[文章编号] 1003－4684(2025)04-0113-08

基于出血性脑卒中的临床智能诊断和治疗模型

江 霞[1]，赵 斌[2]

(1 湖北工业大学校医院，湖北 武汉 430068；2 湖北工业大学理学院，湖北 武汉 430068)

[摘 要] 针对出血性脑卒中这一脑血管破裂引发的严重医疗事件，深入分析急性期患者神经功能损害与高死亡率的影响因素；通过收集和分析出血性脑卒中患者的临床数据，探讨了脑水肿变化、治疗条件与改良兰金量表(MRS)评分之间的相互关系，以提供更加精准的临床建议。为了解决血肿扩张预测、水肿体积变化规律和患者预后轨迹预测这三个关键问题，创建了 HemExPred、EdemaVolReg 和 PredictisPred 三种模型。利用机器学习技术，确定了与血肿扩大事件紧密相关的特征，并通过弹性网络方法验证了这些特征的有效性；同时，应用多项式回归和层次聚类分析揭示了水肿体积变化的复杂动力学过程；最终，借助树外回归模型的先进分析能力，成功预测了患者的预后情况，并确认了血肿体积、水肿体积和年龄对患者预后的关键影响。

[关键词] 出血性脑卒中；树外回归模型；弹性网络；多项式回归；血肿体积

[中图分类号] TP391.4 [文献标识码] A

出血性脑卒中是脑血管疾病的一种类型，其复杂性和严重性是医学上面临的一大难题[1]。这种特殊的中风表现通常以快速发作为特征，会导致治疗延误的后果，严重危及患者的生命[2]。出血性中风的治疗手段一直在解决许多关键问题，包括血肿扩张、水肿和预后评估等[3]。

在出血性中风发生后的 48 h 内，患者可能会遇到令人痛苦的血肿扩张和水肿量波动[4]，这给临床治疗带来了巨大的挑战。其中，血肿扩张和水肿是影响这种疾病发展的关键因素，对患者的预后产生了影响。血肿扩张可能会加剧神经损伤，而水肿通过加大颅内压，使脆弱的脑组织进一步受损。因此，对血肿扩张和水肿进行有效的监测和管理是当务之急，以便改善出血性脑卒中患者的预后。为此，医生和医疗保健人员采用改良兰金量表评分，全面分析患者的神经损伤和自我护理能力[5]。该评分具有十分重要的意义，是衡量患者整体预后的关键指标。

近年来，医学成像和大数据分析相融合，引起了中风研究和治疗干预方面的变革。其中，Xu 等[6]利用深度学习的方法，对计算机断层扫描结果进行识别，从而细化诊断工作。同样，Mattila 等[7]的研究揭示了一种用于急性出血性中风的开创性筛查方法，这种方法提高了出血性中风诊断的准确性，并提供了宝贵的治疗建议。通过融合医学成像技术和大数据分析技术，医疗保健人员能够精心评估血肿扩张和水肿的发展倾向[8]。这一发现促进了个性化治疗的形成[9]，并在患者预后方面有了极大的改善。

机器学习是大数据分析领域的一个关键工具[10]，在出血性中风研究和治疗干预方面有着重要的作用[11]，能够从大量的医学临床数据中收集出宝贵的信息[12]，揭示肉眼无法发现的模式和相关性。通过研究一系列关键变量，包括出血性中风患者的临床表现、影像学特征[13]、治疗模式[14]和预测数据，机器学习使医疗人员能够更好地理解这种疾病的本质，为他们在治疗时提供有力的依据。医学成像技术和大数据分析技术相结合，提高了对出血性中风患者预后的处理能力，以及准确的诊断能力，并能提供个性化的治疗方案[15]，进而有效减轻血肿扩张和水肿的影响，提高患者的生活质量，使患者早日康复。

机器学习和大数据分析[16]的结合为医疗保健人员提供了准确性、精确性和针对性的治疗建议，帮助发现这种疾病复杂的细微差别，为出血性中风患者提供更完善的治疗方案。

[收稿日期] 2023－10－27

[第一作者] 江 霞(1976－)，女，湖北郧阳人，湖北工业大学校医院医生，研究方向为脑卒中的临床诊断、中医药学。

1 诊断和预测模型

1.1 HemExPred 模型

血肿扩展事件发生的预测(HemExPred)模型能够识别和预测血肿扩展事件发生的可能性。HemExPred 结合了一系列患者的因素,包括个人和疾病史、发病时间、相关治疗和影像学等,巧妙地计算了 48 h 内出现血肿扩展事件(HEE)的概率。在本研究中,运用主成分分析、随机森林和套索回归等特征提取方法,提取了 10 个最关键的特征,并将它们联合和相交,形成一个庞大的特征集。随后,将这些识别出的特征作为训练支持向量机、随机森林、逻辑回归模型的基础,在指定的 48 h 内预测血肿扩展发生的概率。为每个患者分配 n 个随访,$n \in [1,+\infty]$。为了确定每个患者从随访检查到发病的时间间隔 t ,使用:

