

知识与数据融合驱动的可解释原发性头痛辅助诊断模型

于海洋¹, 李恒¹, 韩珣², 董钊², 尹梓名¹

(1. 上海理工大学健康科学与工程学院, 上海 200093; 2. 解放军总医院第一医学中心神经内科, 北京 100039)

摘要: 原发性头痛是神经内科常见的慢性疾病, 诊断过程依赖症状特征识别与临床经验。国际头痛疾病分类第3版 ICHD-3 为头痛诊断提供了标准, 但很多条目是定性描述, 边界不清, 难以直接建模。传统机器学习难以同时处理临床规则和复杂特征; 深度学习虽能提高分类效果, 但解释性不足, 临床使用受限。为此, 提出一种知识与数据融合驱动的可解释原发性头痛辅助诊断模型。首先, 模型由分层模糊逻辑和 KAN 网络两部分组成: 利用模糊规则对 ICHD-3 标准进行可解释映射; 通过 KAN 网络的函数逼近能力高效建模多维症状特征间的非线性关系。其次, 引入 SHAP 方法对模型特征重要性进行分析, 优化诊断关键变量的学习过程。最后, 将医学规则知识嵌入深度学习训练流程, 实现诊断规则与数据驱动的协同优化与动态权重调整。实验表明, 所提融合模型在准确率(0.958 4)、灵敏度(0.951 3)、特异度(0.961 8)和 Kappa 值(0.888 1)等指标上均优于比较方法, 诊断性能优越。对模型结构与推理过程的可视化分析表明, 模型在临床可解释性方面优秀, 具有较强的实用价值与推广前景。

关键词: 原发性头痛; 可解释人工智能; 分层模糊逻辑; KAN 网络; 知识与数据融合

DOI: 10.11907/rjdlk.261032

扫描二维码阅读全文: 

中图分类号: TP18

文献标识码: A

文章编号: 1672-7800(2026)007-0012-10

An Interpretable Primary Headache Auxiliary Diagnostic Model Driven by Knowledge and Data Fusion

YU Haiyang¹, LI Heng¹, HAN Xun², DONG Zhao², YIN Ziming¹

(1. School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

2. Department of Neurology, The First Medical Center of Chinese PLA General Hospital, Beijing 100039, China)

Abstract: Primary headache is a common chronic disease in neurology, and the diagnostic process relies on symptom feature recognition and clinical experience. The International Classification of Headache Diseases, 3rd edition ICHD-3, provides criteria for the diagnosis of headaches, but many entries are qualitative descriptions with unclear boundaries, making it difficult to model directly. Traditional machine learning is difficult to simultaneously process clinical rules and complex features; Although deep learning can improve classification performance, its interpretability is insufficient and its clinical use is limited. To this end, a knowledge and data fusion driven interpretable primary headache auxiliary diagnosis model is proposed. Firstly, the model consists of two parts: hierarchical fuzzy logic and KAN network: interpretable mapping of ICHD-3 standard using fuzzy rules; Efficiently modeling the nonlinear relationship between multidimensional symptom features through the function approximation ability of KAN network. Secondly, the SHAP method is introduced to analyze the importance of model features and optimize the learning process of diagnosing key variables. Finally, embedding medical rule knowledge into the deep learning training process enables collaborative optimization and dynamic weight adjustment of diagnostic rules and data-driven approaches. The experiment shows that the proposed fusion model outperforms the comparison method in terms of accuracy (0.958 4), sensitivity (0.951 3), specificity (0.961 8), and Kappa value (0.888 1), demonstrating superior diagnostic performance. The visualization analysis of the model structure and

收稿日期: 2025-07-28

基金项目: 国家自然科学基金项目(81801797)

作者简介: 于海洋(1997-), 男, 上海理工大学健康科学与工程学院硕士研究生, 研究方向为医学人工智能; 李恒(1999-), 男, 上海理工大学健康科学与工程学院硕士研究生, 研究方向为临床决策支持; 韩珣(1987-), 女, 博士, 解放军总医院第一医学中心神经内科副主任医师, 研究方向为头痛的诊疗; 董钊(1976-), 男, 博士, 解放军总医院第一医学中心神经内科主任医师, 研究方向为头痛的诊疗; 尹梓名(1986-), 男, 博士, 上海理工大学健康科学与工程学院副教授、博士生导师, 研究方向为医学人工智能。本文通讯作者: 韩珣、尹梓名。

reasoning process shows that the model has excellent clinical interpretability and strong practical value and promotion prospects.

Key Words: primary headache; explainable artificial intelligence; hierarchical fuzzy logic; Kolmogorov–Arnold network; knowledge–data fusion

0 引言

原发性头痛是一类常见的神经系统疾病,主要包括偏头痛(Migraine)、紧张型头痛(Tension-type Headache, TTH)和丛集性头痛(Cluster Headache, CH)^[1,2]。该类疾病在全球范围内具有极高的发病率和致残率,会显著损害患者生活质量^[3,4]。如何实现对原发性头痛的精准和规范化诊断,已成为神经内科临床实践中的重要问题^[5,6]。

国际头痛疾病分类第3版(ICHD-3)为原发性头痛诊断提供了较完整的诊断标准,内容覆盖头痛发作频率、持续时间、疼痛性质、疼痛部位、伴随症状、诱发与缓解因素等信息,但很多指标为定性描述,例如“中度至重度疼痛”“偶有恶心”等,表述边界不清,在症状不典型或特征交叉时容易导致误诊或漏诊^[7-9]。

随着人工智能在医学领域的不断发展,数据驱动模型已成为头痛辅助诊断的重要研究方向^[10,11]。传统机器学习方法(决策树、支持向量机等)可提升诊断效率,但对临床知识的表达能力有限,在症状边界模糊、语义不清时性能不稳定^[12,13]。深度学习模型虽具备较强的特征表征能力,但普遍存在“黑箱”问题,难以为临床决策提供可追溯的依据,限制了其在医疗场景中的应用价值^[14-18]。

为此,本文采用医学知识与数据融合驱动的方法,提出具有较高诊断性能和可解释性的智能诊断模型^[19,20]。首先,引入模糊逻辑(Fuzzy Logic, FL)对 ICHD-3 中的模糊描述(“轻度疼痛”“较频繁发作”等)进行建模,通过隶属度函数和规则将临床语言转化为可计算形式,以提升模型可解释性^[21-23];其次,采用分层模糊系统(Hierarchical Fuzzy Systems, HFS)分解复杂规则,处理多维症状之间的交互关系^[24];最后,采用科尔莫戈洛夫——阿诺尔德定理启发的神经网络(Kolmogorov–Arnold Network, KAN)作为特征融合与非线性建模模块^[25-28]。其中,KAN 是一种用于建模非线性关系的网络结构,相较于多层感知机(Multilayer Perceptron, MLP)而言,在参数规模较小时也能较好地刻画症状之间的交互关系,适合于头痛数据建模任务。

1 数据集与特征描述

1.1 数据来源与清洗

本文研究数据来自于 2014 年 5 月—2019 年 10 月全国 53 家头痛中心连续招募的门诊患者,共计 12 532 例。解放军总医院国际头痛中心医生基于问卷整理病例资料,并进行统一清洗。清洗时,剔除了关键信息缺失(缺失率>30%)且无法诊断的病例、继发性头痛及其他不纳入本文

