

出血性脑卒中智能诊疗建模

姚宇朕

上海理工大学机械工程学院, 上海

收稿日期: 2024年8月4日; 录用日期: 2024年8月28日; 发布日期: 2024年9月4日

摘要

本文针对血肿周围水肿的发生及进展进行建模, 并深入探讨治疗干预和水肿进展的关联关系。首先基于100名患者的水肿体积与检查时间点数据, 运用二次多项式、三次多项式以及一维高斯函数模型进行数据拟合, 以找出最佳拟合模型。结果表明, 一维高斯函数模型在拟合效果上表现出色, 与实际数据高度吻合。因此, 由此建立全体患者水肿体积随时间进展的曲线。接下来, 通过K均值聚类将患者分为四个亚组, 并分别构建了水肿体积随时间的曲线模型, 分析患者群体之间的水肿体积变化趋势, 为个体化治疗提供了重要线索。结果显示, 不同亚组之间存在显著差异, 这可能对治疗策略的制定和优化产生积极影响。最后, 通过对患者的治疗数据进行多因素方差分析, 来探究不同治疗方法对血肿体积、水肿体积的影响。结果表明脑室引流和降颅压治疗对血肿和水肿体积具有显著影响。这为医生在治疗选择上提供了有力的依据。同时, 本文还通过Spearman相关系数计算发现血肿体积和水肿体积之间存在显著的正相关关系。

关键词

出血性脑卒中, 一维高斯函数模型, 多因素方差分析

Intelligent Diagnosis and Treatment Modeling of Hemorrhagic Stroke

Yuzhen Yao

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

Received: Aug. 4th, 2024; accepted: Aug. 28th, 2024; published: Sep. 4th, 2024

Abstract

This study aims to model the occurrence and progression of edema around the hematoma, and explore the relationship between therapeutic interventions and edema progression. First, the data of the first 100 patients with edema volume and time points were fitted using quadratic, cubic

polynomial, and one-dimensional Gaussian function models to find the best fitting model. The results show that the one-dimensional Gaussian function model performed well in terms of fitting effect and was highly consistent with actual data. Therefore, a curve of edema volume over time for all patients was established. Next, the patients were divided into four subgroups using K-means clustering, and edema volume over time curves were constructed for each subgroup to analyze the trend of edema volume changes among patient groups, providing important clues for individualized treatment. The results show that there are significant differences between subgroups, which may have a positive impact on the formulation and optimization of treatment strategies. Finally, a multivariate analysis of variance was conducted on the patient's treatment data to investigate the effects of different treatments on hematoma and edema volume. The results show that ventricular drainage and decompressive craniectomy have a significant effect on hematoma and edema volume. This provides strong evidence for doctors to make treatment choices. Additionally, this study found a significant positive correlation between hematoma volume and edema volume using Spearman's rank correlation coefficient.

Keywords

Hemorrhagic Stroke, One-Dimensional Gaussian Function Model, Multivariate Analysis of Variance

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

出血性脑卒中是一种脑出血性疾病，由非外伤性脑血管破裂所引起。该病的发病率约占全部脑卒中的 10%~15%。出血性脑卒中的病因多种多样，常见原因包括脑动脉瘤破裂和脑动脉异常等。当血管破裂时，血液会流入脑组织，导致机械性损伤并引发一系列生理反应。出血性脑卒中具有急性起病、快速进展以及预后差等特点[1]。疾病急性期的病死率约为 45%~50%，大约 80% 的患者将遗留严重的神经功能障碍，对患者家庭和社会造成了沉重负担。因此，了解出血性脑卒中的发病风险、整合影像学特征和患者临床信息，精确预测患者的预后，并根据预测结果进行临床决策，具有重要意义。

出血性脑卒中后，血肿范围扩大是预后不良的重要危险因素之一。在出血发生后的短时间内，由于脑组织受损和炎症反应等原因，血肿范围可能会逐渐增大，导致颅内压增高，进一步损伤神经功能，甚至威胁生命。因此，监测和控制血肿的扩张成为临床关注的焦点之一。伍发等人通过列线图预测模型来探究原发性脑出血患者发生早期血肿扩大的影响因素[2]；许启仲等通过深度学习模型对脑血肿扩大预测模型研究[3]；宋承东通过多因素 Logistic 回归分析法对血肿扩大影响因素进行建模[4]。此外，血肿周围的水肿作为脑出血后继发损伤的指标，近年来引起广泛关注。李曦通过列线图模型对恶性脑水肿风险预测模型进行建立与验证[5]，血肿周围的水肿可能导致脑组织受压，影响神经元的功能，加重神经功能损伤。因此，早期识别和预测出血性脑卒中后这两个关键事件对于改善患者预后和生活质量具有重要意义。