$$t = t_m + t_i - t_0, i \in [1,n] \tag{1}$$

其中,t_m 代表第一次就诊时发病与检查之间的时间间隔。t_i 对应于该患者第 i 次随访时的检查时间点,t_0 表示该患者第一次随访时的检查时间点。

患者发病 48 h 内血肿体积是否发生变化的事件发生必须满足以下两个条件中的一个:1)绝对体积增加大于或等于 6 毫升;2)相对体积增加大于或等于 33%,采用:

$$F(d_0,d_i)_{HEE} = \begin{cases} 1 & (d_i - d_0) \geqslant \dfrac{6\ or\ (d_i - d_0)}{d_0} \geqslant 0.33 \\ 0 & else \end{cases} \tag{2}$$

其中,d_0 表示第一次检查时的血块体积,d_i 对应于第 i 次随访检查时的血块体积。

HemExPred 模型的流程(图 1)可以作为上述步骤的可视化表示。

随机森林算法[17],是一种基于机器学习原理的分类技术。它通过构建一组决策树并合并它们的预测来进行分类任务。首先,输入包含 N 个样本的数据集,每个样本上有 M 个特征,并附有各自的分类标签;其次,为了构造每棵决策树,需要从原始数据集中随机抽取 N 个样本通过特别的采样过程。在每个子样本集中可能会遇到样本重复的情况,而特定的样本可能不会出现在训练集中;最后,在构建决策树的过程中,从整个特征组中只随机选取一个特征子集,保证了每个决策树中的特征选择过程都具有一定的随机性。决策树分支的递归构造将继续进行,直到停止条件出现,例如节点内的样本数量低于特定阈值或满足特定条件的节点纯度。

在决策树的节点划分过程中,基尼指数假定其

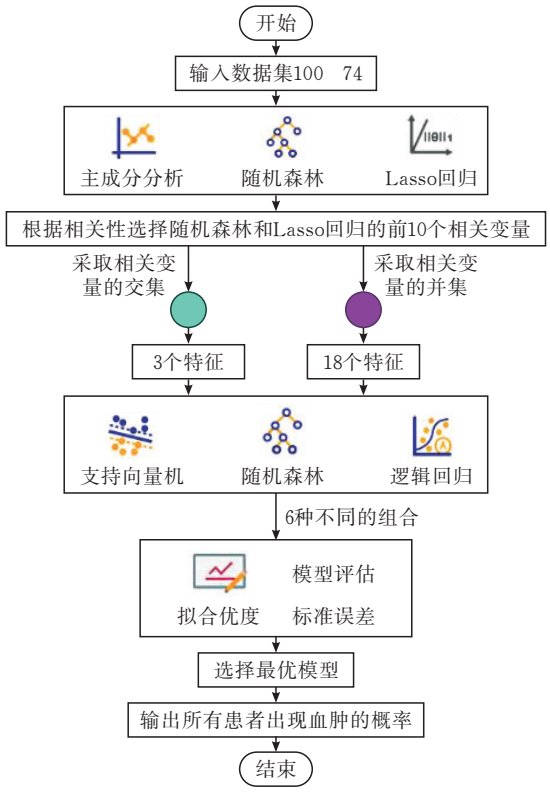

图 1 HemExPred 模型流程

正确的位置作为特征选择标准。对于回归问题,每棵决策树依次遍历其分支,根据特征选择规则进行判断,直到到达叶节点,得到最终的预测结果。在回归领域,最终的回归结果是通过平均或加权平均算得,这取决于模型中每个决策树的性能。

随机森林算法是预测建模中的一种实体,它通过构造和合成多棵决策树来体现它的优势,如参考文献[18]所述。正如图 2 所描述的,随机森林算法过程可以作为一个视觉指南,阐明构建多棵决策树和综合它们的预测所涉及的步骤。

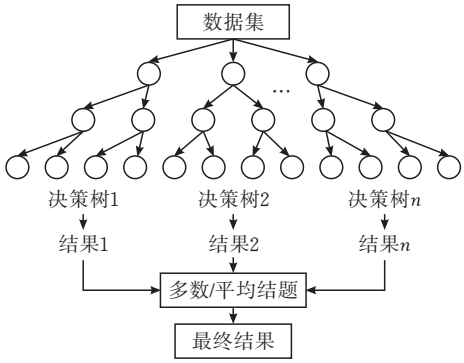

图 2 随机森林算法流程

1.2 EdemaVolReg 模型

EdemaVolReg 模型能够分析水肿体积随时间的变化[19],利用拟合回归模型构建回归曲线捕捉水肿体积的动态[20]。通过探索血肿体积、水肿体积和治疗之间的关系,发现这些因素之间的相互作用。

