IJASCSE, Volume 5, Issue 1, 2016

Feature Selection from Brain Stroke CT Images based on Particle Swarm Optimization

Homiera Kabudvand
Faculty of Computer and Information Technology Engineering
Qazvin Branch, Islamic Azad University
Qazvin, Iran

***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 92 features and using heuristic algorithms to cope with complexity of feature selection. Particle swarm 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; Particle Swam Optimization;

I. INTRODUCTION

Suddenly losing function of nerves more than twenty four hours , that causes interruption in blood flow to brain is called brain stroke .[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.Volocity 2.price, 3 availability .[3] 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 big 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 Particle Swarm Optimization based Binary, like most evolutionary algorithms it begins with population, searching is done in a parallel form. Finally classification is tested by SVM. Then we discuss about experimental results.

II. PROPOSED METHODS

Our methods are based on three stages: 1- Preprocessing 2- features extraction 3- finding features by binary Particle Swarm Optimization, 4- classification by SVM.

In proposed method data set of test and train are surveyed in separated step. In train step totally, we find 92 features of each image. Then , we Select features by using NBPSO algorithm ,and save related parameters in SVM classifier .for computing fitness function at test step.(fig 1)

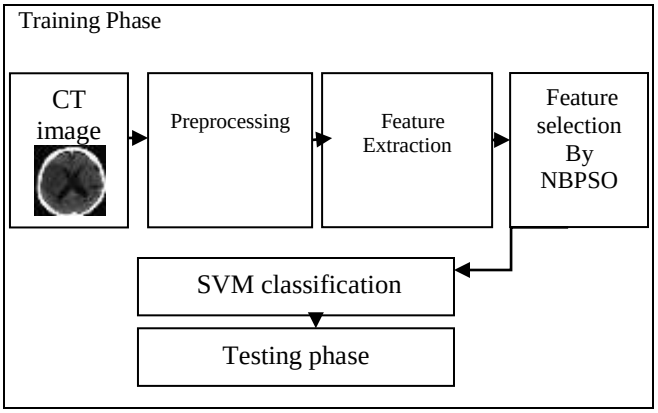


Fig 1 method diagram

1.Preprocessing

1.1 Finding a symmetry

Finding a symmetry of the brain and selecting suspicious points of the CT image where a stroke can be found.

1.2. Surface of skull is determind

In this method first Skull is determined by “Skull-detect “Function. First this function fills empty Pixels of image which don't make a complete area. This work is done by mean technique and “in-fill” Function. Then image is converted to binary image by threshold method. And skull is extracted. But area of skull is not assignable. for determining areas of skull , different areas of image are labeled by Bow -label . number of labels are put in "N" matrix ; by this method image is converted

to "N" areas . After running the following pseudo code, the biggest determined area is considered to be skull.

```
For i=1: N,
  a=sum (sum (l = = i));
    If a>max a,
     Max a=a;
     Max i = i;
   End
End
```

1.3 Image rotation

First we get an image and a random angle from input. then , by using conversion C matrix which is multiplied by primary A matrix the result will be B matrix , new amounts will be made and result is stored in "Z" variable.

B=C×A and Z= NEW AMOUNT

and empty area is created because of rotation; and noise is made . We eliminate the noise by Mean method.

Steps of performance:

1- we find surrounding points of middle pixel which have amounts that are more than zero, by using 3×3 mask .

2- We sum these amounts and then we divide it by number of pixels. finally mean will by obtained .and we put mean in the center of each pixel .

2. Feature extraction

Set of features is very important for recognition.

2.1 Cohesive Rate (CR)

We can obtain CR of suspicious pixels by the following formula: [1]

$$P \epsilon \ T_V = \sum_{l=1}^{TV} 1/DISTANSE(P, Pi) \qquad (1)$$

Value P is set as maximum cohesive rate from Both TV subsets.

