JOURNAL OF PURE AND APPLIED MICROBIOLOGY, Aug. 2015. Vol. **9**, p. 215-226
*SPECIAL ISSUE ON RECENT RESEARCH CHALLENGES IN BIO-MEDICAL APPLICATIONS*

# Independent Component Analysis With Genetic Algorithm Feature Selection for Ischemic Stroke Classification

**C. Amutha Devi[1] and S.P. Rajagopalan[2]**

[1]Peri Institute of Technology, Department of Computer Science and Engineering, Chennai, Tamilnadu, India.
[2]GKM college of Engineering and Technology, Department of Computer Science and Engineering, Chennai, Tamilnadu, India.

(Received: 10 March 2015; accepted: 10 May 2015)

**Stroke is one of the principal causes of death, since it has limited treatments. Early detection of stroke at acute stage is essential for the emergency treatment to the stroke affected patients. This work proposes a novel feature selection using evolutionary algorithm and Hybrid Multi-Layer Perception (HMLP) classifier to label the input brain Diffusion Weighted Imaging (DWI) image into Stroke and Non-Stroke. Features are obtained from the Region of Interest (ROI) using Independent Component Analysis (ICA). Itis proposed to use Genetic Algorithm (GA) to choose the best set of features for classification. HMLP classifier is trained using the selected features to label the input images. Theweights of MLP are optimized by the combination of GA and Local Search (LS). The performance of the proposed HMLP classifier is evaluated usingclassification accuracy, precision and recall. Results show improvements in the classification accuracy using the proposed method.**

**Key words:** Cerebral Infraction, Stroke Classification, Independent Component Analysis, MLP Classifier, Genetic Algorithm, Hybrid Optimization.

Infraction is tissue death which occurs by the disturbance of blood flow in the blood carrying vessels and it leads to lack of oxygen and nutrition in the surrounding cells. Cerebral infraction is also called as stroke. Major types of brain stroke are ischemic and hemorrhagic strokes. Ischemic stroke occurs due to the blood clots which either wander or block the arteries. Haemorrhagic stroke happens when there is a sudden burst in the blood vessel and bleeds into the surrounding cells. Around 80% of the strokes are ischemic type. Stroke affected persons may have problems ranging from emotional issues or short term memory loss to severe problems such as coma and death.

Clinically, non-contrast CT images are used to diagnose and treat or exclude the thrombotic strokes and CT images give positive results within 6 to 8 hours after the occurrence of stroke. Perfusion CT images are sensitive than non-contrast CT images in finding stroke at earlier stage (within 6 hours) and can be used for emergency thrombotic treatments. But perfusion and Diffusion Weighted Magnetic Resonance images (DW-MRI) are the highest sensitive images and widely used to identify the ischemia stroke regions at acute stage. Stroke can be detected within very few minutes or hours after the occurrence of stroke[1].

The analysis and recognizing the lesions from the brain images is an expensive, time consuming task and sometimes result in intra and inter observer variability[2]. The accuracy of detection depends on the domain knowledge of the operator. An automated analysis and classification will be helpful to assist the clinicians

* To whom all correspondence should be addressed.
E-mail: camuthadevi.2011@gmail.com

in the medical diagnostic process. Pre-processing and segmentation tasks are useful to maintain only the necessary detail in the images and improve the accuracy of detection. The classification system may use either unsupervised or supervised learning. Unsupervised learning does not use the feedback from the radiologists. Fuzzy C means (FCM), K-means and Expectation Maximization (EM) algorithms are unsupervised classification algorithms widely used for MRI classification. Supervised learning uses the experience from previous case histories to train the classifiers. Support Vector Machines (SVM) and Artificial Neural Networks (ANN) classifiers use supervised learning methods and widely used for MRI classification.

Chawla *et al*[3] proposed an automated method to classify the CT images to detect acute stroke in the non-contrast CT slices. Image enhancement was done for pre-processing and mid-line symmetry was calculated. Windowing operation was used for enhancing the ROI. Features were extracted using wavelet transform. Dataset with 15 patients having 347 slices were taken for experiments. Proposed method gave 90 % of accuracy with 100 % recall to detect the stroke at patient level. In the slice level average precision and recall were achieve by 91 % and 90 % respectively.

Feature Selection is also termed feature subset selection, variable selection, or attribute reduction[4] and aims to choose a subset of input variables by eliminating irrelevant features or of no predictive information. Larger number of features is time costly to classify and the efficiency of the classifier also reduces. Feature Selection proves theoretically and practically to be effective in improving learning efficiency, reducing complexity of learned results and increasing predictive accuracy[5]. Popular feature selection techniques are filter, wrapper and embedded methods.

**Filter method**

In this method, a univariate metric is used to rank the features and the highest ranking features are chosen and the lower ranking features are eliminated. The general characteristics of data is utilized for feature selection without using any learning algorithm and thus, the filter model's[21] result will not affect any classification algorithm. The main advantages of filter methods is that it is computationally less expensive and can be scaled to large data.

**Wrappers Method**

In wrapper method, a subset evaluator uses technique such as Best First Search (BFS), Linear Floating forward Selection to create all possible subsets. A classifier is used to evaluate each feature subset. This method is multi-step procedures testing different combinations of features and is a subject of research presented such as a sequential algorithm with low computational costs, probably being an example of general family of Forward Feature Selection algorithms[22,23].