首先进行数据清理,确保数据集的完整性和可靠性;其次,构建特征矩阵作为训练线性回归模型和多项式回归模型的基础[21];最后使用线性和多项式回归模型进行实验,通过对实验结果的分析,找出最优的模型和参数,确保回归建模的准确性和有效性。

线性回归能够发现两个变量之间的相互作用,其中一个为因变量,通常用符号 y 表示,而另一个则承担自变量的角色,记为 x 。

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{3}$$

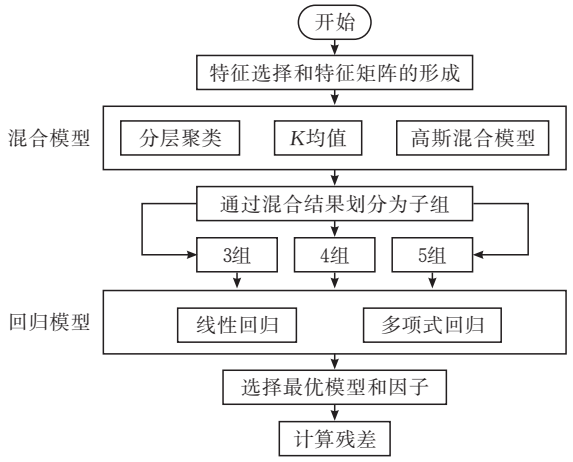

图 3　EdemaVolReg 模型思想流程

2　实验结果

2.1　数据集

该数据集来自第 20 届“华为杯”全国研究生数学建模竞赛 E 题的数据。数据集由 5 张不同的表组成,每个表提供了关于出血性脑卒中患者信息的不同视角:一方面显示了患者的全面概况,涵盖了患者的详尽概况,包括个人信息、病史、接受的治疗方案等;另一方面,准确概括了患者的图像信息,包括血肿和水肿体积及其对应的位置。最后揭示了血肿和水肿患者图像信息中形状和灰度分布的差别,从而为每个图像提供了一个全面的概括。具体来说:

第一张表包含了患者的详细个人信息,例如年龄、性别以及个人和家族的健康历史,这有助于分析患者发病的风险因素;

第二张表记录了患者的病史,可能包括既往疾病、诊断时的临床症状和体征,这对了解患者的基础状况至关重要;

第三张表提供了患者所采用的治疗方案的细节,比如药物治疗、手术治疗等,这对于评估治疗效果和预后具有重要意义;

第四张表准确地总结了患者的图像信息,包括血肿和水肿体积及其在脑部的精确位置,这对于理解病变的严重程度和影响范围非常关键;

第五张表揭示了通过影像技术得到的血肿和水肿的形状及灰度分布差异,从而为医生提供了从影像学角度对病变进行定性和定量分析的依据。

整体上,这些表格提供了多维度的数据便于综合运用统计学、数据挖掘和机器学习等方法,来探索水肿演变规律、预测血肿扩张风险以及评估治疗措施的效果,最终改善出血性脑卒中患者的临床治疗策略和预后评估。

2.2　模型训练结果

首先对 PreOne 数据集进行了特征提取,这是数据分析的关键步骤,旨在从原始数据中识别出对预测任务最有帮助的信息。接下来,进行特征选择的过程,通过这一过程筛选出最重要的特征变量,以减少模型复杂度并提高预测性能。特征选择之后,采用了两种策略来进一步提炼特征:特征交集和特征联合。特征交集方法涉及找出在不同特征集之间共同出现的特征,这通常意味着这些特征既稳定又重要。而特征联合则是将不同特征组合在一起,以捕捉可能被单一特征忽略的复杂模式。通过这种方法,研究收集了 18 个显著特征,提供了更丰富的信息用于构建预测模型。接着,运用了多种机器学习技术(包括弹性网络、支持向量机、随机森林、贝叶斯岭回归等算法,每种算法都有其特定的优势和适用场景)进行模型的构建、验证和预测。

实验结果表明,基于联合特征提取方法时,弹性网络模型展现出了优于其他模型的性能。弹性网络是一种正则化线性回归模型,它能够有效地处理高维数据,并且当数据特征之间存在多重共线性时,仍然能够保持良好的预测性能。而在基于交集中的特征提取方法中,贝叶斯岭模型表现更佳。贝叶斯岭回归是一种概率模型,它通过引入贝叶斯推断为线性回归系数赋予先验分布,从而在数据稀疏或者存在噪声的情况下仍能保持稳定的预测效果[22-24]。

最后,对这些模型进行验证,以确认它们的泛化能力和可靠性。这一系列的分析为出血性脑卒中的临床决策提供了科学依据,有助于医生制定更精确的治疗策略,改善患者的预后。

2.2.1　水肿体积建模　水肿体积的曲线轨迹在整个时间轴被拟合为每个病人的水肿动态变化[25]。研究发现,当采用多项式回归计算时,简单线性模型虽然表现出较高的均方误差,但优于其他模型,如:

$$f(x) = 0.0001x^5 - 0.0001x^4 + 0.0001x^3 - 0.2102x^2 + 112.9559x + 20164.9722 \tag{4}$$

