

Feature Selection from Brain Stroke CT Images based on Binary Ant Colony

Homiera Kabudvand¹ and Saeid Fazli²

¹Faculty of Computer and Information Technology Engineering Qazvin Branch, Islamic Azad University Qazvin, Iran
Kabudvand_h@yahoo.com

²Research Institute of Modern Biological Techniques, University of Zanjan

fazli@znu.ac.ir

ABSTRACT

This paper proposes an automatic method to detect stroke region using Computed Tomography (CT) images. selecting a subset of features from a huge set for segmenting ischemic brain stroke in CT images, is a challenging problem. The proposed method consists of extracting 182 features and using heuristic algorithms to cope with complexity of feature selection. Ant colony optimization as one of the recent heuristic approaches is applied to the 98 brain CT scans from stroke subjects and tested by Support vector machines (SVM) classification. The results show a good improvement in accuracy of stroke diagnosis.

KEYWORDS

Feature Extraction, Classification, Touring Ant Colony Optimization(TACO), Support Vector Machine.

I. INTRODUCTION

Brain stroke occur when the blood supply to one's brain is interrupted or reduced. [2] Reasons of brain stroke are different, totally there are two types, first may be caused by a blocked artery(Ischemic stroke) or the lacking or bursting of a blood vessel (Hemorrhagic stroke).CT is a method that is usually used. Advantages of using it are 1.Velocity 2.price, 3 availability. [3,4] thus we use CT more than MRI. in images, a hemorrhage appears as a bright region (hyper _ intense),an Ischemic stroke appears as a dark region (hypo_ intense).finding a subset of features from a huge set is a challenging problem in fact those features are chosen that have maximum power in predicting output. until now many methods for selecting feature have been studied we use Ant Colony Optimization based Binary. like most evolutionary algorithms, this algorithm begins with population, searching is done in a parallel form and then suitability of population's members is determined according to cost function; and until algorithm is converged this process will be continued. Finally classification is tested by SVM and we discuss about experimental results.

II. METHODS

Our methods are based on three stages: 1- feature extraction, 2- finding feature vectors by binary Ant Colony, 3- classification by SVM.

1.Feature extraction

In this method we extract fourteen features (mean, Standard deviation, six features of GLCM and six features of GLRLM).

1.1 Mean

$$\mu = \frac{1}{a^2} \sum_{i=0}^{a-1} \sum_{j=0}^{a-1} x_{ij} \quad (1)$$

a- width of sliding window, x_{ij} - value of pixel in i - th row and j- th column of sliding window [5,6]

1.2 Standard deviation

$$\sigma^2 = \frac{1}{a^2-1} \sum_{i=0}^{a-1} \sum_{j=0}^{a-1} x_{ij}^2 - \mu^2 \quad (2)$$

1.3 Gray level co-occurrence matrix(GLCM)

Gray level co-occurrence matrix is a statistical method that considers location correlation. between

pixels according to distance and angle between them. GLCM can be formed for the direction of 0,45,90, 135, the mean of these four matrices creates a new matrix (totally 5 matrices) at following distances : d=1 ,5,7 pixels ; totally 15 matrices (5matrices ×3distances) will be made .

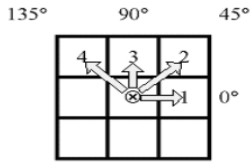


Fig1direction of GLCM

The reason of using four different directions, is that defective areas don't have special direction, and by this selection we can destroy the effect of rotation in GLCM. We use Six features of gray level co-occurrence matrix (GLCM, features are Entropy, Contract, Homogeny, Energy, Variance, Correlation) [7]

$$\text{Entropy} = \sum_{i=1}^n \sum_{j=1}^m c_{ij} \log c_{ij} \quad (3)$$

$$\text{Contract} = \sum_{i=1}^n \sum_{j=1}^m |i - j|^2 c_{ij} \quad (4)$$

$$\text{Energy} = \sum_{i=1}^n \sum_{j=1}^m c_{ij}^2 \quad (5)$$

$$\text{Homogeny} = \sum_{i=1}^n \sum_{j=1}^m \frac{c_{ij}}{1+|i-j|} \quad (6)$$

$$\text{Variance} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m ((i - \mu)^2 c_{ij} + (j - \mu)^2 c_{ij}) \quad (7)$$

$$\text{Correlation} = \sum_{ij} \frac{(i-\mu)(j-\mu)c_{ij}}{\sqrt{\text{var}(i)\text{var}(j)}} \quad (8)$$

Six features for 15 gray level co-occurrence matrix are computed (totally 90 features) .