**Embedded methods**

In this method search mechanisms is built into the classifier model and are specific to a given learning algorithm[24,25].

The optimal feature selection to distinguish between classes is difficult. The evaluation of feature subsets is a tedious task because large amount of computational effort is required. FS problems are NP-hard, ûnding an optimal feature subset is a combinatorial problem. Traditional optimization algorithms offer an attractive approach to find near-optimal solutions to an optimization problem[6].

ICA is a blind source separation method used for extracting the features in many imaging applications. Masood and Brijlal[7] proposed ICA for functional Magnetic Resonance Images (fMRI). ICA in fMRI of each subject was extracted and represented as a features matrix. Then FAST-ICA algorithm was used to aggregate the ICA matrix of each subject by fusion. Chen *et al*[8] proposed ICA extracted features based classifier to classify the mild cognitive impairment (MCI) subjects and Healthy Controls (HC). Information criteria were used to select 9 components and classifier was trained by using the components. The accuracy of proposed classifier was 82.7 % and it was greater than clinical measures. Yang *et al*[9] used ICA to separate the Alzheimer's disease bio markers in the brain MRI. The separated coefficients are given as inputs to the classifier to classify the subjects into diseased subjects and controlled subjects. T1 weighted, T2 weighted, Proton Density Images (PDI), Fluid-Attenuated Inversion Recovery (FLAIR) and DWI brain images are used for abnormality tissue detection. Slice by slice analysis to find the abnormal tissue is a tedious process and ICA gets the global features from the images. ICA fails to collect features from very small lesions and from noisy data. To overcome this, Sindhumol *et al*[10] proposed a Multi resolution Independent Component Analysis (MICA) algorithm for extracting features which was used for micro array classification by Support Vector Machines (SVM). DWIs with 0%, 1% and 3% noise were taken for the experiments. The performance of the SVM classifier was appraised by the parameters such as accuracy, sensitivity and specificity. The accuracy of classifier SVM-MICA was 2.5 times greater than SVM-ICA.

Yang *et al*.,[11] used an ICA based technique for classifying MRI image into Alzheimer's Disease, Mild Cognitive Impairment (MCI), and normal subjects. The proposed method used FastICA for extracting MR image features and classified them using SVM. Jian-Bing Xia-Hou and Kun-Hong Liu[12] proposed classification and regression model called as Penalized Independent Component Regression (P-ICR) by using ICA feature extraction and SVM based classification. Large number of features generated by ICA was used by the GA to select features and to avoid over fitting by using fast convergence stopping criteria. Two classification models were created, one used ICA-GA generated features and another one used only ICA generated features. Experiments were conducted with micro array data and the results proved that classification accuracy using GA based ICA features was better than simple ICA features.

Kong *et al*.,[13] applied ICA on DNA microarray gene expression to extract higher-order statistical structures. The proposed method was used to extract the features and gene clustering. Experimental results showed that the proposed method was able to find significant genes with low expression levels.

In this work, Multi-Layer Perceptron (MLP) Neural Network based classifier is trained to classify the Diffusion Weighted Imaging (DWI) images into stroke and non-stroke classes. Genetic Algorithm (GA) is used to train the network using a novel fitness function.The input features to the proposed MLP NN is obtained using Watershed segmentation and ICA. Once the feature is extracted GA is used to identify the best feature set in the solution space.

## MATERIALS AND METHODS

The proposed system's architecture is shown in Figure 1.

Proposed framework is implemented using the following steps.

1. Median Filter is used for removing impulsive noise in the DWI brain images.
2. To enhance the contrast between different tissues histogram equalization is applied.
3. Watershed segmentation is applied to divide the DWI slices into various ROI.
4. ICA is applied to extract the statistically independent components from ROI.
5. GA is applied to select possible set of features to minimize the objective function.
6. HMLP classifier is trained by using the features selected by step 5. The weights of HMLP classifier are tuned by the combination of GA and LS.

### Pre-Processing Steps

The pre-processing steps consist of median filter and histogram equalization. Median filter is a Gaussian filter belongs to edge preserving and smoothing non-linear filter. It is used to remove impulsive noise in the image. This filter slides a window over the raw input image and performs convolution operation. Simple median filter has the size of 3 x 3. At each position of the window, values of nine pixels are taken and value of center pixel is traded by the median value of these 9 pixels[14]. This operation preserves the sharp and small details when smoothing the regions. Median filter is a central indicator and stronger than average or mean[15].

