Combined Artificial Neural Network and Genetic Algorithm for Cloud Classification

V.Preetha Pallavi[#1], V.Vaithiyanathan, Ph.D[*2]

[#]M.Tech Advance Computing, School of Computing, SASTRA University, Thanjavur, Tamil Nadu, India
[*]Associative Dean, ICT, SASTRA University, Thanjavur, Tamil Nadu, India
[1]preethapallavi@gmail.com
[2]vvn@it.sastra.edu

***Abstract*- Weather forecasting needs multiple features to predict in that cloud classification is the one of the feature. Cloud Classification system through satellite images which are taken from multiple bands. The classification is done by combining the Artificial Neural Network (ANN) and Genetic Algorithm (GA). The spectral ratio value of the image is given as input because it is more efficient to detect the spectral properties of cloud and it enhances the image pixel by pixel. Non convex decision handling and misclassifications of classes were predicts using textural features in the Artificial Neural Network. The optimal weight value is finds using Genetic Algorithm that's combining ANN and GA increases the efficiency of classification and weather prediction. The cloud is separated as eight different classes and the weather temperature is defined using this classification. The combined algorithm of ANN and GA increase the accuracy percentage up to 3% then separate classification algorithm ANN. The overall accuracy of this algorithm is 87.5% for cloud classifications.**

KEYWORDS: Artificial Neural Network, Cloud Classification, Genetic Algorithm, Ratio Image, Weather Forecast.

I. INTRODUCTION

People have to prepare for their day or a vacation so they need to know weather forecasting. Weather forecasting plays a major role in people day today life. To predict weather forecasting we have to know the cloud temperature, cloud type, wind speed, direction, etc. Cloud types are one of the important things for weather forecasting. There are different cloud types which formed in the troposphere, so weather also forms there only. Different clouds means different types of weather so which is easily to forecast the weather. Clouds are classified by their physical properties like temperature, phase and thickness. Cloud classification places a vital role in many atmospheric and environment studies. The images of cloud can be taken from the satellite Metosat-8 SERVIRI MSG-2. The meteorological satellite observes weather changing in large scale. Generally classifications have four steps, pre-processing, training, decision and assessing accuracy. There are several classification algorithms. From simple to neural, thus discuss in the related works. Ratio images of clouds are used to improve accuracy because which enhance the minor differences in surface [1].

Neural Network doesn't have single definitions; it determines complex global behaviour by the simple processing of connections between element parameters and processing element. To produce desire output, this algorithm updates its weight value to strength the connections. Multi-layer Percepton (MLP) is the widely using algorithm which updates weight value using back propagation (BP). For each network the mean square error (MSE) value is calculated finding difference in between the obtained output and the expected target. The BP uses the MSE in gradient descent algorithm (GD) to update weight value, this process done for n iterations. This process will strength the connection and produces expected output. There are several drawbacks in GD which depends on the parameter and MSE. The desired output didn't obtain easily [2]-[4].

Genetic Algorithm produces optimized search output for complex problems. A set of feature vector are defined form raw data pixels plane. The raw data input are taken from the multi spectral image. The finial output plane is produced from supervised classification. Linear chromosomes are used for this GA operation. One or more outputs are produced by the input gene. This process produces the efficient output. Thus if the GA is combined with any other algorithm, it will produce more efficient result then separate process [5], [6].

II. RELATED WORK

The simplest cloud classification approach is classified in the basis of brightness temperature (BT), BT difference and thresholds on reflectance. If the classes have same spectral signatures then this approach cannot fond the cloud difference [7].

The artificial neural network method uses features of BT, reflectance and radiance. This will produce better separation between classes. Increase the speed of ANN probabilistic-NN method is introduced. PNN implemented in SVD principal component analysis, singular value decomposition (SVD) and modified. SVD reveals the spatial and spectral features of clouds, it measures pixels over a block, and this decreases the separation of classes' accuracy, if clouds are heterogeneous [8]. The MSVD method overcomes this error and the error can't reach the pixel level. PCA has the capability of reducing the spectral data dimension. PCA is data dependent, so they are unstable when differentiating the cloud type and difficult to interrupt [9].

