

# 出血性脑卒中患者水肿体积进展建模研究

徐颖

上海理工大学机械工程学院, 上海

收稿日期: 2024年2月5日; 录用日期: 2024年3月21日; 发布日期: 2024年3月28日

## 摘要

本文利用Gauss非线性拟合, 将患者的水肿体积与每一次检查点的间隔时间联系起来, 得到全部患者的水肿体积与时间进展的关系曲线。同时, 通过FCM模糊聚类算法将所有患者依据个人特征划分为四个亚组, 分别利用Gauss非线性拟合对每个亚组的散点图进行拟合, 再根据拟合曲线分析每个亚组的特征。此外, 利用主成分分析(PCA)探究不同疗法对患者水肿体积进展的影响, 分析表明止血治疗对于水肿体积的影响显著, 通过降颅压治疗也可以有效降低患者的水肿体积。最后探究水肿体积与血肿体积以及治疗方法之间的关系, 首先通过PCA分析得到血肿体积与治疗方法之间的关联度, 再联系水肿体积与治疗方法之间的关系, 得到血肿体积与水肿体积是正相关, 以及止血治疗、降颅压治疗对于降低水肿、血肿体积有着较好的治疗效果的结论。

## 关键词

Gauss非线性, FCM模糊聚类算法, 主成分分析, 出血性脑卒中

# Study on Modeling of Edema Volume Progression in Patients with Hemorrhagic Stroke

Ying Xu

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 5<sup>th</sup>, 2024; accepted: Mar. 21<sup>st</sup>, 2024; published: Mar. 28<sup>th</sup>, 2024

## Abstract

This article uses Gauss nonlinear fitting to connect the patient's edema volume with the interval between each checkpoint, and obtains the relationship curve between the edema volume and time

progression of all patients. At the same time, all patients were divided into four subgroups based on personal characteristics through the FCM fuzzy clustering algorithm. Gauss nonlinear fitting was used to fit the scatter plot of each subgroup, and then characteristics of each subgroup were analyzed based on the fitting curve. In addition, principal component analysis (PCA) was used to explore the impact of different treatments on the progression of edema volume in patients. The analysis showed that hemostatic treatment has a significant impact on edema volume, and intracranial pressure lowering treatment can also effectively reduce the patient's edema volume. Finally, the relationship between edema volume, hematoma volume and treatment methods was explored. First, the correlation between hematoma volume and treatment methods was obtained through PCA analysis, and then the relationship between edema volume and treatment methods was combined to obtain the positive correlation between hematoma volume and edema volume, as well as the conclusion that hemostatic treatment and intracranial pressure lowering treatment have a good therapeutic effect on reducing edema and hematoma volume.

## Keywords

Gaussian Nonlinearity, FCM fuzzy Clustering Algorithm, Principal Component Analysis, Hemorrhagic Stroke

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

脑卒中作为一种严重的神经系统疾病，对患者的生命和生活质量造成了极大的威胁。其中，出血性脑卒中作为脑血管疾病的一种重要类型，由于其急性、突发的特点，常常呈现出高度危险性和复杂性。出血性脑卒中主要包括蛛网膜下腔出血(SAH)和脑实质内出血(ICH)，其病理生理机制涉及血管破裂、血肿和水肿等多个关键环节[1]。

近年来，利用影像学技术，研究人员可以定量地测量脑部出血和水肿的体积，并利用计算机辅助技术进行数据处理和分析。通过整合临床数据、影像学特征和分子生物学标志物等信息，来预测出血性脑卒中患者水肿体积的进展。这些模型可以帮助医生更准确地评估患者的风险，并采取相应的治疗措施。通过对水肿进展的建模研究，可以更好地理解水肿的发生机制，并探索新的治疗策略。同时越来越多的研究关注个体差异对水肿进展的影响，通过分析患者的遗传背景、生活方式、疾病历史等因素，可以更精准地预测水肿的进展，并制定个性化的治疗方案。

本文旨在通过真实临床数据，结合临床和患者影像相关信息，预测出血性脑卒中患者的临床预后和各种治疗方法对该疾病的效果，主要围绕血肿周围的水肿相关信息进行建模，并找到不同治疗措施和水肿发展之间的关系。首先通过数据集中前 100 名患者的水肿体积(ED\_volume)和检查时间构建所有患者水肿体积关于时间的函数。并将前 100 名患者的真实值和拟合函数之间的残差进行记录。其次将患者根据个体差异(年龄、病史等)划分为 3~5 个组，分别构建不同组的患者水肿体积关于时间的函数，并将前 100 名患者的真实值和拟合函数之间的残差进行记录。然后根据每个患者所对应的不同治疗方法，再结合该患者的水肿体积变化，分析每种治疗方法对水肿体积的影响。最后，分析血肿体积、水肿体积及治疗方法三者之间的关系。

