

基于可解释人工智能的临床决策支持系统:孟超肝病外脑

方国旭¹ 郭鹏飞² 范鉴慧³ 丁宗仁¹ 张清华⁴ 魏光亚⁴ 李海涛¹ 刘景丰⁴

¹福建医科大学孟超肝胆医院肝胆胰外科,福州 350025;²福建医科大学孟超肝胆医院东南肝胆健康大数据研究所,福州 350025;³福建医科大学孟超肝胆医院肝内科,福州 350025;⁴福建医科大学肿瘤临床医学院 福建省肿瘤医院肝胆胰外科,福州 350014

通信作者:李海涛,Email:lht45182@163.com;刘景丰,Email:drjingfeng@126.com

【摘要】 近年来,人工智能机器学习与深度学习技术有了飞跃的进步,利用临床决策支持系统进行辅助诊断与治疗是智慧医疗发展的必然趋势。临床医务工作者通常在追求模型高准确性的同时忽略模型的可解释性,导致使用者对系统缺乏信任感,阻碍临床决策支持系统的落地应用。笔者团队从可解释人工智能的角度出发,在构建肝病领域的临床决策支持系统上进行初步探索,在追求模型高准确性的同时,采用数据治理技术、内在可解释性模型、对复杂模型的事后可视化、设计人机交互和提供基于临床指南的知识图谱与数据来源等方法赋予系统可解释性。

【关键词】 肝脏疾病; 人工智能; 临床决策支持系统; 数据治理; 可解释

基金项目:福建省卫健委医学创新课题(2021CXA030);福州市科技局科技创新平台项目(2021-P-055)

Clinical decision support system based on explainable artificial intelligence—brain of Mengchao liver disease

Fang Guoxu¹, Guo Pengfei², Fan Jianhui³, Ding Zongren¹, Zhang Qinghua⁴, Wei Guangya⁴, Li Haitao¹, Liu Jingfeng⁴

¹Department of Hepatobiliary and Pancreatic Surgery, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou 350025, China; ²The Big Data Institute of Southeast Hepatobiliary Health Information, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou 350025, China; ³Department of Hepatology, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou 350025, China; ⁴Department of Hepatobiliary and Pancreatic Surgery, Fujian Medical University School of Clinical Oncology, Fujian Cancer Hospital, Fuzhou 350014, China

Corresponding authors: Li Haitao, Email: lht45182@163.com; Liu Jingfeng, Email: drjingfeng@126.com

【Abstract】 In recent years, the artificial intelligence machine learning and deep learning technology have made leap progress. Using clinical decision support system for auxiliary diagnosis and treatment is the inevitable developing trend of wisdom medical. Clinicians tend to ignore the interpretability of models while pursuing its high accuracy, which leads to the lack of trust of users and hamper the application of clinical decision support system. From the perspective of explainable artificial intelligence, the authors make some preliminary exploration on the construction of clinical decision support system in the field of liver disease. While pursuing high accuracy of the model, the data governance techniques, intrinsic interpretability models, post-hoc visualization of complex models, design of human-computer interactions, providing knowledge map based on clinical guidelines and data sources are used to endow the system with interpretability.

DOI: 10.3760/cma.j.cn115610-20221102-00679

收稿日期 2022-11-02

引用本文:方国旭,郭鹏飞,范鉴慧,等.基于可解释人工智能的临床决策支持系统:孟超肝病外脑[J].中华消化外科杂志,2023,22(1):70-80. DOI: 10.3760/cma.j.cn115610-20221102-00679.



【Key words】 Liver disease; Artificial intelligence; Clinical decision support system; Data governance; Interpretability

Fund programs: Medical Innovation project of Fujian Health Commission (2021CXA030); Fuzhou Science and Technology Bureau Science and Technology Innovation Platform Project (2021-P-055)

随着计算机科学与技术的迅猛发展,人工智能在近几年受到持续关注。国家在《“健康中国2030”规划纲要》和《国务院办公厅关于促进“互联网+医疗健康”发展的意见》中都明确要求研发基于人工智能的临床决策支持系统(clinical decision support system, CDSS)。CDSS是一种可以为临床医务工作者等提供特定知识,以实现辅助临床决策的计算机系统。它不仅可以提高诊断与治疗行为的规范性、安全性,而且可以减少漏诊、误诊,减轻医师工作负担,降低社会医疗成本。它有助于实现临床工作的自动化、智能化和规范化,从而促进医疗事业的发展。

近年来,机器学习、深度学习相关技术在计算机视觉、自然语言处理、语音识别等多个领域取得巨大成功,机器学习模型也被广泛应用于部分重要的现实任务中,如人脸识别、自动驾驶、恶意软件检测和智慧医疗分析等^[1-6]。在某些场景中,机器学习模型的表现甚至超过人类。尽管机器学习、深度学习在许多有意义的任务中胜过人类,但由于缺乏可解释性,其表现和应用饱受质疑^[7]。对于普通用户而言,深度学习神经网络模型如同1个黑盒,给它一个输入,只反馈一个决策结果,但决策的依据、规则以及过程是否可靠未知。这种模型黑盒将有可能给现实应用带来严重威胁^[8]。此外,医疗领域中的CDSS除了上述模型黑盒,还存在数据黑盒与临床应用黑盒问题,导致CDSS的临床决策对医务工作者不可解释,严重阻碍人工智能在医疗领域的落地应用。

近年来学术界和工业界进行广泛和深入的研究并且提出可解释的人工智能(explainable artificial intelligence, XAI)概念。XAI是指人类能够理解计算机做出某种决策原因的程度^[9]。模型的可解释性越高,人们就越容易理解模型做出某种决策或预测的原因。如果一个模型的决策比另一个模型的决策更容易理解,那么前者就比后者具有更好的可解释性^[10]。笔者团队从可解释的人工智能角度出发,开发基于XAI的CDSS——孟超肝病外脑。孟超肝病外脑的核心功能是自动实时提取医院信

息系统(hospital information system, HIS)产生的多模态数据中的临床维度,组合成临床证据集合,将临床证据集合输入权威知识图谱,得出高水平的人工智能结论。运用人工智能模拟高水平专家思维,自动收集临床证据,并给出肝病的辅助诊断与治疗意见、依据和决策过程。研究CDSS的可解释性,急需解决阻碍人工智能在医疗领域实现落地应用的三大黑盒问题:数据黑盒、模型黑盒、临床应用黑盒。笔者重点从上述3个方面进行阐述并介绍解决经验,为发展XAI提供新方法、新思路,进一步推进CDSS在医疗领域中的落地应用。