在观察水肿体积进展的时间内,对所有患者进行建模。在各种建模技术中,发现简单线性模型在进行多项式回归计算时,虽然表现出相对较高的均方误差,但模型性能较好(表 1)。聚类结果如图 4

所示。

表 1 使用 PreOne 数据构造 HemExPred 模型记分

模型	R_1^2		R^2		RMSE	
	特征并集	特征交互	特征并集	特征交互	特征并集	特征交互
随机森林分类	−1.90970	−0.45485	−0.30435	−0.30435	0.48305	0.48305
SVR	−1.45173	−0.22602	−0.09905	−0.09919	0.44341	0.44343
贝叶斯岭	−1.20436	−0.09855	0.01184	0.01509	0.42044	0.41975
内核岭	−1.48451	−0.07416	−0.11375	0.03696	0.44636	0.41506
XGB 分类器	−2.32537	−1.07836	−0.49068	−0.86335	0.51640	0.57735
弹性网络	−1.19677	−0.10034	0.01524	0.01349	0.41972	0.42009
AdaBoost 分类器	−4.40373	−0.66269	−1.42236	−0.49068	0.65828	0.51640
树外分类	−2.32537	−0.45485	−0.49068	−0.30435	0.51640	0.48305

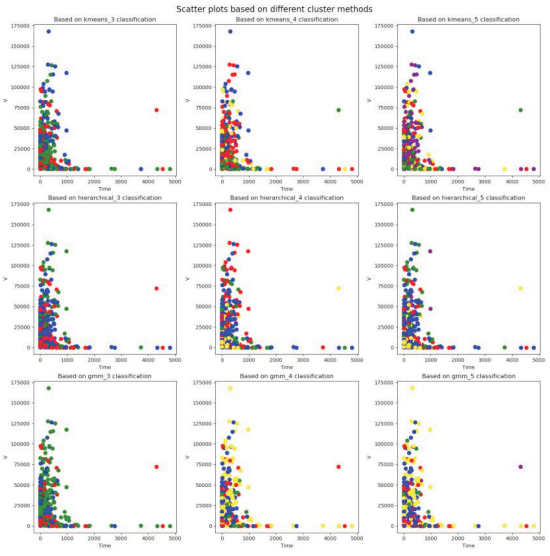

图 4 聚类结果

在这项研究中，通过实现线性回归模型和多项式回归模型来评估聚类分组的有效性[26-27]，进而通过对实验结果的分析，选择了最合适的模型及其对应参数。

2.2.2 迭代实验 为了准确地评价这些回归模型的性能，采用了一系列的迭代实验。在每一轮迭代中，多项式回归模型的参数是变化的，范围从 2 到 10，同时记录两个模型所相应的均方误差和决定系数。从这些实验中得到的发现如表 2 所示。

2.2.3 水肿体积随时间变化而发展的复杂动力学

接下来，研究了病人水肿体积随时间变化而发展的复杂动力学，重点辨别了他们之间的个体差异。通过多种聚类方法揭示出不同人群中浮肿体积轨迹的不同模式，从而促进对这一现象的理解。经过比较和分析，多项式回归作为一种更适合的模型在很多方面都优于线性回归模型。其中，中等度数（如 4 阶或 5 阶）可以更准确地表示数据趋势。然而，关键是要取得平衡，因为过高的程度可能会导致模型过于复杂而降低有效性[30]。聚类方法的选择对模型性能也有着重要影响。层次聚类在不同的环境中取得了较好的结果，特别是与子群 5 的五阶多项式回归相结合时，概括了本研究中的最佳拟合。相反，高斯混合模型在本实验设置下的性能更好[31-33]。

表 2 多簇多项式拟合尺度

聚类方法		回归方法		均方误差	R^2	综合排名
K 均值	3	线性		643339820.4	0.053124323	83
		多项式	8 阶	610101594	0.102044764	53
			9 阶	614438670.8	0.0956614	57
			10 阶	617643624.7	0.090944308	60
	4	线性		630418209.6	0.072142513	73
		多项式	8 阶	599273730.1	0.117981351	42
			9 阶	602786378.1	0.112811391	46
			10 阶	609464708	0.102982141	52
	5	线性		630833745	0.071530922	74
		多项式	8 阶	586499048.7	0.136783289	31
			9 阶	592108089.4	0.128527832	33
			10 阶	601183283.1	0.115170847	44
层次聚类	3	线性		642212930.4	0.054782894	80
		多项式	8 阶	611344587	0.100215311	55
			9 阶	613161462.6	0.097541211	56
			10 阶	614478536.3	0.095602725	58
	4	线性		621664969.5	0.085025642	63
		多项式	8 阶	595083733.5	0.124148241	37
			9 阶	596729084.3	0.121726593	40
			10 阶	601754653.2	0.114329897	45
	5	线性		618931565.6	0.0890487	62
		多项式	8 阶	578589656	0.148424433	24
			9 阶	580376038.6	0.145795212	25
			10 阶	585449211.4	0.138328453	30
高斯混合模型	3	线性		668913900.9	0.015484068	90
		多项式	8 阶	643873288.7	0.052339157	84
			9 阶	646574037	0.048364162	85
			10 阶	650405641.3	0.042724758	87
	4	线性		655347371.3	0.035451458	88
		多项式	8 阶	627641546.3	0.076229241	68
			9 阶	629906741.3	0.072895298	70
			10 阶	631291328	0.070857446	76
	5	线性		649234077.5	0.044449081	86
		多项式	8 阶	625113847.2	0.07994954	65
			9 阶	627809466.2	0.075982094	69
			10 阶	630389420.1	0.072184885	72