## 2. 血肿周围水肿的发生及进展建模

### 2.1. 全体患者水肿体积随时间进展曲线拟合模型

基于数据集中前 100 名患者的水肿体积(ED\_volume)以及每一次检查的时间(即每一次随访), 进行全体患者水肿体积随时间进展曲线拟合模型的建立。其难点在于虽然水肿体积与检查时间有一一对应关系, 但 100 名患者的水肿体积与检查时间是独立的, 只有自相关性。利用散点图将患者首次检查至随访 5 的检查时间与水肿体积的参数存放在二维空间内如图 1 所示。由图可知从日期出发散点图略显杂乱, 不好拟合且难以反映全体患者的水肿体积随时间的变化趋势。为了探究全体患者水肿体积随时间进展曲线, 初步考虑将患者的水肿体积与每一次检查点的间隔时间联系起来, 即横坐标为患者发病至每一次影像检查的时间, 纵坐标为水肿体积[2], 其散点图如图 2 所示。

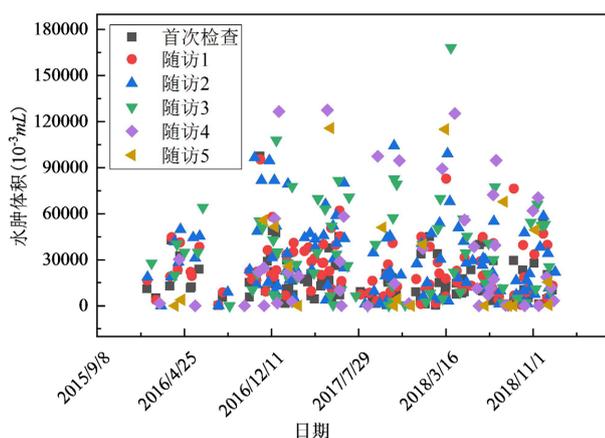


Figure 1. Scatter plot of edema volume and corresponding date for the first 100 patients

图 1. 前 100 名患者水肿体积与对应日期散点图

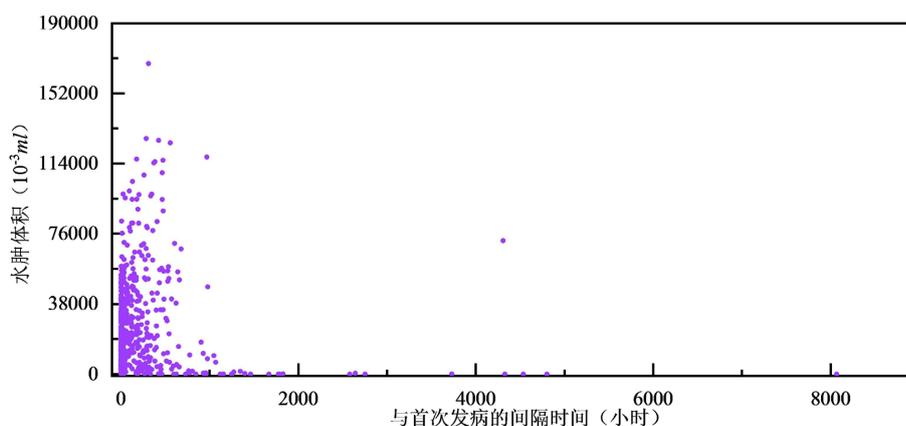


Figure 2. Scatter plot of edema volume versus time between first imaging

图 2. 水肿体积随首次影像间隔时间散点图

由图 2 可知, 这组待拟合数据明显存在异常值, 故首先进行数据清理, 以便拟合, 将图 2 中明显超出界限的散点剔除。将患者水肿体积数据和检查时间等数据汇总到一张表内, 通过做差得到患者每一次检查距离首次发病的时间, 得到水肿体积与时间间隔一一对应的表格。之后可以采用 Gauss 非线性曲线