研究的头痛类型,如图 1 所示。例如,对于多重诊断病例,仅保留“最困扰的头痛”;对于复诊病例,仅保留首次就诊记录;对于“药物过度使用性头痛合并慢性头痛”病例,按规则归为慢性头痛。最终,保留了 9 139 例样本,包括紧张型头痛 1 527 例、偏头痛 7 011 例、丛集性头痛 601 例。

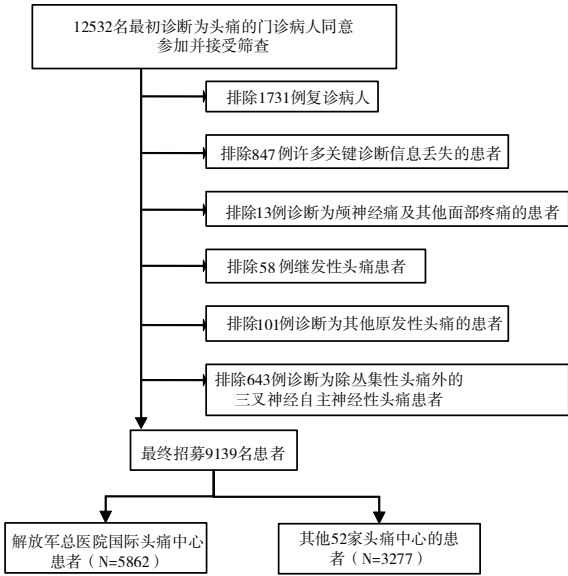


Fig. 1 Data cleaning process
图1 数据清洗过程

1.2 特征构建与编码

在特征工程阶段,依据 ICHD-3 标准和头痛领域专家经验,构建了一个包含 31 个头痛变量特征的结构化问卷。问卷内容包括人口学信息、头痛发作特征(部位、性质、程度、持续时间和发作频率等)、诱发与缓解因素、伴随症状(恶心、呕吐、畏光、畏声等)、自主神经症状、先兆、生活方式、家族史、前驱症状和教育程度等。

诊断标签由头痛专科专家根据临床资料给出,讨论后确定最终结果。本文采用了离散编码与数值编码相结合的方式:性别编码女=1、男=2;头痛部位编码空=0、单侧=1、双侧=2;头痛程度采用 VAS 评分(1~10 分);部分变量采用二元编码(否=1,是=2);教育程度和诊断标签采用分级编码,其中诊断标签为 TTH=0、Migraine=1、CH=2。

1.3 样例数据

为了说明数据的结构和编码方式,展示了部分脱敏后的原始样例数据(见表 1)。其中,数据已去除了身份识别信息,保留内容为建模所需的结构化变量和诊断标签。

1.4 变量分布可视化

为了观察 3 类头痛的特征差异,对年龄、病程和头痛程度等变量进行可视化,以期为后续建模和特征筛选提供参考,如图 2 所示。

Table 1 De-identified original sample data

表 1 脱敏后原始样例数据

ID	性别	年龄	病程/ month	部位	头痛 程度	发作 时长/h	是否 恶心	...	Diagno- sis
S001	2	48	184.0	2	4	12.0	1	...	0
S002	2	31	128.0	2	6	24.0	1	...	0
S003	1	33	88.0	2	8	48.0	2	...	1
S004	1	28	59.0	1	8	48.0	2	...	1
S005	1	16	36.0	1	10	1.0	2	...	2
S006	2	33	228.0	1	8	1.5	2	...	2
...	...	...	...	...	...	...	...	...	...

2 本文模型

本文模型由分层模糊逻辑层、KAN 特征融合层和分类器 3 个部分组成,如图 3 所示。

2.1 分层模糊逻辑层

分层模糊逻辑层用于将 ICHD-3 中的临床知识转化为可计算的模糊规则。为了明确夏普利加性解释(SHapley Additive exPlanations, SHAP)的使用阶段,在规则初始化阶段引入归一化 SHAP 特征权重^[29-31]。