表 3　使用 pre1 和 pre2 数据构建一个预后预测模型评分

模型	数据	R_1^2		R^2		RMSE	
		特征并集	特征交互	特征并集	特征交互	特征并集	特征交互
随机森林回归器	Pre One	−0.82903	0.03048	0.11702	0.13078	1.97107	1.95566
	Pre Two	0.89758	0.87831	0.90471	0.88081	0.52916	0.59182
支持向量回归机	Pre One	−1.40779	−0.29457	−0.16238	−0.16065	2.26152	2.25983
	Pre Two	−0.14480	−0.08640	−0.06504	−0.06414	1.76908	1.76833
贝叶斯回归	Pre One	−0.88932	−0.22612	0.08792	−0.09928	2.00329	2.19928
	Pre Two	0.45086	0.27600	0.48912	0.29084	1.22525	1.44356
内核岭	Pre One	−0.57710	−0.06927	0.23864	0.04134	1.83029	2.05380
	Pre Two	0.30232	0.25885	.35093	0.27404	1.38105	1.46056
XGB 回归器	Pre One	−1.11766	−0.15833	−0.02232	−0.03851	2.12090	2.13762
	Pre Two	0.90231	0.89304	0.90911	0.89524	0.51679	0.55484
弹性网络	Pre One	−0.58297	−0.15011	0.23581	−0.03113	1.83370	2.13002
	Pre Two	0.31874	0.27491	0.36620	0.28977	1.36470	1.44465
AdaBoost 回归器	Pre One	−0.87111	−0.08376	0.09671	0.02835	1.99361	2.06767
	Pre Two	0.67697	0.48022	0.69948	0.49087	0.93973	1.22314
树外回归器	Pre One	−0.92804	0.02573	0.06922	0.12652	2.02372	1.96044
	Pre Two	0.94172	0.90365	0.94578	0.90562	0.39916	0.52662

两个实验都建立了初始检查的 MRS 评分、后续随访和患者治疗之间的相关性。这些发现超过了通过配对数据获得的结果。初始检查结果与血肿体积的相关性比 MRS 评分更高(表 3)。对于患有相关疾病的个体来说，需要时刻观察自己的健康状况，并在必要时寻求医疗照顾。

从 ProTwo 数据集提取特征之后，生成了如图 5 所示的热图，其中，“HM_ACA_L_Ratio_1”“年龄”“原始形状延伸率”“原始形状短轴轴长度”和“胆固醇低”等变量在随机森林的特征重要性排名中名列前茅。这些结果与之前基于 Lasso 回归的特征提取结果一致。不仅如此，皮尔逊相关热图还揭示了某些有趣的关联。具体来说，“原始形状最短轴长度 2”这一特征在不同样本或测量条件下表现出显著的正相关关系，这表明该特征在描述出血性脑卒中病变形态时具有一定的稳定性和一致性；而另一方面，“NCCT 原始一阶最小值 2”特征在不同情境下则呈现出负相关关系，这反映了该特征在反映病变内部结构或密度变化时的反向变化趋势。“原始形状主轴长度 1”“原始形状最大直径切片 1”和“原始形状最大直径行数 1”存在正相关关系，这些发现强调了变量的相互作用。

患者发病后 90 天的预后 MRS 评分与疾病和治疗的各个方面之间存在相互作用。通过分析数据，找出提升患者预后最重要的因素，以便进行更有效的临床诊断和治疗。通过计算，得出特征并集的重要指数，证明上述模型的严密性和准确性(图 6)。

