hintaya^[53]针对肝炎诊断问题,使用朴素贝叶斯算法、朴素贝叶斯可更新算法、FT算法、Kstar算法、J48算法、LMT算法和NN算法进行实验并对结果进行比较分析。实验选用UCI的肝炎疾病数据集。分类结果根据准确率和运行时间衡量,结果表明,朴素贝叶斯算法运行时间0 s,其准确率为96.52%;朴素贝叶斯可更新算法运行0 s达到了84%的准确率;FT树运行0.2 s,其准确率为87.10%;Kstar运行0 s达到83.47%的准确率;J48在0.03 s内达到83%的准确率;LMT运行0.6 s,其准确率为83.6%;神经网络准确率为70.41%。粗糙集技术中应用最广泛的朴素贝叶斯分类算法在该数据上表现也很好,在最短的时间内具有高精度,而神经网络的方法得到的结果较低。

Karlik^[54]使用朴素贝叶斯和反向传播分类器诊断肝炎,并进行了比较分析。选择这两个分类器是因为它们只需要少量的数据就可以进行分类,分类类别根据肝炎病毒的类型可分为A型、B型、C型、D型和E型5类。结果显示,使用朴素贝叶斯分类器,从UCI肝炎数据集的20个特征组成的集合中提取15个,准确率能达到97%;而采用3层前馈神经网络,用反向传播算法对数据集进行训练,最终的准确率能达到98%。

Sathyadevi^[55]采用C4.5、ID3和CART算法诊断肝炎。研究同样使用UCI的肝炎患者数据集,使用WEKA进行分析。CART算法生成的二元决策树(DT)只有两个或没有子树,由C4.5和ID3组成的决策树可以有两个或更多的子结点。结果显示,CART算法的分类准确率

最高,为83.2%,这是因为CART在处理缺失值方面有很好的性能;ID3算法的准确率为64.8%;C4.5算法的准确率达到71.4%。在该肝炎数据集上,CART算法在准确率和时间复杂度方面表现最好。

肝纤维化的早期诊治可以有效抑制肝硬化的病变速度,从而降低肝癌的患病风险。Wang^[56]等人研究一个基于CNN的弹性成像影像组学模型,用以对患者进行肝纤维化分期。选取398例病人,共1990幅二维SWE(超声剪切波弹性成像)图像作为输入数据,按照2比1的比例划分训练集与测试集。结果显示F4期(肝硬化)、≥F3期(晚期纤维化)、≥F2期(显著纤维化)的AOC分别为0.97、0.98、0.85,对肝纤维化病人的无创诊治具有很大的实际意义。

通过观察表6发现,在肝炎疾病的检测中,基于反向传播算法的前馈神经网络准确率最高。在该模型中,采用误差反向传播算法训练了3个前馈神经网络。基于增量学习规则的反向传播训练是一种迭代梯度算法,旨在减少多层前馈神经网络实际输出与期望输出之间的均方根误差。朴素贝叶斯算法的准确率稍低,但在运行时间上比神经网络更快,这反映了神经网络需要大样本量以及更强计算能力的缺点。

表6 肝脏疾病诊断汇总表

作者(年份)	疾病种类	使用算法	数据集	准确率/%
Vijayarani、 Dhayanand (2015)	非病毒性 肝脏疾病	SVM	UCI的	79.66
		Naive Bayes	ILPD	61.28
		J48		70.67
		MLP		70.84
Gulia等 (2014)	非病毒性 肝脏疾病	Random Forest	UCI	71.87
		SVM		71.36
		Bayesian Network		69.13
		Naive Bayes		96.52
Ba-Alwi、 Hintaya (2013)	肝炎	Updateable Naive Bayes		84.00
		FT		87.10
		Kstar	UCI	83.47
		J48		83.00
		LMT		83.60
		NN		70.41
Karlík (2011)	肝炎	Naive Bayes	UCI	97.00
		FF NN with BP		98.00
Sathyadevi (2011)	肝炎	C4.5		71.40
		ID3	UCI	64.80
		CART		83.20
Wang等 (2018)	肝硬化	CNN	1990幅 SWE图像	—