1.4 Gray level run length matrix (GLRLM)

We use Six features of gray level run length matrix (GLRLM) [8] (features are SRE, GLNU, LRE, FARC RLNU, RP), GLRLM is the same as GLCM; which can be formed for the following directions : 0,45,90, 135, the mean of these four matrices creates a new matrix (totally 5 matrices) at following distances : d=1 ,5,7 pixels ; totally 15 matrices (5matrices *3distances) will be made .

$$\text{SRE} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \frac{c(i,j)}{j^2} \quad (9)$$

$$\text{LRE} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m j^2 c(i, j) \quad (10)$$

$$\text{GLNU} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m c_{ij}^2 \quad (11)$$

$$\text{RLNU} = \frac{1}{n} \sum_{j=1}^m (\sum_{i=1}^n c_{ij}) \quad (12)$$

$$\text{RP} = n \sum_{i=1}^n \sum_{j=1}^m \frac{1}{j \times c_{ij}} \quad (13)$$

$$\text{FRAC} = \frac{n}{\sum_{i=1}^n \sum_{j=1}^m j c_{ij}} \quad (14)$$

Six features for 15 gray level run length matrix are computed (totally 90 features) .

Totally, we find 182 features from images. After extracting features, we find feature vectors by Touring Ant Colony Optimization (TACO)

2. Finding feature vectors by TACO .

2.1 Ant Colony Optimization

Ant Colony Optimization was first proposed by Dorigo and his colleagues. This algorithm has been created by revelation of ants' behavior in nature. Ants are able to find the shortest path from a food source to their nest [9]; while going from the nest to the food source and vice versa. This ability of ants is a result of pheromone track that they leave from themselves for communication and the result will be quickly discovering good responses and it prevents from hasty and unseasonable converge and presents acceptable responses in searching.

In general, the k^{th} ant moves from state s_i to state s_j with following probability:

$$p_{ij}(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{S_j \in \text{Allowed}} [\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta} & s_j \in A \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

Where τ_{ij} is the amount of pheromone between s_i to s_j , α, β are parameters that have positive amount η_{ij} , is the desirability of state transition s_i, s_j .

The amount of trace by k^{th} ant on the path is shown by following formula:

$$\Delta \tau_{01}^k(t, t+1) = \begin{cases} \frac{Q}{F_k} & \text{if the ant } k \text{ passes path} \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

After the condition that all m ants select their path the amount of pheromone is shown by following formulas:

$$\Delta\tau_{01}(t, t + 1) = \sum_{k=1}^M \Delta\tau_{01}^k(t, t + 1) \quad (17)$$

The pheromone is updated by the following equation:

$$\tau(t + 1) = \rho\tau_{01}^k \Delta\tau_{01}(t, t + 1) \quad (18)$$

ρ is the pheromone evaporation rate ($0 < \rho < 1$), K is the number of ants

2.2 Binary Ant Colony Optimization

A method for resolving continual and Binary discrete, named TACO was presented. two parallel fields of 0 and 1 pass through them and make a binary string, can be seen in fig 2: [10]