Histogram represents the intensity distribution of an image graphically. It enumerates the number of pixels for each considered intensity value. Histogram of an image with intensity value in the range [0 – L-1] is represent by

$$H(r_k) = n_k \quad ...(1)$$

Where $r_k$ represents the $k^{th}$ intensity value and $n_k$ represents the pixel's number in the image with the intensity $n_k$. Histograms are normalized based on the probability of occurrence of $r_k$ in the image[16]. If the original image has M x N pixels, then normalization is represented by,

$$P(r_k) = \frac{n_k}{MN} \quad k=0,1,2,\ldots,L-1 \qquad \ldots(2)$$

Histogram equalization is utilized to improve the contrast of the image by mapping the actual intensity distribution to a wider and uniform distribution. Therefore, the intensity values are spread over the entire range of intensities. For simple histogram equalization, cumulative distribution function is used. If H (i) is the histogram of an image, then cumulative function is,

$$H'(i) = \sum_{0 \le j \le i} H(i) \qquad \ldots(3)$$

Then equalization is done by every pixel at the position (x, y),

$$Equalized(x,y) = H'(src(x,y)) \qquad \ldots(4)$$

**Segmentation - Watershed Transform**

Watershed transform is used for the segmentation of images by using morphology. In geography, a watershed is the ridge which divides the areas drained by different river system[17]. The watershed transform is a morphological and gradient-based segmentation technique. The gradient map of the raw image is taken as a relief map in which different gradient values represents to different heights. If water is poured continuously on this map, the level of water will rise over the Catchment Basins (CB) and gets continued until all points in the map are immersed. Lastly the whole image can bedivided into segments called as watersheds and the segmented regions are referred to as CB. A CB is like the geographical area draining into a river or reservoir. These ideas are applied in the watershed algorithm for gray-scale image processing to solve a variety of image segmentation problems.

Different methods are used for using watershed segmentation. Two basic principle methods mentioned are used. 1) Local minima of the image gradient is computed and chosen as a marker. This leads to the problem of over segmentation. After selecting a marker region, merging will be done in the second step. 2) Watershed transformation using markers uses the specifically defined marker positions. These positions are either defined clearly by a user or by

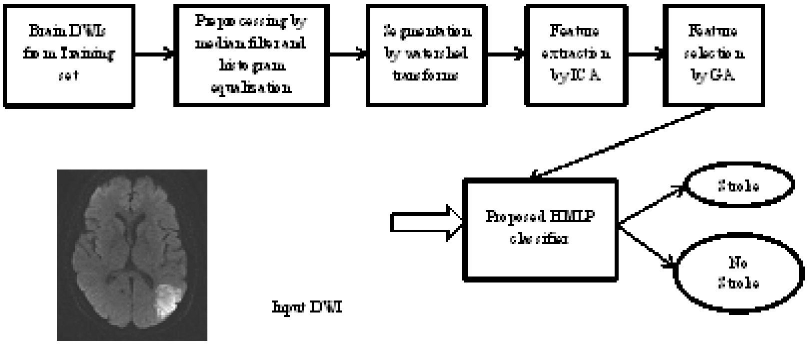


**Fig. 1.** Proposed HMLP Classifier

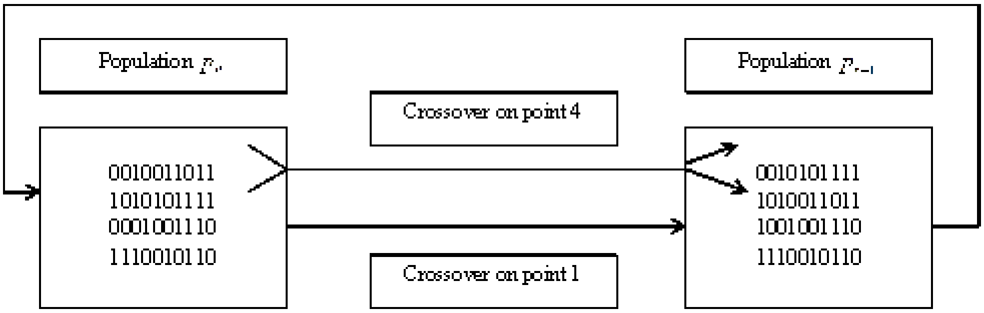


**Fig. 2.** Example of GA operators (Crossover and Mutation) using binary encoding

using morphological tools, they can be determined automatically. 3) Morphological Operators. The purpose of morphological operators is to separate the stroke affected region of the image. Now only the stroke affected portion of the image is visible. Usually, this portion has the highest intensity than other regions of the image[16].

The watershed transform include the following steps:

**Gradient Value Calculation**

Consider let $u(x,y)$ where $(x, y) \in R^2$, be a scalar function which describes an image *I*. The morphological gradient of *I* can be defined as

$$\delta_D u = (u \oplus D) - (u \ominus D) \qquad ...(5)$$

where $(u \oplus D)$ and are the elementary dilation and erosion of u by the structuring element D respectively. The Morphological Laplacian is given as

$$\Delta_D u = (u \oplus D) - 2u + (u \ominus D) \qquad ...(6)$$

**Threshold Method**

Thresholding at level $\lambda$ is defined in two different steps:

1. The set denoted of all points x of such that

$$X_\lambda = \{x \in R^2 : f(x) \leq \lambda\} \qquad ...(7)$$

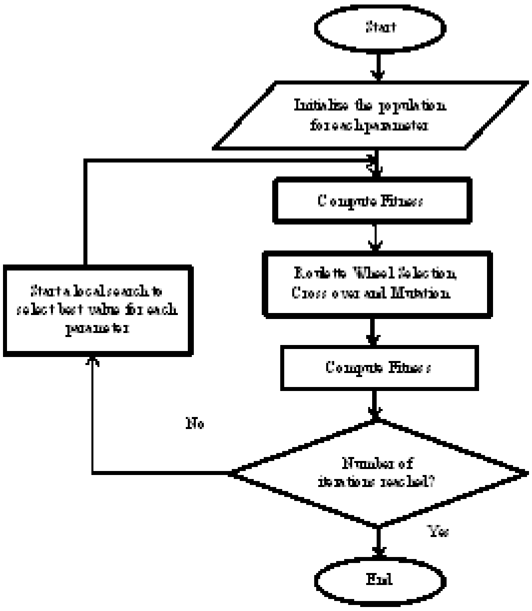


**Fig. 3.** Flow chart of the hybrid optimization

2. The set denoted of all points x of such that

$$Y_\lambda = \{x \in R^2 : f(x) < \lambda\} \qquad ...(8)$$

The family for $0 \leq \lambda$ perfectly defines the function f, so

$$\forall x \in R^2,\ f(x) = \inf(\lambda \mid X_\lambda) \qquad ...(9)$$

**Zones of Influence**

X is considered as a part of $R^2$. The distance between two points x and y of X is defined as the smallest length of the arcs, if they exist, enclosed in X and joining x to y. if there is no arc like that then the which is named as a geodestic distance.

Given a point x of X and a subset y of X, so the geodesic distance between x and y is defined as given in equation (10):

$$d_X(x,y) = \inf_{y \in Y} d_X(x,y) \qquad ...(10)$$

Let Y be the subset of X consisting of n set $k_1$,........$k_n$ and disjoined pairwise:

A zone of influence , consisting of the set of all points of X at a finite distance from and closer to than to any other , can be associated to every . The set of three latter points is named as "Skeleton by zone of influence" of Y with respect to X. it is denoted as S(Y:X) and it is to prove that it is a finite union of simple arcs, locally.

**Feature Extraction**

Independent Component Analysis (ICA), a statistical data processing method, which finds a linear representation of non-Gaussian data so that the components are statistically independent[18]. Datavariables may be linear or non-linear mixtures of some unknown latent variables and the system of mixing is unknown. ICA is based on the hypothesis that latent variables are statistically

**Table 1.** GA Parameters for Feature Selection

| Parameters | Setting |
|---|---|
| Initial Population size | 50 |
| Number of generations | 500 |
| Encoding | Binary |
| Crossover | Two-point |
| Crossover rate | 0.8 |
| Mutation Rate | 0.05 |

independent and non-Gaussian[19]. For a set of observations of random variables where t is the time which are generated from a linear mixture of sources that are statistically independent. This can be expressed as in equation (13):

$$[(x_1(t), x_2(t), \ldots\ldots, x_n(t))]^T = A[(s_1(t), s_2(t), \ldots\ldots, s_n(t))]^T \quad \ldots(13)$$

Where A is unknown mixing matrix and T stands for the transpose operator of a matrix. ICA estimates both A and if only explanations are known.

**Feature Selection**

Feature selection has proven to be effective in enhancing learning, increasing predictive accuracy and reducing complexity of learned results [20]. Feature selection in supervised learning has been well studied, and its goal is to find a feature subset that produces higher classification accuracy. But, the traditional approaches to feature selection with single evaluation criterion have shown only limited capability about knowledge discovery and decision support. In this work, each feature subset is evaluated based on multiple objectives. Solutionis associated with an evaluation vectorwhere $C$ is the number of quality criteria. One solution is said to *dominate* another solution $S_2$ if

$$\forall c : FC(s_1) \geq F_C(s_2) \text{ and } \exists c : F_C(s_1) > F_C(s_2). \quad \ldots(11)$$

where is the $c$-th criterion, . Neither solution dominates the other if

$$\exists c_1, c_2 : F_{C1}(s_1) > F_{C2}(s_2), F_{C2}(s_2) > F_{C2}(s_1) \quad \ldots(12)$$

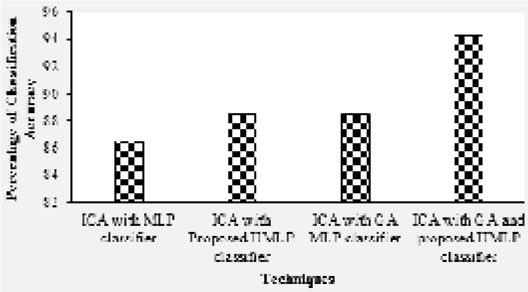


**Fig. 4.** Classification Accuracy

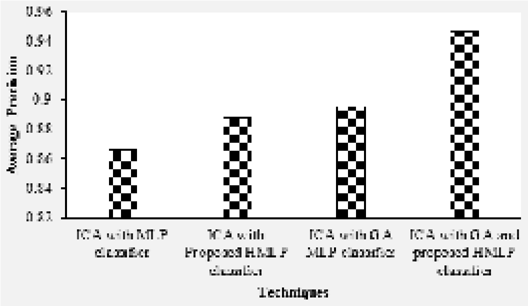


**Fig. 5.** Precision

**Table 2.** Experiment setup

| | |
|---|---|
| Architecture of MLP | |
| Number of neurons in input layer | Number of features selected by GA+LS |
| Number of neurons in hidden layer | Half of the number of neurons in input layer |
| Number of neurons in input layer | 2 |
| Learning rate | 0.1 |
| Momentum | 0.2 |
| Activation Function | Sigmoidal function in both the hidden and output layer |
| Number of epochs | 500 |
| GA parameters | |
| Initial Population | 20 |
| Encoding | Real Encoding |
| Selection Mechanism | Roulette Wheel |
| Cross over | Two point cross over |
| Cross over probability | 0.9 |
| Mutation Probability | 0.001 |
| Minimum value of learning rate | 0.001 |
| Maximum value of learning rate | 0.8 |
| Minimum value of momentum | 0.1 |
| Maximum value of momentum | 0.98 |