首先,在训练集上训练数据驱动基线模型,计算各特

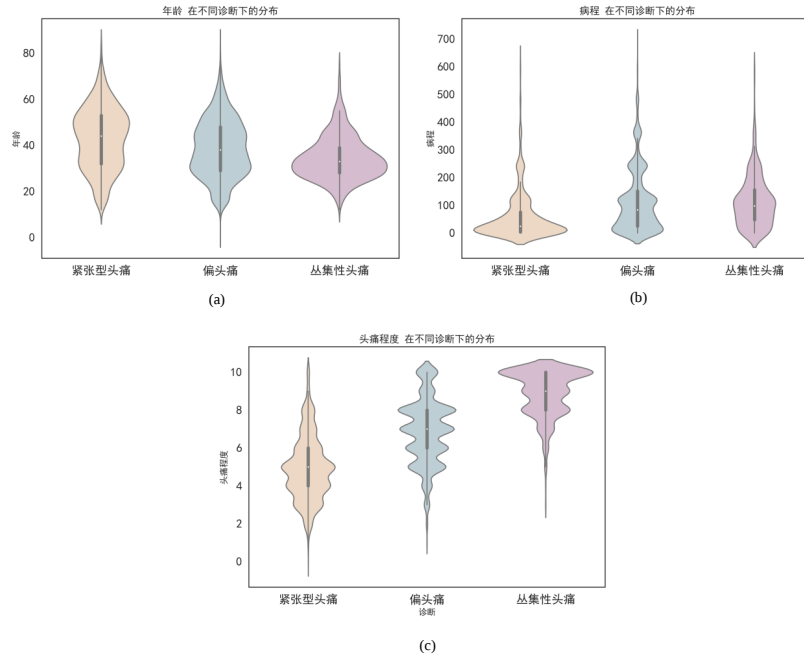


Fig. 2 Distribution of partial attribute values

图 2 部分属性数值分布情况

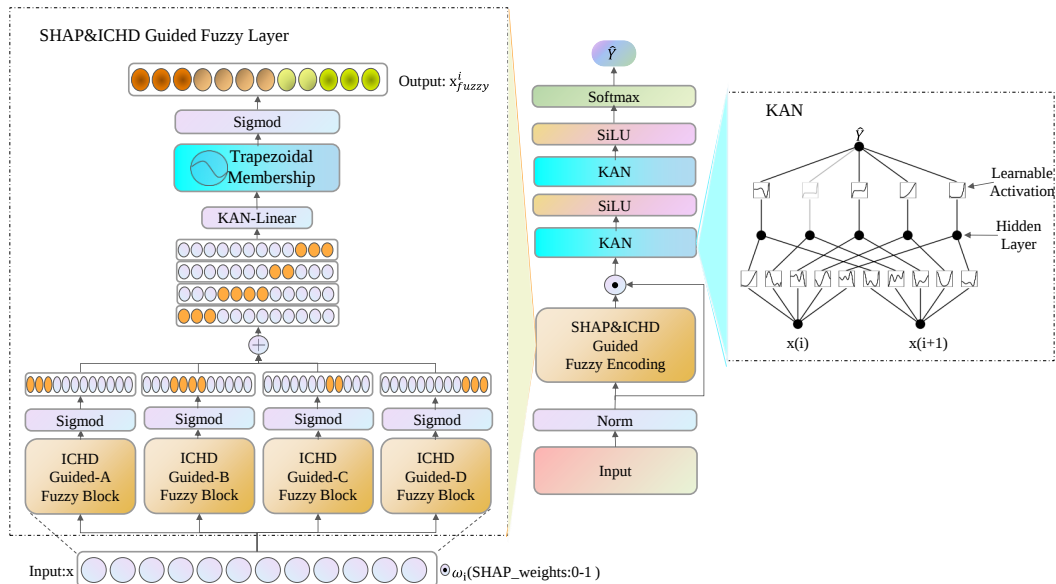


Fig. 3 Overall network architecture

图 3 网络总体架构

征的SHAP值并归一化;其次,将该权重用于模糊规则权重分配和输入加权;最后,采用梯形隶属度函数构建细粒度模糊表示,并按不同头痛类型的关键指标进行差异化初始化^[32]。

(1)偏头痛常伴恶心、畏光等症状,且此类症状出现后加重较快^[33]。因此,对恶心、畏光等指标设置陡峭上升的隶属度区间,使其在达到阈值后快速提高隶属度,以突出偏头痛的识别特征。

(2)紧张型头痛症状通常较稳定,头痛强度变化较缓^[34]。因此,采用在核心区间变化平稳、边界过渡平滑的对称梯形隶属函数,描述紧张型头痛的症状变化。以头痛强度为例,当强度处于一定范围内时,隶属度随强度缓慢变化,即可避免不必要的跳变。

(3)丛集性头痛发作时常伴鼻塞、流泪等自主神经症状,且起病急、程度重^[35]。因此,对相关指标设置敏感响应区间,使其在轻微变化时能快速触发较高的隶属度,以提升模型对丛集性头痛的识别能力。

在具体实现过程中,对于每个特征指标 x ,其梯形隶属度函数可表示为:

$$\mu(x; a, b, c, d) = f(x) = \begin{cases} 0, & x \leq a \\ (x-a)/(b-a), & a < x \leq b \\ 1, & b \leq x \leq c \\ (d-x)/(d-c), & c < x < d \\ 0, & x \geq d \end{cases} \quad (1)$$

式中: a 、 b 、 c 、 d 为根据ICHD-3标准和不同头痛类型的特点进行调整的参数。例如,对于偏头痛的恶心指标, b 、 c 的取值会使得隶属度在恶心症状较为明显时迅速达到1,以体现其陡峭上升区间的特点。

接下来,网络层对原始数据进行细粒度加权,并通过KAN线性层进行特征映射得到网络层输出。

$$x_{fine-grained}^i = x_i * \mu_i \Rightarrow x_{fine-grained} = x \odot \mu \quad (2)$$

$$\tilde{x}^{base} = \tilde{W}^{base} \cdot \text{SiLU}(x_{fine-grained}) \quad (3)$$

$$\tilde{x}^{spline} = \sum_{i=0}^{30} \sum_{j=0}^n \tilde{W}_{:,i,j}^{spline} \cdot B_{i,j} \quad (4)$$

$$x_{fuzzy} = \tilde{x}^{base} + \tilde{x}^{spline} \quad (5)$$

式中: x_i 表示单条数据的第 i 个特征的特征值; μ_i 表示单条数据的第 i 个模糊隶属度值; $x_i * \mu_i$ 表示每条数据的模糊细粒度特征融合; $\tilde{x}^{base}$ 、 $\tilde{x}^{spline}$ 分别表示数据经过KAN线性层处理过程中的SiLU激活值与B样条特征投影后的样条基函数; x_{fuzzy} 为模糊特征提取层的输出。

2.2 KAN特征融合层

KAN特征融合层主要融合分层模糊逻辑层输出的模糊隶属度特征与原始特征,以增强模型对数据特征的表达能力^[32]。该网络基于Kolmogorov-Arnold定理,能高效处理多元非线性关系。模块输入包括:原始特征 X 、分层模糊逻辑层输出的模糊隶属度特征 X_{fuzzy} 。为了实现二者融合,引入GroupFeatures函数,通过KAN线性变换将组级隶

属度映射回特征空间。

$$X_{enhanced} = X * \text{GroupFeatures}(X_{fuzzy}) \quad (6)$$

$$\text{GroupFeatures}(X_{fuzzy}) = W_{group} * X_{fuzzy} + b_{group} \quad (7)$$

式中: W_{group} 为权重矩阵; b_{group} 为偏置向量。

通过上述线性变换,模型可将模糊隶属度特征转换为与原始特征具有相同维度的特征表示,与原始特征进行逐元素相乘便能实现特征增强。

随后,KAN通过独特的结构对增强后的特征 $X_{enhanced}$ 进行处理,即通过学习不同特征之间复杂的非线性组合关系,挖掘数据中的潜在信息,为后续分类任务提供更具判别性的特征表示,KAN特征提取过程如式(3)—(5)所示,整体输出结果如式(8)所示。

$$X_{kan} = \text{KAN}(X_{enhanced}) \quad (8)$$

式中:KAN()为多层的KAN-Linear层组成; X_{kan} 为KAN网络提取后的分类特征,供分类器进行分类。

2.3 分类器设置

本文采用三分类结构,输出层使用Softmax函数,以KAN特征融合层的输出特征为输入,输出头痛类型。首先,通过线性变换将特征映射到三维输出空间;其次,经过Softmax函数得到三类头痛的预测概率。假设线性变换的权重矩阵为 W ,偏置向量为 b ,则分类器输出 P 为:

$$P(y_i | X_{kan}) = \text{Softmax}(X_{kan}) \quad (9)$$

$$\text{Softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^3 e^{z_k}} \quad (10)$$

式中: y_i 表示第 i 种头痛类型($i=1,2,3$ 分别对应紧张型头痛、偏头痛、丛集性头痛); X 表示输入的特征; z_j 为线性变换后的第 j 个输出值,模型将概率最大的类别作为最终分类结果。

3 实验环境与性能指标

实验在Windows 10环境下运行,硬件配置为64 GB内存和NVIDIA RTX 3090 GPU。软件环境为Python、PyTorch 1.12、CUDA 12.1,以及NumPy、Pandas和matplotlib等常用库。模型训练过程中采用Adam优化器,训练轮数(epoch)设为100,学习率设为 1×10^{-3} ,批处理大小(batch size)设为32,并引入早停机制降低模型过拟合风险。

由于三类头痛样本量不平衡,所有实验均采用分层抽样,单次实验采用分层8:2划分,消融实验采用分层十折交叉验证,以提升实验结果的稳定性和可重复性。研究以头痛专科专家的临床诊断结果为金标准,比较模型输出与专家诊断的一致性,评价指标包括机器学习领域及诊断模型一致性常用的敏感性(Sensitivity)、特异性(Specificity)、阳性预测值(PPV)、阴性预测值(NPV)、Youden指数和Kappa。其中,敏感性、特异性反映漏诊与误诊风险;PPV、NPV反映临床判别价值;Youden指数用于综合评价诊断效能;Kappa用于评估与专家诊断的一致性。