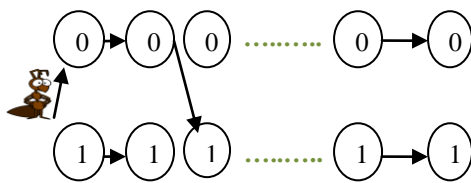


Fig 2 create binary string

Example: 01101000111

For choosing a bit, ants only use the information from pheromone. when ant traverses a complete field this will be a solution for problem; which can be estimated by function cost; then different solutions are calculated this is accompanied by vaporizing a little of previous pheromone and placing new pheromone. in this method the probability of choosing 1st condition for k^{th} ant which is in 0 condition at t^{th} time is shown by following formula:[11]

$$P_{01}^k(t) = \frac{\tau_{01}}{\tau_{01} + \tau_{00}} \quad (19)$$

In this formula τ_{01}, τ_{00} (in sequence) are trace of pheromone on the 0 to 1 path and trace of pheromone on the 0 to 0 path. The amount of trace by k^{th} ant on the path is shown by following formula:

$$\Delta\tau_{01}^k(t, t + 1) = \begin{cases} \frac{Q}{F_k} & \text{if the ant } k \text{ passes path } 0 \rightarrow 1 \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

After the condition that all m ants select their path, the amount of pheromone is shown by following formula:

$$\Delta\tau_{01}(t, t + 1) = \sum_{k=1}^M \Delta\tau_{01}^k(t, t + 1) \quad (21)$$

$$\tau(t + 1) = \rho\tau_{01}^k \Delta\tau_{01}(t, t + 1) \quad (22)$$

In 2004 in reference [12]

A method was presented for improving TACO algorithm. In this method frequencies of repeating all sub-paths are calculated, If a sub-path is selected more than other sub-paths by ants, the probability of that sub-path will be 1.

2.3 Proposed method

We Select features by using TACO algorithm, and save related parameters in SVM classifier. we extract features then obtain a template to use in train step (selecting features). finally, we use parameters to obtain the train step; and obtain percent of detection from test set. (fig 3)

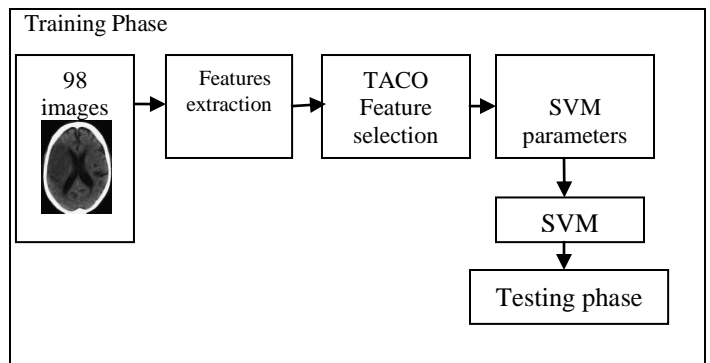


Fig 3 method diagram

IF $(a, b) = (1, 1)$ happens then result is true

And until algorithm is converged, this process will be continued. (fig 4)



Fig4: creating feature vectors

01101000111

2.4 Experiment

Totally we consider 2 fields of 0, 1 like figure 2, that their length is equal to the number of features. Then Ant Colony algorithm is run for finding binary string of 0,1 in the gained field. Existence of 1 expresses acceptability of feature, and existence of 0 expresses elimination of feature.

2.5 Feature selection

Finding a subset of features from a huge set is a challenging problem in fact those features are chosen that have maximum power in predicting output.

We use two classes (n, m) In this method, class n is the number of features without class and class m is the number of selected features. The parameters of the algorithm are defined in the following table 1.

Number of ants :50 ; the pheromone evaporation rate: 0.6 ; number of iteration :200; The Mutation Probability 0.04 .

Table 1: the parameters of the algorithm

Parameter	Values	Parameter description
RHO	0.75	Correlation (my data)
p	500	Size of population
K	50	number of ants
α	1	Pheromone influence
β	1	Control s discovered information
Q	0.04	The Mutation Probability
n	30	The number of features without class
m	3	The number of selected features
na	200	The end iteration
cc	1	The start iteration
counter	1	The first iteration
ρ	0.6	the pheromone evaporation rate

we consider parameters of algorithm as shown in table1. Classes are defined by using following pseudo code . in proposed method data set of test and train are surveyed in separated two steps .in train step , we extract fourteen features . Totally ,we find 182 features of each image. (Table 2).

```

for i=1:na
for j=i+1:na
if (s(i,1:m)==s(j,1:m))
Select = rand perm(n);
s(j,1:m)= Select(1:m);
end;
end;
end ;
data = dataset(my data(:,1:n),mydata(:,n+1));
[Train ,Test]=gen data(data,0.6);
for i=1:m
A = [A ,Train(:,s(j ,i ))];
B = [B ,Test(:,s(j ,i ))];
End
    
```