The superior algorithm for machine learning is support vector machine (SVM). SVM separates the classes using boundaries. This algorithm locates the optimal boundaries.

There are several algorithms to detect the information from images such as skeletons, contours and edges etc. To define properties of an image, fuzzy uses different stochastic relationships. The possible different types of stochastic properties are measured by fuzzy. Fuzzy measure is used when the region is more related to the fuzzy property. It function particularly describes the distribution of gray values. The fusion of fuzzy measures and fuzzy function is fuzzy integral i.e. fusion of two stochastic properties [10].

The combination of fuzzy and SVM are proposed to overcome the misclassification regions in multiclass. The n-class problem of training in SVM was converted into two n classes' problem. The ability is maximized by the decision function. The quadratic optimal separating hyper plane separates two classes to solve optimization problem [11].

From maximum likelihood to neural network technique all depends on intensity value of spectral channel for the pixel for each one. The spatial content information also is added into the feature vector of pixels. There n number of feature vectors which makes classification better. Genetic algorithm is used to choose thus feature vector automatically [12].

The texture based image classification increase the accuracy the NN [13]. The kernel type and parameter which affects the decision boundaries and this would reduce the performance of SVM. The fuzzy overcomes the problem of SVM by its generalization ability. The combination of GA with any other classifier method improves the performance and accuracy rate [14], [15].

III. CLOUD SPECTRAL PROPERTIES

The spectral properties are calculated using the reflectance and emission of the clouds. The content of the cloud like water, ice or water or ice is refers the phase property of spectral. The radiation transparency refers the thickness of the cloud. The cloud height refers the temperature of the clouds. The thin clouds transmit more and reflect less, thick clouds reflect more and transmit less. The height of cloud increases vice versa the coolness of the cloud also increases. There are three types of clouds they are high, low and middle level clouds. Stratus, cumulus type cloud having water content, they are low level clouds. Altocumulus type cloud also having water content but thicker than stratus and cumulus, it is middle level cloud. Cirrus stratus and cirrus cumulus type cloud are having ice content, which are high level cloud. The variation in cloud properties are reveals by the visible (VIS), infrared (IR), near-infrared (NIR), thermal infrared (TIR) channels. The thickness of the cloud is reveal by VIS channel. Thin cloud looks darker than then the thin clouds. The phase property of the cloud is reveal by the NIR channel. Here the water cloud appears brighter than ice clouds. The height of the cloud is reveals the IR channel. The low cloud looks darker than middle and high clouds in IR invert channel [16], [17], [18].

IV. PROPOSED METHOD

The proposed block diagram is shown in below Fig.1. Modules of the proposed system are band ratio images, identifying cloud types, Genetic algorithm training, back propagation (BP) training. The following sections describe these modules.