考虑到本文研究为多分类任务,敏感性、特异性、PPV、NPV和Youden等指数均采用one-vs-rest方式按类别进行计算。

$$\text{Sensitivity}(v_{\text{class}}) = \frac{\sum_{i=1}^n TP(\delta_i)}{\sum_{i=1}^n TP(\delta_i) + \sum_{i=1}^n FN(\delta_i)} \quad (11)$$

$$\text{Specificity}(v_{\text{class}}) = \frac{\sum_{i=1}^n TN(\delta_i)}{\sum_{i=1}^n TN(\delta_i) + \sum_{i=1}^n FP(\delta_i)} \quad (12)$$

$$\text{PPV} = \frac{\sum_{i=1}^n TP(\delta_i)}{\sum_{i=1}^n TP(\delta_i) + \sum_{i=1}^n FP(\delta_i)} \quad (13)$$

$$\text{NPV} = \frac{\sum_{i=1}^n TN(\delta_i)}{\sum_{i=1}^n TN(\delta_i) + \sum_{i=1}^n FN(\delta_i)} \quad (14)$$

$$\pi(v_{\text{class}}) = \frac{\sum_{i=1}^n TP(\delta_i) + \sum_{i=1}^n TN(\delta_i)}{\sum_{i=1}^n TP(\delta_i) + \sum_{i=1}^n TN(\delta_i) + \sum_{i=1}^n FP(\delta_i) + \sum_{i=1}^n FN(\delta_i)} \quad (15)$$

$$\text{Youden Index} = 1 - \alpha - \beta; \alpha = 1 - \text{Specificity}, \beta = 1 -$$

$$\text{Sensitivity} \quad (16)$$

式中: v_{class} 表示可由模型诊断的原发性头痛疾病;TP、FN、TN、FP分别表示真阳性、假阴性、真阴性和假阳性的数量;Sensitivity为在“金标准”判断下确实患病的人群中,该检测方法判定为阳性的比例;Specificity为在“金标准”判断下确实未患病的人群中,该检测方法判定为阴性的比例;PPV表示被诊断为阳性的个体中,实际患病的概率;NPV表示被诊断为阴性的个体中,实际未患病的概率;总符合率 π 表示所评估的诊断方法与金标准诊断方法结果的一致性程度;Youden指数反映诊断方法区分患者与非患者的能力;误诊率 α (假阳性率)表示健康个体被所评估的诊断方法误判为患者的概率;漏诊率 β (假阴性率)表示患者被所评估的诊断方法误判为健康个体的概率。

本文统计分析使用SPSS 16.0工具,通过Kappa值评估模型诊断结果与专家参考标准之间的一致性程度(见式(17)),使用各评价指标95%置信区间(CI)来报告估计结

果的不确定性范围。

$$\text{Kappa} = \frac{P_A - P_e}{1 - P_e} \quad (17)$$

式中: P_A 为实际观察到的一致率; P_e 为预期的一致率;Kappa ≥ 0.85 则一致性非常好,0.6 $\leq$ Kappa < 0.85 则一致性良好,0.45 $\leq$ Kappa < 0.6 则一致性中等,Kappa < 0.45 则一致性较差。

4 实验结果与分析

4.1 训练结果

在单次分层8:2的训练集和测试集中,Fuzzy-KAN模型的损失曲线如图4所示。由此可见,模型在早期学习速度较快,在第10个epoch后验证集损失逐渐趋于稳定,表明模型在当前验证集上的表现较稳定。当验证集损失连续多个epoch未继续下降时,早停机机制被触发,有助于降低过拟合风险。训练后期(第15个epoch后),验证集损失下降减慢,模型逐步进入收敛平台期。

图5为模型混淆矩阵及各类别的ROC曲线,详细性能

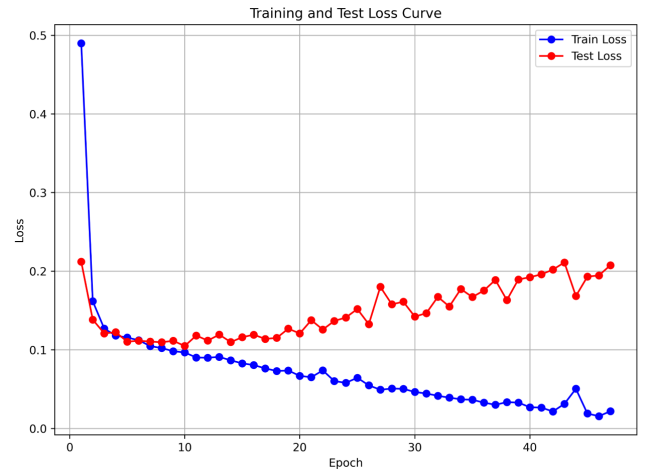


Fig. 4 Loss curve of the Fuzzy-KAN model

图4 Fuzzy-KAN模型损失曲线

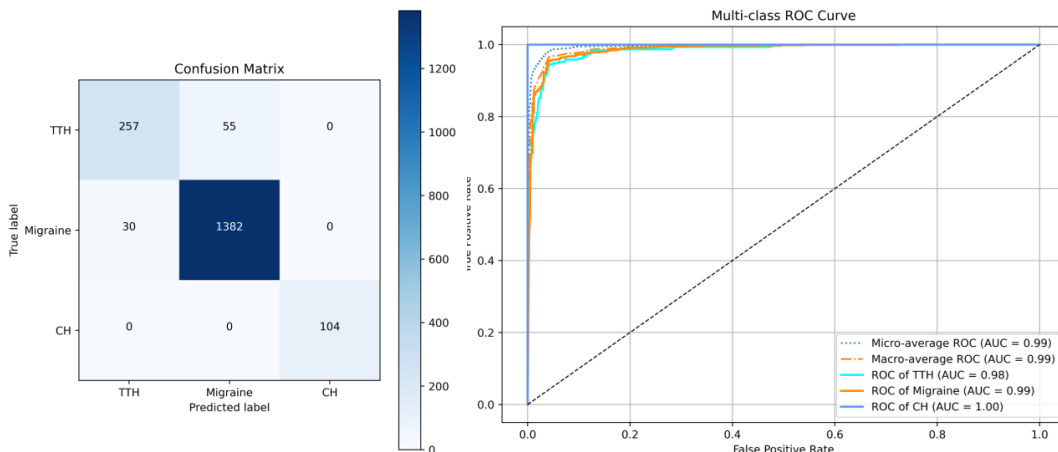


Fig. 5 Confusion matrix and ROC curve of Fuzzy-KAN model

图5 Fuzzy-KAN模型混淆矩阵与ROC曲线

指标如表 2 所示。其中,图 5(a)为模型混淆矩阵,图 5(b) 8:2 的独立测试集中,模型准确率为 0.962 6,Kappa 值为 0.898 4,模型预测结果与专家诊断具有较高的一致性。

Table 2 Overall performance of the model(95% CI)
表 2 模型总体性能(95% CI)