Table 2: Selecting features by TACO

Feature	Entropy	Contract	Energy	homogeny	variance	correlation	SRE	LRE	GLNU	RLNU	RP	FRAC	Mean	standard division	14
Feature extraction	15	15	15	15	15	15	15	15	15	15	15	15	1	1	182
Selecting Feature By TACO	9	13	8	6	7	8	6	7	9	8	6	8	1	1	97

3. Support vector machine

The structured support vector machine is a machine learning algorithm that generalizes the Support Vector Machine (SVM) classifier . In this step , by having obtained features from previous step and using SVM classifier we do the classification . SVM is a binary classification that separates 2 classes by a linear border. Classifying information based on minimizing the error of test set information is one of the important features of SVM . while in other classifications like neural network the function is based on minimizing the error of educational set information . so in SVM local minimums don't exist. We classify instances by using SVM classification. And We can obtain percentage of classification's accuracy by using " Class pro".

3.1 linear kernel

Given some training data D, a set of N points of the form

$$D=\{(x_i, y_i)|x_i \in R^p, y_i \in \{-1,1\}\}_{i=1}^n(23)$$

$$f(x) = \text{sign}(\langle w, x \rangle + b) \quad , w \in R^p, b \in R(24)$$

where the y_i is either 1 or -1, indicating the class to which the point x_i belongs. Each x_i is a dimensional real vector. We want to find the maximum margin hyper plane that divides the points having $y_i = 1$ from those having $y_i = -1$. Any hyper plane can be written as the set of points x satisfying.

$$W \cdot x - b = 0 \tag{25}$$

These hyper planes can be described by the equations:

$$W \cdot x - b = 1 \tag{26}$$

$$\text{and } w \cdot x - b = -1 \tag{27}$$

By using geometry, we find the distance between these two hyper planes is $\frac{2}{\|w\|}$, so we want to minimize $\|w\|$.

$$y_i(x_i \cdot w_i + b) \geq 1 \quad \text{for } i = 1, 2, \dots, n \tag{28}$$

These margin maximum can be described by the equation:

$$M = \frac{w \cdot (x^+ - x^-)}{\|w\|} = \frac{2}{\|w\|} \tag{29}$$

A linear classifier that is defined by a 2-dimensional weight vector and a threshold value. The decision boundary is perpendicular to the weight vector, offset from the origin by an amount proportional to $\frac{1}{\|w\|}$.

The window displays some options and a plot of a 2-D input space $[-5, 5]^2$ with a blue weight vector and red decision boundary.

Data appears on the plot as '+' signs, either red or blue. (Figure 8)[17]

This method, by using a kernel transfers data to a space in which we can separate data 2 pieces by a hyper plane; than in that data space it finds a hyper plane which does the best separation and has the same distance from 2 categories. we use SVM (linear kernel model). Experimental tests are performed to verify the efficiency of the algorithm for the various values of parameters TP, TN, FP, and FN Result from tables 3, 4.

III. EXPERIMENTAL RESULTS

We use a data set including 98 images. size of images is 256×256 and they are in jpeg format (images are related to brain stroke) in this paper we have used Mean, Standard deviation, Six feature GLCM and GLRLM. In preprocessing stage, images are converted to gray images. 14 dimensions of features are extracted from each image; (totally 182 features) and for classifying images we use SVM (linear kernel model) [17]. 70% of data is for training (68 images) and 30% of data is for testing (30 images). In testing form if brain stroke happens, result is 1; else it will be 0, and will be net result 0 or 1.