2.2 Probability of a stroke

For selected $k \epsilon (0, 1)$ calculate number of pixels U_V which cohesive rate is from range ($k P_{max}$; P_{max}) and are in set TV, where V={R, L}. K is a type of a TPR. A larger k causes a larger probability of a stroke. Stroke is only on one hemispheres of the brain. The probability of a stroke for left and right hemispheres can be calculated by formula (2)

$$PV=UV/(UR+UL) \qquad (2)$$

General probability of a stroke for a series of CT images is defined by a formula:

$$p=\frac{Pavgr-Pavgl}{Pavgr+Pavgl} \qquad (3)$$

Calculating average values of them gives the probability of a stroke for left P_{AVGL} and right P_{AVGR} side of the brain.

2.3 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 . 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).

[15, 16]

$$\text{Entropy}= \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} log c_{ij} \quad (4)$$

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

$$\text{Energy}=\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}^2 \qquad (6)$$

$$\text{Homogeny}= \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{cij}{1+|i-j|} \quad (7)$$

$$\text{Variance}= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left((i-\mu)^2 c_{ij} + (j-\mu)^2 c_{ij}\right) (8)$$

$$\text{Correlation} = \sum_{ij} \frac{(i-\mu)(j-\mu)c_{ij}}{\sqrt{var(i)var(j)}} \qquad (9)$$

Six features for 15 gray level co-occurrence matrix are computed (totally 90 features). table2

3. Finding Features by binary Particle Swarm Optimization

3.1 Particle Swarm Optimization

PSO is a population-based stochastic approach for solving continuous and discrete optimization. It was introduced by Dr. Beernaert and Dry Kennedy in 1995. PSO is a powerful and random method for evolutional calculation based on intelligent transfer of a group of animals for finding food. In comparison to Genetic algorithm , especially in problems containing continuous design variable , PSO has more efficiency in finding Global optimum answers . Some advantages of PSO : implementation becomes easier and we don't need to use gradient inform action . This algorithm follows evolutional calculation method. A. it begins with random population of possible answers. B. it searches optimum answer by updating generations. C. evaluating population is done based on past generations. In PSO, probable answers (particles) are transferred following current optimal particles. this relocation is done under the influence of a fitness function which evaluates quality of each particle . in

short , let the solution space to be D- dimensional , then the ith particle in the swarm is X_i=($X_{i1}, X_{i2}, \ldots . X_{iD}$) the velocity vector is $V_i = (V_{i1}, V_{i2}, \ldots, V_{iD})$, standard PSO is generated by the following two equations:[6]

$$v_{i,k,t} + x_{i,k,t} = x_{i,k,t+1} \quad (10)$$

$$v_{i,k,t} = w.v_{i,k,t} + c_1 u_1(x_{pbest} - x_{i,k,t}) + c_2 u_2(x_{gbest} - x_{i,k,t}) \quad (11)$$

(10) is the velocity update equation and (11) is the position update equation, d=1,2,…D, i =1,2,…Z is the population size ;and C_1, C_2 are positive coefficients , r_1, r_2 are random numbers distributed uniformly in [0,1], in PSO each particle is searching for optimum answers , they are moving with a velocity . the best result is achieved (p best) and information is changed with other particles to recognize the best particle (g best) between the swarm .

3.2 Binary Particle Swarm Optimization

Binary version of PSO was introduced by Dr. Elbe hart and Dry Kennedy in 1997. Unlike standard PSO , it can optimize in discrete spaces By considering a binary field for each particle (fig 2)

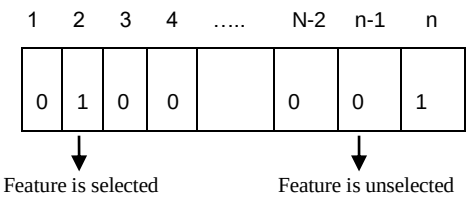


Fig2 a binary field for each particle

BPSO begins in this field; 0 represents elimination of feature and 1 represents selecting a feature . the only difference which is known in comparison with standard PSO is updating equation (12,13) .

$$S(V_{id}^{k+1}) = \frac{1}{1+exp(-v_{id}^{k+1})} \quad (12)$$

IF

$$S(V_{id}^{k+1}) > rand \;\; then \;\; x_{id}^{k+1} = 1 \; else \; x_{id}^{k+1} = 0 \quad (13)$$