4.5 登革热

登革热(dengue)是登革病毒经蚊媒传播引起的一

种急性虫媒传染病,在热带、亚热带较为流行,如泰国、印度尼西亚和马来西亚等气候潮湿的国家都面临着严峻的挑战。机器学习技术的发展给登革热的疫情检测、区域控制注入了新的活力。

文献[57]利用WEKA构建了一个数据挖掘模型对马来西亚登革热疫情进行预测分析。分类算法使用决策树(DT)、人工神经网络(ANN)和粗糙集理论(RS)。数据集来源于马来西亚雪兰莪州的公共卫生部门,对数据集的划分使用了10折交叉验证和百分比分割两种测试方法。通过使用10折交叉验证,决策树的准确率为99.95%,ANN的准确率为99.98%,RS的准确率为100%;相比之下,百分比分割后决策树和人工神经网络都达到99.92%的准确率,RS达到99.72%的准确率。

类似地,Ibrahim等^[58]也提出了一种使用人工神经网络(ANN)对登革热患者进行预测的系统,用MATLAB实现。区别在于该系统不是分类问题,是个回归问题,即预测的是登革热患者发热天数、退烧需要的天数。实验采用多层前馈神经网络(MFNN)算法,仅使用临床体征和症状进行检测。数据来源于某医院的实际病例,共有252例住院患者,其中4例为DF(登革出血热),248例为DHF(登革出血热)。实验结果显示,MFNN算法对DF和DHF患者的发热、退烧所需时间都能做出准确预测,准确率达到90%。

在登革热的诊断研究中,如表7所示,RS将准确率提升到最高,它能够有效处理不确定性、噪声和丢失的数据;从时间上看,RS也是最好的;与DT相比,RS不需要任何关于数据的初始和附加信息。

表7 登革热诊断汇总表

作者(年份)	疾病种类	使用算法	数据集	准确率/%
Ahmad Tarmizi等 (2013)	登革热	DT	雪兰莪州	99.95
		ANN	公共卫生部	99.98
		RS		100.00
Ibrahim等 (2005)	登革热	MFNN	252个真实 病例	90.00

4.6 脑梗死

高血脂症是患者的血脂水平过高,而脑梗死的病理生理过程实质上是在动脉粥样硬化基础上发生的局部脑组织缺血坏死过程,因此高血脂症患者拥有很大的风险发生脑梗死。

Hwang和Chen^[59]基于5种机器学习方法:NN、贝叶斯分类器、KNN、决策树和SVM对3个医疗实例进行疾病风险预测分析,并以此对比不同机器学习方法的性能指标,以选择最适合的机器学习模型;最后给出了5种机器学习方法的优缺点。针对患者是否属于高血脂高危

人群的风险预测模型,输入患者特征:甘油三酯、总胆固醇、高密度脂蛋白和低密度脂蛋白等,得到患者属于高危人群或低危人群的二分类结果,并计算精确度、查全率、查准率、F1 度量。结果显示,单考虑准确率,SVM 和决策树的准确率都在92%左右,而其他3种方法的准确率都在90%左右;分析精确度,发现NN 和决策树都优于KNN 的80%;在召回率方面,KNN 方法的召回率最高,其余4种算法的召回率保持在90%;在F1 度量方面,决策树获得最高的F1 度量,达到了95%,其他方法在90%左右。在模型训练时间上,KNN 的训练时间相较其他方法要长得多。基于这些结果,认为在该数据集上决策树方法的表现最好,KNN 方法表现最差。

Chen^[60]等人提出了一种新的基于CNN 的结构化与非结构化数据多模态疾病风险预测(CNN-MDRP)算法,将结构化与非结构化特征相结合,得到疾病风险模型,对利用统计学知识确定的该地区主要慢性病类别——脑梗死进行预测。使用潜在因子模型重建缺失的数据,对于结构化数据,询问医生以获得有用的特征,对于非结构化文本数据,提取文本中的单词来学习嵌入,再利用CNN 进行独立特征提取。如表8所示,该方法在某些数据上有效提高了准确率。

表8 脑梗死诊断汇总表

作者(年份)	疾病种类	使用算法	数据集	准确率/%
Hwang,Chen (2017)	高血 脂症	NN	—	90.00
		Bayes Classifier		90.00
		KNN		90.00
		DT		92.00
		SVM		92.00
Chen 等 (2017)	脑梗死	CNN-MDRP	2013—2015 年 华中地区医院 实际数据	94.80

