# To explore intracerebral hematoma with a hybrid approach and combination of discriminative factors

Hui-Chu Chiu[1], Deng-Yiv Chiu[2,*], Yao-Hsien Lee[3] , Chih-Cheng Wang[2] , Chen-Shu Wang[4] , Chi-Chung Lee[2], Ming-Hsiung Ying[2], Mei-Yu Wu[2], Wen-Chih Chang[2]

1 PhD. Program of Technology Management, Chung-Hua University, Hsinchu, Taiwan.

[2] Department of Information Management, Chung-Hua University, Hsinchu, Taiwan.

[3] Department of Finance, Chung-Hua University, Hsinchu, Taiwan.

[4] Department of Information and Finance Management, National Taipei University of Technology, Taipei, Taiwan.

* Corresponding author: Deng-Yiv Chiu, Professor, Department of Information Management, Chung- Hua University, Hsinchu, Taiwan, ROC, E-mail: chiuden@chu.edu.tw

## Supplementary material

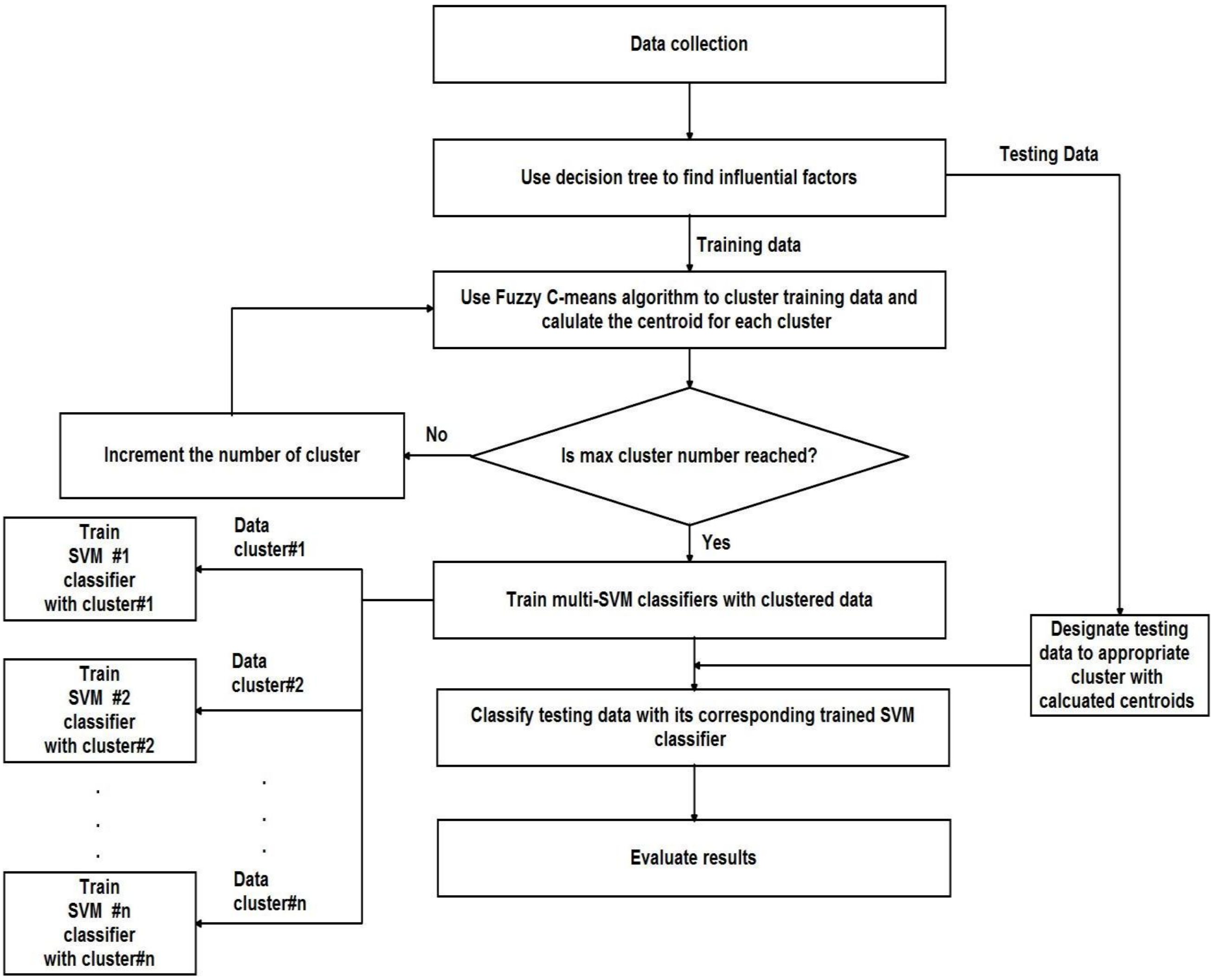


Fig. 1. The structure chart of the proposed approach. The chart demonstrates the procedure of the proposed approach, including decision tree, Fuzzy C-means algorithm, multi-SVM classifier, and evaluation.