来拟合水肿体积与首次发病时间间隔点的曲线。初步建立 Gauss 模型如下所示：

$$y = y_0 + \frac{A}{w\sqrt{\pi/2}} e^{-\frac{2(x-x_c)^2}{w^2}} \quad (1)$$

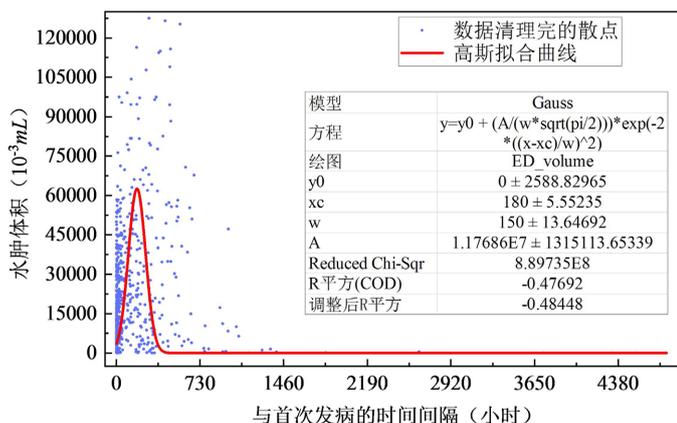
式中  $A$  为振幅； $x_c$  为平均值； $w$  为标准差； $y_0$  为偏移量。

将表格数据代入高斯模型进行拟合，其拟合各项数据如表 1 所示。拟合曲线如图 3 所示，然后将拟合曲线与真实水肿体积做差得到残差，残差与原始数据的比较如图 4 所示。对于残差的标准，因为每位患者有过多次检查结果，故结果文件中的残差计算方法为取患者所有检查的平均值再做差。

**Table 1.** Summary of various error data obtained through Gauss fitting

**表 1.** 通过 Gauss 拟合得到的各种误差数据汇总

	$y_0$		$x_c$		$w$		$A$	
	值	标准误差	值	标准误差	值	标准误差	值	标准误差
水肿体积	0	2588.82965	180	5.55235	150	13.64692	1.17686E7	1315113.65339
	<b>sigma</b>		<b>FWHM</b>		<b>Height</b>		<b>统计</b>	
	值	标准误差	值	标准误差	值	标准误差	Reduced Chi-Sqr	调整后 R 平方
	75	6.82346	176.6115	16.06802	62599.93865	4032.01145	8.89735E8	-0.48448