### Proposed ICA-GA Feature Selection

Genetic Algorithms (GA) are stochastic optimization algorithms used to search the wide solution spaces and to avoid local minima. Avoiding local minima makes more possibilities of finding an optimal or near-optimal solution. Multiple solutions are handled by the GA simultaneously and with the help of random elements, getting trapped in the suboptimal solutions is avoided. GA operations are:

i. Initial population generation: The initial population is generated randomly which represent the solutions.

ii. Evaluation of fitness: Fitness of each individual is determined ongeneration of the initial population. It is a numeric index which measures the effectiveness of each individuals of population as a solution.The fitness value forms the basis of the selection of members for reproduction.

iii. Selection Operation: A pair of individuals is selected as parents from the current population using tournament selection.

iv. Crossover Operation: A multipoint crossover is applied to parentsselected for generating two offspring.

v. Mutation Operation: Mutation operator is applied randomly to the newly generated offspring in order to prevent from premature convergence into local minima[26].

ICA is a statistical method and GA is an optimization algorithm. These two techniques are combined to improve on the feature selection, the GA-ICA consists of two stages:

- First stage: ICA is applied on the population to find the independent components.
- Second stage: GA evolves the population to find the solution.

A novel fitness function for feature selection is proposed. The proposed fitness functions considers the misclassification rate along with the number of features chosen. As the number of features increase the misclassification rate decreases till a threshold and starts increasing. Equation (14) shows the proposed fitness function.

$$f(c) = \frac{\alpha \times \textit{Number of features}}{(1-\alpha) \times \dfrac{e^{\delta - 1}}{e - 1}} \quad \text{...(14)}$$

Where $\delta$ is given by $\dfrac{\textit{wrongly classified instances}}{\textit{Total number of training instances}}$ and $\alpha$ is a constant between 0 and 1.

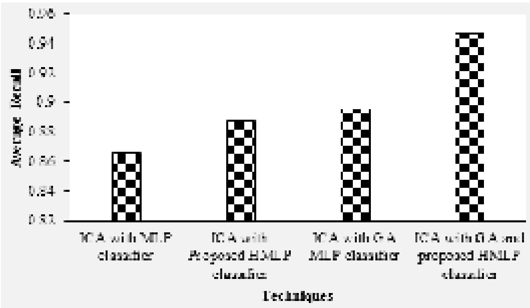

**Fig. 6.** Recall

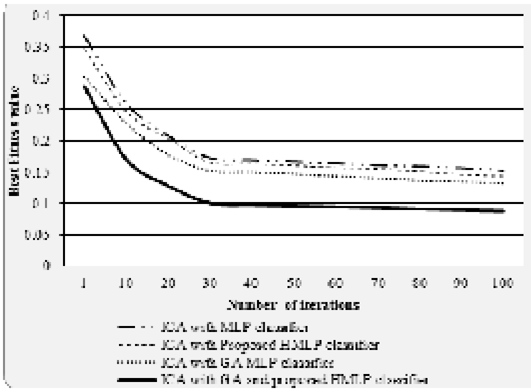

**Fig. 7.** Best Fitness

**Table 3.** Comparison of Performance

| Data set | Author | Methods Used | Classification Accuracy in Percentage |
|---|---|---|---|
| 150 CT images | Renugadevi and Thangaraj | Modified PSO + SVM | 92.67 |
| 91 CT images | Pramod Bhat and Mandeep Singh | SVM, Random Forest, C 4.5, K-NN and CART | 85.39 |
| 52 DWI | Divya and Janet | MLP with GA | 94.23 |
| 52 DWI | Proposed work | GA-ICA + Hybrid MLP | 94.23 |

GA parameters for feature selection are mentioned in table 1.

The features obtained using ICA are encoded into genes and chromosomes as binary values. The '0' represents a feature not being selected and '1' represents the feature as selected. For binary encoding figure 3 shows the crossover and mutation operation.The initial population of N individuals in the GA-ICA is generated randomly in the range of each parameter. The fitness of each individual is determined based on the kurtosis. Tournament selection is used to select pair of individuals for mating.

**Proposed Hybrid MLP classifier**

MLP is a feed forward neural network developed to simulate the automatic learning behavior of human biological neurons. A typical MLP architecture consists of input layer, hidden layer and output layer and each layer has number of neurons connected with adjacent layer of neurons by a weighted directed link. The weights represent the strength of synaptic connections. The weights of the link are learned from the input and output mappings by the error propagation and activation function withBack Propagation (BP) algorithm. The activation function used in the hidden layer and output layer neurons is the following sigmoidal function.

$$f(x) = \frac{1}{1+\exp(-x)} \quad ...(15)$$

Neuronsat the output layer perform summations of inner products between the output layer weights and outcome vector of hidden layer. Then the output values are compared with the actual values to compute the error. Then a backward pass is use to update the weights of output and hidden layer. This process is repeated for much iteration until the error value converged to a small value. Then the learned weights are frozen[27].