Rand is a random number which has been spread the same in [0, 1]. BPSO 's innovators recommend restricting of velocity in [-4,4] for preventing from sigmoid function's saturation . proposed system selects a subset of features which leads to the best result by using BPSO . The other challenging problem is diagnosing optimization of subset. in non-inclusive methods , it should be a balance between smallness or desirability of selected subset . in this problem accuracy of identification is more important than smallness of selected subset. However, if 2 subsets have the same accuracy, we will have the following function:

$$Fitness = \alpha.Acc\text{Uracy} + \beta.\frac{|n|-|s|}{|n|} \quad (14)$$

$|n|$: Number of all features

$|S|$: Number of selected features

First term is a factor that shows the accuracy.

Second term is a factor that shows the rate of feature's reduction.

We consider the sum of α, β constant (100)

3.3 Feature selection

Finding a subset of features from a big 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.

BPSO starts working by some random subsets. The parameters of the algorithm are defined in the following (table1).

BPSO'S introduced features are used in SVM. "accuracy of SVM " and " number of selected features " are evaluated .If we have "m" features , BPSO does 2^m searches for finding optimum answer . Maximum number of repeating is considered as algorithm's ending condition.

Table 1: the parameters of the algorithm

parameter	Parameter description	values
N	the number of the particles	50
K	the number of iteration	1000
VAR	number of variables	5
L	the length for each variable	15
W-min	the inertia factor	0.1
W-max	the inertia factor	0.6
C1=C2	Are learning factor	2
P-best ,G best	-	rand(N,D) >0.5
Fitness	-	0
D	(the dimension of each particle)	L*VAR

Following flowchart shows obtaining selecting template (fig 3).

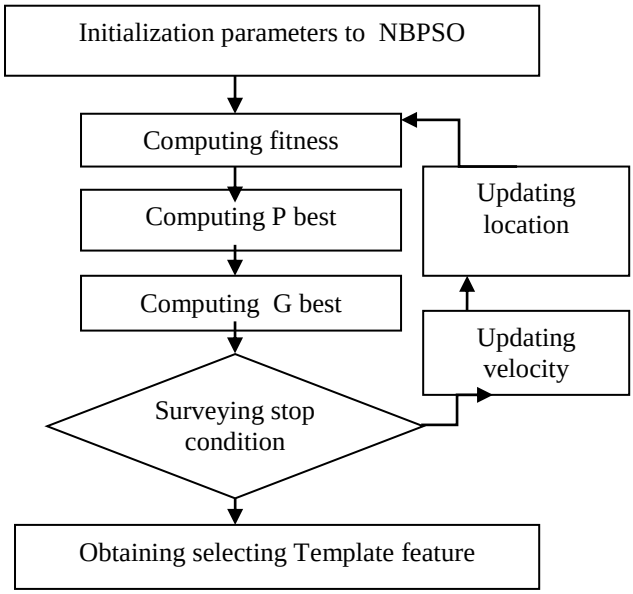


Fig3 obtaining selecting template

Table2: feature selecting by NBPSO

Feature	Feature extraction	Feature selection
Entropy	15	9
contract	15	13
energy	15	8
homogeny	15	6
variance	15	8
correlation	15	8
probability	1	1
Cohesive Rate(CR)	1	1
8 features	92	54

4. 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 . 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 classifier. And we can obtain percentage of classification's accuracy by using Class pro. [14]

4.1 linear kernel

$$x_i^T x_j \qquad (15)$$

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 theta/|w|^2.

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 [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). Result from table 3.