一、数据黑盒问题

数据黑盒问题主要体现在数据缺乏标准化、结构化、透明化。构建高质量的CDSS,让数据驱动人工智能的落地应用,必须从源头解决数据黑盒问题。若没有成功的数据治理,会出现数据标准化程度低、结构化程度低、不透明以及可解释性差的问题,最终会导致基于知识图谱的决策模型和基于机器学习、深度学习的决策模型可靠性和准确率极大降低,导致医务工作者对CDSS缺乏信任感。因此,各级医疗机构和医疗管理部门应建立有效的数据治理机制,解决数据黑盒问题,是构建基于XAI CDSS的前提和基础。笔者团队围绕数据的标准化、结构化、透明化3个方面进行探索,拟解决数据黑盒问题。

(一)数据标准化

笔者团队在国际肝胆胰协会中国分会的指导下,经各领域专家评审后制订《原发性肝癌和慢性乙型病毒性肝炎临床标准化术语集》(图1)。2021年7月,由笔者团队制订的《原发性肝癌临床数据集》地方标准被批准发布,与国际上已经形成的医学系统化命名(临床术语、医学主题词表、临床药品标准命名术语表、国际疾病分类与代码)等医学术语标准实现科学衔接,实现数据治理的标准化,解决数据定义不清、概念混淆的黑盒问题^[11]。

(二)数据结构化

基于上述医学术语标准,团队自主研发前结构化的电子病历、影像报告、病理报告的人工智能数

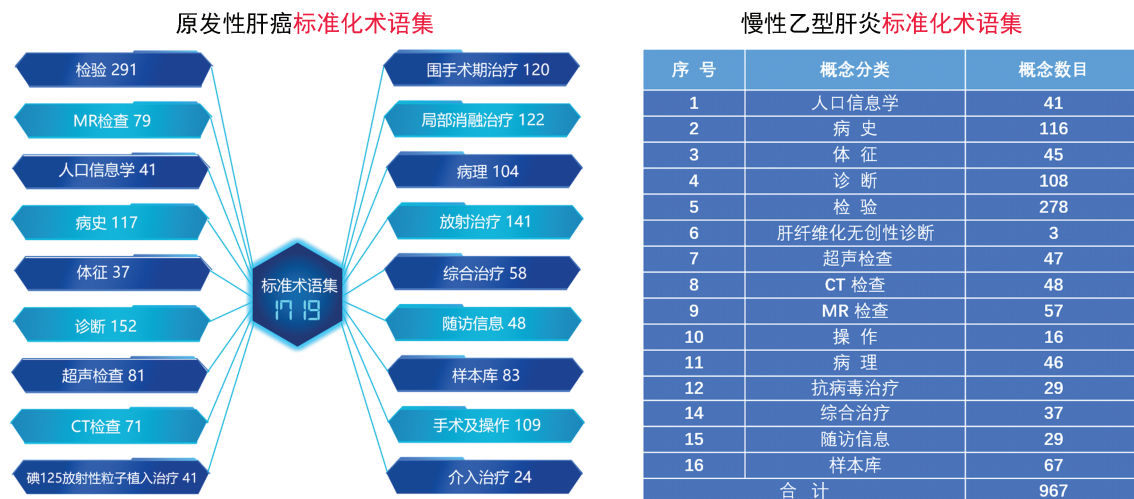


图1 原发性肝癌和慢性乙型肝炎临床标准化术语集

Figure 1 Clinical standardized glossary of primary carcinoma of liver and chronic viral hepatitis B

据采集系统(图2),便于CDSS能够快速、精确抓取CDSS决策所需维度。前结构化的人工智能数据采集系统有助于医务工作者理解CDSS提取临床数据来源,实现数据治理的结构化,解决数据结构混乱的黑盒问题。

(三)数据透明化

科研工作者开展临床研究,需要筛选特定的人群。研究人员可以在笔者团队前期建立的肝病和肝癌大数据平台设置纳入标准和排除标准,原数据从大数据平台经过还原、清洗、转换、融合、人工校验后,得到符合入组和排除标准的数据。每一步处理过程造成的数据损失均有记录(图3),提取的数据可核查、溯源,实现数据治理的透明化,避免数据提取的黑盒问题。

二、模型黑盒问题

孟超肝病外脑不仅是一种基于知识图谱和HIS的CDSS,而且还是一种基于机器学习、深度学

习和大数据的CDSS。孟超肝病外脑可以实时自动采集HIS产生的多模态数据(医疗结构化数据、医疗文本数据、医疗影像数据),组成标准特征集合,输入到基于知识图谱的模型或基于大数据的机器学习和深度学习算法模型,得出高水平的人工智能结论(图4)。模型是CDSS的核心技术和重要组成部分。由于模型存在黑盒问题,会导致医务工作者对CDSS缺乏信任。

(一)基于知识图谱和HIS的CDSS

此类型CDSS透明度不足,会导致医务工作者无法理解其采用的算法模型。因此,需要对CDSS算法模型进行可视化。HIS不仅贮存医疗档案,还包含大量的结构化数据、医疗文本数据、医学影像数据等。孟超肝病外脑能够将这些多模态数据汇总后与笔者团队自主研发的基于知识图谱的模型进行实时数据共享和对接,为患者提供实时的辅助决策。基于知识图谱和HIS的CDSS以临床指南为

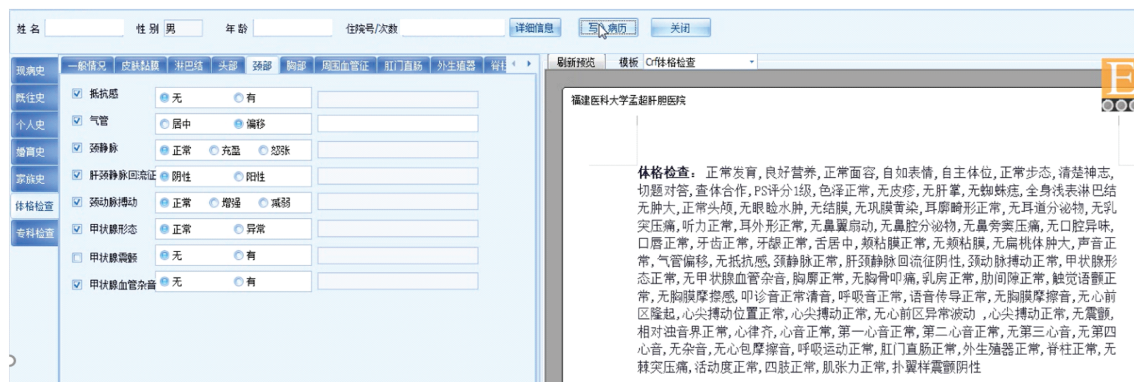


图2 前结构化电子病历、影像报告、病理报告的人工智能数据采集系统

Figure 2 Artificial intelligence data acquisition system of pre-structured electronic medical records, imaging reports and pathological reports