2. 研究思路

首先通过构建一条全体患者水肿体积随时间进展曲线 $y = f(x)$ ，为后文研究做准备。本文考虑将 100 个患者的水肿体积与对应检查的时间点提取出来，经过数据处理，以发病至影像检查时间为 x 轴，水肿体积为 y 轴。将每个患者每一次影像检查的水肿体积与对应检查的时间转化为一个点，将这些点拟合成一条曲线。本文选用二次多项式、三次多项式、一维高斯函数模型分别去拟合这些点，进行计算。

接下来通过不同人群的水肿体积随时间进展曲线，探索患者水肿体积随时间进展模式的个体差异。首先需要收集可靠且准确的数据，包括患者的水肿体积和对应的时间信息。对收集到的数据进行清洗和整理，处理可能存在的缺失值、异常值或错误数据，并确保数据格式的一致性。然后使用 K-mean 聚类将患者分成不同的亚组，对每个亚组的患者，构建水肿体积随时间变化的曲线模型，描绘出水肿体积随时间的趋势。

随后对不同治疗方法对血肿体积、水肿体积的影响情况进行分析，并探究血肿体积、水肿体积之间的关系。首先需要搜集患者的水肿体积测量数据和所接受的各种治疗方法信息。接下来，对数据进行整理和清洗，以确保数据的准确性和一致性，并根据需要将数据分组处理。结合前文的模型参数，运用多因素方差分析来探索治疗方法与水肿体积，血肿体积各模型参数之间的关系，评估不同因素的显著性，并计算血肿体积与水肿体积各模型参数之间的皮尔逊相关系数，以评估它们之间的相关性。以便更全面地理解血肿体积、水肿体积和治疗方法之间的关系。

3. 智能诊疗建模

3.1. 水肿体积变化曲线拟合

函数逼近论是函数论的重要组成部分，它的基本原理表示为在选定的某一类函数中寻找某个函数 g ，使它是已知函数在一定意义下的近似表示，并求出用函数 g 近似表示而产生的误差。在函数逼近问题中，用来逼近已知函数的函数类可以有不同的选择，通常有拉格朗日插值、牛顿插值、分段插值、样条插值等。本题采用二次多项式、三次多项式、一维高斯函数分别进行拟合。

基于以上的三类模型进行数据拟合分析，首先由原始数据可以得到前 100 个患者的水肿体积随重复检查时间点的变化情况，如图 1 所示。

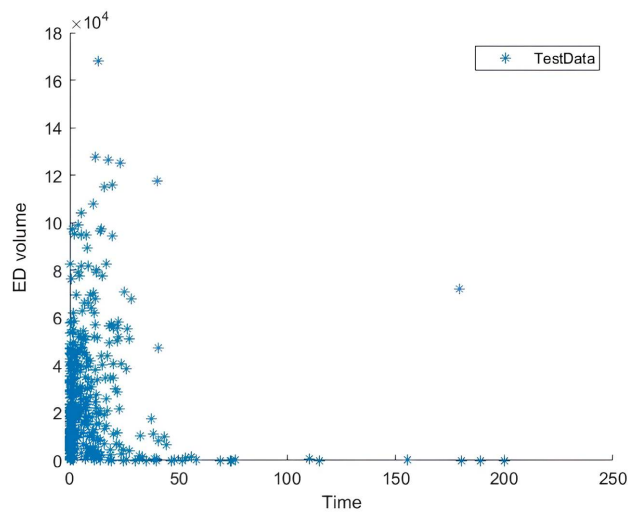


Figure 1. Scatter plot of edema volume in the first 100 patients
图 1. 前 100 名患者时间水肿体积散点图

对于图 1，x 轴表示为发病至影像检查时间的时间间隔，单位为 h，y 轴表示为水肿体积的大小，单位为 10^4 ml。观察图 1 可知，在水肿体积的大小主要分布在时间间隔小于 250 小时内。