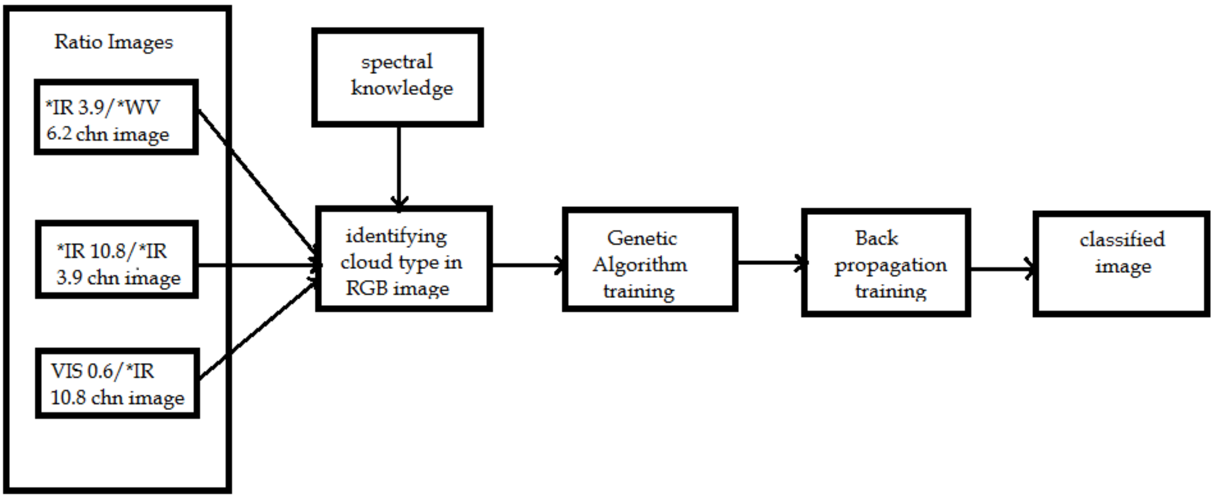

Fig.1. Proposed Method for Classification

A. BAND RATIO IMAGE

Band ratio image was calculated by dividing images of two, which are taken from different channels. The image is divided on the basis of pixel by pixel. It will produce a floating point value of pixels. An enhancement method is applied to produce colour image. The division of two images from different channel reveals the minor changes in spectral property of clouds. Here, four different channel images are taken which differentiate the cloud properties clearly. They are IR 3.9, VIS 0.6, WV 6.2, and IR 10.8. The * symbol represents the invert of the channel [8].

B. IDENTIFYING CLOUD TYPES

On the combination of three ratio images an RGB colour image is formed. This will differentiate cloudy and non cloudy area by visually. Each colour in the image represents the different types of cloud. Select the different colours of cloud images by interrupting on the image to pass the different colour cloud image for classification training. This will reduce the timing taken for classification. The whole image takes more time to classify.

C. NEURAL NETWORK TRAINING

The three layer back propagation feed forward neural network method is used for training. Here, p inputs, q outputs and g hidden nodes. Sigmoid function f(x) is activation function for both of layers hidden and output

$$H_h = f(W^T X) = f\left(\sum_{i=1}^{p} w_i x_i - \delta_h\right) \quad (1)$$

$$O_u = f(V^T H) = f\left(\sum_{i=1}^{q} v_i H_i - \delta_u\right) \quad (2)$$

Where, T represents transpose of the vector. The output neuron of hidden layer is h ($1\leq h\leq g$) and the output neuron of output layer is u ($1\leq u\leq q$). W = $[w_1, w_2, \dots w_i, \dots w_p]$, V = $[v_1, v_2, \dots v_i, \dots v_q]$ are weights for input and hidden layers, X = $[x_1, x_2, \dots x_i, \dots x_p]$ is the input vector for training, Z = $[z_1, z_2, \dots z_i, \dots z_p]$ is the hidden node output. δ_h *and* δ_u are the bias of the hidden and output nodes. The equation for sigmoid function is

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

The mean square error (ε) is calculated by

$$\in (net) = \frac{1}{n.k} \sum_{i=1}^{n} \sum_{j=1}^{k} (t_{ij} - O_{ij})^2 \quad (4)$$

Where t is the targeted output i.e. T= $\{t_1, t_2, \dots t_j, \dots t_k\}$ are the expected output samples. O is the obtained output from the NN training. If the O_{ij} is represents the t_{ij} class then the MSE is calculated for this two vectors [19].

D. GENETIC ALGORITHM

The weight vector training of neural network is done by the genetic algorithm (GA). Here there are four steps to evaluate the weight connections [20]. The steps are

1. **Gene production:** Each weight connection and bias vectors are considered as chromosome genes in the GA. Real coded genotype is used for gene productions for calculation [21].
2. **Evaluation process:** The neural network MLP was evaluates the weight connections and calculates its mean square error (MSE) using the above equations (1), (2), (3), (4). The high error is proportional to the low fitness value. To increase fitness of weight connection is done by the equation

$$\epsilon^*(\mathrm{net_i}) = \frac{\epsilon(\mathrm{net_i} - \min(\varepsilon(\mathrm{net_i}))}{\max(\epsilon(\mathrm{net_i})) - \min(\epsilon(\mathrm{net_i}))} \quad (5)$$

$$\delta(\mathrm{net_i}) = e^{-\psi - \epsilon^*(\mathrm{net_i})} \quad (6)$$

Where δ is the fitness function, ψ is the positive constant. $\varepsilon(\mathrm{net_i})$ is the normalization of MSE in each MLP.