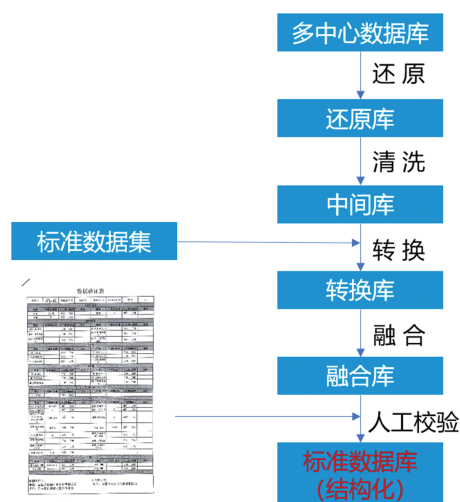


图3 透明化的数据处理过程

Figure 3 Transparent data processing

过 程	例 数	损 失%
1. 源数据—巨大肝细胞癌	10 283	
2. 基础资料缺失	220	2.14
3. 剔除复发性肝细胞癌	530	5.15
4. 剔除有癌栓	605	5.88
5. 剔除有转移	321	3.12
6. 剔除合并胆管细胞癌	55	0.53
7. 剔除合并混合癌	33	0.32
8. 剔除既往恶性肿瘤史	182	1.77
9. 剔除术前介入治疗	81	0.78
10. 剔除围手术期死亡病人	16	0.15
11. 剔除复发日期小于 2 个月	360	3.50
12. 最终	7 880	



图4 孟超肝病外脑的核心框架:基于全局可解释人工智能的临床决策支持系统

Figure 4 Core framework of the brain of Mengchao liver disease: clinical decision support system based on globally explainable artificial intelligence

理论支撑,临床指南包括疾病的诊断、分期分级、功能评分、治疗方案推荐,图像识别诊断等。目前临床指南大部分以自由文本的形式存在,医务工作者通过阅读与学习,进而开展临床诊断与治疗工作。由于我国患者基数大和医疗健康大数据的与日俱增,自由文本形式的临床指南已经不能满足临床实际需求。随着计算机在临床实践和医院信息化过程中的逐步应用,能够将自由文本形式的临床指南转化成常用的可解释算法模型(如线性回归、决策树、决策规则等),利用计算机进行建模,构建可被医务工作者理解并具有可解释性的CDSS,使其在实际临床工作中发挥重要作用。

在肝功能 Child-Pugh 分级评分知识决策模型上,将肝功能评分表转换成机器学习中内在可解释的决策树模型(图5)。此类模型自身结构简单,使用者可以清晰地看到模型内部结构,且模型结构具有解释效果,模型设计时就已具备可解释性(如朴

素贝叶斯、线性回归、决策树、基于规则的模型)^[12]。决策树模型是一种基于 IF-THEN 规则、自顶向下对数据进行分类的树形结构模型,由节点与有向边组成。样本从根节点被划分到不同的子节点中,子节点进行特征选择,直到满足结束条件^[13-17]。

在中国分期知识决策模型上,笔者团队是基于现有的《中国肝癌分期方案》,利用 Oracle 内在可解释的 CASE-WHEN 条件判断语句自动判断分期结果^[18]。Oracle 分期判断的建模过程如图6所示。若全身情况卡氏评分为 3~4 分,则为Ⅳ期。若肝功能 Child-Pugh 分级为 C 级,则为Ⅳ期。若有肝外转移,则为Ⅲb 期。若有血管侵犯,则为Ⅲa 期。若肿瘤数目≥4 个,则为Ⅱb 期。若肿瘤数目为 2~3 个且肿瘤最大径>3 cm,则为Ⅱa 期。若肿瘤数目为 2~3 个且肿瘤最大径≤3 cm,则为Ⅰb 期。若肿瘤数目为 1 个且肿瘤最大径>5 cm,则为Ⅰb 期。若肿瘤数目为 1 个且肿瘤最大径≤5 cm,则为Ⅰa 期。

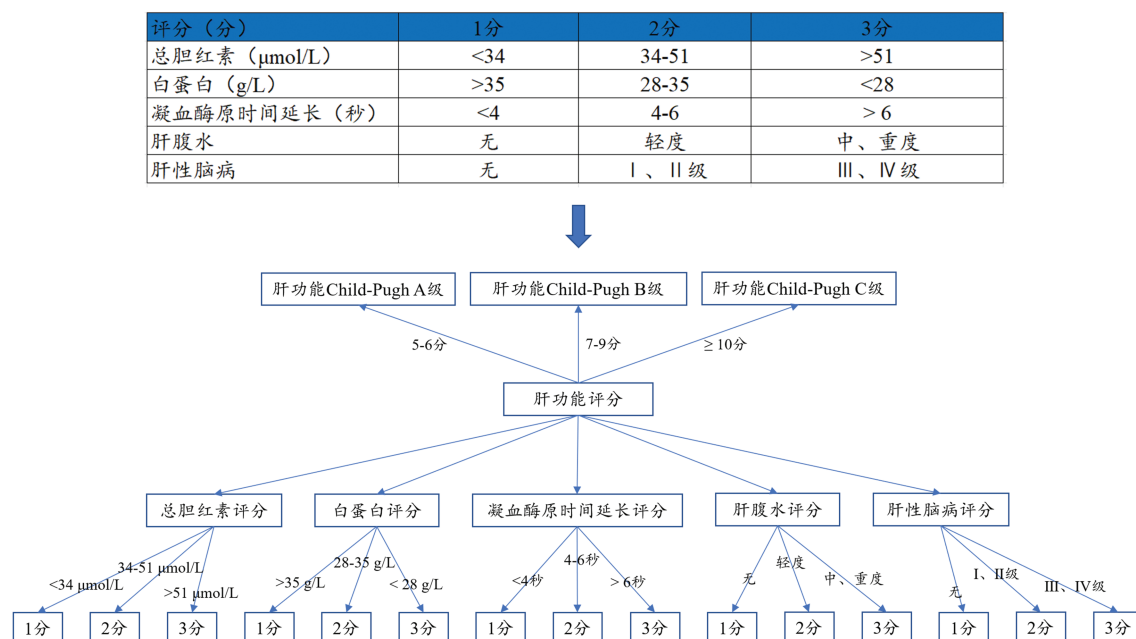
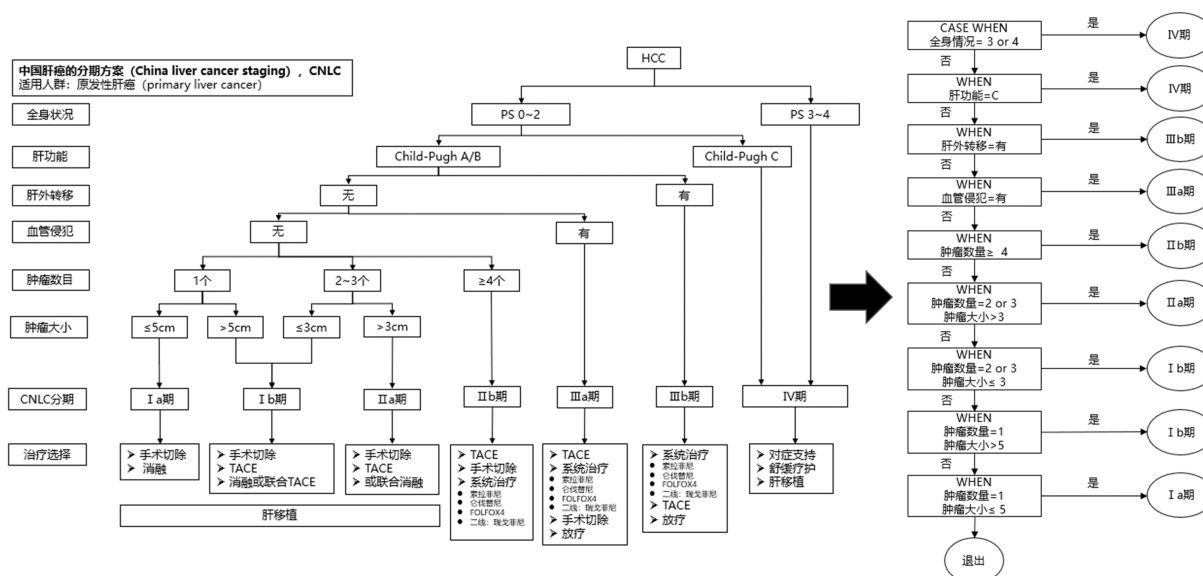