二次多项式拟合是一种常见的数据拟合方法，它通过一个二次函数来逼近一组数据点。假设有一组自变量 x 和对应的因变量 y 的数据。使得该函数能够近似地通过这些数据点。

二次多项式拟合模型形式:

$$y = ax^2 + bx + c \quad (1)$$

在进行二次多项式拟合时, 本文使用最小二乘法来确定拟合曲线的系数 a 、 b 和 c 。最小二乘法的目标是使得每个数据点到拟合曲线的距离平方和最小化。

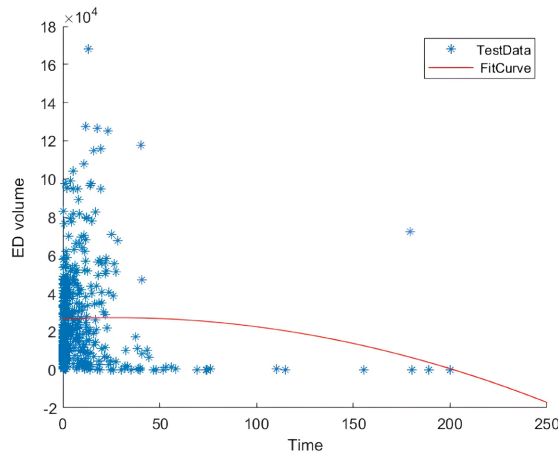


Figure 2. Quadratic polynomial fitting graph
图 2. 二次多项式拟合图

得二次多项式拟合图, 见图 2, 拟合曲线与数据点存在较大的偏差、拟合曲线的形状不符合预期或拟合曲线无法很好地适应数据的分布。此种模型拟合效果不理想。

三次多项式拟合是通过一个三次函数来逼近一组数据点。与二次多项式拟合类似, 它也是一种常见的数据拟合方法, 可以用于寻找数据的趋势或提取数据的特征。

与二次多项式拟合不同的是, 三次多项式拟合的模型形式为:

$$y = ax^3 + bx^2 + cx + d \quad (2)$$

其中 a 、 b 、 c 和 d 是待求的系数, x 和 y 分别表示自变量和因变量。需要求解的系数数量也相应地增加了一个。拟合结果如图 3 所示:

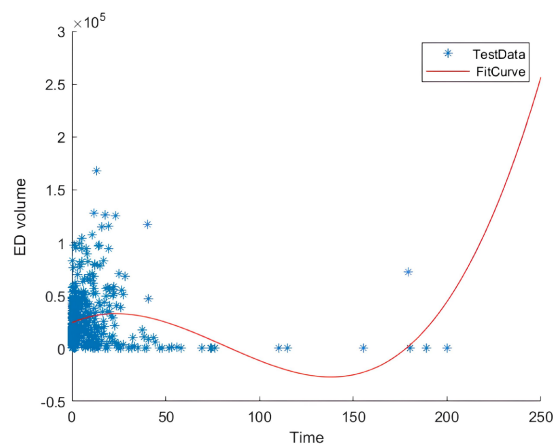


Figure 3. Cubic polynomial fitting graph
图 3. 三次多项式拟合图

观察拟合图，将拟合曲线与原始数据进行对比，拟合曲线未能有效地捕捉到数据的特征或趋势。此种模型拟合效果不理想。

一维高斯函数是一种常见的数学函数，用于描述连续变量的概率分布。它的数学形式为：

$$f(x) = ae^{-\frac{(x-b)^2}{2c^2}} \quad (3)$$

式中：

a ——幅度，表示高斯函数的峰值或最大值。

b ——偏移，表示高斯函数在 y 轴方向上的平移。

c ——常数，表示高斯函数的基准值或背景值。

一维高斯函数在很多领域中广泛应用，例如统计学、信号处理、机器学习等。它具有很好的数学性质和良好的模型拟合能力，因此被广泛用于对数据进行建模和分析[6]。

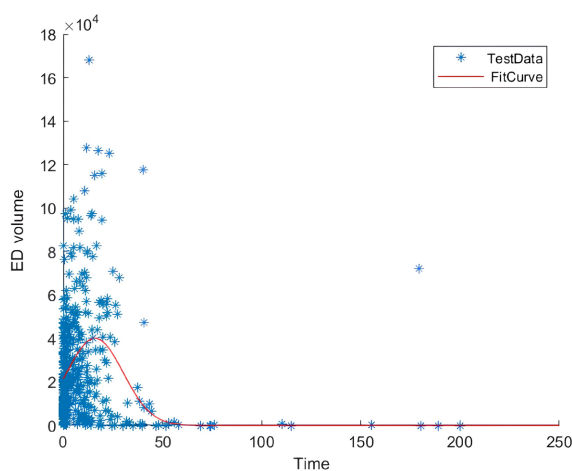


Figure 4. One-dimensional Gaussian function fitting graph

图 4. 一维高斯函数拟合图

得一维高斯函数拟合图，见图 4。由各拟合图可得，一维高斯函数拟合图，拟合曲线与数据点非常接近、拟合曲线很好地适应了数据的分布或拟合曲线能够很好地捕捉到数据的特征和趋势。此模型的拟合效果最为理想。

全体患者水肿体积随时间进展曲线为：

$$f(x) = 40080e^{-\frac{(x-16.03)^2}{20.12^2}} \quad (4)$$

3.2. K-Mean 聚类分组

为了更好地分析患者水肿体积随时间进展模式的个体差异，从而科学合理的反映患者水肿体积随时间进展模式的整体的变化趋势，本文对 9 个不同特征类型变化趋势进行 K-mean 聚类，将患者人群整体划分为四个亚组。

聚类分析是一种无监督学习方法，用于将相似的样本分组或聚类到同一个类别中。K 均值聚类(K-mean 聚类)是一种基于中心的聚类算法，在迭代过程中使用样本与中心点之间的距离来确定样本的类别，并通过优化中心点位置来最小化样本与中心点距离之和[7]。相比于其他聚类算法，K 均值聚类是按照样

本进行聚类，而不需要事先定义字段或特征。选择适当的聚类个数 K 是关键，可以采用启发式方法、经验或评估指标来选择最佳值[8]。总而言之， K 均值聚类能够将相似的样本聚集在一起，有助于发现隐藏的模式和结构。

患者的特征分别为年龄、性别、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、高血压九种不同类型，通过 K 均值聚类算法将患者整体聚类成四个聚类类别。

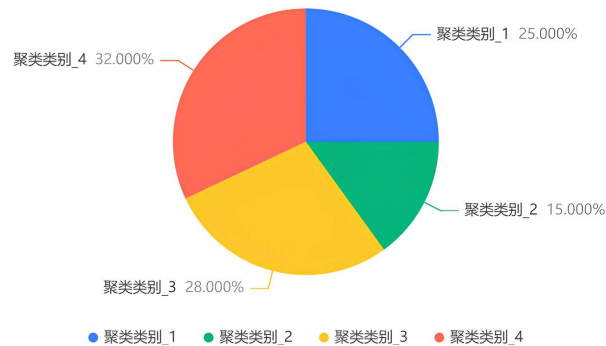


Figure 5. Cluster summary diagram

图 5. 聚类汇总图

图 5 以可视化的形式展示了模型聚类的结果，包括频数、所占百分比。构建好四个亚组，聚类类别_1、聚类类别_2、聚类类别_3、聚类类别_4 分别对应亚组 1、亚组 2、亚组 3、亚组 4。后面的模型与问前文相同，采用一维高斯函数模型进行拟合求解。