Fuzzy C-means algorithm

Input:

$P = \{P_1, ..., P_N\}$ //set of input nodes

$N$ //total number of input nodes

$c$ //total number of clusters

$\varepsilon$ //convergence threshold

Output:

$V = \{\mathbf{v}_1, ..., \mathbf{v}_c\}$ //set of cluster centroids of all clusters

1. initialize membership degree of node $P_j$ belonging to cluster $i$ at time $t$, $\mu_{ij}^{(t)}$, satisfying

$$\sum_{i=1}^{c} \mu_{ij}^{(t)} = 1, \quad j = 1, 2, ..., N$$

$$0 < \sum_{j=1}^{N} \mu_{ij}^{(t)} < N, \quad i = 1, 2, ..., c$$

//the sum of all membership degrees of a node must be 1.

//the sum of membership degrees of all nodes belonging to same cluster must be between 0 and the total number of input nodes.

2. initialize performance index

$J^{(0)} = 0$

//performance index is used to estimate the distances among nodes and centroids.

3. find centroid for each cluster at time $t$

$$\mathbf{v}_i^{(t)} = \frac{\sum_{j=1}^{N} (\mu_{ij}^{(t)})^2 \cdot P_j}{\sum_{j=1}^{N} (\mu_{ij}^{(t)})^2}, \quad i = 1, ..., c$$

// $\mathbf{v}_i^{(t)}$ is the centroid of cluster $i$ at time $t$.

// $\mu_{ij}^{(t)}$ is the membership degree of node $P_j$ belonging to cluster $i$ at time $t$.

4. update membership degree for each node at time $t$

$$\mu_{ij}^{(t)} = \left[ \sum_{k=1}^{c} \left( \frac{\left\| P_j - \mathbf{v}_i^{(t)} \right\|^2}{\left\| P_j - \mathbf{v}_k^{(t)} \right\|^2} \right) \right]^{-1}$$

5. calculate performance index at time $t$, $J^{(t)}$:

$$J^{(t)} = \sum_{i=1}^{c} \sum_{j=1}^{N} \left[ (\mu_{ij}^{(t)})^2 \cdot \left\| P_j - \mathbf{v}_i^{(t)} \right\|^2 \right]$$

6. testify if convergence threshold is satisfied

If $\left| J^{(t)} - J^{(t\text{-}1)} \right| \geq \varepsilon$, $t = t + 1$; go to step 3

7. $V = \{\mathbf{v}_1^t, ..., \mathbf{v}_c^t\}$

Fig. 2. Fuzzy C-means algorithm. It combines the automatic clustering feature of K-means algorithm and the membership degree of fuzzy theorem to appropriately cluster objects and calculate the membership degrees among objects and clusters.

# Support Vector Machine

SVM defines the input variable supposition space with linear function and introduces the learning deviation to learn the mapping between input and output. As the linear fitting machine is operated in the feature space of the kernel function for learning, when the applied field has high dimension feature space, SVM shall effectively avoid overfitting problem and pose excellent learning performance [11].

The main concept is to transfer the mapping of input space kernel to high dimension feature space before re-classification and then find separation limit to classify data. To begin, SVM selects several support vectors from the training data to represent the entire data.

In the linear analysis, an optimal hyperplane can completely separate the sample into two classifications.

For SVM, finding an optimal hyperplane has to solve the following optimization problem:

$$\min_{w,b} \frac{1}{2}\|W\|^2 \tag{1}$$

Constraint:

$$y_i(w \cdot x + b) \geq 1$$

Where w: weight vector

b: bias

$x_i$ : input vector

$y_i$ : output vector

In the linear separation, it is a typical quadratic programming problem. Lagrange formula can be used to find the solution, where $\alpha$ is a Lagrange multiplier.

$$L(w,b,\alpha) = \tfrac{1}{2}\|w\|^2 - \sum_{i=1}^{p} \alpha_i [y_i (w \cdot x_i + b) - 1] \tag{2}$$

Where w: weight vector

b: bias

$x_i$ : input vector

$y_i$ : output vector

$\alpha_i$ : a Lagrange multiplier

In the optimization linear separation for nonseparable data, Lagrange formula also can be used to find the solution.

$$\max \ W(\alpha) = \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{3}$$

Constraint:

$$\sum_{i=1}^{p} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1,2,\ldots,p$$

Where w: weight vector

$x_i$ : input vector

$y_i$ : output vector

$\alpha_i$ : a Lagrange multiplier

p: count of total vectors

The Non-liner separation question can be solved using a mapping function $\Phi$, which called kernel function, can map input space of training data into a higher-dimensional feature space. The inner product is replaced by kernel function as Eq. (4).

$$(\Phi(x_i)\Phi(x_j)) := k(x_i, x_j) \tag{4}$$

Where $x_i, x_j$ : input vector

Any functions that meet Mercer's condition can be used as kernel functions.