图5 基于决策树的肝功能Child-Pugh分级评分

Figure 5 Child-Pugh grading score of liver function based on decision tree



注:TACE为肝动脉化疗栓塞术;CNLC为中国肝癌分期

图6 基于CASE-WHEN条件判断语句的中国肝癌分期模型

Figure 6 China liver cancer staging model based on CASE-WHEN conditional judgment sentence

(二) 基于机器学习和深度学习以及大数据的 CDSS

此类型 CDSS 通常缺乏医疗健康大数据和标准化、结构化、透明化的数据处理流程, 缺乏对临床特征及其相关权重(或者重要性程度)的可视化, 存在算法模型黑盒。近年来, 随着人工智能机器学习与深度学习的发展和突破, 基于医疗健康大数据的算法模型有望弥补基于知识图谱算法模型的局限性, 从而可应用于更为广泛、精确的 CDSS。构建具有

可解释性的基于机器学习和深度学习以及大数据的 CDSS, 需要解决算法模型黑盒问题和建立标准化、结构化、透明化数据处理流程, 以及对机器学习与深度学习筛选出的临床特征及其相关权重进行可视化。

构建符合循证医学, 具有可解释性的算法模型通常需要多中心、大规模的临床数据, 笔者团队前期搜集海军军医大学第三附属医院、福建医科大学孟超肝胆医院及 92 家肝病医联体单位的肝癌临床

数据,采用大数据治理技术,构建标准化、结构化、透明化的原发性肝癌大数据(primary liver cancer big data, PLCBD)。PLCBD 的建设保证了数据的准确性和规范性,为构建可解释的算法模型提供数据支撑平台^[19]。

笔者团队基于医疗健康大数据,采用内在可解释的算法模型——逻辑回归模型,构建术前微血管侵犯风险预测模型。该模型有 2 个可能结果的概率分类问题,它是针对分类问题的线性回归模型扩展。逻辑回归模型不仅是一个分类模型,而且还给出概率,同时对模型中的计算公式进行可视化(图 7A)。笔者团队基于医疗健康大数据,采用复杂模型(随机森林、支持向量机、LightGBM、CatBoost)构建肝癌早期复发风险预测模型。内在可解释性机器学习模型(如决策树、决策规则、逻辑回归等)虽然具有较好的解释性,但其在精度上依旧无法与复杂模型相媲美。

事后可解释是模型训练完成后,增强模型的可解释性,挖掘模型学习到的信息。有的模型自身结构比较复杂,使用者很难从模型内部知道结果的推理过程,模型结果也不带有解释的语言,通常只是给出预测值。复杂模型的事后解析方法有可视化、扰动测试、代理模型等^[8]。在基于随机生存森林算法预测肝癌早期复发的模型中,笔者团队采用可视化方法让模型进行事后可解释,展示每项特征对复发结果的贡献程度。模型权重直接反应临床特征的重要性。权重绝对值越大,则该特征对肝癌早期复发贡献越大,反之则越小。如果权重值为正,则该特征与肝癌早期复发呈正相关,反之呈负相关。重要程度由高到低分别为肿瘤大小、大血管侵犯、微血管侵犯、肿瘤数目、AFP、卫星结节、肿瘤包膜、Edmondson 分级、Alb、性别、病因学、肝硬化程度、PLT、TBil、年龄(图 7B)。对复杂机器学习算法模型的可视化能够让医师理解模型输入的特征及权重,削弱复杂机器学习算法模型的黑盒效应^[20]。在基于深度学习图像识别肝癌的模型中,基于国际诊断标准,笔者团队自主研发四期联动的图像配准及标注方法,通过捕捉有无快进、快出和包膜强化特征,判断病灶区域是否发生癌变,并对快进、快出以及包膜强化的识别结果进行可视化,最后对肿瘤大小和位置进行描述。笔者团队采用深度学习算法中的特征金字塔网络对多尺度特征信息进行融合,

进而提高目标检测的精度。特征金字塔网络是 ResNet 或 DenseNet 等通用特征提取网络的附加组件,可以将经典网络组合提升原网络效果^[21]。笔者团队先通过深度学习特征金字塔网络目标检测算法识别快进、快出、包膜强化等肝癌典型特征,接着基于快进、快出、包膜强化的目标检测识别结果,采用广义加性模型判断病灶区域是否发生癌变^[22]。广义加性模型与传统基于深度学习单纯输出分类结果的图像识别模型比较,其优势主要体现在快进、快出、包膜强化特征与有无肝癌结果以加性的方式呈现,符合国际诊断标准和医师阅片诊断思维。该模型具有较强的可解释性(图 7C)。

三、临床应用存在黑盒问题

临床应用是 CDSS 实现落地应用最为关键的环节。若单纯输出人工智能结论,无法给出其结论所需的临床证据和决策过程,会导致使用者无法理解模型的内部规则,从而引起使用者对 CDSS 辅助决策结果的质疑。因此,促进 CDSS 的落地应用,需要解决临床应用的黑盒问题。