Backpropagation (BP), a training algorithm finds a set of weights in a reasonable amount of time[28]. BP is a variation on gradient search that generally uses least-squares optimality. Aim of BP is to calculate the gradient of the error with respect to weights using propagating error backwards through the network for a given input. But some drawbacks of this training algorithm exist such as "scaling problem". This is because backpropagation works well on simple training problems but when the problem complexity, the performance of backpropagation degrades rapidly. The performance degradation occurs at stem because complex spaces have nearly global minima which are sparse among the local minima. Gradient search techniques tend to get trapped at local minima. With a high enough gain (or momentum), backpropagation can made these local minima to escape but without knowing about next one (better or worse)it leaves the current one. When the nearly global minima are well hidden among the local minima, backpropagation can able to end bouncing thatoccurred between local minima without much overall improvement. This leads to slow training [29].The performance of the MLP classifier depends on the ideal weight. In this work it is proposed to use GA to find the optimal weights[31]. Proposed Hybrid MLP classifier uses local search algorithm on the GA, because GA cannot guarantees all the times for global optimal solutions. Therefore, hybrid optimization is applied to select best solution among the solutions created by the GA.

Local Search is a meta heuristic technique and used for solving the hard optimization problems. LS algorithms move from one solution to the next solution by applying some local changes until an optimal solution occurred within the time lapses. Local search is an iterative algorithm that moves from one solution *S* to another *S'* based on some neighborhood structure. Procedure of LS is mentioned in following steps:

i. Initialization: An initial schedule S is chosen and considered as the current solution to compute the value of objective function F(S).
ii. Neighbour Generation: Neighbour S' of the current solution S is selected and F(S') is computed.
iii. Acceptance Test: Testing is done to identify whether move is accepted from S to S' or not. If the move is accepted, then S' replaces S as the current solution; otherwise S is retained as the current solution.
iv. Termination Test: This is done to know whether the algorithm is to be terminated. If it terminates, output the best solution generated; otherwise, return to the neighbour generation step.

Step i explains a starting solution that obtained or it can be specified as a random feature permutation. For several times, if a local search procedure is applied, then it is reasonable to use random initial schedules. To generate a neighbour *S'* in Step ii, a neighborhood structure must be specified. Consider the various types of neighborhoods:

- Transpose neighborhoodoccupies two featurethat is adjacent positions in the sequence can be interchanged.For example (1, 2, 3, 4, 5, 6, 7) - (1, 3, 2, 4, 5, 6, 7),
- Swap neighborhoodinterchange the two arbitrary features such as (1, 2, 3, 4, 5, 6, 7) - (1, 6, 3, 4, 5, 2, 7),
- Insert neighborhoodremoves one feature from its current position and inserted elsewhere such as (1, 2, 3, 4, 5, 6, 7) - (1, 3, 4, 5, 6, 2, 7). Neighbours are generated randomly, systematically, or by some combination of the two approaches.

The proposed GA-LS optimizes the weights of the MLP. The chromosomes are encoded using real encoding where the gene represents the weights of the MLP. The initial population is chosen randomly. The fitness of the population is evaluated and iterated through the algorithm till termination is achieved. A novel fitness function is designed with correlation and mean square error. Equation (16) gives the proposed fitness function f.

$$f = \alpha\beta(1-\alpha)(1-\beta)MSE \quad ...(16)$$

The values of α, β, and γ are calculated by the equations (17 and 18),

$$\alpha = \frac{TP+FN}{N} \quad ...(17)$$

$$\beta = \frac{TP+FP}{N} \quad ...(18)$$

where, TP, FP and FN are the number of True positives, the number of false positives and the number of false negatives achieved in the classification of images respectively. Here, N is the number of instances in the classification.

## EXPERIMENTS AND RESULTS

For the experiments 52 DWI images are collected from the Vijaya Health Centre, India. In this data set, 25 are stroke images and the remaining are non-stroke images. This precise classification was done by the experts before training the proposed classifier. Classifier was trained by the features selected by the proposed method.

Table 2 shows the experimental set up. The four classifiers with the different combinations of ICA, GA and Hybrid MLP are implemented to evaluate the performance. The classification accuracy, precision and recall of the classifiers are shown as graphically from the figure 3 to 5. In figure 6, the best fitness values against the number of iterations are plotted for the several classifiers.

Figure 5 shows the comparisons of classification accuracy with different classifiers. Result graph show that the accuracy of proposed ICA with GA and HMLP classifier increases by 8.89% than ICA with MLP classifier, by 6.52% than ICA with proposed HMLP classifier and by 6.52% than ICA with GAMLP classifier.

Figure 6 shows the comparisons of precision with different classifiers. Result graph show that the precision of proposed ICA with GA and HMLP classifier increases by 9.28% than ICA with MLP classifier, by 6.63% than ICA with proposed HMLP classifier and by 5.68% than ICA with GAMLP classifier.

Figure 7 shows the comparisons of recall with different classifiers. Result graph show that the recall of proposed ICA with GA and HMLP classifier increases by 9.25% than ICA with MLP classifier, by 6.96% than ICA with proposed HMLP classifier and by 7.14% than ICA with GAMLP classifier.