3. **Selection Process:** Roulette wheel selection scheme is used for parent chromosome selection. The selection process takes the current generation, and selects the fitness value from it and reproduces it new generation.
4. **Crossover and Mutation process:** The selected chromosomes applied for the next operation of mutation and crossover which produces next generation of weight value.

***Crossover operator*:** There are two types of crossover operator is applied to produce new chromosome offspring. They are asexual and arithmetical crossover operation. In asexual, the first best 10% of the chromosome does not change. In arithmetical, the following process have take place.

$B_i^k = (b_1^k, \dots b_i^k, \dots b_n^k)$, where k=1, 2.

$$\mathrm{b_i^1} = \lambda \mathrm{c_i^1} + (1-\lambda)\mathrm{c_i^2}$$
$$\mathrm{b_i^2} = \lambda \mathrm{c_i^2} + (1-\lambda)\mathrm{c_i^1} \quad (7)$$

Where B represents the new offspring chromosomes, c represents each gene and λ is positive constant, set as 0.28.

***Mutation operator*:** Here, also there are two types of mutation operator is applied single point mutation and non uniform mutation operation. In single point, randomly one chromosome is chosen and applies mutation operation for that gene. $\mathrm{c_i^1}$ gene value only change in this operation. In non uniform mutation operation the equation is

$$\mathrm{c_i^1} = \begin{cases} \mathrm{c_i} + \Delta(\mathrm{t}, \mathrm{d_i} - \mathrm{c_i}) & \text{if } \tau = 0 \\ \mathrm{c_i} - \Delta(\mathrm{t}, \mathrm{c_i} - \mathrm{a_i}) & \text{if } \tau = 1 \end{cases}$$

$$\Delta(\mathrm{t}, \mathrm{y}) = \mathrm{y}\left(1 - \mathrm{r}^{\left(1 - \frac{\mathrm{t}}{\mathrm{g_{max}}}\right)^{\mathrm{d}}}\right) \quad (8)$$

Where τ is a binary number is randomly chosen, d is set as 0.5, g_{max} is the maximum generation, [0, y] will be the value given by the function.

These steps increase the weight fitness value and reduce the MSE error. The method produces the efficient result in finding the result. The genetic algorithm finds the better weight value for MLP neural network.

E. CONJUGATE GRADIENT ALGORITHM FOR WEIGHT CONNECTIONS OPTIMIZATION

GA can find near optimal set of connections without use gradient algorithm. But some time it does not find directly so here incorporate the conjugate gradient method also used to find the optimal value of weight. Finally an optimal weight value and efficient result will produce with low MSE.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A set of metosat-8 MSG2-SEVI satellite images including IR3.9, IR10.8, VIS0.6 and WV6.2 channels is used for training in Fig.2. The clouds classified as 8 different types they are high ice, very cold water, middle water and low water clouds, and thin clouds over sea, high water vapour, sea, cold land surface, hot land surface. To differentiate land surface and its weather condition using differentiate cloud is the aim of this system.