Class	Accuracy	Sensitivity	Specificity	PPV	NPV	Kappa	Youden
TTH	0.953 5 (0.943 1–0.962 2)	0.823 7 (0.782 6–0.863 3)	0.980 2 (0.972 5–0.986 8)	0.895 5 (0.858 1–0.929 3)	0.964 3 (0.954 8–0.973 1)	0.830 3 (0.799 3–0.864 4)	0.803 9 (0.763 9–0.845 6)
Migraine	0.953 6 (0.944 7–0.962 8)	0.978 8 (0.970 4–0.985 7)	0.867 8 (0.836 5–0.899 5)	0.961 7 (0.951 8–0.971 1)	0.923 3 (0.895 2–0.948 5)	0.864 9 (0.837 7–0.892 9)	0.846 5 (0.815 4–0.879 9)
CH	1.000 0 (1.000 0–1.000 0)	1.000 0 (1.000 0–1.000 0)	1.000 0 (1.000 0–1.000 0)	1.000 0 (1.000 0–1.000 0)	1.000 0 (1.000 0–1.000 0)	1.000 0 (1.000 0–1.000 0)	1.000 0 (1.000 0–1.000 0)
Overall	0.962 6 (0.944 7–0.962 8)	0.982 5 (0.970 4–0.985 7)	0.949 3 (0.909 0–0.989 7)	0.952 4 (0.951 8–0.971 1)	0.962 5 (0.926 9–0.998 1)	0.898 4 (0.845 8–0.898 8)	0.883 5 (0.842 5–0.924 5)

图 6、图 7 分别为分类性能柱状图和雷达图。由图 6、图 7 可见,从集性头痛在测试集中均被正确识别,但考虑到该类样本量较少,结果仍需进一步验证;偏头痛的敏感性较高(97.88%),特异性相对较低(86.78%),证明模型对偏头痛的识别能力较强,但部分样本会被误分为紧张型头痛;紧张型头痛的特异性较高(98.02%),但敏感性较低(82.37%),表明该类头痛存在漏诊。混淆矩阵表明,紧张型头痛与偏头痛之间存在双向误判,可能与两类样本在头痛强度和持续时间上的重叠程度相关。

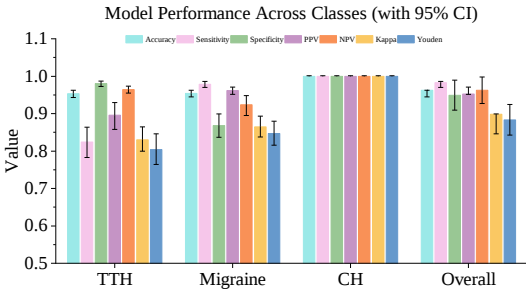


Fig. 6 Classification performance of Fuzzy-KAN model(95% CI)
图 6 Fuzzy-KAN 模型分类性能(95% CI)

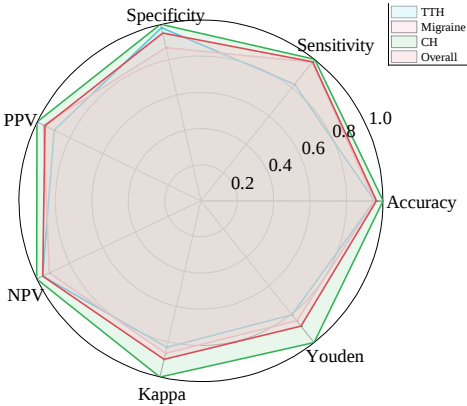


Fig. 7 Visualization of classification performance of Fuzzy-KAN model

4.2 消融与比较实验

将 MLP、KAN、Fuzzy-MLP、RandFuzzy-KAN、Fuzzy-KAN 模型进行消融与比较实验,实验结果如表 3 所示。其

中,实验采用分层十折交叉验证,既减小了单次划分带来的随机性,又解决了样本类别不平衡,可使各折中的类别比例保持相对稳定。

由表 3 可知,Fuzzy-KAN 的 Accuracy、Sensitivity、Specificity 均优于其他模型,Youden 指数、Kappa 值也处于较高水平,说明该模型不仅分类效果较好,还与专家诊断的一致性较高。MLP 各项指标相对较低,凸显了基础神经网络在处理复杂头痛分类任务时的局限性。KAN 与 Fuzzy-MLP 虽在性能上相较于 MLP 有一定提升,但在关键指标上仍不及 Fuzzy-KAN 模型。随机初始化的 Fuzzy-KAN 模型部分指标略低,表明合理的初始化对模型性能具有积极作用。

Table 3 Results of ablation experiment
表 3 消融实验结果

Models	Accuracy	Sensitivity	Specificity	PPV	NPV	Youden	Kappa
MLP	0.944 2	0.930 1	0.945 6	0.935 6	0.949 6	0.875 7	0.848 5
KAN	0.955 1	0.934 0	0.949 4	0.956 3	0.965 5	0.883 4	0.875 8
Fuzzy-MLP	0.955 1	0.938 2	0.952 3	0.952 7	0.962 8	0.890 5	0.876 9
Rand-Fuzzy-KAN	0.956 8	0.928 9	0.946 0	0.965 5	0.972 1	0.874 9	0.878 4
Fuzzy-KAN	0.958 4	0.951 3	0.961 8	0.949 9	0.960 8	0.913 0	0.888 1

图 8 为消融与比较实验的可视化结果。

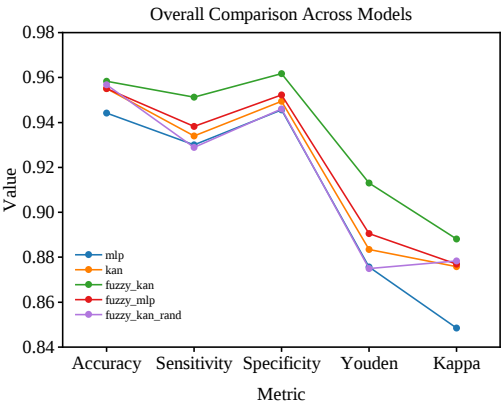


Fig. 8 Visualization results of ablation and comparative experiments
图 8 消融与比较实验可视化结果

由图 8 可知, Fuzzy-KAN 模型在 Accuracy、Specificity 等指标上位居前列, 尤其是 Specificity 指标的优势明显, 这对减少误诊(将其他头痛误判为丛集性头痛)具有重要意义, 证明了将模糊逻辑与 KAN 结合, 可同时提高分类性能和模型解释性, 为头痛辅助诊断提供了新方法。

4.3 可解释性分析

图 9 为模型预测部分关键特征的模糊隶属度分布情

况, 以观察不同头痛类型在关键特征上的差异。由此可知, 不同头痛类型在关键特征的隶属度分布上存在明显差异: 在偏头痛组中, 恶心隶属度高于 0.8 的样本占比较高, 符合偏头痛常伴随恶心的临床特征; 在丛集性头痛组中, 结膜充血在 0.7 以上区间出现明显集中, 证明模型对该类自主神经症状较敏感; 紧张型头痛组的头痛强度隶属度分布较平缓, 符合其症状变化相对稳定的特点。