5 结束语

本文介绍了基于机器学习的医疗决策支持,并对涉及的机器学习方法和在不同疾病模型上的应用进行综述分析。医疗决策支持主要指疾病智能诊断和疾病风险预测两方面,通过对相关医学数据的挖掘分析,帮助医生做出在全医疗流程范围内的诊断决策。

机器学习技术在许多应用程序中发挥着至关重要的作用,例如图像检测、数据挖掘、自然语言处理和疾病诊断。在所有这些领域中,机器学习都提供了具有极大可能性的解决方案。在智慧医疗领域,从诸多学者的研究中可以清楚地看到,利用人工智能建立的诊断模型,在各种疾病诊断和预测上都有助于提高决策的速度和质量,降低假阳性率。医疗决策支持的发展极大推动了智慧医疗的发展,将为人们提供更丰富的医疗资源和更便捷、智能的健康服务,更有效地满足人们对医疗保健

服务以及增进健康的普遍需求。智慧医疗的进一步发展普及,还需要在以下几方面取得突破。

(1)数据质量和标准化:医疗数据的采集还没有形成行业内整体的统一标准,造成数据的不兼容、结构复杂,大大影响了数据的质量。当处理存在异构数据的医疗任务时,势必会消耗模型的计算能力,形成资源浪费;智慧医疗涉及传感网、通信网、健康信息处理、医疗终端和应用等各个环节,各个医疗机构、研发机构以及各个环节的数据及系统有各自实施的标准,因此,如何将所有数据统一标准化,是智慧医疗面临的瓶颈之一。

(2)智能诊断算法的创新:从机器学习角度来看,疾病诊断属于强非线性分类问题,虽然已经有多种机器学习方法应用于智能医疗决策支持系统,但这些方法往往只能适用于某种疾病、某些具有特定特征的数据集,并严重依赖于特征工程、模型的泛化能力与精度难以取得平衡。深度学习对特征的自动提取能力和强非线性映射的表达能力使得其成为未来智能诊疗的最具潜力的方法,当前其在疾病诊断、预测领域的研究目前仍处于起始阶段,针对不同的疾病选择何种网络模型,疾病样本的时空特征如何表达,模型参数如何训练和优化,如何提高学习方法的效率等等大量的科学问题亟待解决。

参考文献:

[1] Kumar G.A survey on machine learning techniques in health care industry[J].International Journal of Recent Research Aspects,2016,3(2):128-132.

[2] Liu Y,Kohlberger T,Norouzi M,et al.Artificial intelligence-based breast cancer Nodal metastasis detection[J].Archives of Pathology and Laboratory Medicine,2018(1).

[3] Esteva A,Kuprel B,Novoa R A,et al.Dermatologist-level classification of skin cancer with deep neural networks[J].Nature,2017,542:115-118.

[4] Adlassnig K P.Fuzzy set theory in medical diagnosis[J].IEEE Transactions on Systems,Man and Cybernetics,1986,16(2):260-265.

[5] Malmir B,Amini M,Chang S I.A medical decision support system for disease diagnosis under uncertainty[J].Expert Systems with Applications,2017,88:95-108.

[6] Baiju B V,Remy Janet R J.a survey on heart disease diagnosis and prediction using naive Bayes in data mining[J].International Journal of Current Engineering and Technology,2015,5(2):1034-1038.

[7] 赵学彤,杨亚东,渠鸿竹,等.组学时代下机器学习方法在临床决策支持中的应用[J].遗传,2018,40(9):693-703.