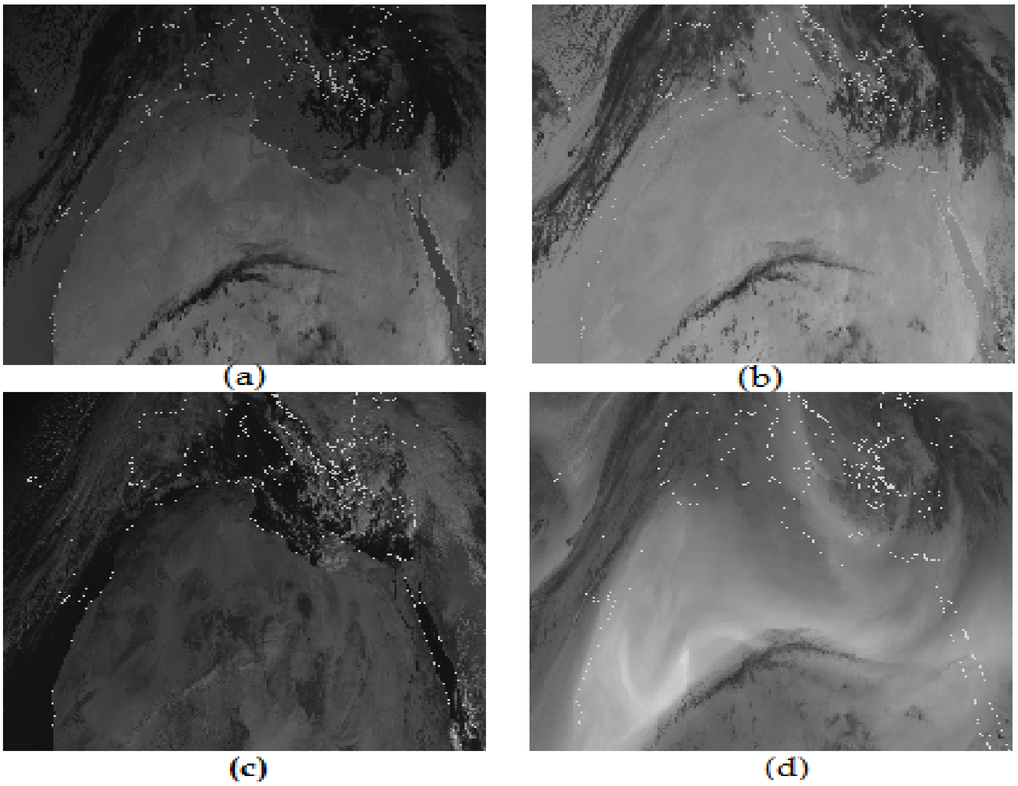

Fig.2. Four MSG channels (a) IR3.9 channel, (b) IR10.8 channel, (c) VIS0.6 channel, (d) WV6.2 channel.

The three band ratio image forms the RGB colour image in Fig.3. The band ratio images are VIS0.6/*IR10.8, *IR10.8/*IR3.9, *IR3.9/*WV6.2 the * represents the invert of the image. Here, visually interpret with this image and select the clouds by different colours. Interactively select the training data and save it for training. The selected data is sends for training. Here, a MLP three layer BP-NN with consisting of hidden layer of one having 10 hidden nodes. In GA, using two generations and each generation having population size 100. The probability of asexual reproduction is 0.1, 0.8 set as arithmetical crossover probability. 0.05 is set for non uniform mutation and single point random mutation probability. - 5.0 are set for a_i and 5.0 are set for b_i. The maximum generation is set as 300. The learning rate for is BP is set as 0.01, 1.03 is set for incremental of learning rate.

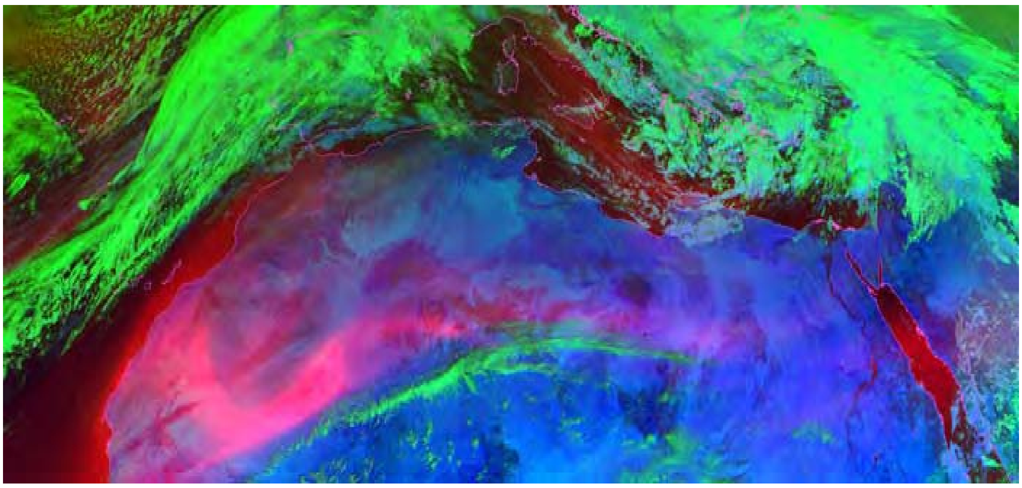

Fig.3.RGB Band Ratio Image

The three band ratio images help to select the input by visual interprets. The gray level of the band ratio is shown in below table-1. The range is slated as low (L), in between low and middle (LM), middle (M), in between middle and high (MH), high (H). This type slated input selection improves the accuracy of selection.