在基于知识图谱和 HIS 的 CDSS 临床应用中,笔者团队在孟超肝病外脑界面中对决策论据的数据来源、基于知识图谱的推理流程图和人工智能结论进行可视化。在肝功能 Child-Pugh 分级评分模型中,血清 TBil、血清 Alb、PT 采集于 2021-08-24 的检验报告,腹水采集于 2021-08-25 的超声检查报告,肝性脑病采集于 2021-08-24 病例的病史信息。基于肝功能评分知识图谱,通过内在可解释性的决策树模型判定该患者肝功能 Child-Pugh 分级评分为 5 分, A 级(图 8)。

在中国肝癌分期中,全身情况采集于 2021-08-20 病例的病史信息,肝功能 Child-Pugh 分级评分由上述模型人工智能计算获取,有无肝外转移、有无血管侵犯、肿瘤数目、肿瘤最大径采集于 2021-08-20 的 MRI 检查报告。基于中国肝癌分期知识图谱,通过基于决策规则 CASE-WHEN 条件判断语句自动评估分期结果(图 9)。

在 RFA 治疗方案推荐的临床应用中,笔者团队还将决策论据分为“自动处理型”和“医师处理型”2 类,“自动处理型”的决策论据由 CDSS 自动采集 HIS 临床数据;针对“医师处理型”决策论据,团队设计人机交互选项,让医务工作者参与 CDSS 辅助决策过程(图 10)。

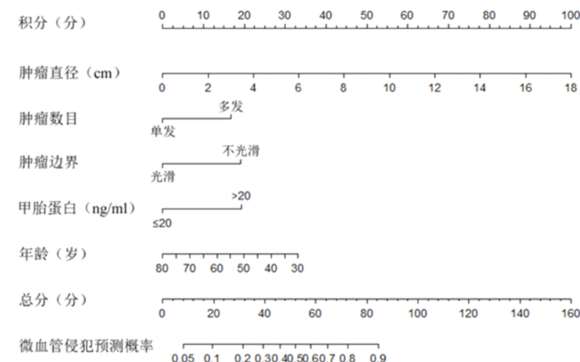
肝细胞癌MVI术前预测模型

适用人群:

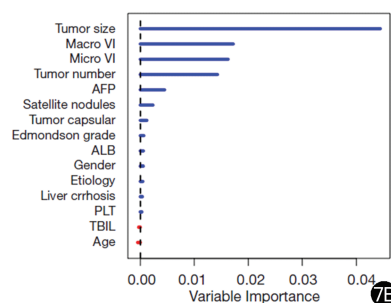
1. 术前肝功能Child A或Child B级。
2. 计划行肝癌切除术。
3. 术前1月内在院行MRI增强检查。
4. 无术前其他抗肿瘤治疗。
5. 肝静脉、门静脉、胆管及下腔静脉未见肉眼癌栓。
6. 无肝外远处转移。

计算公式:

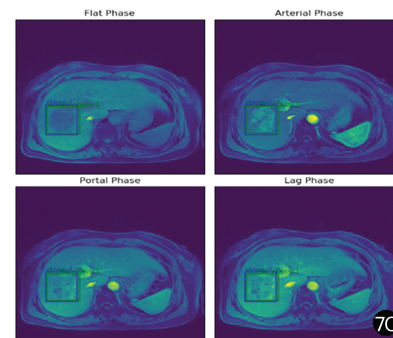
$$Y = 0.0747 + 0.3740 \times (\text{tumor size}) + 1.1314 \times \text{Tumor number (0: 单发; 1: 多发)} + 1.2886 \times \text{Tumor margin (0: 光滑; 1: 不光滑)} + 1.3030 \times \text{AFP (0: } \leq 20; 1: > 20) - 0.0447 \times (\text{Age})$$

微血管侵犯概率: $P = 1/(1 + e^{-y})$ 

7A



7B



7C

注: MVI为微血管侵犯; MRI为磁共振成像; Tumor size为肿瘤直径(肿瘤最大径); Tumor number为肿瘤数目; Tumor margin为肿瘤边界; AFP为甲胎蛋白; Age为年龄; Satellite nodules为卫星结节; Tumor capsular为肿瘤包膜; Edmondson grade为Edmondson分级; Alb为白蛋白; Gender为性别; Etiology为病因; Liver crrhosis为肝硬化程度; PLT为血小板; TBil为总胆红素

图7 算法模型的可解释性 7A: 基于逻辑回归模型的微血管侵犯风险预测模型; 7B: 基于随机生存森林算法预测肝癌早期复发的模型; 7C: 基于深度学习图像识别肝癌模型(广义加性模型)

Figure 7 The interpretability of algorithm model 7A: Risk prediction model of microvascular invasion based on logistic regression model; 7B: Prediction model of early recurrence of liver cancer based on random survival forest algorithm; 7C: Liver cancer model based on depth learning image recognition (generalized additive model)

AI 报告

AI结论:

CHILD-PUGH评分为5分, A级。

诊断历史

复制

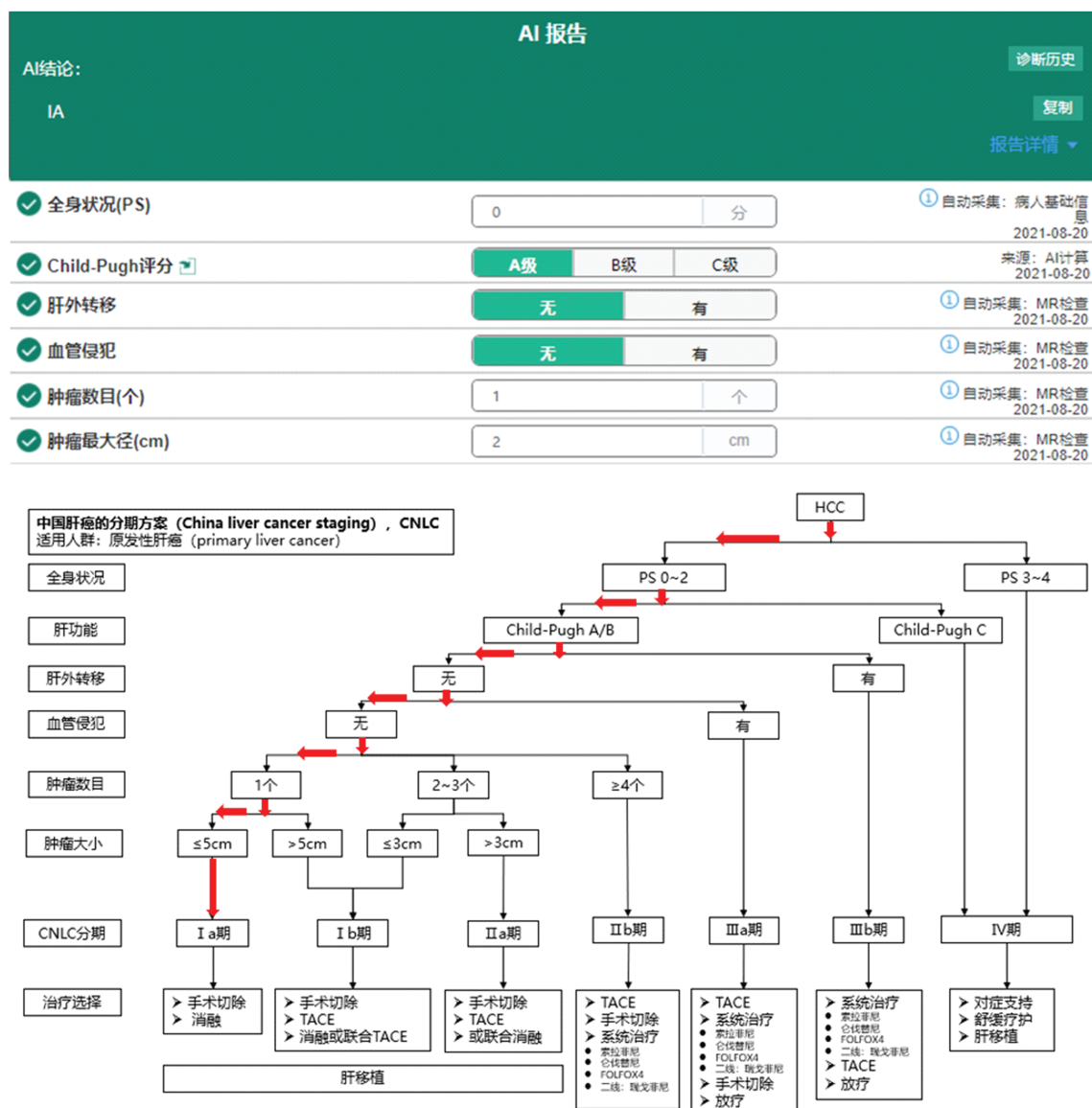
报告详情

✓ 血清总胆红素(μmol/L)	17.7	μmol/L	自动采集: 检验报告 2021-08-24
✓ 血清白蛋白(g/L)	38	g/L	自动采集: 检验报告 2021-08-24
✓ 凝血酶原时间延长(秒)	2	秒	自动采集: 检验报告 2021-08-24
✓ 腹水	无 轻度 中、重度		① 自动采集: 超声检查 2021-08-25
✓ 肝性脑病	无 1 2 3 4		① 自动采集: 病人基础信息 2021-08-24

评分 (分)	1分	2分	3分
总胆红素 (μmol/L)	<34	34-51	>51
白蛋白 (g/L)	>35	28-35	<28
凝血酶原时间延长 (秒)	<4	4-6	> 6
肝腹水	无	轻度	中、重度
肝性脑病	无	I、II级	III、IV级

图8 基于决策树的肝功能Child-Pugh分级评分临床应用界面

Figure 8 Clinical application interface of Child-Pugh grading score of liver function based on decision tree



注: TACE 为肝动脉化疗栓塞术; CNLC 为中国肝癌分期; PS 为体力活动状态; HCC 为原发性肝癌; FOLFOX4 为奥沙利铂+亚叶酸钙+氟尿嘧啶

图9 基于决策规则 CASE-WHEN 条件判断语句的中国分期临床应用界面

Figure 9 Clinical application interface of China liver cancer staging model based on CASE-WHEN conditional judgment sentence

在基于机器学习和大数据的 CDSS 临床应用中,该模型开发了个体化在线预测工具,让医师理解特征对决策的贡献程度与模型内部规则(图 11)。

在基于深度学习图像识别肝癌的模型中,该模型能够对快进、快出以及包膜强化的识别结果进行可视化,最后对肿瘤大小和位置进行描述(图 12)。

在临床应用中,笔者团队采用内在可解释性模型;对复杂模型进行可视化,并设计个体化在线预测工具;针对“医师处理型”的决策论据,设计人机交互选项等方法,削弱临床应用黑盒问题。

四、结语

近年来,随着大数据产业的蓬勃发展,围绕人

工智能技术的机器学习和深度学习取得突破性进展^[23]。越来越多的 CDSS 开始逐步应用机器学习集成学习模型和深度学习模型,与传统内在可解释模型比较,上述 2 类模型更能够充分地利用大数据资源,模型的准确性能实现质的飞跃。

机器学习模型的准确性越好,是否就能够越信任此类模型而忽略决策的理由呢?显然,答案是否定的。虽然机器学习集成学习模型(如 XGBoost、LightGBM),深度学习模型(如神经网络模型)在许多行业中取得良好效果,但是它们具有 1 个共同特点:模型内部结构非常复杂,其运行机制犹如黑盒,难以用人类可以理解的语言去表达,模型输出的



图 10 基于人机交互的射频消融治疗方案的临床应用界面

Figure 10 Clinical application interface of radiofrequency ablation based on human-computer interaction