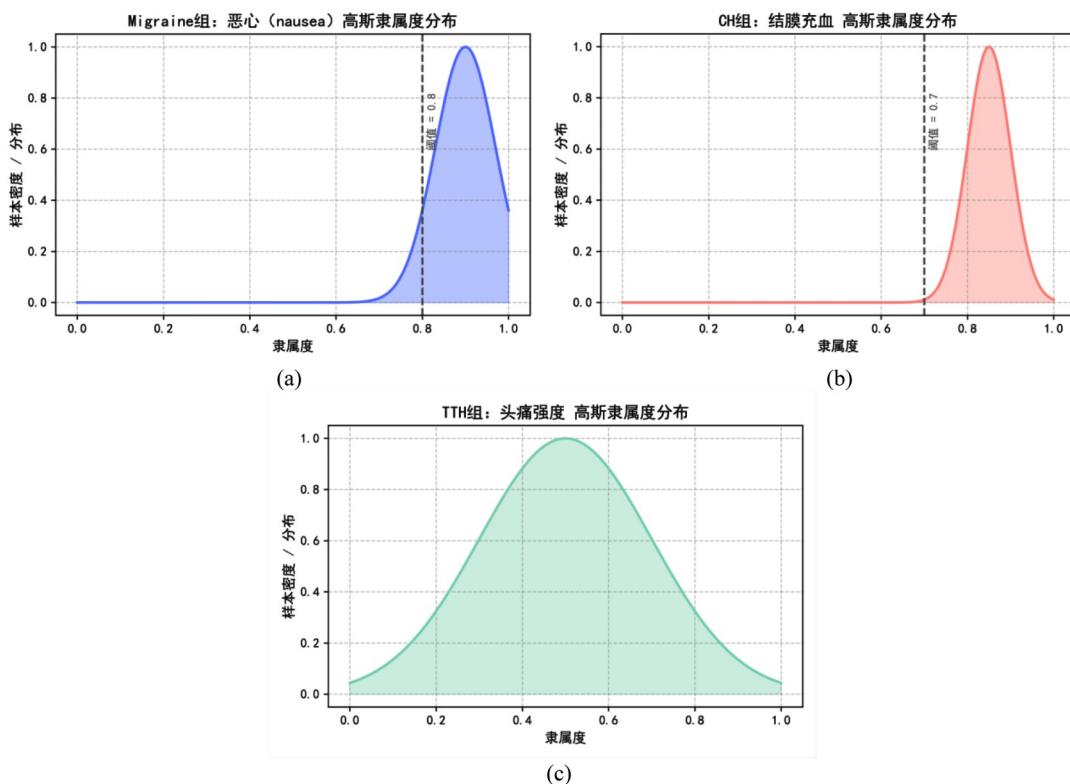


Fig. 9 Fuzzy membership degree distribution of key features predicted by the model

图 9 模型预测部分关键特征的模糊隶属度分布情况

由于 KAN 的每一层均通过 B 样条函数对输入特征进行非线性变换, 因此可分析各特征对输出的响应过程, 如图 10、图 11 所示。

由图 10、图 11 可见, 模型在第一层即对头痛程度、恶心、呕吐、畏寒和日常活动加重等特征产生较强响应, 能较早捕捉关键诊断信息, 且与临床经验基本一致。

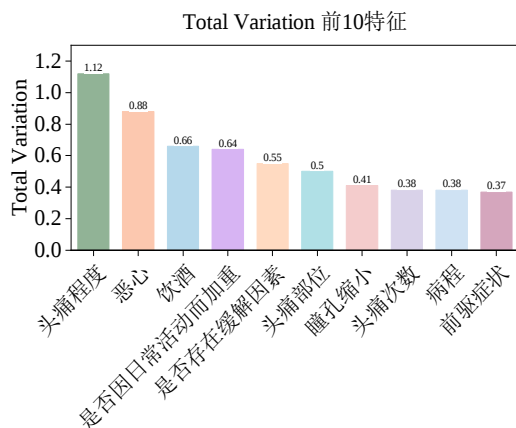


Fig. 10 Top 10 features with the largest response degrees in the model's KAN network

图 10 模型 KAN 网络中响应度前 10 个最大的特征

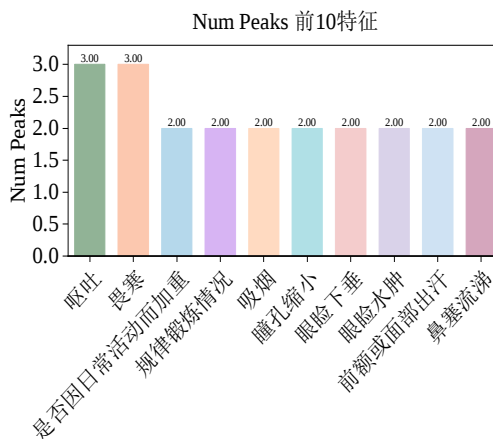


Fig. 11 Top 10 features with peak response degrees in the model's KAN network

图 11 模型 KAN 网络中响应度波峰前 10 的特征

模型训练完成后,采用SHAP对最终模型进行解释分析,如图12、图13所示。由此可见,排名前5的重要特征依次为头痛持续时间、恶心、呕吐、头痛程度和畏光,该排序与临床经验和ICHD-3诊断要点一致,说明模型在分类时重点使用了关键症状特征,具有较好的临床可解释性。

5 讨论

Fuzzy-KAN模型在真实门诊数据上取得了较好的分类效果,模型解释结果与ICHD-3关键诊断规则基本一致,说明该方法在原发性头痛辅助分型方面具有应用潜力。从分类结果而言,不同头痛类型的识别效果存在差异:丛集性头痛识别效果最好,在流泪、鼻塞、结膜充血等自主神经症状方面具有较强的区分度,但该类样本量较少,结论仍需通过外部队列和前瞻性研究验证;偏头痛表现为高敏感性、较低特异性,证明模型对偏头痛识别较强,但存在误分现象;紧张型头痛表现为高特异性、较低敏感性,表明该

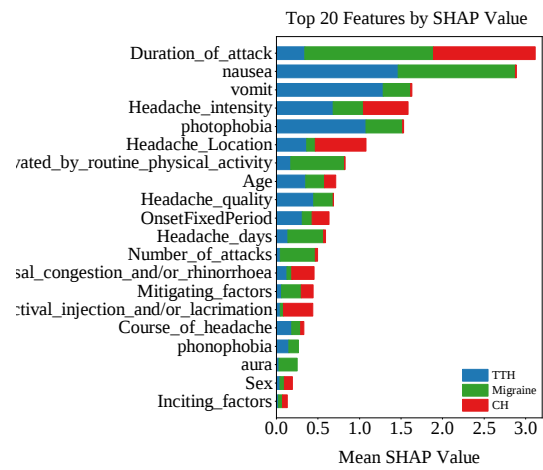


Fig. 12 The distribution of key factors for SHAP in different headaches predicted by the model

