Hybrid Emotional Neural Network for Facial Expression Classification

K.V. Krishna Kishore
Department of C S E,
Vignan University,
Guntur, A.P, India.

G.P.S.Varma
Department of IT,
S.R.K.R. Engineering College,
Bhimavaram, A.P, India

ABSTRACT

This paper describes a novel Hybrid Emotional Neural Network (HENN) for classification of emotions from Facial expressions. The novelty of this work is that along with the parameters of the feed forward neural network, i.e. the learning rate, momentum, new parameters such as anxiety and confidence is taken as emotional parameters are from Gabor Wavelet and are used to update emotional parameters of the network. An improved Back propagation algorithm is used for training of the proposed neural network. Features are extracted from facial expressions by applying Gabor wavelet and Discrete Cosine Transform (DCT). Both the feature sets are high dimensional so Principle Component Analysis (PCA) are used to reduce the dimensionality of features. Then Wavelet fusion technique is used to fuse the features. The fused features are used to train the neural network. The classification efficiency of the proposed method was tested on static images from the Cohn-Kanade database. The results of the proposed network were compared with the standard Feed Forward Neural Network and Radial Basis Neural Network. We also make a detailed comparison of different fusion techniques along with wavelet fusion, as well as different Neural Network classifiers. Extensive experimental results verify the effectiveness of our approach outperforms most of the approaches.

Keywords

Hybrid Emotion Neural Network (HENN), Gabor Wavelet, DCT, PCA, Wavelet fusion, RBF, FFNN.

1. INTRODUCTION

Human beings possess and express emotions in everyday interactions with others. Emotions are often reflected on the face, and all agreed definition of emotion does not exist, it is undeniable that emotions are an integral part of our existence, as one smiles to show greeting, frowns when confused, or raises one's voice when enraged. The fact is to understand emotions and know how to react to other people's expressions greatly enriches the interaction. There is a growing amount of evidence shows that emotional skills are part of what is called “intelligence". Computers today, on the other hand, are still quite “emotionally challenged.” They neither recognize the user's emotions nor possess emotions of their own. The “universal facial expressions” are happiness, sadness, anger, fear, surprise, and disgust. Automatic face recognition is based on two approaches. One is Geometrical based approach and the other is Holistic based approach. The whole face is used to extract features in the Holistic based approach. Neural Networks have been successfully used as classifiers in face recognition and as well as Facial expression classification.

A neural network [1] is represented by weighted interconnections between processing elements (PEs). These weights are the parameters that actually define the non-linear function performed by the neural network. The process of determining such parameters is called training or learning, relying on the presentation of many training patterns. Back-propagation algorithm [1, 2] defines a systematic way to update the synaptic weights of Multi-layer Perceptron (MLP) [2]. The supervised learning is based on the gradient descent method, minimizing the global error on the output layer. The learning algorithm is performed in two stages: feed-forward and feed-backward. In the first phase the inputs are propagated through the layers of processing elements, generating an output pattern in response to the input pattern presented. In the second phase, the errors calculated in the output layer are then back propagated to the hidden layers where the synaptic weights are updated to reduce the error. This learning process is repeated until the output error value, for all patterns in the training set, are below a specified value. The definition of the network size (the number of hidden layers and of neurons in each layer) is a compromise between generalization and convergence. Convergence is the capacity of the network to learn the patterns on the training set and generalization is the capacity to respond correctly to new patterns. The idea is to implement the smallest network possible, so it is able to learn all patterns and, at the same time, provide good generalization. The Back Propagation, however, has two major limitations: a very long training process, with problems such as local minima and network paralysis; and the restriction of learning only static input-output mappings. To overcome these restrictions, new model has been proposed called Hybrid Emotional Neural Network (HENN).

10 subjects, 10 samples for each of six expressions from Cohn-Kanade database [3] are used as training and testing inputs. In experiments, all the training sets are completely different from that of test dataset. In the proposed method histogram equalization is used to enhance the brightness and contrast of the database. Then Gabor wavelets and Discrete Cosine Transform (DCT)[4,5] are applied as feature extraction algorithms on the preprocessed database. Now the reduced features are fused to form the Training feature set. Architecture of the proposed method is shown in Fig 1.

The remaining of this paper provides details on the proposed methods, experiments and results. Section 2 provides feature extraction using Gabor Wavelets and DCT. Section 3, describes the proposed architecture Hybrid Emotional Neural Network, and the modified Back propagation algorithm. Section 4 presents the experimental results and comparative study of the methods, and Section 5 contains the conclusion.