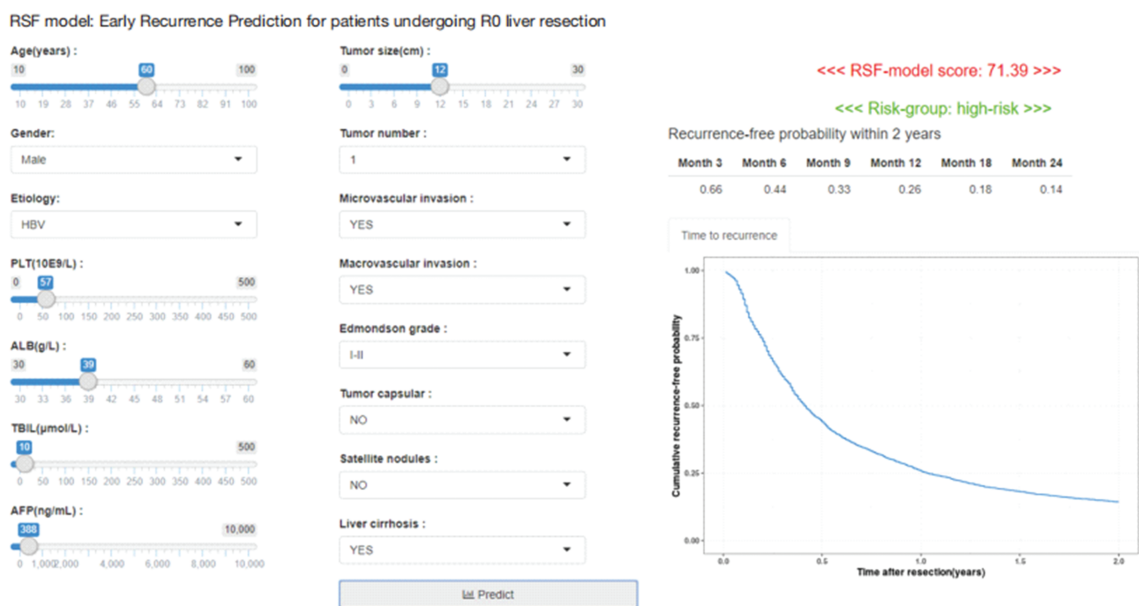


图 11 个体化的在线预测工具

Figure 11 Individualized online prediction tool

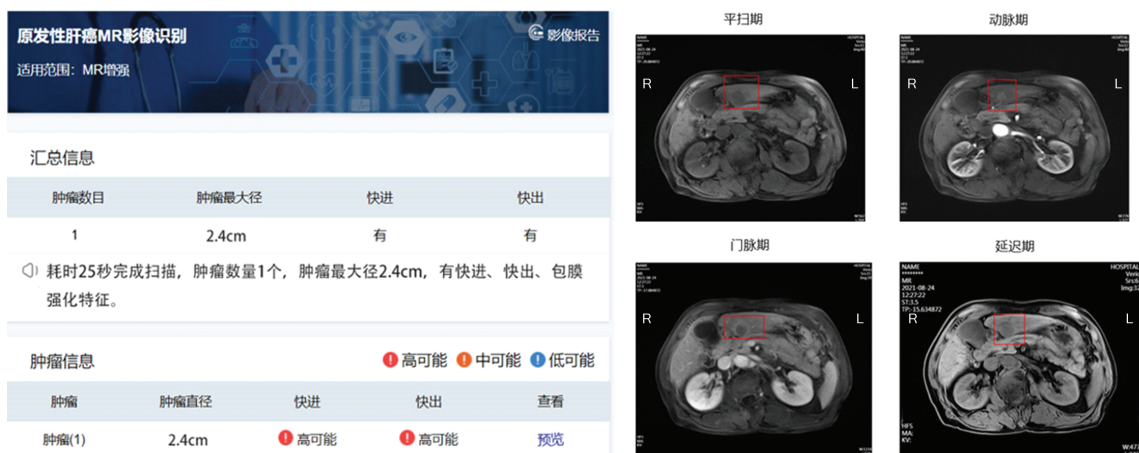


图 12 深度学习图像识别模型的临床应用界面

Figure 12 Clinical application interface of depth learning image recognition

结果也难以被解释,使得其在有关生命安全的应用中受到巨大挑战。因此,需要赋予模型“表达能力”,这样才能够让使用者更加理解和信任模型。除了追求模型的准确性之外,还应增加可解释性作为模型评价指标。模型设计者在设计模型时,应该从模型的准确性与可解释性 2 个角度进行考虑。香港大学张爱军教授提出:横坐标代表模型的可解释性,越往正方向,代表模型的可解释性越好,纵坐标代表模型的准确性,越往正方向,代表模型的准确性越高(图 13)^[24-25]。因此,对于传统内在可解释模型(广义加性模型、决策树模型、逻辑回归模型、线性回归模型),模型可解性较强,在部分信噪比高的领域,拟合效果通常没有机器学习集成学习模型和深度学习效果好。应在保证其可解释性的前提下,改良模型结构,提高模型准确性。对于机器学习集成学习模型(XGBoost 模型、LightGBM 模型),深度学习模型(神经网络模型),其内部结果非常复杂,在部分信噪比低的领域,很容易将噪声拟合进去不易被发现,导致模型可解释性差。对于此类情况,应该降低模型的复杂度,

以“牺牲”模型的准确性换取模型的可解释性;在保证模型准确度的基础上,采用事后可解释方法及可视化工具,获得模型可解释性。可解释学习研究的最终目的是保证模型高准确性的同时赋予模型可解释性。

CDSS 应用于临床实践尚存在许多问题,包括如何集成电子病历,实现结构化数据的自动获取;如何构建标准化的数据集,提供基于证据的临床实践指南;如何与临床医师的诊断与治疗工作相结合,实现人机交互;如何保证模型高准确性的同时赋予模型可解释性;如何在 CDSS 界面中对证据来源与模型内部结构进行可视化,增强临床医师对模型的信任感等。尽管诸多问题有待解决,但 CDSS 是医院信息化、智能化发展的必然趋势,对全面提升临床医师诊断与治疗能力、提高医疗质量具有重要意义。笔者相信:在未来,CDSS 必将成为医疗管理的一个重要辅助工具。