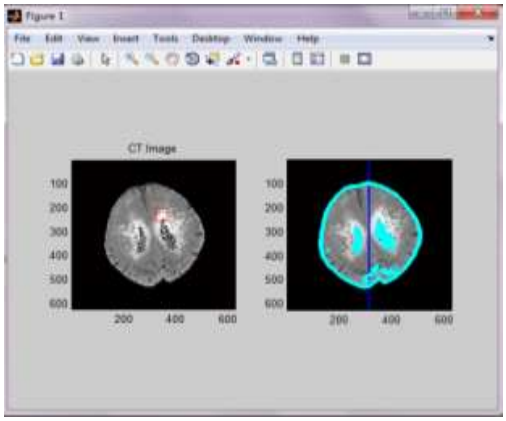
III. EXPERIMENTAL STUDY

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 cohesive rate, probability, Six feature GLCM. In preprocessing stage, images are converted to gray images. 8 dimensions of features are extracted from each image; (totally 92 features) and for classifying images we use SVM (linear kernel model) [17]. 70% of data is for training (68images) and 30% of data is for testing (30 images). In testing form if brain stroke happens, result is 1; else it will be 0,[0,1] , and will be net result 0 or 1 . experimental tests are performed to verify the efficiency of the algorithm for the various values of parameters TP,TN,FP, and FN.
ACC – precision factor,
TPR – sensitivity property,

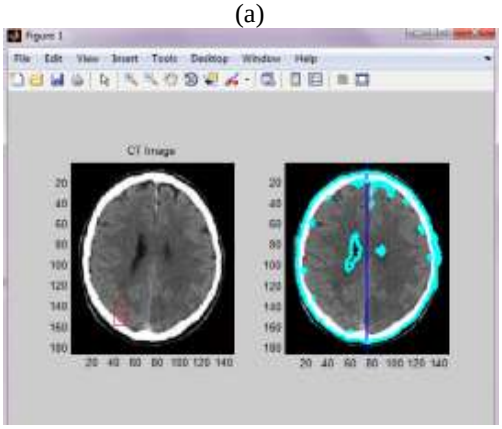
In our work mean isotropy of fitness is improved by increasing the number of repetition in particle swarm optimization algorithm. By NBPSO and defining fitness function , we select features ; and by SVM classifier obtained features classification is done and results are obtained according to table 3 .

Table3: result methods

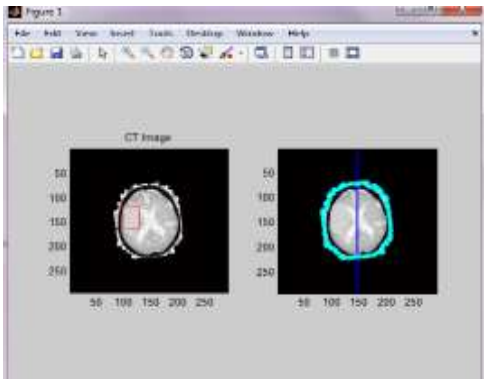
Method	TP	TN	FN	FP	ACC	TPR	SPC
paper [1]	82	8	6	2	91%	94%	80%
Proposed method NBPSO/SVM (linear kernel)	85	6	2	5	92%	97%	54%



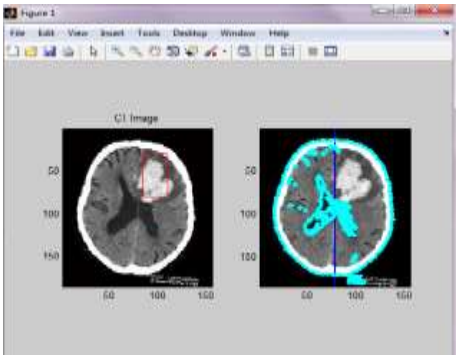
(a)



(b)



(c)



(d)

Fig4 results algorithm(a) image rotation (b) surface of skull is determined (c) ,(d)stroke detection

In fig 5 :drawing mean of fitness based reapetation number for traning data and testing data.

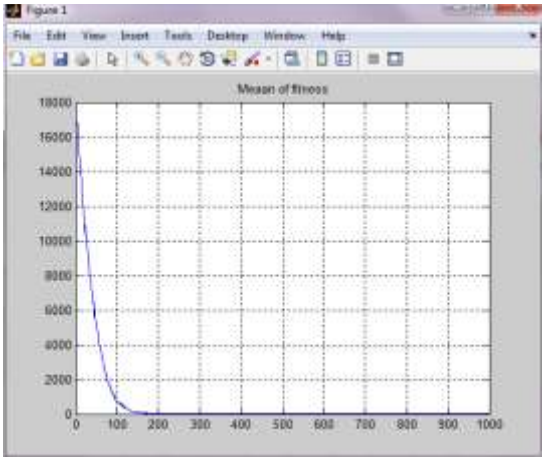


Fig5 mean of fitness

In fig 6 :drawing best of fitness based reapetation number for traning data and testing data

