DOI: 10.23977/medsc.2023.040820
ISSN 2616-1907 Vol. 4 Num. 8

Intelligent diagnosis and treatment of hemorrhagic stroke patients

Chang Cai[1,*], Yanwen Wang[1], Zijian Zhang[1], Feng Liu[1]

[1]*School of Electronic Information, Xijing University, Xian, China*
[*]*Corresponding author*

Keywords: Intelligent prediction, KNN algorithm, LSTM neural network, K-Means clustering algorithm

Abstract: In recent years, artificial intelligence technology in the medical field has developed rapidly, providing revolutionary possibilities for deep mining of large amounts of image data. Hemorrhagic stroke is a serious disease in which enlargement of hematoma is one of the important risk factors for poor prognosis. Therefore, monitoring and controlling hematoma expansion is one of the key clinical concerns. In addition, edema around the hematoma, as a sign of secondary injury after cerebral hemorrhage, has attracted widespread clinical attention in recent years. Monitoring and controlling the expansion of the hematoma is one of the focuses of clinical attention. This study is a model based on a large amount of medical data, aiming to analyze and intelligently predict key clinical events of hemorrhagic stroke. Based on the patient's personal history, disease history, clinical information related to the onset and hematoma and edema data obtained from images, we use The KNN algorithm is used to obtain the probability of hematoma expansion in patients, the K- means clustering algorithm is used to classify and analyze the edema volume changes in different patients, and finally the LSTM neural network is used to complete the comparison. The work of predicting patient functional status and disability provides a reference for intelligent medical diagnosis and treatment models.

1. Introduction

Hemorrhagic stroke refers to cerebral hemorrhage caused by non-traumatic intraparenchymal blood vessel rupture, accounting for 10-15% of all stroke. The cause of the disease is complex, usually due to factors such as ruptured cerebral aneurysm and abnormal cerebral arteries, which cause blood to flow into the brain tissue from the ruptured blood vessels, causing mechanical damage to the brain and triggering a series of complex physiological and pathological reactions. Hemorrhagic stroke has an acute onset, rapid progression, and poor prognosis. The mortality rate in the acute period is as high as 45-50%. About 80% of patients will have severe neurological dysfunction, which brings heavy consequences health and financial burden. Therefore, it is of great clinical significance to explore the risk of hemorrhagic stroke, integrate imaging characteristics, patient clinical information and clinical diagnosis and treatment plans, accurately predict patient prognosis, and optimize.

After hemorrhagic stroke, the expansion of hematoma is one of the important risk factors for poor prognosis. In a short period of time after the hemorrhage occurs, the scope of the hematoma may gradually expand due to factors such as brain tissue damage and inflammatory response, resulting in a rapid increase in intracranial pressure, which may further worsen neurological function and even endanger the patient's life. Therefore, monitoring and controlling hematoma expansion is one of the focuses of clinical attention. In addition, edema around the hematoma, as a marker of secondary injury after cerebral hemorrhage, has attracted widespread clinical attention in recent years. The edema around the hematoma may cause compression of the brain tissue, which in turn affects the function of neurons and further damages the brain tissue, thereby aggravating the patient's neurological damage. In summary, early identification and prediction of two important key events after hemorrhagic stroke, namely hematoma expansion and the occurrence and development of peripheral hematoma edema, are of great significance to improving patient prognosis and improving their quality of life.

2. The Establishment Of Intelligent Model

2.1. Hematoma Expansion Discriminant Model

First identifies whether 100 patients will experience hematoma expansion within the next 48 hours of disease onset based on the corresponding fields and data based on the serial number of the first impact examination on admission and the time interval from onset to first imaging examination. The data used are the serial number of the first impact examination from admission and the time interval from the onset to the first imaging examination, the serial number of each time point and the corresponding hematoma volume. Therefore, it is necessary to construct a relative model to determine the expansion of hematoma, and the basis for establishing this model is mainly divided into the following two parts:

Whether imaging examinations were performed within 48 hours of onset of illness.

Whether the hematoma volume has increased by ≥6ML in subsequent examinations compared with the absolute volume of the hematoma in the first examination or whether the relative volume has increased by ≥33 %.

If the above two points are met at the same time, it means that the patient has expanded the hematoma, and the expansion time needs to be recorded. If the above two points do not occur, based on this, it can be judged that the patient does not have hematoma expansion.