[8] Fatima M,Pasha M.Survey of machine learning algorithms for disease diagnostic[J].Journal of Intelligent

- Learning Systems and Applications, 2017, 9(1): 1-16.
- [9] Dinu A J, Ganesan R, Joseph F, et al. A study on deep machine learning algorithms for diagnosis of diseases[J]. International Journal of Applied Engineering Research, 2017, 12(17): 6338-6346.
- [10] Razia S. A review on disease diagnosis using machine learning techniques[J]. International Journal of Pure and Applied Mathematics, 2017, 117(16): 79-85.
- [11] Celebi M E, Kingravi H A, Vela P A. A comparative study of efficient initialization methods for the k-means clustering algorithm[J]. Expert Systems with Applications, 2012, 40(1): 200-210.
- [12] 陈学松, 杨宜民. 强化学习研究综述[J]. 计算机应用研究, 2010, 27(8): 2834-2838.
- [13] 郑金华. 多目标进化算法及其应用[M]. 北京: 科学出版社, 2007.
- [14] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521: 436.
- [15] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012.
- [16] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Proceedings of International Conference on Neural Information Processing Systems, 2015.
- [17] Girshick R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision, 2015.
- [18] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2014, 18(7): 1527-1554.
- [19] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313: 504-507.
- [20] 王耀玮, 唐伦, 刘云龙, 等. 基于多任务卷积神经网络的车辆多属性识别[J]. 计算机工程与应用, 2018, 54(8): 21-27.
- [21] 李博, 赵翔, 王帅, 等. 改进的卷积神经网络关系分类方法研究[J]. 计算机科学与探索, 2018, 12(5): 697-707.
- [22] 喻涛, 罗可. 利用动态多池卷积神经网络的情感分析模型[J]. 计算机科学与探索, 2018, 12(7): 1182-1190.
- [23] Gopala K A, Josh C, Chang E F. Speech synthesis from neural decoding of spoken sentences[J]. Nature, 2019, 568: 493-498.
- [24] Babak A, Andrew D, Matthew T W, et al. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning[J]. Nature Biotechnology, 2015, 33: 831-838.
- [25] Zhou J, Olga G T. Predicting effects of noncoding variants with deep learning-based sequence model[J]. Nature Methods, 2015, 12: 931-934.
- [26] Ledley R S, Lusted L B. Reasoning foundations of medical diagnosis[J]. Science, 1959, 130: 9-21.
- [27] Ootom A F, Abdallah E E, Kilani Y, et al. Effective diagnosis and monitoring of heart disease[J]. International Journal of Software Engineering and Its Applications, 2015, 9(1): 143-156.
- [28] Chaurasia V, Pal S. Data mining approach to detect heart disease[J]. International Journal of Advanced Computer Science and Information Technology, 2013, 2(4): 56-66.
- [29] Vembandasamy K, Sasipriya R, Deepa E. Heart diseases detection using naive Bayes algorithm[J]. International Journal of Innovative Science, Engineering and Technology, 2015, 2: 441-444.
- [30] Parthiban G, Srivatsa S K. Applying machine learning methods in diagnosing heart disease for diabetic patients[J]. International Journal of Applied Information Systems, 2012, 3(7): 25-30.
- [31] Tan K C, Teoh E J, Yu Q, et al. A hybrid evolutionary algorithm for attribute selection in data mining[J]. Expert Systems with Applications, 2009, 36(4): 8616-8630.
- [32] Iyer A, Jeyalatha S, Sumbaly R. Diagnosis of diabetes using classification mining techniques[J]. International Journal of Data Mining and Knowledge Management Process, 2015, 5(1): 1-14.
- [33] Sen S K, Dash S. Application of meta learning algorithms for the prediction of diabetes disease[J]. International Journal of Advance Research in Computer Science and Management Studies, 2014, 2(12): 396-401.
- [34] Kumari V A, Chitra R. Classification of diabetes disease using support vector machine[J]. International Journal of Engineering Research and Applications, 2013, 3(2): 1797-1801.
- [35] Ephzibah E P. Cost effective approach on feature selection using genetic algorithms and fuzzy logic for diabetes diagnosis[J]. International Journal on Soft Computing, 2011, 2(1): 1-10.
- [36] 张巧丽, 赵地, 迟学斌. 基于深度学习的医学影像诊断综述[J]. 计算机科学, 2017(S2): 11-17.
- [37] Wang S, Yin Y, Cao G, et al. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning[J]. Neurocomputing, 2015, 149: 708-717.
- [38] 曹文哲, 应俊, 陈广飞, 等. 基于 Logistic 回归和随机森林算法的 2 型糖尿病并发视网膜病变风险预测及对比研究[J]. 中国医疗设备, 2016, 31(3): 33-38.
- [39] Chandrakumar T, Kathirvel R. Classifying diabetic retinopathy using deep learning architecture[J]. International Journal of Engineering and Technical Research, 2016, 5(6): 19-24.
- [40] 余绍德. 卷积神经网络和迁移学习在癌症影像分析中的研究[D]. 北京: 中国科学院大学(中国科学院深圳先进技术研究院), 2018.