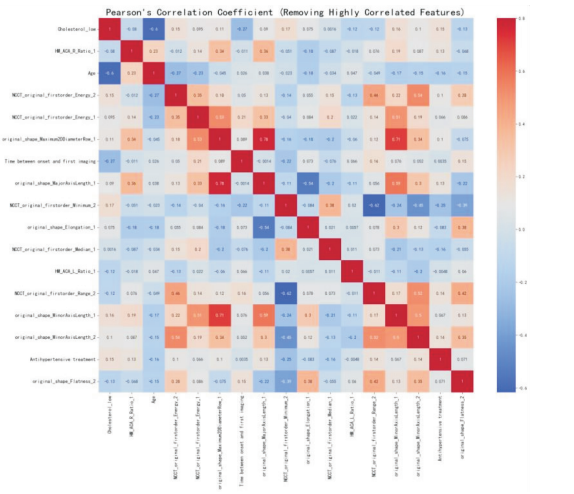

图 5　特征提取的 ProTwo 数据热图和相关性计算

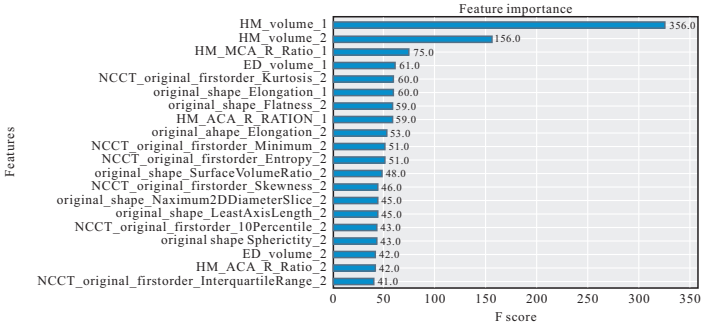

图 6　ProTwo 数据特征重要性指数排行图

对患者发病后90天预后MRS评分与大量疾病相关信息、治疗相关细节和多种影像学检查结果之间进行了深入探讨，取得了满意的结果。其中，“HM_volume_1”“HM_volume_2”“HM_MCA_R_Ratio_1”“ED_volume_1”和“HM_ACA_R_Ratio”就像前面提到的分数一样与MRS高度相关。这一发现验证了影响因素的准确性。因此，在对出血性中风患者进行影像学评估时，必须将注意力转向“HM_volume”“ ED_volume”“ HM_MCA_R_Ratio”和“HM_ACA_R_Ratio”等关键指标。其中，“HM_volume”的重要性超过了其他特征，强调了其与患者发病90天MRS评分的重要关系。基于皮尔逊相关系数，“HM_volume”和MRS得分之间存在正相关关系。简单来说，“HM_volume”值越大，MRS评分越高，表明血肿体积越大，预后越差。

虽然“HM_MCA_R_Ratio”和“HM_ACA_R_Ratio”影响出血性中风的发展、严重程度和预后，但其具体的病因和机制需要进一步研究。在临床诊断中，重要的是监测血肿体积的变化，并提供治疗方式或药物以控制血肿恶化。

3 探讨与结论

该研究对两个出血性脑卒中数据集，即PreOne和PreTwo，进行了细致的分析。在数据预处理和特征工程阶段，采用了多种方法来优化模型的训练过程。具体来说，通过特征交集和特征联合这两种策略，分别提取了重要且稳定的特征以及能够捕捉复杂模式的联合特征。

在模型选择方面，树外回归模型（如随机森林、梯度提升树及其变体等）在PreTwo数据集上的表现尤为突出，这可能归因于这类模型处理缺失值和离群值时的高度稳健性。这种稳健性使得树外回归模型在多个评估指标上均优于其他模型，例如在预测精度、召回率、精确度等方面都取得了较好的分数。

在出血性脑卒中的研究中，选择的特征包括患者的人口统计信息、病史、临床测量指标以及影像学特征等。这些特征的医学意义在于它们可以帮助医生了解患者的风险状况，制定个性化的治疗计划，并预测治疗的可能结果。例如，一个具有高风险病史特征和较大血肿体积的患者可能需要更紧急的干预和更密切的监测。通过这些特征的综合分析，医生能够更好地管理患者的治疗进程，从而改善患者的预后。

3.1 优势

1）模型选择的多样性：通过采用多种机器学习模型，研究可以比较不同模型在处理同一问题上的表现，从而找到最适合该特定数据集的方法。这种比较有助于深入理解每种模型的优势和局限性。

2）特征工程的深度：研究不仅使用了单一特征，还通过特征交集和特征联合等策略进行了深度特征提取，这可能有助于揭示更复杂的数据模式，从而提高预测的准确性。

3）模型稳健性：树外回归模型通常对于数据的缺失值和离群值具有较强的鲁棒性，这使得它们在实际应用中更加可靠。

3.2 不足

1）内核岭模型的挑战：研究表明内核岭模型可能面临欠拟合或过拟合的风险，尤其是在数据量有限的情况下，如支持向量机（SVM）与径向基函数（RBF）内核，通常对数据集的大小较为敏感。

2）数据集规模的限制：虽然研究建议使用更大的数据集进行验证，但实际操作中可能存在数据收集的难度和成本问题。如果数据集较小，模型可能会遇到欠拟合问题，无法捕捉数据的全部复杂性；而如果数据集过大，则可能导致过拟合，模型学习到的决策边界过于复杂，泛化能力下降。为了确保模型结果的有效性和可靠性，需要寻求更大规模的数据集进行验证。