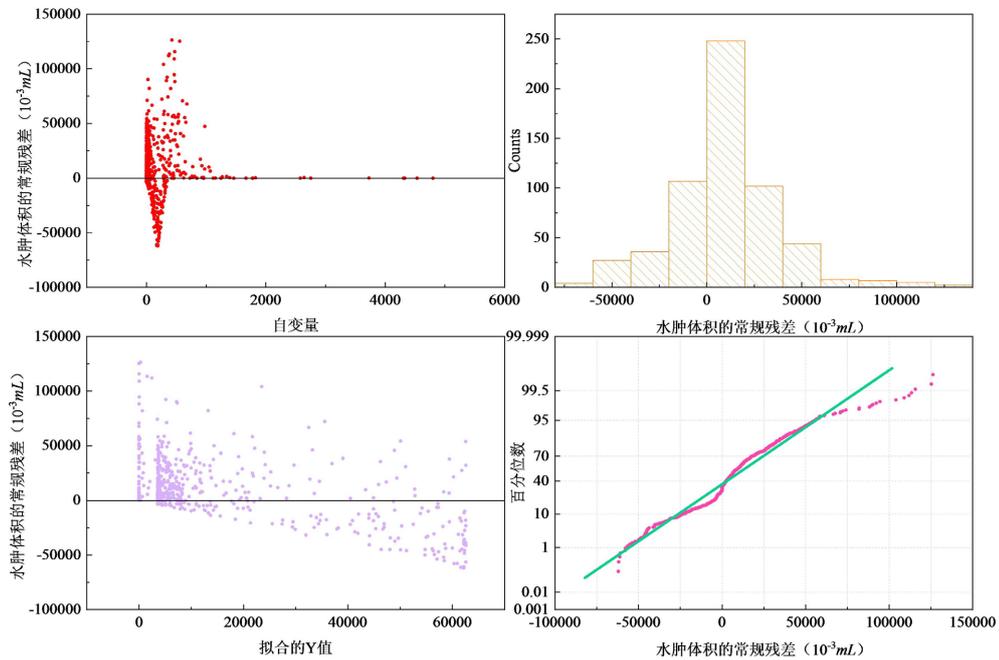
**Figure 3.** Gauss fitting curve

**图 3.** Gauss 拟合曲线

## 2.2. 基于个体差异的患者水肿体积进展曲线模型

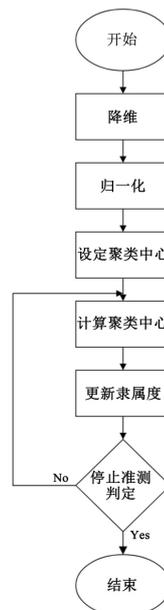
为探索不同临床信息以及医疗特征的患者对于水肿体积与时间进展的关系。在 2.1 节中我们求得了全体患者水肿体积与时间进展的关系，有助于医学的整体研究，但在实际治疗中，需要考虑不同个体的差别，有针对性的进行治疗。故构建基于个体差异的患者水肿体积进展曲线模型是很有必要的。

本文有意将患者分为 3~5 个亚组，初步考虑可以按照患者年龄、性别、疾病史等特征进行分组。按照数据进行聚类分析，可以采用 K-means、K-means++ 或 FCM 聚类等算法先对患者进行亚组的划分，把相似特征的患者划分到一个组里。再按照 2.1 节类似的思路对每一个亚组进行拟合。但是患者的个人差异有 12 个参数，如果单纯以年龄或者个别病史进行分类，难以表达整体数据之间的关联性。故可以考虑个体差异参数的权重，再按照权重进行分类。



**Figure 4.** Residual plot from Gauss fit  
**图 4.** 通过 Gauss 拟合的残差图

FCM 聚类算法是一种基于模糊理论的聚类算法，它在运行过程中不需要人为干预，只需要给定聚类中心的数量，FCM 算法就可以自动进行权重的计算。其是通过迭代计算聚类中心和更新隶属度的步骤来不断优化聚类结果，从而确定聚类中心，使得每个数据点都能够在多个聚类中心之间有一定的隶属度。这样的处理方式使得 FCM 聚类算法在处理高维数据时相较于 HCM 算法(K-means 等)具有较好的效果。故本题选用 FCM 算法对个体特征不同的患者进行分类[3]。



**Figure 5.** Basic process of FCM clustering algorithm  
**图 5.** FCM 聚类算法基本流程

在本文 12 维的数据之中，高血压病史一列与高低血压的实际数值两列实际上存在重复之处，故可以对该组数据进行降维处理，将所有患者的高低血压的实际数值删去，得到 10 维的数据，这样可以有效提高算法的准确率。对于 10 维数据的 FCM 聚类算法其寻找聚类中心的基本原理如下：首先，随机选择  $k$  个聚类中心( $k$  是预先设定的聚类数量)，并为每个数据点分配初始的隶属度(在 0 到 1 之间)。其次根据当前的隶属度，计算每个聚类中心的权重向量。权重向量由所有数据点的 10 维特征向量和对应的隶属度加权平均得到。然后根据当前的聚类中心和权重向量，计算每个数据点对于每个聚类中心的隶属度。隶属度的计算基于每个数据点与每个聚类中心之间的欧氏距离，距离越小，隶属度越高。重复前两步，直到满足迭代次数达到预定值或者聚类中心的变化小于某个阈值。最终根据最终的隶属度，将每个数据点划分到具有最高隶属度的聚类中心。本题采用 FCM 聚类算法基本流程如图 5 所示。

由于输入数据的量纲各有不同，首先需要对输入的 10 维数据进行归一化处理，然后根据 FCM 聚类算法流程，选择 4 个聚类中心，得到聚类中心以及聚类中心选取时对于 10 维输入的权重如表 2 以及图 6 所示。分析表 2 和图 6 可知，FCM 聚类算法将患者进行分类的主要依据是高血压病史，次要依据为年龄。第一亚组中的主体为具有高血压病史且年龄较大脑出血前 mRS 评分(mRS, Modified Rankin Scale 是一种用于评估中风患者功能状况和残疾程度的标准，mRS 评分范围从 0 到 6，评分越高代表患者身体状况越严重)较高的男性患者；第二亚组的主体为具有高血压病史且年龄较小的女性患者；第三亚组的主体为具有高血压病史且年龄较小的患者，且他们中的大多数还具有吸烟史和饮酒史；第四亚组的主体为具有高血压病史且年龄较大的男性患者。