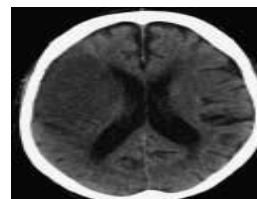


Fig5 input image

fig 5, shows input image and fig 6, Output image. then from run algorithm in matlab 2011. The output image includes more gray levels than input image. (fig 6).

Table3: results of other papers

Method	TP	TN	FN	FP	Accuracy	Sensitivity
ANDRIUS.USINSKAS ET AL.,[1]	85	7	4	2	93.88%	95.51%
Proposed Method Taco/ svm Linear kernel	87	4	3	5	94.90%	96.67%

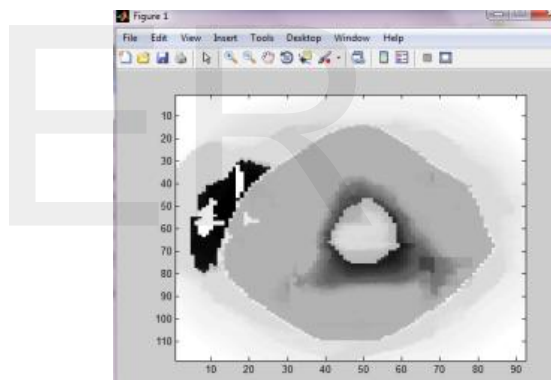


Fig6 output image (Gray level and stroke region)

Fig7 Shows stroke region by threshold procedure.

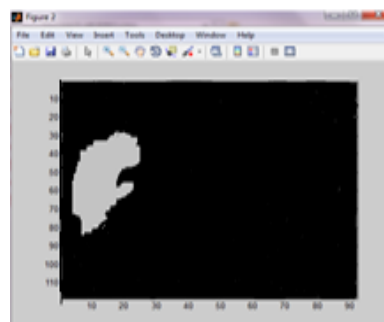


Fig7: segmenting output image by threshold procedure

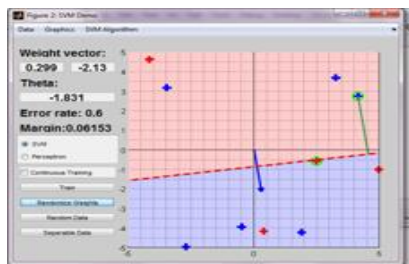


Fig8 classification by SVM

Fig 9 shows histogram of input image:

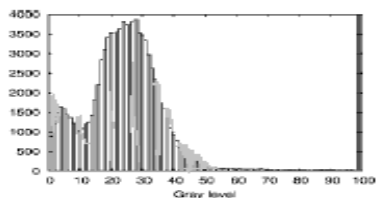


Fig 9 histogram of input image

Fig 10 shows histogram of output image:

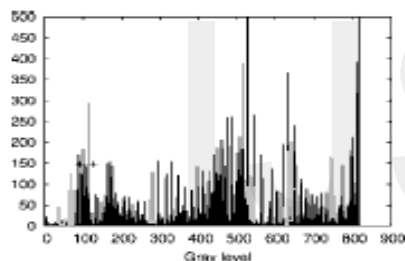


Fig10 histogram of output image

IV. CONCLUSION

Our method is tested successfully on the proposed dynamic optimization problems. We present a method by TACO and we obtain area of brain stroke; the extracted features are so useful and the optimal parameters are chosen for classification. The recognition results of this paper prove that the method we use, combined with SVM classifier can result well. The accuracy was 93.88%, result of our work is the accuracy was 94.90%.

In the future the authors' aim will be extolling the system and increasing the accuracy of diagnosis, the aim is increasing certainty factors for clinical specialist and patients. Our work is a guarantee for system that both clinical specialist and patient can rely on software and its output.