利益冲突 所有作者均声明不存在利益冲突

作者贡献声明 方国旭:撰写论文;郭鹏飞、范鉴慧、丁宗仁、张清华、魏光亚:讨论并修改论文;李海涛、刘景丰:拟定写作思路,指导撰写文章并最后定稿

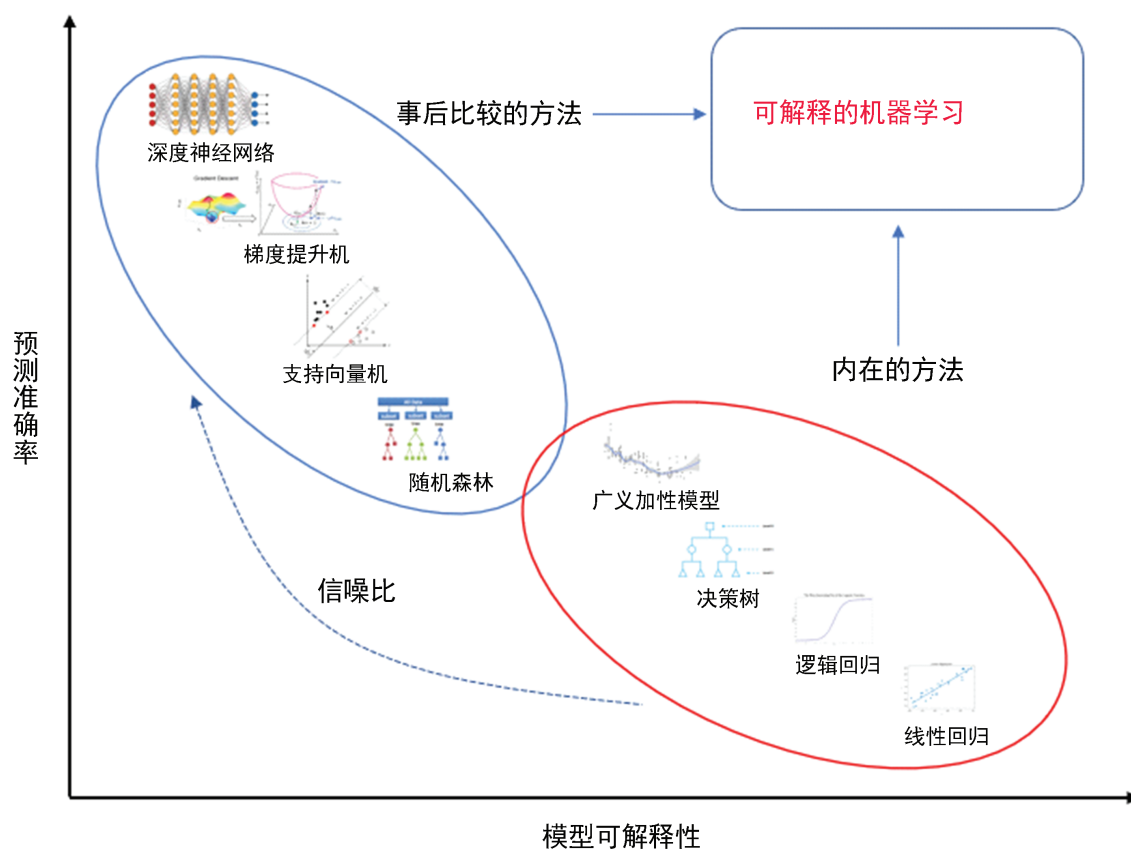


图 13 模型可解释性与准确性的关系

Figure 13 Relationship between model interpretability and accuracy

参 考 文 献

- [1] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. [S.L.]: IEEE, 2015. DOI: 10.1109/cvpr.2015.7298682.
- [2] Sun Y, Liang D, Wang X, et al. Deepid3: face recognition with very deep neural networks[EB/OL]. (2015-2-3)[2022-11-20]. <https://arxiv.org/abs/1502.00873>. DOI: 10.48550/arXiv.1502.00873.
- [3] Taigman Y, Yang M, Ranzato MA, et al. Deepface: closing the gap to human-level performance in face verification: 2014 IEEE Conference on Computer Vision and Pattern Recognition[C]. [S.L.]: IEEE, 2014. DOI: 10.1109/cvpr.2014.220.
- [4] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite: 2012 IEEE conference on computer vision and pattern recognition[C]. [S.L.]: IEEE, 2012. DOI: 10.1109/cvpr.2012.6248074.
- [5] Tobiyama S, Yamaguchi Y, Shimada H, et al. Malware detection with deep neural network using process behavior: 2016 IEEE 40th annual computer software and applications conference (COMPSAC)[C]. [S.L.]: IEEE, 2016. DOI: 10.1109/compsac.2016.151.
- [6] Rajpurkar P, Irvin J, Zhu K, et al. Chexnet: radiologist-level pneumonia detection on chest X-rays with deep learning: 2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)[C]. [S.L.]: IEEE, 2020. DOI: 10.1109/icspcs47537.2019.9008684.
- [7] Ibrahim M, Louie M, Modarres C, et al. Global explanations of neural networks: Mapping the landscape of predictions AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society[C/OL]. [S.L.]: [s.n.], 2019[2022-11-20]. <https://dl.acm.org/doi/10.1145/3306618.3314230>. DOI: 10.1145/3306618.3314230.
- [8] 纪守领, 李进锋, 杜天宇, 等. 机器学习模型可解释性方法、应用与安全研究综述[J]. 计算机研究与发展, 2019, 56(10): 26. DOI: 10.7544/issn1000-1239.2019.20190540.
- [9] Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)[J]. IEEE Access, 2018, 6: 52138-52160. DOI: 10.1109/access.2018.2870052.
- [10] Antoniadi AM, Du Y, Guendouz Y, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review[J]. Applied Sciences, 2021, 11(11): 5088. DOI: 10.3390/app1115088.
- [11] 刘景丰, 周伟平, 刘红枝. 原发性肝癌临床标准化术语[J]. 临床外科杂志, 2020, 28(8): 27-29. DOI: CNKI: SUN: LCWK. 0. 2020-08-008.
- [12] Alvarez-Melis D, Jaakkola TS. Towards robust interpretability with self-explaining neural networks//Proceedings of the 32nd International Conference on Neural Information Processing Systems[C]. New York: Curran Associates Inc., 2018: 7786-7795.
- [13] Huysmans J, Dejaeger K, Mues C, et al. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models[J]. Decision Support Systems, 2011, 51(1): 141-154. DOI: 10.1016/j.dss.2010.12.003.
- [14] Breslow LA, Aha DW. Simplifying decision trees: a survey [J]. The Knowledge Engineering Review, 1997, 12(1): 1-40. DOI: 10.1017/s0269888997000015.
- [15] Frank E, Witten I H. Generating accurate rule sets without global optimization: proceedings of the fifteenth international conference on machine learning[C]. [S.L.]: Morgan Kaufmann, 1998.
- [16] Quinlan JR. Generating production rules from decision trees: proceedings of the 10th international joint conference on Artificial intelligence[C]. [S.L.]: [s.n.], 1987.
- [17] Deng H. Interpreting tree ensembles with in trees[J]. J Data Sci, 2019, 7(4): 277-287. DOI: 10.1007/s41060-018-0144-8.
- [18] 杨云飞. 基于 Oracle 的数据可视化工具开发与实现[D]. 苏州: 苏州大学, 2016. DOI: 10.7666/d.Y3164032.
- [19] 王垒, 郭鹏飞, 杨远, 等. 原发性肝癌大数据建设初步探索[J]. 中华肝胆外科杂志, 2019, 25(9): 695-698. DOI: 10.3760/cma.j.issn.1007-8118.2019.09.014.
- [20] Zeng J, Zeng J, Lin K, et al. Development of a machine learning model to predict early recurrence for hepatocellular carcinoma after curative resection[J]. Hepatobiliary Surg Nutr, 2022, 11(2): 176-187. DOI: 10.21037/hbsn-20-466.
- [21] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. [S.L.]: IEEE, 2017. DOI: 10.1109/cvpr.2017.106.
- [22] Wood SN. Generalized additive models: an introduction with R[M]. New York: Chapman and Hall/CRC, 2006. DOI: 10.1201/9781420010404.
- [23] 刘景丰, 刘红枝, 陈振伟, 等. 肝病和肝癌大数据平台建设体系及其初步应用[J]. 中华消化外科杂志, 2021, 20(1): 46-51. DOI: 10.3760/cma.j.cn115610-20201126-00742.
- [24] Yang Z, Zhang A, Sudjianto A. Enhancing explainability of neural networks through architecture constraints[J]. IEEE Trans Neural Netw Learn Syst, 2021, 32(6): 2610-2621. DOI: 10.1109/TNNLS.2020.3007259.
- [25] Yang Z, Zhang A, Sudjianto A. GAMI-Net: an explainable neural network based on generalized additive models with structured interactions[J]. Pattern Recognition, 2021, 120: 108192. DOI: 10.1016/j.patcog.2021.108192.