After the selection of samples it was store as an input data. Here 3200 samples are taken from different class. These samples are sending to the MLP NN and the weight vector is update using the GA algorithm. The result of the output classes are differentiate using colour different is shown in the below fig.4.

TABLE-I
NUMERICAL RANGE OF PIXELS FOR BAND RATIO GRADE

GRAY GRADE	L	LM	M	MH	H
Numerical Range	0-50	51-100	101-140	141-180	181-255

TABLE-II
INTERPRETATION RESULT

COLOR	SCENE TYPE(ABBREVATION)	*IR3.9/*WV6.2 (RED)	*IR10.8/*IR3.9 (GREEN)	VIS0.6/*IR10.8 (BLUE)
Red	High Ice clouds(HIC)	M	MH	LM
Light Green	Very Cold Water Clouds(VCWC)	M	MH	M
Light Blue	Middle Water Clouds(MWC)	M	M	M
Dark Blue	Low Water Clouds(LWC)	MH	LM	MH
Brown	Thin Clouds over Sea(TCOS)	H	LM	L
Dark Green	Areas of High Water Vapour(AHWV)	L	L	L
Pink	Sea Surface(SS)	MH	LM	L
Yellow	Land Surface(LS)	L	M	MH

VI. COMPARISON

The combined ANN and GA algorithm is efficient than the other individual algorithms. Here, comparing the BP-MLP and GA-MLP. The percentage of efficiency in GA is higher than the percentage of BP. The comparison table of confusion matrix is given below in Table-3. The table having eight different classes classified pixels.

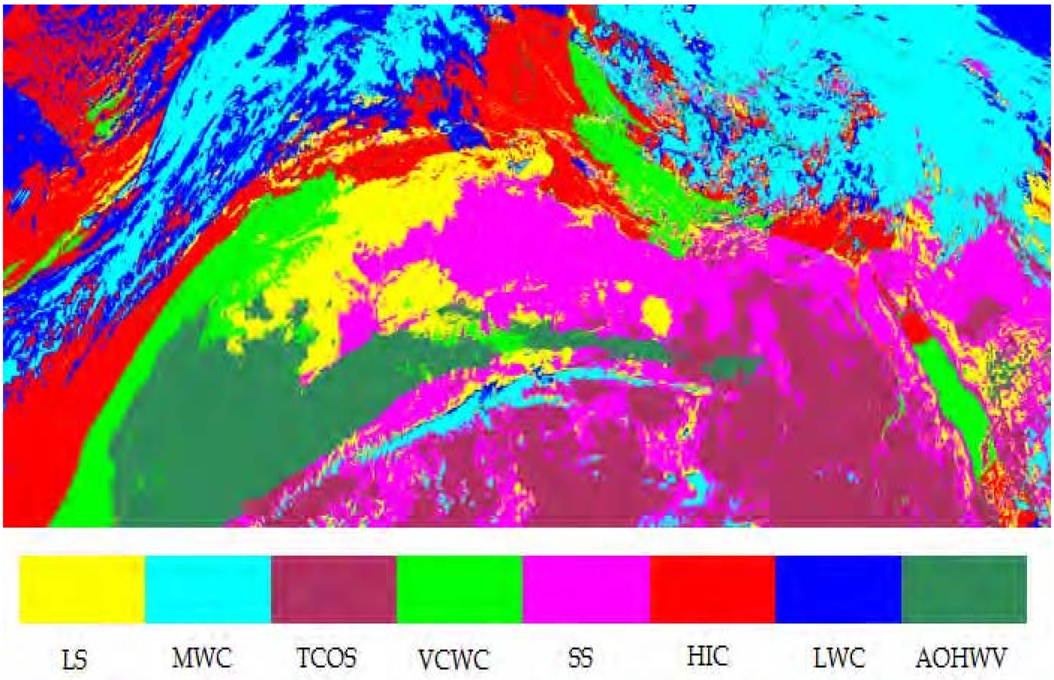

Fig.4. Result of the cloud classification

The MSE error reducing and its corresponding performances are compared between the BP-MLP and GA-MLP. The plot of MSE error value and the number iteration are shown in below fig.5. The MSE value reduces in less iteration in GA-MLP but in BP-MLP it takes large number of iterations.