Table 2. Clustering centers of 10-dimensional data for four subgroups

表 2. 四个亚组 10 维数据的聚类中心

组别	年龄	性别	脑出血前 mRS 评分	高血压病史	卒中病史	糖尿病史	房颤史	冠心病史	吸烟史	饮酒史
第一亚组	73.182	1.329	0.226	0.712	0.316	0.107	0.079	0.132	0.129	0.076
第二亚组	56.864	1.218	0.111	0.898	0.197	0.138	0.009	0.028	0.212	0.150
第三亚组	57.160	1.181	0.009	0.994	0.077	0.179	0.009	0.038	0.358	0.246
第四亚组	69.533	1.488	0.209	0.945	0.240	0.137	0.071	0.110	0.145	0.084

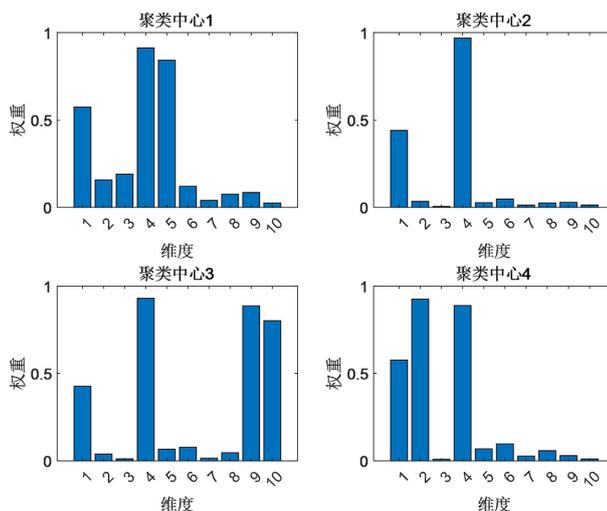


Figure 6. The weight of four cluster centers for 10-dimensional data

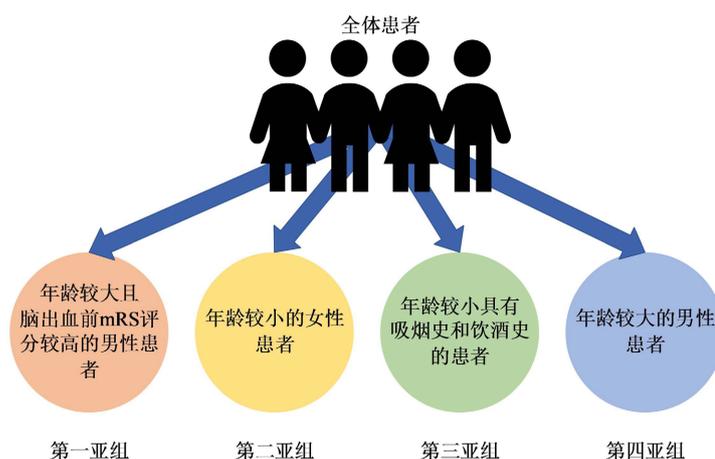
图 6. 四个聚类中心对于 10 维数据的权重

结合数据集原始数据,再分析上述 FCM 分类数据,根据统计学估计不难发现具有高血压病史的出血性脑卒中患者大约占全体患者的 88.125%,也就是说出血性脑卒中患者大概率存在高血压的情况或曾经得过高血压,所以如果按照高低血压或者高血压病史去区分患者显然结果并不理想。结合患者个人特征数据,根据 FCM 模糊算法,自动迭代出来的权重,可以将本题中全体患者划分为如图 7 所示的四个亚组。每个组的人数分别为表 3 所示。