3.3. 各亚组水肿体积变化曲线拟合

将患者分成四个亚组，分别对亚组进行一维高斯模型拟合，如图 6。

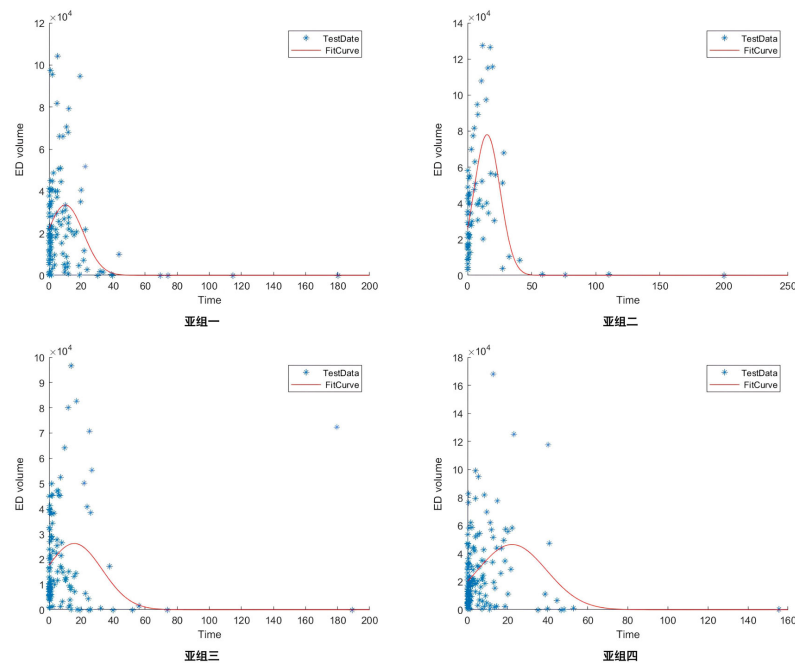


Figure 6. Patient time-edema volume fit diagram (4 subgroups)

图 6. 患者时间 - 水肿体积拟合图(4 种亚组)

由各拟合图可知，不同亚组之间存在显著差异，这可能对治疗策略的制定和优化产生积极影响。基于一维高斯函数拟合的曲线与数据点非常接近，拟合曲线很好地适应了数据的分布。

4. 血肿体积、水肿体积及治疗方法三者关系分析

4.1. 治疗方法对血肿体积、水肿体积的影响

对于治疗方法对血肿体积、水肿体积的影响问题，本文考虑使用多因素方差分析进行分析。

多因素方差分析(ANOVA)是一种常用的统计方法，用于分析多个因素对观测变量的影响程度，以及各因素之间是否存在交互作用。它可以帮助确定不同因素对于观测变量的主要影响，并评估这些影响的显著性[9]。

在多因素方差分析中，通常有两个或更多的自变量(也称为因素)以及一个因变量(也称为观测变量)。自变量可以是分类变量(如性别、治疗方式等)或连续变量(如年龄、剂量等)，而因变量则是我们感兴趣的主要测量指标。

通过进行方差分析，可以比较不同因素水平之间的均值差异，并判断这些差异是否显著。如果发现在至少一个因素上存在显著差异，则说明该因素对于观测变量的影响是显著的。此外，多因素方差分析还可以检验不同因素之间是否存在交互作用。如果存在交互作用，意味着不同因素的组合对于观测变量的影响不仅仅是各因素单独的总和，而是相互作用的结果。

在多因素方差分析中，F值(F-value)和P值(P-value)是用来评估不同因素对观测变量的影响是否显著的统计指标。

F值是通过计算方差比来度量不同因素之间的均值差异。它是组间方差与组内方差的比值，反映了观测变量的变异程度是否由于不同因素的影响而显著增加。F值越大，表示组间差异相对于组内差异的比例越大，意味着不同因素对观测变量的影响越显著。

P值则用于衡量F值的显著性。P值表示在原假设成立的情况下，观察到的统计量(如F值)或更极端结果的概率。在多因素方差分析中，P值通常与显著性水平(如0.05或0.01)进行比较。如果P值小于显著性水平，通常是0.05，我们可以拒绝原假设，即认为不同因素对观测变量的影响是显著的。

表1为各治疗项目与水肿拟合高斯模型中参数a的方差分析表，其中***、**、*分别代表1%、5%、10%的显著性水平，表2为各治疗项目与水肿血肿各高斯模型参数的P值汇总。

Table 1. ED_Gaussian_a variance table

表 1. ED_Gaussian_a 方差分析表

项	平方和	自由度	均方 F	F	P	R ²
截距	13186724346.406	1	13186724346.406	7.847	0.006***	
脑室引流	766719741.853	1	766719741.853	0.456	0.501	
止血治疗	2786047742.607	1	2786047742.607	1.658	0.201	
降颅压治疗	3603545138.815	1	3603545138.815	2.144	0.147	
降压治疗	403607174.351	1	403607174.351	0.24	0.625	0.048
镇静镇痛治疗	295303877.107	1	295303877.107	0.176	0.676	
止吐护胃	98391384.976	1	98391384.976	0.059	0.809	
营养神经	25354079.907	1	25354079.907	0.015	0.903	
误差	144525328398.169	86	1680527074.397			