From the figure 8, it is observed that the proposed classifier fits for the classification in earlier number of iterations when comparing to the ICA-MLP, ICA-HMLP and ICA-GA-MLP classifier.

## DISCUSSION

The results of the proposed method are comparable with the works available in the literature. Renugadevi and Thangaraj[31] proposed SVM classifier based on the RBF kernel to classify CT images. Features from the images were extracted by using Coiflet wavelets. The selection of parameters of the RBF kernel function such as Gamma and C were influencing the accuracy of the proposed SVM classifier. Particle Swarm Optimization (PSO) was used to select suitable values of these parameters. To solve the problem of premature convergence and improve the convergence speed GA was used with PSO. 150

CT images were taken for the experiments and the results proved that the classification accuracy of the proposed classifier was much better than non-optimized SVM classifier.

Divya and Janet[32] proposed a segmentation based retrieval of MRI in telemedicine. The preprocessed images were segmented by Haar wavelet transform and the features were extracted by Fast Fourier Transform (FFT). The extracted features were ranked by Information Gain (IG). The top ranked features were inputs to a Genetic Optimized MLP for classification. Also, the same feature extraction, selection and classification steps were done on compressed images. The results proved that compression improved 35 % of the savings in the bandwidth with no reduction in the accuracy of the classification. Amini *et al.*,[33] used c4.5 classification and k-Nearest Neighborhood algorithm for the classification of risk factors of the stroke affected patient data. Experiment results by the factors Accuracy, Precision and Specificity proved that KNN was better than C4.5 algorithm.

Bhat and Mandeep Singh[34] analyzed the performance of different classifiers with textural features extracted from the CT images. 91 CT images with ROI marked by the experts were used for the experiments. Textural features were extracted in the ROI and properties such as mean, variance and Fisher's Discrimination Ration (FDR) for every feature were found for every class. The stages of stroke such as acute, chronic and hemorrhagic were used as the target class. The classifiers such as SVM, C 4.5, K-NN, Random Forest, and CART (classification using regression) were combined and used to classify the CT images. Results prove accuracy of classifier ensemble output was 85.39% and the area under Receiver Operating Characteristic (ROC) was found to be about 93 % for all the classes.

Table 3 shows the comparison of classification accuracy of the proposed method with SVM classifier designed by Renugadevi and Thangaraj [30], classifiers with textural features designed by Pramod Bhat and Mandeep Singh[34] and Genetic Optimized MLP designed by Divya and Janet[32].

## CONCLUSION

In this work, ANN based classifier is trained to classify the DWI images into stroke and non-stroke images. This precise classification done by experts and then was taken for training the proposed classifier. ICA algorithm is used to extract the features from the ROI and GA is used to select best set of features for ANN classifier. The learning parameters of ANN are tuned by the hybrid optimization. From the numerical results it is observed that the classification accuracy, precision and recall of the proposed classifier are improved from 6.52 % to 8.89 % when comparing to the simple MLP and GA-MLP classifiers with extracted features by ICA.

The goal of the current investigation is to verify the performances of the proposed optimized feature selection based on ICA and proposed Hybrid MLP classifier for classifying stroke and non-stroke images. Further investigations are required to check the performance of the proposed methods using Open Access Series of Imaging Studies (OASIS) and Alzheimer's disease Neuro imaging Initiative (ADNI). The proposed method in this study was limited to classifying 2 class problems. Future direction of work is to address multi-class classification of the medical images.

## REFERENCES

1. Mehmet Teksam, Banu Cakýr, Mehmet Coskun "CT perfusion imaging in the early diagnosis of acute stroke", *Diagn Interv Radiol* 2005; **11**: 202-205.
2. W.Wells, W.Grimson, R. Kikinis, F.A. Jolesz, "Adaptive Segmentation of MRI Data", *IEEE Transaction on Medical Imaging*, 1992; **15**(4), pp. 429-442.
3. Chawla, M.; Sharma, S.; Sivaswamy, J.; Kishore, L. T., "A method for automatic detection and classification of stroke from brain CT images," Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, vol., no., pp.3581,3584, 3-6 Sept. 2009.
4. Roselin, R., Thangavel, K., & Velayutham, C. Fuzzy-Rough Feature Selection for Mammogram Classification. *Journal of Electronic Science and Technology*

*(JEST)(International)*, 2011; **9**(2), 124-132.