3）解释性：虽然树外回归模型在实践中表现良好，但它们的解释性不如一些参数模型，如线性回归。在医学领域，模型的解释性是一个重要的考虑因素。此外，进一步的研究可以考虑集成学习方法，结合多个模型的优势来提高整体预测性能。

总之，本研究强调了在选择和应用机器学习模型时考虑数据特性和模型假设的重要性，并通过实证分析展示了树外回归模型在处理特定医疗数据集时的潜力。同时，它也提出了对于内核岭模型需要谨慎处理的观点，并推荐了使用更广泛数据进行模型验证的建议。未来的工作将探索出血性脑卒中患者发病机制、诊断和治疗过程的基本特征。通过综合信息，可以对模型进行细化，更好地服务于医疗领域，为医生和患者提供有价值的诊断和治疗见解，最终促进患者康复。

[参 考 文 献]

[1] DORIA J W, FORGACS P B. Incidence, implications, and management of seizures followingischemic and hemorrhagic stroke[J]. Curr Neurol Neurosci Rep, 2019, 19(07):37.

[2] TOYODA K, YOSHIMURA S, NAKAI M, et al. Twenty-year change in severity and outcome of ische-

mic and hemorrhagic strokes[J]. JAMA Neurol, 2022, 79(01):61-69.

[3] CHAUDHARY N, PANDEY AS, WANG X, XI G. Hemorrhagic stroke-Pathomechanisms of injury and therapeutic options[J]. CNS Neurosci Ther, 2019, 25(10):1073-1074.

[4] GU X, LI Y, CHEN S, et al. Association of lipids with ischemic and hemorrhagic stroke: a prospective cohort study among 267500 Chinese[J]. Stroke, 2019, 50(12):3376-3384.

[5] BAMBHROLIYA A B, DONNELLY J P, THOMAS E J, et al. Estimates and temporal trend for US nationwide 30-Day hospital readmission among patients with ischemic and hemorrhagic stroke[J]. JAMA Netw Open, 2018, 1(04):e181190.

[6] XU Y, HOLANDA G, SOUZA LFABRiCIO DE F, et al. Deep learning-enhanced internet of medical things to analyze brain ct scans of hemorrhagic stroke patients: a new approach[J]. IEEE Sensors Journal, 2021, 21(22):24941-24951.

[7] MATTILA OS, ASHTON NJ, BLENNOW K, et al. Ultra-early differential diagnosis of acute cerebral ischemia and hemorrhagic stroke by measuring the prehospital release rate of GFAP[J]. Clinical Chemistry, 2021, 67(10):1361-1372.

[8] LI X, HUANG X, TANG Y, et al. Assessing the pharmacological and therapeutic efficacy of traditional Chinese medicine Liangxue Tongyu prescription for intracerebral hemorrhagic stroke in neurological disease models[J]. Frontiers in Pharmacology, 2018, 9: 1169.

[9] ALEX MATOS RIBEIRO J, FERNANDA GARCÍA-SALAZAR L, Regina Saade-Pacheco C, et al. Prognostic molecular markers for motor recovery in acute hemorrhagic stroke: A systematic review[J]. Clinica Chimica Acta, 2021, 522:45-60.

[10] GU J, SHI Y, CHEN N, WANG H, et al. Ambient fine particulate matter and hospital admissions for ischemic and hemorrhagic strokes and transient ischemic attack in 248 Chinese cities[J]. Science of The Total Environment, 2020, 715:136896.

[11] AHANGAR A A, SAADAT P, HEIDARI B, et al. Sex difference in types and distribution of risk factors in ischemic and hemorrhagic stroke[J]. International Journal of Stroke, 2018, 13(01):83-86.

[12] WELTEN S J G C, ONLAND-MORET N C, BOER J M A, et al. Age at menopause and risk of ischemic and hemorrhagic stroke[J]. Stroke, 2021, 52(08):2583-2591.

[13] ZHANG T, ZHANG W, LIU X, et al. Multifrequency magnetic induction tomography for hemorrhagic stroke detection using an adaptive threshold split bregman algorithm[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71:1-13.

[14] LEE E C, HA T W, LEE D H, et al. Utility of exosomes in ischemic and hemorrhagic stroke diagnosis and treatment[J]. International Journal of Molecular Sciences, 2022, 23(15):8367-8374.

[15] RIBEIRO J A M, GARCIA-SALAZAR L F, SAADE-PACHECO C R, et al. Prognostic molecular markers for motor recovery in acute hemorrhagic stroke: A systematic review[J]. Clinica Chimica Acta, 2021, 522:45-60.