Figure 2. Summary of P values of each parameter**表 2.** 各参数 P 值汇总

	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
HM_Gaussian_c	0.285	0.962	0.791	0.738	0.812	0.557	0.509
HM_Gaussian_b	0.25	0.913	0.306	0.919	0.454	0.85	0.576
HM_Gaussian_a	0.001***	0.886	0.046**	0.623	0.412	0.529	0.709
ED_Gaussian_c	0.027**	0.111	0.081*	0.826	1	0.253	0.533
ED_Gaussian_b	0.022**	0.977	0.327	0.478	0.298	0.077*	0.398
ED_Gaussian_a	0.501	0.201	0.147	0.625	0.676	0.809	0.903

由多因素方差分析可见，脑室引流对血肿体积与水肿体积影响具有较大显著性水平，降颅压治疗与血肿体积、水肿体积也有显著性水平。由此可见对于脑内水肿血肿的抑制和治疗，脑室引流，降颅压治疗或是更有效的方法。

4.2. 血肿体积、水肿体积相关性分析

血肿体积、水肿体积的量化通过高斯拟合数学模型的三个参数来反映，常见的相关性分析有 Pearson 相关系数计算和 Spearman 相关系数计算，不满足线性关系，且不满足正态分布的数据应当使用 Spearman 相关系数计算，故需对全部患者的各高斯拟合参数进行正态分布判别。

是否服从正态分布可以通过正态性检验来判别。

对于正态性检验，常用的方法有 Shapiro-Wilk 检验和 Kolmogorov-Smirnov 检验。Kolmogorov-Smirnov 检验适用于大样本(一般样本数 5000 以上)，原假设也是数据符合正态分布。但是本题不适用。Shapiro-Wilk 检验适用于小样本(一般样本数 5000 以下)，它的原假设是数据符合正态分布。如果 Shapiro-Wilk 检验的 P 值大于显著性水平(通常是 0.05)，即 $P > 0.05$ ，原假设接受，认为数据近似服从正态分布，反之则拒绝原假设，说明数据不符合正态分布。

各参数 Shapiro-Wilk 检验结果如表 3 所示，可以看出各参数的 S-W 检验 P 值基本都 < 0.05 ，可以确定其均不服从正态分布。

Table 3. Table of Shapiro-Wilk test results for each parameter**表 3.** 各参数 Shapiro-Wilk 检验结果表

变量名	S-W 检验 P 值
ED_Gaussian_a	0.000
ED_Gaussian_b	0.025
ED_Gaussian_c	0.003
HM_Gaussian_a	0.000
HM_Gaussian_b	0.061
HM_Gaussian_c	0.004

综合 Q-Q 图和正态性检验可以得出全部患者的各高斯拟合参数均不服从正态分布，因此本文选择更适合处理非正态分布数据的 Spearman 模型[10] [11]进行相关性检验。

斯皮尔曼相关系数被定义成等级变量之间的皮尔逊相关系数。

对于样本容量为 n 的样本， n 个原始数据被转换成等级数据， ρ 相关系数为：

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (5)$$

其中, x , y 为两个待分析变量取值, \bar{x} , \bar{y} 为两个变量的平均值。相关系数的绝对值越大, 则代表两个变量之间的相关性越强。部分变量的相关性如下图所示, 其中, 对角线变量表示自相关, 其他位置的变量表示互相之间的相关性, 颜色越深, 接近于红色表明相关性越强, 反之变量之间的相关性越弱。

在 SPSSPRO 中计算各参数相关性系数, 得相关系数表, 见表 4, 其中***、**、* 分别代表 1%、5%、10% 的显著性水平, 和相关系数热力图, 见图 7。

Table 4. Table of correlation coefficient of each parameter

表 4. 各参数相关性系数表

	ED_Gaussian_a	HM_Gaussian_b	HM_Gaussian_c
HM_Gaussian_a	0.486 (0.000***)	0.114 (0.331)	0.266 (0.022**)
HM_Gaussian_b	0.267 (0.021**)	0.285 (0.014**)	0.103 (0.384)
HM_Gaussian_c	0.348 (0.002***)	0.468 (0.000***)	0.521 (0.000***)