图 12 模型预测不同头痛 SHAP 关键因素的分布情况

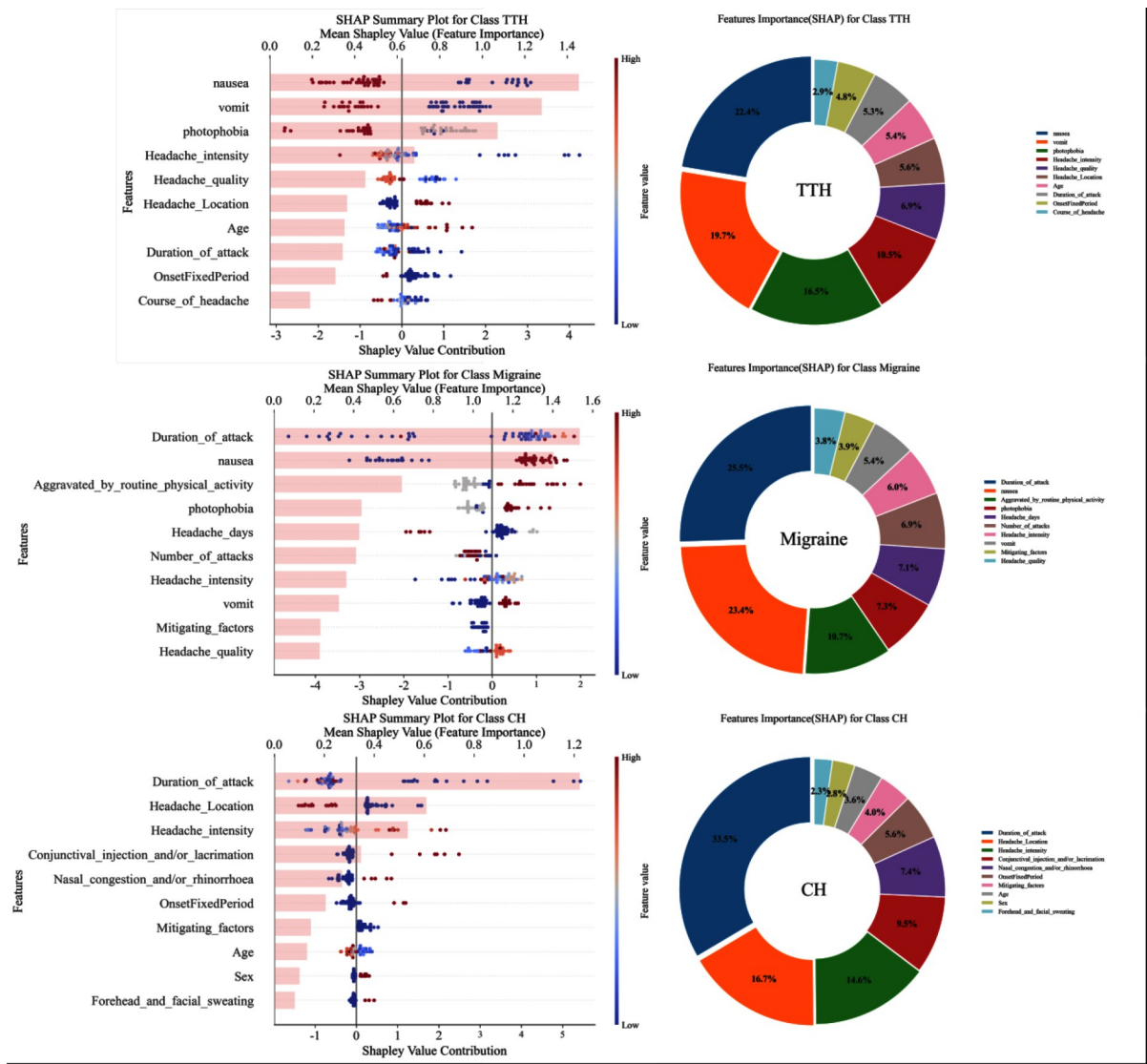


Fig. 13 Detailed distribution of SHAP key factors predicted by models for various types of headaches

图 13 模型预测各类头痛 SHAP 关键因素的详细分布情况

类头痛存在漏检风险。

混淆矩阵显示,紧张型头痛与偏头痛之间存在双向混淆,这与两类疾病在疼痛强度、持续时间及部分伴随症状上的重叠程度有关。若用于门诊初筛,应重点关注紧张型头痛漏检和偏头痛误报问题。模型在单次分层8:2的测试集中的准确率为0.962 6,在分层十折交叉验证中的平均准确率为0.958 4,数值存在小幅差异的原因是统计口径不同,属于正常现象。

此外,研究从3个方面验证了模型的可解释性:①模糊隶属度分布与各类头痛的典型临床特征基本一致,即偏头痛组中恶心等伴随症状的隶属度较高,丛集性头痛组中自主神经症状分布更集中,紧张型头痛组中头痛强度分布较平缓;②KAN响应度分析表明,模型对头痛程度、恶心/呕吐、畏声/畏光等特征响应较快,说明模型内部表征与临床诊断要点一致;③SHAP排名前列的特征主要为头痛持续时间、恶心、呕吐、头痛程度和畏光,与ICHD-3诊断要点一致。综上,模型不仅给出了头痛分类结果,还能提供解释依据,便于临床使用和医患沟通。

6 结语

本文提出一种融合分层模糊逻辑与KAN的原发性头痛可解释辅助诊断模型。首先,将ICHD-3中的描述性规则转化为可计算的隶属度函数和规则权重;其次,将隶属度函数、规则权重与数据驱动特征学习相结合。

消融与比较实验表明,MLP对复杂症状组合和边界模糊病例的表征能力有限;KAN能更好地建模非线性关系,但缺少显式临床先验约束;加入模糊层后,Fuzzy-KAN模型可更稳定地表示症状的模糊语义,也更容易将ICHD-3规则转成可计算的先验信息;基于ICHD-3差异化初始化的Fuzzy-KAN优于随机初始化的RandFuzzy-KAN,说明合理的临床先验初始化不仅能提高解释性,还能提升模型性能。

然而,研究仍存在一些局限性:①本文模型仅能用于辅助决策,无法替代临床医生独立诊断,输出结果仍需结合医生进行决断;②研究尚未开展系统地亚组分析,模型在不同人群中的公平性仍需验证;③真实门诊问卷数据存在回忆偏差、填写错误和症状表述差异等噪声,模型对噪声的鲁棒性仍需专人评估;④当前任务仅覆盖紧张型头痛、偏头痛和丛集性头痛等3类常见原发性头痛,后续需纳入更多原发性头痛亚型及混合表型;⑤研究评估主要基于内部分割数据集和交叉验证,缺乏严格的外部验证、前瞻性评估及真实临床工作流下的人机协同效果评估。后续,将在多中心开展临床试用,以逐步解决上述问题。

参考文献:

[1] ROBBINS M S. Diagnosis and management of headache: a review[J]. JAMA, 2021, 325(18): 1874-1885.

[2] International Headache Society. The international classification of headache disorders, 3rd edition (ICHD-3)[J]. Cephalalgia, 2018, 38(1): 1-211.

[3] GBD 2021 Nervous System Disorders Collaborators. Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: a systematic analysis for the global burden of disease study 2021[J]. Lancet Neurol, 2024, 23(4): 344-381.