2.2. KNN Algorithm Model

According to the subsequent analysis, it can be concluded that the problem of dealing with the probability of hematoma expansion is essentially a two-classification problem. The so-called two-classification problem means that there are two classifications, in "is" or "is" and "has" Or choose between "none", and KNN can be used to calculate this probability problem. This algorithm, also known as the nearest neighbor method, is mainly used in classification and regression, but it can also be used for probability estimation through the results obtained. It is assumed that there is a sample and most of the K- similar samples in the feature space are in the same category. This algorithm is based on this[1]. We can use Euclidean distance to calculate similarity. Euclidean distance measures are the absolute distance between two points in multi-dimensional space. In n- dimensional space, the formula of Euclidean distance is:

The steps of using this algorithm model to calculate probability can be roughly divided into the following three steps: Step 1: Among all samples, the training samples found are the K closest ones.

Step 2: Record the number of occurrences of K neighbors in all categories.

Step 3: Calculate the estimated probabilities among all categories.

Corresponding probabilities will be generated in all categories during the execution steps.

Generally, the category with the highest probability is selected as the final result.

Using MATLAB software, the personal history, disease history, incidence-related data, and other parameter variables used by 100 patients are used as training input, and the patient's hematoma judgment results are used to train the KNN model, then the patient data used were input into the model, and the accuracy was obtained at 73 %.

2.3. Curve Fitting of the Volume and Time of Edema

According to the scatter distribution shown in Figure 1, it can be observed that there is no linear relationship between the edema volume and the time from onset to imaging examination in all patients. This suggests that in this patient population, the increase in edema volume is not solely dependent on the time from onset to imaging. There may be other factors, such as pathophysiological processes, individual differences, etc., that play an important role in the formation and development of edema volume. In order to more fully understand the relationship between edema volume and time from onset to imaging examination, we decided to use curve fitting analysis. This fitting method can better reveal possible nonlinear trends and underlying correlation patterns. Through curve fitting, we can more accurately assess the relationship between edema volume and time from onset to imaging and further explore factors associated with edema development. After curve fitting, it can be observed that all points are basically evenly distributed on both sides of the curve, which means that the curve can better capture the overall trend of the data and can better adapt to changes in the data. This uniform distribution indicates that curve fitting can better summarize the overall characteristics of the data. Finally, the fitting curve is obtained through the MATLAB model as

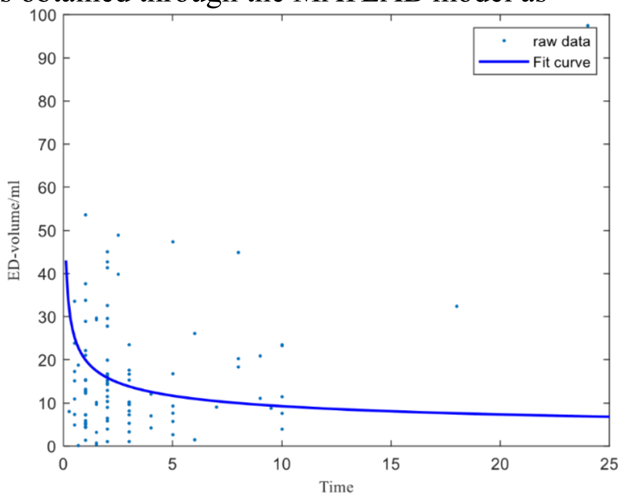

Figure 1: Fitting curve of volume and time

2.4. Curve Fitting of Edema Volume Progression over Time

Through typical clustering problem research, the curve fitting of the progression of edema volume in patients over time was performed. To achieve this goal, we consider the K- means clustering algorithm. When exploring the characteristics of patients' edema volume changes over time, the data set was analyzed and divided by applying the K- means clustering algorithm. K- means clustering algorithm is a simple and effective iterative clustering algorithm. It divides the data set into K non-overlapping clusters based on Euclidean distance as a similarity measure, and describes each cluster by the cluster center. We use the patient's edema volume changes over time as the input data set. Through the K- means clustering algorithm, we are able to divide these data points into clusters with similar patterns of change. Each cluster represents a group of patients who exhibit similar trends and patterns in changes in edema volume over time[2]. By analyzing the clustering results, we can obtain the characteristics and patterns of edema volume changes among different groups of people.

Step 1: Choose k (the number of clusters) and determine how many clusters you want to divide the data into.

Step 2: Initialization: Select the number K of clusters to be divided, and randomly select a few from the data as the initial midpoints of clustering. You can also optimize this initial clustering midpoint.