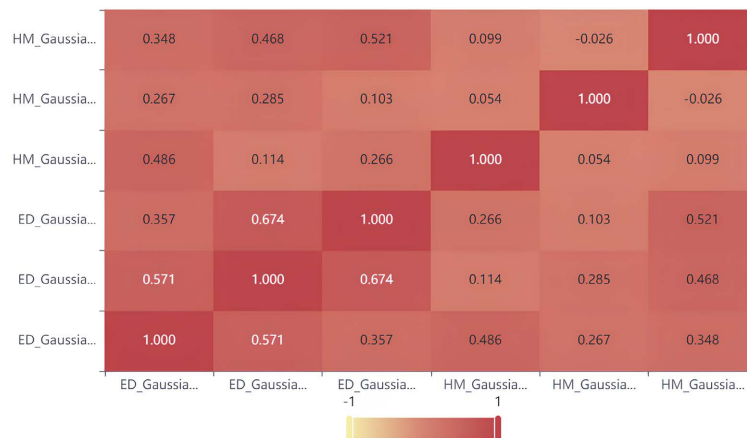


Figure 7. Thermal map with correlation coefficient of each parameter

图 7. 各参数相关性系数热力图

根据各参数相关性系数表和各参数相关性系数热力图可以分析出水肿高斯模型与各参数与血肿高斯模型各参数具有很强的正相关, 由此可见血肿周围水肿或会伴随血肿的变化而变化。

5. 结论

本文针对出血性脑卒中智能诊疗建模, 首先通过一维高斯函数拟合构建一条全体患者水肿体积随时间进展曲线。随后通过 K-mean 聚类将患者分成不同的亚组, 对每个亚组的患者, 构建水肿体积随时间变化的曲线模型, 根据结果发现不同亚组之间存在显著差异, 进行患者分类分组可能可以对治疗策略的制定和优化产生积极影响。

针对不同治疗方法对血肿体积、水肿体积的影响情况, 与血肿体积、水肿体积之间的关系。通过多因素方差分析, 其结果可见, 脑室引流对血肿体积与水肿体积影响具有较大显著性水平, 降颅压治疗与血肿体积、水肿体积也有显著性水平。由此可见对于脑内水肿血肿的抑制和治疗, 脑室引流, 降颅压治疗或是更有效的方法。同时通过皮尔逊相关系数, 可以分析得知水肿体积与血肿体积存在着很强的正相关关系。

参考文献

- [1] 孙艳, 李龙侗, 杨宝义. 出血性脑卒中患者家属 ICU 后综合症临床分析[J]. 湖北医药学院学报, 2022, 41(3): 304-307.
- [2] 伍发, 杨钰林, 伍婷婷, 等. 原发性脑出血早期血肿扩大的影响因素及列线图预测模型[J]. 解放军医学杂志, 2024, 49(5): 504-510.
- [3] 许启仲, 陈义均, 尹游兵, 等. 基于 CT 图像的高精度血肿检测与脑血肿扩大预测模型研究[J]. 深圳中西医结合杂志, 2024, 34(9): 1-5+137-138+141.
- [4] 宋承东, 周聪, 李昌勇, 等. 脑出血患者血肿扩大的 CT 征象、参数表现及其相关的影响因素模型构建[J]. 中国 CT 和 MRI 杂志, 2024, 22(3): 28-31.
- [5] 李曦, 黄鹞, 谭昌亚, 等. 急性缺血性脑卒中患者机械取栓后恶性脑水肿风险预测模型建立与验证[J]. 成都医学院学报, 2023, 18(6): 746-751.
- [6] 张淼, 王琰, 祁燕. 基于高斯一维随机插值算法的分形云建模[J]. 沈阳理工大学学报, 2008, 27(1): 10-13.
- [7] 金巨波. 基于 K-mean 聚类算法的海量数据分析方法[J]. 九江学院学报(自然科学版), 2020, 35(4): 53-55.
- [8] 陈慧, 龙飞, 段智云. 一种基于小波零树编码和 K-mean 聚类的图像压缩的实现[J]. 软件, 2016(2): 33-34, 38.
- [9] 龚江, 石培春, 李春燕. 使用 SPSS 软件进行多因素方差分析[J]. 农业网络信息, 2012(4): 31-33.
- [10] 王晓燕, 李美洲. 浅谈等级相关系数与斯皮尔曼等级相关系数[J]. 广东轻工职业技术学院学报, 2006, 5(4): 26-27.
- [11] 谢文华. Spearman 相关系数的变量筛选方法[D]: [硕士学位论文]. 北京: 北京工业大学, 2015.