[16] FEIGIN V L, KRISHNAMURTHI R V, PARMAR P, et al. Update on the global burden of ischemic and hemorrhagic stroke in 1990-2013: the GBD 2013 study[J]. Neuroepidemiology, 2015, 45(03):161-176.

[17] SCHONLAU M, ZOU R Y. The random forest algorithm for statistical learning[J]. The Stata Journal, 2020, 20(01):3-29.

[18] WAGER S, ATHEY S. Estimation and inference of heterogeneous treatment effects using random forests[J]. Journal of the American Statistical Association. 2018, 113(523):1228-1242.

[19] REZA M, HAQUE M A. Photometric redshift estimation using ExtraTreesRegressor: Galaxies and quasars from low to very high redshifts[J]. Astrophysics and Space Science, 2020, 365(03):50.

[20] MU Y, LIU X, WANG L. A Pearson's correlation coefficient based decision tree and its parallel implementation[J]. Information Sciences, 2018, 435:40-58.

[21] LIN J, ADJEROH D A, JIANG B H, et al. FastWKendall: an efficient algorithm for weighted Kendall correlation[J]. Computational Statistics, 2018, 33:1823-1845.

[22] 王陇德,王金环,彭斌.《中国脑卒中防治报告 2016》概要[J].中国脑血管病杂志, 2017,14(04):217-224.

[23] 刘莉君,余其贵,杨琳琳.老年衰弱缺血性脑卒中合并亚临床甲减患者的临床研究[J].国际老年医学杂志, 2023, 44(03):289-292.

[24] 吕洋. 80 岁及以上急性缺血性卒中患者重组组织型纤溶酶原激活剂静脉溶栓有效性和安全性分析[D]. 青岛: 青岛大学, 2021.

[25] 尹霄. 缺血性脑卒中患者入院时血清甘油三酯水平与远期死亡风险的相关性研究[D]. 镇江: 江苏大学, 2021.

[26] 杨丹丹,蒋琳芝,李丽燕.预后营养指数对老年急性缺血性脑卒中患者静脉溶栓后短期功能预后的预测价值[J]. 中华老年心脑血管病杂志, 2023, 25(01):5-11.

[27] 林吕, 李松林, 张大丁. 中枢整合疗法治疗出血性脑卒中患者运动障碍预后效果的预测模型[J]. 西部医学, 2023, 35(06):881-884.

[28] 赵婧,王佩,卢丹丹. 出血性脑卒中患者血清 miR-

4443,TRAF4 相对表达量变化的临床意义[J].检验医学, 2023, 38(05): 413-418.
[29] 金意, 鞠忠, 乔大伟. 缺血性与出血性脑卒中危险因素对比分析[J]. 中国公共卫生, 2008, 24(03): 2-9.
[30] 陈璐, 郑思琪, 马晴. 改良关注和解释疗法对脑卒中偏瘫病人康复护理效果的影响[J]. 护理研究, 2023, 36(24): 4388-4393.
[31] 高希春, 叶贵生, 张建波. 基于脑卒中影像规范化检查的科室管理方案的构建与应用[J].医学诊断, 2023, 13(03): 5-12.
[32] 刘利静, 程琳, 王爱杰. 急性缺血性脑卒中动脉取栓术后动态颅脑 CT 图像上高密度影的鉴别[J]. 中国中西医结合影像学杂志, 2023, 21(03): 295-299.

Clinical Intelligent Diagnosis and Treatment Modeling of Hemorrhagic Stroke

JIANG Xia[1], ZHAO Bin[2]
(1 *Hospital, Hubei Univ. of Tech., Wuhan* 430068, *China*;
2 *School of Science, Hubei Univ. of Tech., Wuhan* 430068, *China*)

Abstract: This study aims to analyze in depth the factors influencing neurological dysfunction and high mortality rate in patients with hemorrhagic stroke, a serious medical event caused by cerebral vascular rupture. By collecting and analyzing clinical data of patients with hemorrhagic stroke, we explored the relationship between changes in cerebral edema, treatment conditions, and the modified Rankin Scale (MRS) score to provide more accurate clinical recommendations. In order to address the three key issues of predicting hematoma expansion, edema volume changes, and patient prognosis trajectory, we developed three models: HemExPred, EdemaVolReg, and PredictsPred. Using machine learning techniques, we identified features closely related to hematoma enlargement events and validated their effectiveness through elastic network methods. Meanwhile, polynomial regression and hierarchical clustering analysis were applied to reveal the complex dynamic process of edema volume changes. Finally, with the advanced analytical ability of out of tree regression models, we successfully predicted the prognosis of patients and confirmed the key effects of hematoma volume, edema volume, and age on patient prognosis. The findings of this study provide valuable experience for clinical decision-making in hemorrhagic stroke, helping doctors develop more precise and effective treatment plans, which is expected to significantly improve the survival rate and quality of life of patients.

Keywords: hemorrhagic stroke; extra trees regressor model; elastic net; polynomial regression; hematoma volume

[责任编校: 闫 品]