- [41] Zhen L, Chan A K. An artificial intelligent algorithm for tumor detection in screening mammogram[J]. IEEE Transactions on Medical Imaging, 2001, 20(7): 559-567.
- [42] Jhajharia S, Varshney H K, Verma S, et al. A neural network based breast cancer prognosis model with PCA processed features[C]// Proceedings of International Conference on Advances in Computing, 2016.
- [43] Spanhol F A, Oliveira L S, Petitjean C, et al. Breast cancer histopathological image classification using convolutional neural networks[C]// Proceedings of International Joint Conference on Neural Networks, 2016.
- [44] 张鹏, 徐欣楠, 王洪伟, 等. 基于深度学习的计算机辅助肺癌诊断方法[J]. 计算机辅助设计与图形学学报, 2018(1): 90-99.
- [45] Sun W, Zheng B, Qian W. Computer aided lung cancer diagnosis with deep learning algorithms[C]// Proceedings of SPIE Medical Imaging Conference, 2016.
- [46] Pan H, Xu Z, Huang J. An effective approach for robust lung cancer cell detection[C]// Proceedings of International Workshop on Patch-based Techniques in Medical Imaging, 2015.
- [47] Xu T, Zhang H, Huang X, et al. Multimodal deep learning for cervical dysplasia diagnosis[C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016.
- [48] 刘文广, 谢斯敏, 周雅芳, 等. 机器学习及其在肝脏疾病影像诊断中的研究进展[J]. 国际医学放射学杂志, 2019(1): 16-21.
- [49] Chen Y, Luo Y, Huang W, et al. Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B[J]. Computers in Biology and Medicine, 2017, 89: 18-23.
- [50] Kuppli V, Biswas M, Sreekumar A, et al. Extreme learning machine framework for risk stratification of fatty liver disease using ultrasound tissue characterization[J]. Journal of Medical Systems, 2017, 41(10): 152.
- [51] Vijayarani S, Dhayanand S. Liver disease prediction using SVM and naive Bayes algorithms[J]. International Journal of Science, Engineering and Technology Research, 2015, 4(4): 816-820.
- [52] Gulia A, Vohra R, Rani P. Liver patient classification using intelligent techniques[J]. International Journal of Computer Science and Information Technologies, 2014, 5(4): 5110-5115.
- [53] Ba-Alwi F M, Hintaya H M. Comparative study for analysis the prognostic in hepatitis data: Data mining approach[J]. International Journal of Scientific and Engineering Research, 2013, 4(8): 680-685.
- [54] Karlik B. Hepatitis disease diagnosis using backpropagation and the naive Bayes classifiers[J]. Revista De Saúde Pública, 2011, 29(3): 215-220.
- [55] Sathyadevi G. Application of CART algorithm in hepatitis disease diagnosis[C]// Proceedings of International Conference on Recent Trends in Information Technology, 2011.
- [56] Wang K, Lu X, Zhou H, et al. Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: A prospective multicentre study[J]. Gut, 2018, 68(4): 729-741.
- [57] Ahmad Tarmizi N D, Jamaluddin F, Bakar A A, et al. Malaysia dengue outbreak detection using data mining models[J]. Journal of Next Generation Information Technology, 2013, 4(6): 96-107.
- [58] Ibrahim F, Taib M N, Abas W A, et al. A novel Dengue Fever (DF) and Dengue Haemorrhagic Fever (DHF) analysis using Artificial Neural Network (ANN)[J]. Computer Methods and Programs in Biomedicine, 2005, 79(3): 273-281.
- [59] Hwang K, Chen M. Big data analytics for cloud/iot and cognitive computing[M]. Hoboken, NJ, USA: Wiley, 2017.
- [60] Chen M, Hao Y, Hwang K, et al. Disease prediction by machine learning over big data from healthcare communities[J]. IEEE Access, 2017, 5: 8869-8879.