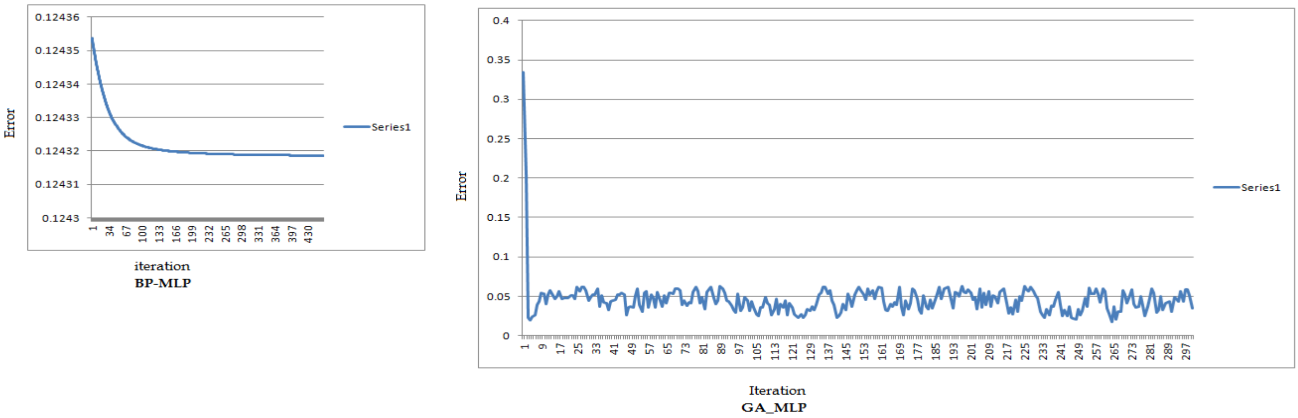

Fig.5. Training Performance

TABLE-3
CONFUSION MATRIX

	HIC	VCWC	MWC	LWC	TCOS	AHWV	SS	LS
HIC	367	0	0	15	0	13	5	0
VCWC	0	400	0	0	0	0	0	0
MWC	30	0	350	0	20	0	0	0
LWC	0	0	12	378	0	0	0	10
TCOS	0	0	0	0	400	0	0	0
AHWV	0	0	0	13	17	361	9	0
SS	4	0	0	5	2	0	389	0
LS	0	5	0	0	4	0	0	391

BP-MLP

	HIC	VCWC	MWC	LWC	TCOS	AHWV	SS	LS
HIC	400	0	5	0	0	0	0	0
VCWC	0	385	0	0	10	0	0	0
MWC	0	0	400	6	0	0	8	10
LWC	0	0	11	400	0	0	0	0
TCOS	0	0	0	0	370	0	0	0
AHWV	0	0	0	0	4	382	0	3
SS	9	0	0	0	0	0	391	0
LS	0	0	0	0	0	0	0	400

GA-MLP

VI. CONCULSION

The proposed work, Genetic algorithm evolves the back propagation neural network for classification. The permutation problem and convergence problem are avoided by optimizing the weight value of neural network by GA. The combined of GA and back propagation NN increases the efficiency of the classification of clouds. The ratio image method reveals minor variations in the spectral properties. This will increase the efficiency of finding the spectral properties. The conjugate gradient algorithm increases the performance of the GA. The drawback of this method is takes more time for processing. The combination of other methods with this GA and finding its optimization and reduce the processing time will be the future work.