[4] STEINER T J, HUSØY A, STOVNER L J, et al. GBD2021: headache disorders and global lost health - a focus on children, and a view forward[J]. The Journal of Headache and Pain, 2024, 25(1): 91.

[5] KANIECKI R G. Migraine and Tension-type headache: an assessment of challenges in diagnosis[J]. Neurology, 2002, 58(6): 15-20.

[6] BUTURE A, BOLAND J W, DI PIETRO F, et al. Systematic literature review on the delays in the diagnosis of cluster headache[J]. Journal of the Neurological Sciences, 2019, 40(1): 25-39.

[7] International Headache Society. 1.1 migraine without aura [EB/OL]. <https://ichd-3.org/1-migraine/1-1-migraine-without-aura/>.

[8] International Headache Society. 2.1 Infrequent episodic tension-type headache [EB/OL]. <https://ichd-3.org/2-tension-type-headache/2-1-infrequent-episodic-tension-type-headache/>.

[9] International Headache Society. 2. Tension-type headache [EB/OL]. <https://ichd-3.org/2-tension-type-headache/>.

[10] STUBBERUD A, LANGSETH H, NACHEV P, et al. Artificial intelligence and headache [J]. Cephalalgia, 2024, 44(8): 3331024241268290.

[11] KATSUKI M, SHIMAZU T, KIKUI S, et al. Developing an artificial intelligence-based headache diagnostic model and its utility for non-specialists' diagnostic accuracy[J]. Cephalalgia, 2023, 43(5): 3331024231156925.

[12] KHAN L, SHAHREEN M, QAZI A, et al. Migraine headache (MH) classification using machine learning methods with data augmentation [J]. Scientific Reports, 2024, 14: 5180.

[13] KWON J, LEE H, CHO S, et al. Machine learning-based automated classification of headache disorders using patient-reported questionnaires [J]. Scientific Reports, 2020, 10: 14062.

[14] LITJENS G, KOOI T, BEJNOORDI B E, et al. A survey on deep learning in medical image analysis [J]. Medical Image Analysis, 2017, 42: 60-88.

[15] ORHANBULUCU F, LATIFOĞLU F, BAYDEMİR R. A new hybrid approach based on time frequency images and deep learning methods for diagnosis of migraine disease and investigation of stimulus effect [J]. Diagnostics, 2023, 13(11): 1887.

[16] TJOA E, GUAN C. A survey on explainable artificial intelligence (XAI): toward medical XAI [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(11): 4793-4813.

[17] AMANN J, BLASIMME A, VAYENA E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective [J]. BMC Medical Informatics and Decision Making, 2020, 20: 310.

[18] LIU S, MCCOY A B, PETERSON J F, et al. Leveraging explainable artificial intelligence to optimize clinical decision support [J]. Journal of the American Medical Informatics Association, 2024, 31(4): 968-974.

[19] RUDIN C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead [J]. Nature Machine Intelligence, 2019, 1(5): 206-215.

[20] SADEGHI Z, ALIZADEHSANI R, CIFCI M A, et al. A review of ex-

- plainable artificial intelligence in healthcare[J]. *Computers and Electrical Engineering*, 2024, 118: 109370.
- [21] ZADEH L A. Fuzzy sets[J]. *Information and Control*, 1965, 8(3): 338–353.
- [22] BELIAKOV G, WARREN J. Fuzzy logic for decision support in chronic care[J]. *Artificial Intelligence in Medicine*, 2001, 21(1–3): 209–213.
- [23] YIN Z, LI H, HAN X, et al. Clinical decision support system using hierarchical fuzzy diagnosis model for migraine and tension-type headache based on International classification of headache disorders, 3rd edition[J]. *Frontiers in Neurology*, 2024, 15: 1444197.
- [24] RAZAK T R, GARIBALDI J M, WAGNER C, et al. Interpretability indices for hierarchical fuzzy systems[C]// Naples: 2017 IEEE International Conference on Fuzzy Systems, 2017.
- [25] LIU Z, WANG Y, VAIDYA S, et al. KAN: kolmogorov-arnold networks[DB/OL]. <https://arxiv.org/abs/2404.19756>.
- [26] ZHANG B, HUANG H, SHEN Y, et al. MM-UKAN++: a novel kolmogorov-arnold network-based U-shaped network for ultrasound image segmentation[J]. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2025, 72(4):498–514.
- [27] LI C, LIU X, LI W, et al. U-KAN makes strong backbone for medical image segmentation and generation[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(5):4652–4660.
- [28] GAO J, LIAO Y, ZHANG Z, et al. MSFKAN: a multi-scale feature prediction network combined with KAN for medical image classification[J]. *Neural Processing Letters*, 2025, 57:67.
- [29] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[C]// Long Beach: 31st International Conference on Neural Information Processing Systems, 2017.
- [30] JI S L, LI J F, DU T Y, et al. A survey on interpretability methods, applications and security for machine learning models[J]. *Journal of Computer Research and Development*, 2019, 56(10): 2071–2096.
- 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): 2071–2096.
- [31] CHEN K R, MENG X F. Interpretability of machine learning[J]. *Journal of Computer Research and Development*, 2020, 57(9): 1971–1986.
- 陈珂锐,孟小峰. 机器学习的可解释性[J]. *计算机研究与发展*, 2020, 57(9): 1971–1986.
- [32] XIE J J, LIU C P. Fuzzy mathematics methods and their applications (3rd ed)[M]. Wuhan: Huazhong University of Science and Technology Press, 2006.
- 谢季坚,刘承平. 模糊数学方法及其应用(第3版)[M]. 武汉:华中科技大学出版社,2006.
- [33] Chinese Society of Neurology, Headache Collaborative Group of Chinese Society of Neurology. Chinese guideline for diagnosis and treatment of migraine (the first edition of Chinese Society of Neurology)[J]. *Chinese Journal of Neurology*, 2023, 56(6): 591–613.
- 中华医学会神经病学分会,中华医学会神经病学分会头痛协作组. 中国偏头痛诊断与治疗指南(中华医学会神经病学分会第一版)[J]. *中华神经科杂志*, 2023, 56(6): 591–613.
- [34] Chinese Society of Neurology, Headache Collaborative Group of Chinese Society of Neurology. Chinese guideline for diagnosis and treatment of tension-type headache (the first edition of Chinese Society of Neurology)[J]. *Chinese Journal of Neurology*, 2023, 56(6): 614–625.
- 中华医学会神经病学分会,中华医学会神经病学分会头痛协作组. 中国紧张型头痛诊断与治疗指南(中华医学会神经病学分会第一版)[J]. *中华神经科杂志*, 2023, 56(6): 614–625.
- [35] Chinese Society of Neurology, Headache Collaborative Group of Chinese Society of Neurology. Chinese guideline for diagnosis and treatment of cluster headache (the first edition of Chinese Society of Neurology)[J]. *Chinese Journal of Neurology*, 2023, 56(6): 626–636.
- 中华医学会神经病学分会,中华医学会神经病学分会头痛协作组. 中国丛集性头痛诊断与治疗指南(中华医学会神经病学分会第一版)[J]. *中华神经科杂志*, 2023, 56(6): 626–636.

(责任编辑:刘嘉文)