Table 1. Demographic and clinical characteristics of ICH patients. The data is collected from a paper published at PLoS One [7]. The data collection consists of patients' vital signs, laboratory data, and CT findings.

| | Variable | Death* (n = 16) | Survival* (n = 90) | Total (n = 106) | p value | Missing data |
|---|---|---|---|---|---|---|
| | Males | 11 (68.8%) | 58 (64.4%) | 69 (65.1%) | 0.74 | 0 |
| | Age (years) | 62.1±16.1 | 63.0±15.2 | 62.8±15.3 | 0.83 | 0 |
| | Age ≧ 80 years | 3 (18.8%) | 16 (17.8%) | 19 (17.9%) | 0.92 | 0 |
| Vital signs | Systolic BP (mmHg) | 185.3±2.3 | 171.6±34.3 | 173.7±35.7 | 0.23 | 0 |
| | Diastolic BP (mmHg) | 96.6±14.3 | 91.3±18.7 | 9.1±18.1 | 0.21 | 0 |
| | Pulse pressure (mmHg) | 88.7±30.7 | 80.3±25.3 | 81.6±26.2 | 0.31 | 0 |
| | Pulse rate (per minute) | 91.8±14.8 | 89.7±12.3 | 90.0±12.7 | 0.60 | 0 |
| | Respiratory rate (per minute) | 20.1±1.7 | 19.7±1.9 | 19.7±1.8 | 0.44 | 0 |
| | Body temperature (◦C) | 36.7±0.3 | 36.8±0.4 | 36.8±0.4 | 0.12 | 0 |
| Laboratory Data | Glucose (mg/dL) | 185.6±89.5 | 145.4±51.3 | 149.5±57 | 0.22 | 18 (17.0%) |
| | WBC count ($\times 10^3/\mu L$) | 11.9±5.3 | 10.5±3.9 | 10.7±4.1 | 0.34 | 3(2.8%) |
| | Hemoglobin (mg/dL) | 12.2±2.8 | 13.3±2.4 | 13.1±2.5 | 0.16 | 1 (0.9%) |
| | Platelet count ($\times 10^5/\mu L$) | 1.76±0.83 | 2.53±0.99 | 2.43±1.0 | 0.003 | 1 (0.9%) |
| | INR | 1.4±0.5 | 1.2±1.0 | 1.2±1.0 | 0.43 | 19 (18.0%) |
| CT findings | Irregular shape | 12 (75%) | 33 (36.7%) | 45 (42.5%) | 0.004 | 0 |
| | Hematoma size >30 mL (ABC/2) | 11 (68.8%) | 24 (26.7%) | 35 (33.0) | 0.001 | 0 |
| | Intraventricular hemorrhage | 11 (68.8%) | 36 (40%) | 47 (44.3%) | 0.032 | 0 |
| | Infratentorial origin | 3 (18.8%) | 10 (11%) | 13 (12.3%) | 0.48 | 0 |
| | GCS score | 4.4±2.1 | 10.8±3.9 | 9.8±4.3 | <0.001 | 0 |
| | ICH score | 3.6±1.0 | 1.5±1.3 | 1.8±1.5 | <0.001 | 0 |

Table 2. The mortality distribution of Glasgow Coma Scale score for ICH patients. Patients with a GCS score of 3 to 4 have a significantly higher 30-day mortality (68.4%) than those with a GSC score of 5 to 12 (6.7%) and a GCS score of 13 to 15 (0%).

| Glasgow Coma Scale Score | Number Of Data | 30-Day Mortality |
|---|---|---|
| 3~4 | 19 | 68.4% |
| 5~12 | 45 | 6.7% |
| 13~15 | 42 | 0.0% |

Table 3. Patient distribution of decision tree. It shows the 30-day mortality for each leaf node of the decision tree.

| Tree Node | Count Of Patients | 30-Day Mortality |
|---|---|---|
| GCS score < 4.5 | 19 | 68.42% |
| GCS score >= 4.5 and Hematoma Size <38.05 | 70 | 1.43% |
| GCS score >= 4.5 and Hematoma Size>=38.05 | 17 | 11.76% |

Table 4. Patient distribution percentage of clusters by Fuzzy C-means algorithm. The percentage distribution consists of percentages for 3 clusters of training data and testing data.

| | Training data | Testing data |
|---|---|---|
| Cluster 1 | 76% | 62% |
| Cluster 2 | 11% | 10% |
| Cluster 3 | 13% | 28% |