Step 3: Assign data points: For each data point in all data, calculate their relationship with each number

8. According to the distance between the cluster midpoints, the data points are assigned to each nearest cluster midpoint.

Step 4: Update the cluster midpoint: Calculate the mean of the data in the cluster and change the cluster midpoint, assign the mean of the data to the cluster midpoint.

Step 5: Determination conditions: Check whether the maximum number of iterations is reached, or check whether the change in the cluster center is less than a certain threshold. If the stop condition is met, proceed to the next step; otherwise, return to the third step and re-determine the allocation data.

Step 6: Output results: The final clustering results include the cluster label to which each data point belongs and the location of the cluster center[3].

Finally, we performed iterative learning on the K- means clustering algorithm. We tried k=3, 4, and 5. Through observation, we found that when k=5, the clustering effect at this time is the best.

The clustering effect is shown in Figure 2 below.

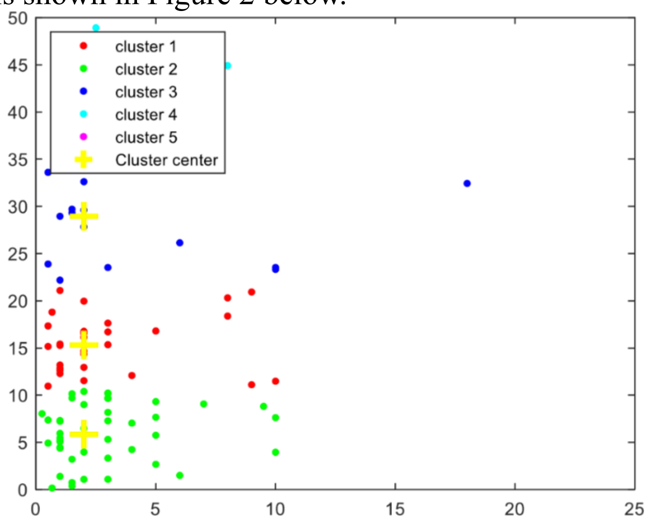

Figure 2: The clustering effect of K-means clustering algorithm

Through the algorithm, the fitting curve of the changes in edema volume and time in different groups of people can be obtained. The equation is as follows.

$$v_1 = 0.1197\, t_1 + 15.1107 \tag{1}$$

$$v_2 = 0.1848\, t_2 + 5.3593 \tag{2}$$

$$v_3 = -0.057\, t_3 + 28.2624 \tag{3}$$

$$v_4 = 0.1927\, t_4 + 44.0603 \tag{4}$$

$$v_5 = 4.0626\, t_5 \tag{5}$$

In the above formula, v represents the volume of edema, and t represents time. From this, we can predict the results of changes in edema volume over time in patients in different clusters.

2.5. Effect of Different Medical Regimens on Edema

First, different patients have different treatment plans. Record the treatment plans of 100 patients. There are a total of seven medical methods we can use: ventricular drainage, hemostatic treatment, hypothermia treatment of intracranial pressure, anti-hypertensive treatment, analgesic treatment, anti-vomiting and stomach protection, and nerve nutrition. Here we use the entropy weight method to conduct modeling analysis and research on 7 treatment options.

The entropy weight method relies on the concept of information entropy for decision analysis, and its purpose is to configure appropriate weights for multiple evaluation indicators. This method comprehensively reflects the objective and subjective considerations of each indicator, so it is often used when making an overall evaluation of multiple indicators. The key point of the entropy weight method is that when an indicator shows greater change or dispersion, it usually contains richer information and should be given a higher weight; while when the indicator changes less, its weight is reduced accordingly entropy weight[4].The operation process of the method is as follows: Data adjustment: Standardize the initial data, often using linear normalization.

Calculate entropy: Use the mathematical formula for information entropy to determine the entropy value for each metric. Determine the information gain value: calculated by subtracting the entropy value from 1. Configuration weight: The weight of an indicator is determined by the sum of its information benefit value and the information benefit value of all indicators. The distinguishing feature of the entropy weight method is its objectivity, because it configures weights according to the specific distribution of data. And avoid subjective judgment as much as possible. However, this method may lead to inaccurate evaluation of weights when there is a correlation between some indicators[5]. Based on the above content, the weight distribution corresponding to the 7 different treatment methods can be obtained. Table 1 of the above results is presented:

Table 1: Weight allocation of treatment methods

Treatment options	Weight assignment
Ventricular drainage	0.6591
Analgesic therapy	0.1113
Hemostasis therapy	0.0901
Treat with cranial pressure	0.0813
Antihypertensive therapy	0.0320
Antiemetic and stomach protection	0.0130
Nutritional nerves	0.0130

After allocating the weights of all different treatment methods, it can be observed that the treatment method of ventricular drainage has the highest proportion of weight, and the proportion of antiemetics, stomach protection, and nerve nutrition has the lowest weight. This shows that antiemetics, stomach protection, and nerve nutrition have the lowest weight. Edema has minimal volume change.