REFERENCES

- [1] Andrius Usinskas, Romualdas A. Brovolskis Ischemic Stroke Segmentation on CT Images Using Joint Features *et al.*, 2004)
- [2] Céline R. Gillebert a, Glyn W. Humphreys a, Dante Mantini a, b, * Automated delineation of stroke lesions using brain CT images, *NeuroImage: Clinical* 4 (2014) 540–548
- [3] Esquevin A¹, Raoult H, Ferré JC, Ronzière T, Stamm A, Perennes M, Bellou A, Gauvrit JY. 2013 Systematic combined noncontrast CT-CT angiography in the management of unexplained nontraumatic coma.
- [4] Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., Karnath, H.O., 2012. Age-specific CT and MRI templates for spatial normalization. *Neuroimage* 61, 957–65. <http://dx.doi.org/10.1016/j.neuroimage.2012.03.020> , 22440645
- [5] Usinskas, A., and V. Mikelaitis. Mean and standard deviation for ischemic stroke segmentation of human, (2008)
- [6] Usinskas, A., and V. Mikelaitis (2002). Mean and standard deviation for ischemic stroke segmentation of human head brain in computed tomography images. In *Proceedings of the 8th Biennial Baltic Electronics Conference BEC 2002*. Tallinn. pp. 371–374.
- [7] Usinskas, A., T. Wittenberg, P. Hastreiter and B.F. Tomandl (2002). Improvements on the gray level cooccurrence matrix technique to compute ischemic stroke volume. In *Bildverarbeitung für die Medizin 2002: Algorithmen-Systeme-Anwendungen, Proceedings des Workshop*. Springer, Leipzig. pp. 267–270.
- [8] Aswini Kumar Mohanty*, Swapnasikta Bebertha**, Saroj Kumar Lenka*** Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram ISSN: 2248-9622 , www.ijera.com Vol. 1, Issue 3, pp.687-693
- [9] M. Dorigo, T. Stutzle, *Ant colony optimization* 2004 MTI Press.
- [10] T. Hiroyasu, M. Miki, Y. Ono and Y. Minami, "Ant colony for continuous functions", *The Science and Engineering Doshisha University*, 2000.
- [11] K. Karaboga, A. Kalinli, and D. Karaboga, "Designing digital IIR filters using ant colony optimization algorithm", *Engineering Application of Artificial Intelligence*, (17), pp. 301-309, 2004.
- [12] M. Dorigo, V. Maniezzo and A. Colomi, "The Ant System: Optimization by a colony of cooperating agents," *IEEE Trans. on Systems, Man, and Cybernetics—Part B*, Vol 2012
- [13] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol.,
- [14] A. Rezaee, "Extracting Edge of Images with Ant Colony," *Journal of Electrical Engineering*, vol.
- [15] J. Tian, W. Yu, and S. Xie, "An Ant Colony Optimization Algorithm for Image Edge Detection," *IEEE Congress on Evolutionary Computation*, 2009.
- [16] Zhaohui L, Xiaoming W, Shengwen G, Binggang Y. Diagnosis of breast cancer tumor based on manifold learning and support vector machine. *Proc. IEEE Int. Conf. Information and Automation* 2008: 703-7.
- [17] Dorigo M., Blum C., *Ant Colony Optimization theory: A Survey*, *Theoretical Computer Science*, 344: 243-278, 2005.
- [18] Eberhart R. C., Kennedy J., *A new optimizer using particle swarm theory*, In *Proceedings of the 6th International Symposium on Micromachine and Human Science*, Nagoya, Japan. pp. 39-43, 1995.

- [19] Fernandes C., Ramos V., Rosa A.C., *Self-Regulated Artificial Ant Colonies on Digital Image Habitats*, International Journal of Lateral Computing, 2(1): 1-8, 2005.
- [20] Guntch M., Middendorf M., *Applying Population Based ACO to Dynamic Optimization Problems*, Proceedings of the 3rd International Workshop ANTS2002, LNCS 2463: 111-122, 2002.
- [21] Meilunas, M., A. Usinskas, R. Kirvaitis and R.A. Dobrovolskis (2003). Automatic countouring of segmented human brain ischemic stroke region of CT images. *Mathematical Modelling and Analysis*, 8(1), 43–50.

IJSER