**Table 3.** Clustering centers of 10-dimensional data for four subgroups

**表 3.** 四个亚组 10 维数据的聚类中心

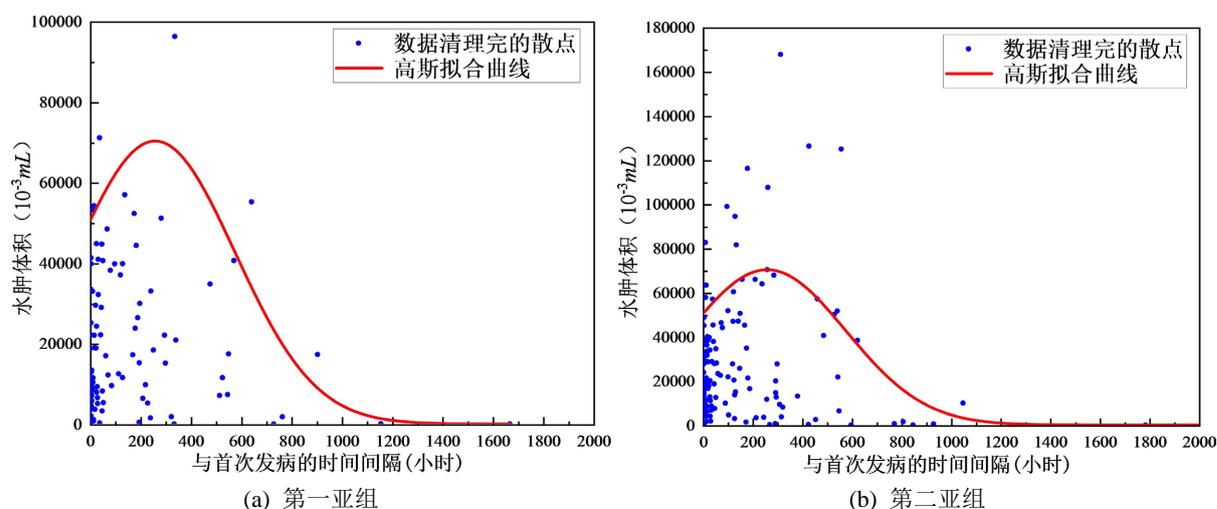
第一亚组	第二亚组	第三亚组	第四亚组
34	55	31	40



**Figure 7.** All patient subgroups

**图 7.** 全体患者亚组划分

划分完亚组以后将每个亚组中的患者单独提取出来绘制散点图,曲线拟合与 2.1 节相似采用高斯拟合对四组散点进行单独的拟合并各自计算残差,这里就不过多赘述具体流程。图 8 为利用 Gauss 非线性拟合得到的不同人群水肿体积进展曲线。



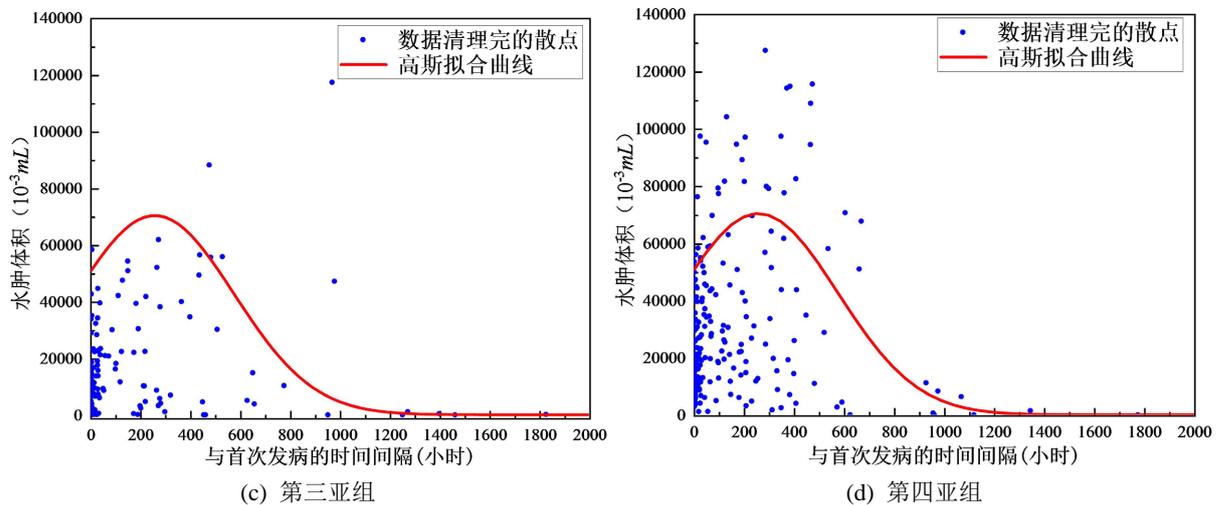


Figure 8. Edema volume progression curve for different groups of people  
图 8. 不同人群水肿体积进展曲线