5. Ramaswami, M., & Bhaskaran, R. A study on feature selection techniques in educational data mining. *arXiv preprint arXiv:0912.3924;* 2009.
6. Kharrat, A., Gasmi, K., Messaoud, M. B., Benamrane, N., & Abid, M. A hybrid approach for automatic classification of brain MRI using genetic algorithm and support vector machine. *Leonardo J. Sci*, 2010; **17**: 71-82.
7. Khan Arjumand Masood and Nagori Meghana Brijlal, " Automated Characterization of Brain Tasks using Fast ICA Feature Extraction Algorithm ", International Conference in Recent Trends in Information Technology and Computer Science (ICRTITCS - 2012).
8. Xinyun Chen, Wenlu Yang and Xudong Huang, " ICA based Classification of MCI vs HC", IEEE Seventh International Conference on Natural Computation, 2011; **3**.
9. Wenlu Yang, Halei Xia, Bin Xia and Lok Ming Lui, " ICA based feature extraction and automatic classification of AD-related MRI data ", Sixth International Conference on Natural computation, 2010; **3**: pages 1261-1265.
10. S. Sindhumol, Anilkumar and Kannan Balakrishnan, " Abnormality Detection from Multispectral Brain MRI using Multi resolution Independent Component Analysis ", International Journal of Signal Processing, *Image Processing and Pattern Recognition,* 2013; **6**(1).
11. Yang, W., Lui, R. L., Gao, J. H., Chan, T. F., Yau, S. T., Sperling, R. A., & Huang, X. Independent component analysis-based classification of Alzheimer's disease MRI data. *Journal of Alzheimer's Disease,* 2011; **24**(4): 775-783.
12. Jian-Bing Xia-Hou and Kun-Hong Liu, "A GA Based Approach to Improving the ICA Based Classification Models for Tumor Classification ", *WSEAS Transactions on Information Science and Applications*, 2011; **1**(8).
13. Kong, W., Mou, X., Liu, Q., Chen, Z., Vanderburg, C. R., Rogers, J. T., & Huang, X. Independent component analysis of Alzheimer's DNA microarray gene expression data. *Molecular neurodegeneration,* 2009; **4**(1): 1-14.
14. Liu, K. An Implementation of the Median Filter and Its Effectiveness on Different Kinds of Images. Thomas Jefferson High School for Science and Technology Computer Systems Lab 2006.
15. Lei, P. E. N. G. Adaptive Median Filtering. In Seminar Report, Machine Vision (Vol. 140).
16. J.C. Russ, "The Image Processing Handbook" , CRC Press, Boca Raton, FL., 1992.
17. Christ, M. J., & Parvathi, R. M. S. Segmentation of medical image using clustering and Watershed algorithms. *American Journal of Applied Sciences,* 2011; **8**(12): 1349.
18. Li, S. Z., Lu, X., Hou, X., Peng, X., and Cheng, Q. Learning multiview face subspaces and facial pose estimation using independent component analysis.Image Processing, *IEEE Transactions on*, 2005; **14**(6): 705-712.
19. Dagher, I., and Nachar, R. Face recognition using IPCA-ICA algorithm. Pattern Analysis and Machine Intelligence, *IEEE Transactions on,* 2006; **28**(6): 996-1000.
20. Ramaswami, M., & Bhaskaran, R. A study on feature selection techniques in educational data mining. *arXiv preprint arXiv:0912.3924;* 2009.
21. Saeys, Y., Inza, I., & Larrañaga, P. A review of feature selection techniques in bioinformatics. *bioinformatics*, 2007; **23**(19): 2507-2517.
22. Pietruszkiewicz, W. Unbiasedness of feature selection by hybrid filtering. *International Journal of Computing*, 2011; **10**.
23. Moosavian, A., Ahmadi, H., & Tabatabaeefar, A. Journal-bearing fault detection based on vibration analysis using feature selection and classification techniques 2012.
24. George, G., & Raj, V. C. Review on feature selection techniques and the impact of SVM for cancer classification using gene expression profile. *arXiv preprint arXiv:1109.1062* ; 2011.
25. Hazra, A., Deb, K., Kundu, S., & Hazra, P. Shape Oriented Feature Selection for Tomato Plant Identification. *International Journal of Computer Applications Technology and Research*, 2013; **2**(4), 449-meta.
26. Acharya, D. P., Panda, G., Mishra, S., and Lakshmi, Y. V. S. (2007, December). Bacteria foraging based independent component analysis. In Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on (Vol. 2, pp. 527-531). IEEE.
27. Khoshhal, J., and Mokarram, M. Model for Prediction of Evapotranspiration Using MLP Neural Network. *International Journal of Environmental Sciences,* 2012; **3**(3).
28. McClelland, J. L., Rumelhart, D. E., & PDP Research Group. Parallel distributed processing. *Explorations in the microstructure of cognition,* 1986; **2**.
29. Montana, D. J., & Davis, L. Training Feedforward Neural Networks Using Genetic Algorithms. In *IJCAI*, 1989; **89**: pp. 762-767.
30. Belavendram. N. Application of Genetic Algorithms for Robust Parameter Optimization. *International Journal of Automative and Mechanical Engineering (IJAME),* 2010; **2**: 211-

220.

31. N.T.Renukadevi and P.Thangaraj, Improvements in RBF Kernel using Evolutionary Algorithm for Support Vector Machine Classifier. *Life Science Journal,* 2013; **10**(7s).
32. Divya Mohandass and J. Janet. Segmentation based Retrieval of Medical MRI Images in Telemedicine, *Journal of Scientific and Industrial Research,* 2012; 72: pp 107-113.
33. Leila Amini, Reza Azarpazhouh, MAohammed Taghi Farazadfar, Sayed Ali Mousavi, Farahnaz Jazaieri, Fariborz Khorvash, Rasul Norouzi and Nofiseh Toghianfar. Prediction and Control of Stroke by Data Mining. *International Journal of Preventive Medicine,* 2013; **4** (Suppl 2), pp 245–249.
34. Pramod Bhat and Mandeep Singh. Classification of Stroke Using Texture Analysis on CT images. *International Journal of Soft Computing and Engineering (IJSCE)* ISSN: 2231-2307, 2012; **2**(3).