2.6. The Relationship between Hematoma Volume, Edema Volume and Treatment Methods.

This time we still use the entropy weight method above to analyze the relationship between hematoma volume, edema volume and treatment methods, and then calculate the weight distribution of the three to provide us with a clear and quantitative perspective. This is shown in Table 2.

Table 2: The weight distribution of the three

Treatment options	Hematoma volume	Edema volume
Hemostasis therapy	0.6060	0.4399
Antiemetic and stomach protection	0.1023	0.0743
Treat with cranial pressure	0.0829	0.0602
Ventricular drainage	0.0805	0.3325
Antihypertensive therapy	0.0747	0.0542
Analgesic therapy	0.0294	0.0213
Nutritional nerves	0.0119	0.0087

A significant finding that can be obtained from the data in table 2 is that there is a very close relationship between hemostatic treatment and hematoma volume and edema volume. Hemostatic treatment accounts for the largest weight of hematoma volume and edema volume. This means that hemostatic treatment has the greatest weight here. On the other hand, the relationship with trophic nerves appears to be weak relative to other factors.

2.7 Prediction of Patients' MRS Scores

Recurrent neural network, also known as RNN, is used to process and model tasks with sequence dependencies, but has insufficient ability to learn long-term dependencies.

The upgraded version of RNN solves the problem of gradient disappearance or explosion when the recurrent neural network processes long sequences. LSTM uses a gating mechanism to control information, and the lack of long-term dependency learning ability has also been supplemented.

Table 3: Setting of network parameters

Parameter name	Parameter value
Enter the number of neurons in the layer	104
The number of neurons in the first hidden layer	450
The number of neurons in the second hidden layer	150
The number of neurons in the third hidden layer	20
The number of neurons in the output layer	1
Training methods	ADAM
Maximum number of rounds	300
Initial learning rate	0.001
Reduce the cycle interval for the learning rate	100
Learning rate reduction factor	0.2
Hardware resources	Single GPU

In stroke-related clinical studies, modified Rank score (MRS), NIHSS score or living status BI score are usually used to quantify patient prognosis and analyze disease progression or treatment efficacy. We use MRS to assess the functional status and disability of stroke patients. The lower the MRS, the milder the patient's symptoms and the lower the risk. Import values from the patient's personal history, disease history, disease-related data, and first imaging results into in LSTM, the patient's MRS score is used as the output and used to form a model, where the training set is 100 patient data, using all patients for testing. There are 104 inputs, and the outputs are divided into 0-6. Whether the training has converged can be observed through the convergence curve of the loss function. When there is a lot of data at the input end, it takes a long time to achieve the convergence effect, which is inconsistent with the demand. Therefore, when a large amount of data is imported, the data should be first Using the normalization method, the data in this way will eventually fall between 0 and 1, which can accelerate the convergence of the loss function[6-7].

The model estimates the training set and sets the corresponding network parameters. The model has three hidden layers. After multiple attempts, other parameters can obtain the best values. This is shown in Table 3.

After debugging, the model achieved the best prediction effect. The training progress and effect of the model are shown in figure 3. The function loss is gradually becoming smaller, and the RMSE curve is gradually smoothing, which means that with continuous training, Parameters were optimized to predict 90- day MRS scores for all patients.

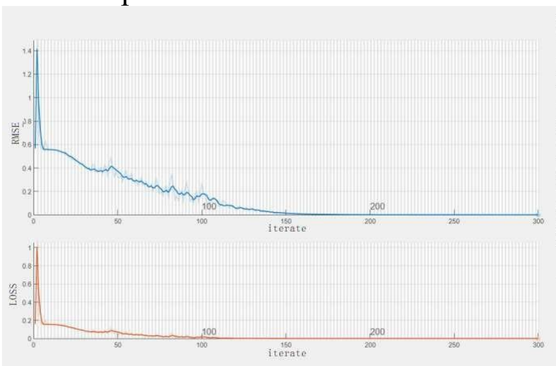

Figure 3: The training effect of the model

This data is added to the follow-up data. Counting the original indicator data, there are 280 in total. 100 patient data are used as the training set, and the remaining patient data are used as the test set. There are 280 inputs and the output is divided into 0-6. The model estimates the training set and sets the corresponding network parameters. The model has three hidden layers. After multiple attempts, other parameters can obtain the best values. This is shown in Table 4.