### 3. 考虑不同疗法对患者水肿体积进展影响的建模

为探究不同治疗方法对于水肿体积的影响，数据集中有 8 种治疗方法分别为脑室引流、止血治疗、降颅压治疗、降压治疗、镇静、镇痛治疗、止吐护胃以及营养神经。治疗方法对于患者的作用往往具有滞后性，如果进行一次治疗后，患者没有再进行随访检查，那也无从探究治疗方法对于患者水肿体积到底有没有影响。也就是说如果患者随访次数过低的话，那么他的数据对于探究不同疗法的患者水肿体积进展的影响就没有意义，所以要对所有患者的医疗数据进行筛选。

对于数据的汇总，在数据处理部分就已经做好。现在需要提取出做过 4 次回访或以上的患者数据，因为他们的数据量更多，与结果具有更高的关联性。为了探究不同治疗方法对于患者水肿体积影响，首先考虑到主成分分析的方法。根据题目要求，实际上就是求解每一个治疗方法对于患者体内水肿体积的影响并输出排序结果；得到每种治疗方法对于水肿体积大小的重要程度。

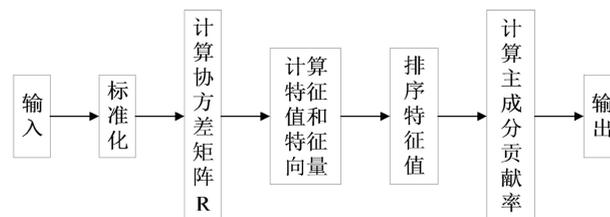


Figure 9. Principal component analysis process  
图 9. 主成分分析流程

主成分分析(PCA)可以有效处理多维数据适用于本题，它通过线性投影将数据映射到低维空间，并保留关键特征，如最大值、最小值、平均值和方差等。在进行 PCA 时，选择更多的特征可以更好地捕捉不同类别之间的差异。通过 PCA 可以消除冗余特征，最终选择保留的特征及其数量将在很大程度上决定模型的准确性。其具体步骤如图 9 所示。在这之中，最重要的是计算协方差矩阵  $R$  的特征值并进行排序，对于已经标准化的矩阵  $B$ ，其协方差矩阵为[4] [5]:

$$R = \frac{1}{x-1} B^T B \quad (2)$$

其中  $x$  为样本集数据，该矩阵中每一个元素  $a_{np}$  为：

$$a_{np} = \frac{\sum_{j=1}^x (b_{nj} - \bar{b}_n)(b_{jp} - \bar{b}_p)}{\sqrt{\sum_{j=1}^x (b_{nj} - \bar{b}_n)^2 \sum_{i=1}^u (b_{jp} - \bar{b}_p)^2}} \quad (3)$$

式中  $b_{nj}$  为标准化矩阵  $\mathbf{B}$  中元素； $\bar{b}_n$ 、 $\bar{b}_p$  为均值。

运用公式(2)和(3)得到协方差矩阵后便可将其特征值进行排序，并通过式(4)计算每个主成分的贡献率。

$$\varphi_i = \frac{\lambda_i}{\sum_{j=1}^k \lambda_k}, i = 1, 2, \dots, n \quad (4)$$

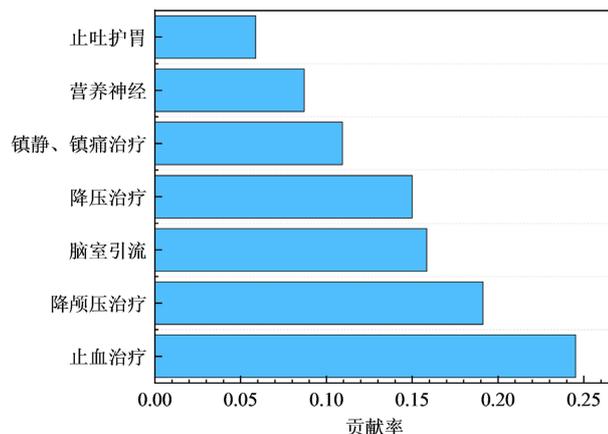
式中  $n$  为主成分的个数。

最终选取主成分较高的进行保留，同时保留该特征值所对应的特征向量  $a_1, a_2, \dots, a_d$ 。也就是输出该矩阵[6]。最终根据 PCA 法，得到 7 种治疗方法对于水肿体积的影响的贡献率如表 4 所示。由表 4 可知，是否进行过止血治疗是影响患者水肿体积的主要因素，其次降颅压治疗也是关键因素之一。将 7 种治疗方法，按对于水肿体积能起到显著抑制作用的高低排序，如图 10 所示。