VII. REFERENCES

[1] Ahmed AK. Tahir, A System Based on Ratio Images and Quick Probabilistic Neural Network for Continuous Cloud Classification, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 49, NO. 12, DECEMBER 2011.

[2] Benediktsson, J. A., Swain, P. H., Ersoy, O. K., 1990. Neural Network approaches versus statistical methods in classification of multisource remote sensing data. IEEE Transaction on Geoscience and Remote Sensing, 28(4): 540-552.

[3] J. Anderson, “An Introduction to Neural Network”.

[4] MathWorks, Inc., 1997. Neural Network Toolbox User’s Guide. MathWorks, Inc., Natick, MA.

[5] D.E Goldberg, *Genetic Algorithms in Search, Optimization*, and Machine Learning. Addison-Wesley, New York, 1989.

[6] Hancock P. J. B., 1992. Genetic algorithms and permutation problems: a comparison of recombination operators for neural net structure specification. Proceedings of the Int. Workshop on Combinations of Genetic Algorithms and Neural Networks (COGANN-92), pp. 108-122. IEEE Computer Society Press, Los Alamitos, CA.

[7] B. Huang, S. A. Ackerman, and W. P.Menzel, “A new method for MODIS cloud classification,” in *Proc. SPIE*, 2005, vol. 5890, pp. 210–216.

[8] R. Kaur and A. Ganju, “Cloud classification in NOAA AVHRR imageries using spectral and textural features,” *J. Indian Soc. Remote Sens.*, vol. 36, no. 2, pp. 167–174, Jun. 2008.

[9] 1687-398XI. S. Bajwa, M. N. Shaheed, A. M. Nadim, and H. S. Irfan, “Feature based image classification by using principal component analysis,” *ICGST-GVIP J.*, vol. 9, no. 11, pp. 11–17, 2009.

[10] H. Hetzheim, “Anwendung der Fuzzy-Logic in der Bildverarbeitung,” Vision Jahrbuch’ 93, Sprechsaal Pub., Coburg, pp. 47-53, 1993.

[11] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition”, *Data Mining and Knowledge Discovery*, **2**, 1998, pp. 121–167.

[12] Michalewicz Z., 1992. Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag, New York.

[13] Hertz J., Krogh A., Palmer R., 1991. A Introduction to the Theory of Neural Computation. Addison-Wesley, Readings, CA.

[14] Pal S. K., Bandyopadhyay S., Murthy C. A., 2001. Genetic classifier for remotely sensed images: comparison with standard methods. International Journal of Remote Sensing, 22(13): 2545-2569.

[15] Zhengjun Liu, Changyao Wang, Aixia Liu, Zheng Niu. Evolving Neural Network Using Real Coded Genetic Algorithm (GA) for Multispectral Image Classification, LARSIS, the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101, China.

[16] J. Li, W. Paul Menzel, Z. Yang, R. A. Frey, and S. A. Ackerman, “Highspatial- resolution surface and cloud-type classification from MODIS multispectral band measurements,” *J. Appl. Meteorol.*, vol. 42, pp. 204–226, 2003.

[17] J. Kerkmann, Applications of Meteosat Second Generation (MSG). RGB Images: Part 1, Overview MSG SEVIRI Channels version 1.0, 2004. (Accessed 15 Feb 2009). [Online]. Available: http://oiswww.eumetsat.org/WEBOPS/msg_interpretation/PowerPoints/ Channels/rgbpart01_20040630.ppt

[18] V. Zwatz-Meise, Introduction Into Absorption Channels, Description of Characteristics and Content of WV Channel, 2004. (Accessed 18 March 2009). [Online]. Available: http://oiswww.eumetsat.org/WEBOPS/msg_ interpretation/msg_channels.php

[19] Rumelhart, D. E., Hinton, G. E., and Williams, R.J., 1986. Learning representations by back propagating errors. Nature, 323: 533-536.

[20] R. K. Sundaram, *A First Course in Optimization Ttheory*, (New York: Cambridge University Press), 1996.

[21] Herrera F., Lozano M., Verdegay J. L., Tackling Real-Coded Genetic Algorithms: Operators and Tools for Behavioural Analysis. NEC Research Index. http://citeseer.nj.nec.com/