Table 4: Setting of network parameters

Parameter name	Parameter value
Enter the number of neurons in the layer	280
The number of neurons in the first hidden layer	550
The number of neurons in the second hidden layer	200
The number of neurons in the third hidden layer	40
The number of neurons in the output layer	1
Training methods	ADAM
Maximum number of rounds	300
Initial learning rate	0.001
Reduce the cycle interval for the learning rate	100
Learning rate reduction factor	0.2
Hardware resources	Single GPU

After debugging, the model achieved the best prediction effect. The training progress and effect of the model are shown in figure 4. The function loss is gradually becoming smaller, and the RMSE curve is gradually smoothing, which means that with continuous training, Parameters were optimized to predict 90- day MRS scores for the remaining patients.

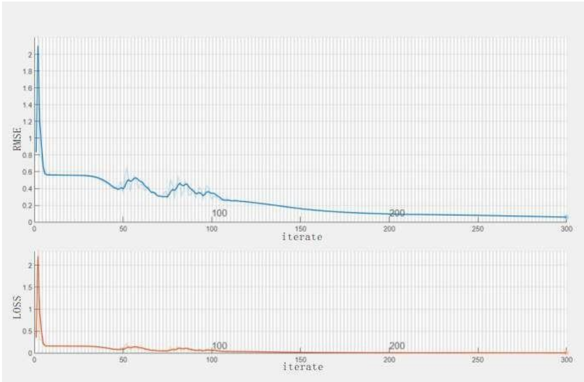

Figure 4: The training effect of the model

3. Conclusion

This article uses the patient's personal history, disease history, disease-related clinical information and hematoma and edema data obtained from images to construct a hematoma expansion discrimination model, a KNN probability prediction model and The MRS score prediction model enables related problems to be fundamentally solved. This article uses KNN to estimate the incidence of hematoma expansion, which is streamlined, popular and complete, and does not require training procedures; it is oriented to cumbersome planning situations:

KNN is particularly flexible in solving nonlinear and cumbersome boundaries. Because it is divided based on neighbor voting, it can perform good classification and prediction on complex surrounding data. KNN makes no assumptions about the distribution form of the data, so it can be suitable for different types of data. It also works effectively when dealing with non-linear, non-normal or imbalanced data. Among them, the LSTM neural network can handle gradient elimination well.

In order to solve the loss / gradient explosion problem, the training speed of the LSTM model will be faster, which greatly improves the computing efficiency.

Acknowledgments

The paper was supported by Key R&D Program Projects in Shaanxi Province, China (Grant No. 2023-YBGY-115) and Special Fund for High-level Talents of Xijing University, China (Grant No. XJ18B01).

References

[1] Linjie Zhu, Guangpeng Zhao, Lianghe Kang. Intrusion detection method based on combining MI feature selection and KNN classifier [J]. Gansu Science and Technology, 2022, 38(15): 33-36.

[2] Junyi Li. Research on network security detection application of K-means clustering algorithm based on big data [J]. Mechanical Design and Manufacturing Engineering, 2021, 50(09): 115-118.

[3] Mao Zhou, Ziruo Li, Xinhe Zhou, et al. Risk assessment of power grid construction projects based on AHP- entropy weight method [J]. Electrical Technology and Economics, 2023(05):192-196.

[4] Guoqing Z .Origin and development of hemorrhagic stroke [J].China Journal of Medical History, 2005.

[5] Xiaodong L, Zhiyi L, Yumin Y, et al. Analysis of the etiology, clinical features and prognosis of 345 cases of youth hemorrhagic stroke[J].Chinese Journal of the Frontiers of Medical Science(Electronic Version), 2017.

[6] Bajenaru O, Tiu C, Moessler H, et al. Efficacy and safety of Cerebrolysin in patients with hemorrhagic stroke[J]. Journal of Medicine and Life, 2010, 3(2).

[7] Meairs, Stephen. Advances in neurosonology – Brain perfusion, sonothrombolysis and CNS drug delivery [J]. Perspectives in Medicine, 2012, 1(1–12):5-10. DOI:10.1016/j.permed.2012.02.038.