**Table 4.** Effects of 7 treatment methods on edema volume obtained by PCA method

**表 4.** PCA 法得到的 7 种治疗方法对于水肿体积的影响

脑室引流	止血治疗	降颅压治疗	降压治疗	镇静、镇痛治疗	止吐护胃	营养神经
0.1584813	0.2451938	0.1911932	0.1499991	0.1093644	0.0587926	0.0869751



**Figure 10.** Ranking of contribution of 7 treatment methods

**图 10.** 7 种治疗方法贡献量排序

上述分析已经求出 7 种治疗方法对于水肿体积的影响的贡献率，即得到水肿体积与治疗方法的关系，为了探究血肿体积、水肿体积及治疗方法三者关系，故只需求得血肿体积与水肿体积的关系或与治疗方法之间的关系就可以。由于上述问题已经实现 PCA 分析，故同样运用 PCA 分析，进行血肿体积对于 7 种治疗方法的贡献量分析，如表 5 所示，血肿体积与治疗方法的关联度与水肿体积类似。根据相关资料显示并结合本文建模数据，血肿体积与水肿体积是正相关的，而 7 种治疗方法都能起到减小血肿或水肿体积的作用，其中止血治疗是治疗效果最好的，其次是降颅压治疗和脑室引流，而止吐护胃和营养神经对于减小血肿或水肿体积作用较小。

**Table 5.** Effects of 7 treatment methods on hematoma volume obtained by PCA method**表 5.** PCA 法得到的 7 种治疗方法对于血肿体积的影响

脑室引流	止血治疗	降颅压治疗	降压治疗	镇静、镇痛治疗	止吐护胃	营养神经	脑室引流
0.199540	0.2817396	0.20769817	0.123372	0.1167713	0.044839	0.026037	0.199540

#### 4. 结论

针对不同患者个体差异,本文利用 FCM 模糊聚类算法将全体患者划分为主体为年龄较大且脑出血前 mRS 评分较高的男性患者;主体为年龄较小的女性患者;主体为年龄较小吸烟史和饮酒史的患者;主体为年龄较大的男性患者四个亚组。并分别拟合四个亚组的曲线,以便区分不同人群的水肿体积随时间进展趋势。四个亚组之中,都有一个共同点就是他们中的大多数都有过高血压病史,可见高血压病史是引起患者水肿体积过大关键之处。在第一亚组内由于脑出血前 mRS 评分较高同时患者自身年龄较大,所以他们的水肿体积往往在峰值处较大,且从峰值的时间点更为提前。在第二亚组内主要为年龄较小的女性患者,该曲线峰值明显较低,但水肿体积缩小的也更慢一些。第三亚组的成员虽然他们年龄较小,但是存在吸烟史和饮酒史,其峰值较第二亚组即年龄较小的女性患者较高,同时他们的恢复速度也较慢。最后一组的特征与上一组类似。综合上述特征,不难发现,除了高血压病史以外,脑出血前 mRS 评分较高以及具有吸烟史和饮酒史或者高龄对水肿体积的影响较大,这也符合临床上得出的结论。

#### 参考文献

- [1] 朱博雅,谭晓东,顾耀华,陈奇,周晴,彭淑珍. 脑卒中危险因素对脑血管功能影响[J]. 中国公共卫生, 2022, 38(10): 1264-1268.
- [2] 朱帅,张一英,向芳,姜欣志,刘峰,陈波,刘星,梁冬妮,周晶,季莹. 脑卒中高危人群缺血性脑卒中 3 年发病影响因素 COX 回归分析[J]. 中国慢性病预防与控制, 2022, 30(2): 134-137.
- [3] 刘佳成,张志勇,周钦渊,等. 基于 FCM 聚类模型约束的二维初至旅行时反演[J]. 石油地球物理勘探, 2023, 58(5): 1115-1123.
- [4] 郭强,江清梅,何洪良,等. 基于主成分分析和聚类分析评价 141 份中蔗优良品系资源[J]. 热带作物学报, 2024, 45(1): 49-59.
- [5] 李思奇,吕王勇,邓桢,陈雯. 基于改进 PCA 的朴素贝叶斯分类算法[J]. 统计与决策, 2022, 38(1): 34-37.
- [6] 韩非,王颖竹,杨鹏,顾俊怡. 血清 PDGF-D 表达水平与老年出血性脑卒中患者颅内血肿扩大的相关性[J]. 中国老年学杂志, 2020, 40(24): 5179-5182.