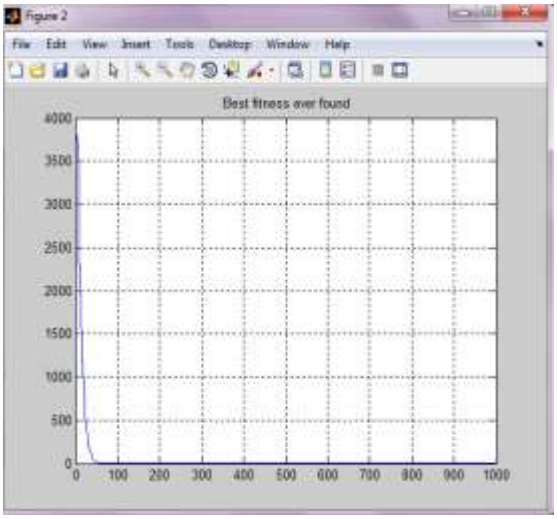


Fig6 best fitness ever found

IV. CONCLUSION

Our model was tested with success on the proposed dynamic optimization problems. We present a method by NBPSO and 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. And results of experiments are presented. The accuracy was 91%, result of our work is the accuracy was 92% . we will change CT images to MRI , FMRI images and will verify another methods for example ant colony , neural network , type methods classification and will compare which is better . In the future will present papers. 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.

REFRENCE

[1] Elżbieta Hudyma, Grzegorz Terlikowski, Computer-aided Detecting of Early Strokes and its Evaluation on the Base of CT Images) Proceedings of the International Multiconference on ISBN 978-83-60810-14-9 Computer Science and Information Technology, pp. 251 – 254 ISSN 1896-7094

[2] C ´line R. Gilbert a, Glyn W. Humphreys a, Dante Martini a, b, * automated delineation of stroke lesions using brain CT images, Neuroimaging: Clinical 4 (2014) 540–548

[3]- Sequin A[1], Raoul H, Ferro 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]. Liu B, Abbass H, McKay B. Classification rule discovery with ant colony optimization.IEEE Computational Intelligence Bulletin 2004; 3(1): 31-5.

[6]. Meng Y. A swarm intelligence based algorithm for proteomic pattern detection of ovarian cancer. IEEE Symp. Computational Intelligence and Bioinformatics and Computational Biology (CIBCB) 2006: 1-7.

[7]. Kennedy J, Eberhart R. Particle swarm optimization. Proc. IEEE Int. conf. Neural Networks 1995: 1942-8.

[8]. Nakamura M. A multi-objective particle swarm optimization incorporating design sensitivities. 11th AIAA/ISSMO multidisciplinary Analysis and Optimization Conf 2006: 70-5.

[9]. Chen Y. A local linear wavelet neural network. Proc. 5th world Cong. Intelligent Control and Automation 2004: 15-9.

[10]. Shi Y, Eberhart R. A modified particle swarm optimizer. Proc. IEEE Int. Conf. Evolutionary Computation 1998: 69-73.

[11]. Eberhart R, Shi Y. Comparing inertia weights and constriction factors in particle swarm optimization. Proc. IEEE Cong. Evolutionary Computations 2000: 84-8.

[12]. Kennedy J, Eberhar t R. A discrete binary version of the particle swarm algorithm. Proc. Conf. Systems, Man, and Cybernetics 1997: 4104-9.

[13]. Franken N, Engelbrecht A. Investigating binary PSO parameter influence on the knights cover problem. IEEE Cong. Evolutionary Computations 2005; 1: 282-9.

[14] 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.

[15] 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 fr die Medizin 2002:Algorithmen-Systeme-Anwendungen, Proceedings des Workshop*. Springer, Leipzig. pp. 267–270.

[16] Aswini Kumar Mohanty*, Swapnasikta Beberta**, 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