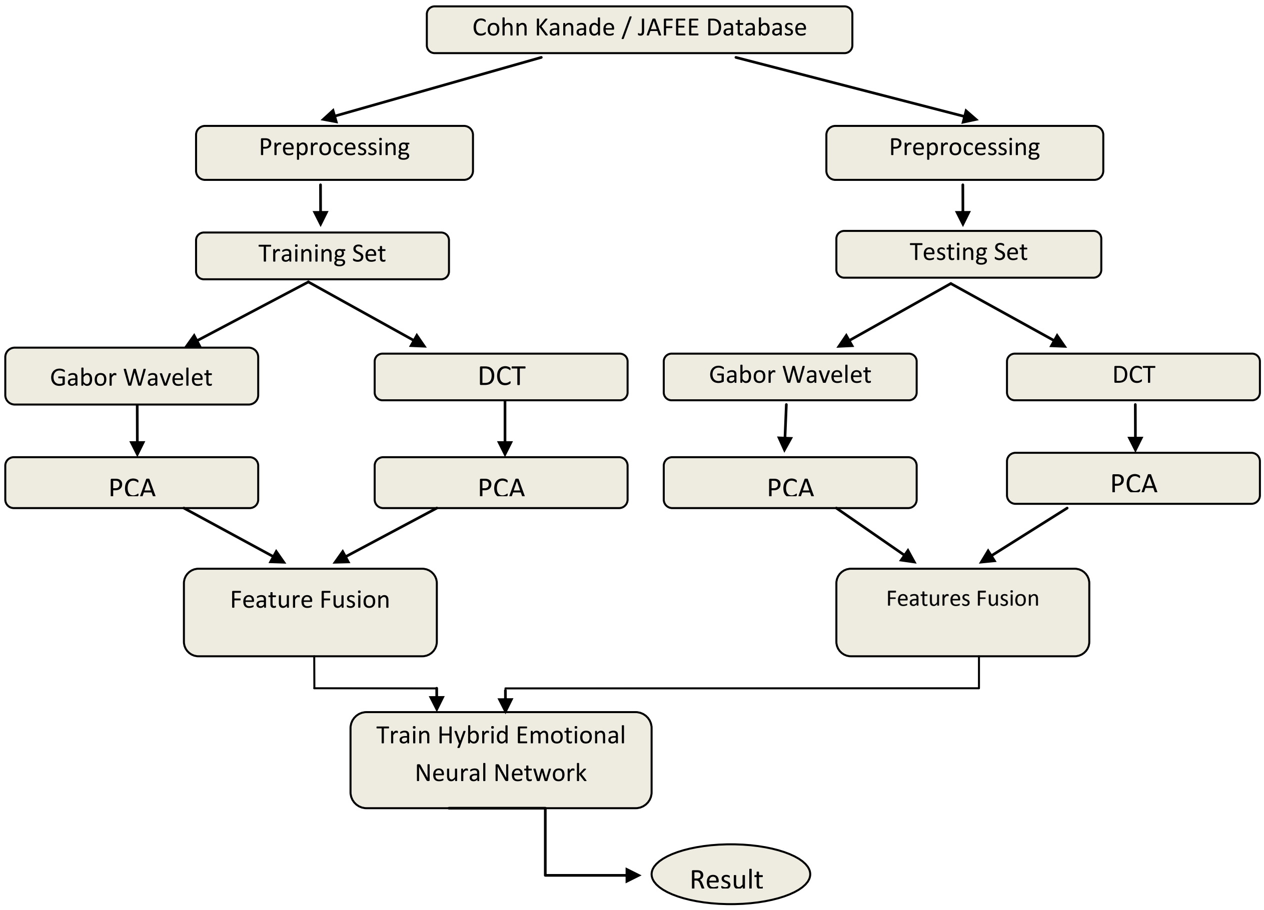

Fig1: Architecture of the Proposed Method

2. FEATURE EXTRACTION

2.1 Gabor Wavelets

The Gabor wavelet [4,6,7] representation captures salient visual properties such as spatial localization, orientation selectivity, and spatial frequency characteristic. Gabor wavelets were introduced to image analysis due to their biological relevance and computational properties. The Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. A complex-valued 2D Gabor function is a plane wave restricted by a Gaussian envelope is defined in following eq.

$$\psi_i(\bar{x}) = \frac{\|K_i\|^2}{\sigma^2} e^{\frac{-\|k_i\|^2 \|x\|^2}{\sigma^2} \left[\!\left[e^{j k_i x} - e^{-\frac{\sigma^2}{2}} \right]\!\right]}$$

The multiplicative factor K^2 ensures that filters tuned to different spatial frequency bands have approximately equal energies. The term $\exp(-\sigma^2)$ is substracted to render the filters insensitive to the overall level of illumination. The Gabor decomposition can be considered as a directional microscope with an orientation and scaling sensitivity. Since such curves correspond to some low-level salient features in an image, these cells can be assumed to form a low level feature map of the intensity image.

2.2 Discrete Cosine Transform:

Discrete Cosine Transform (DCT) [4,5] based image coders perform very well at moderate bit rates, and yields higher compression ratios. DCT attempts to decorrelate the image data. After decorrelation each transform coefficient can be encoded independently without losing compression efficiency. Principal Component Analysis (PCA) [8,9,10] is applied on both Gabor and DCT feature sets as a dimensionality reduction algorithm to reduce the size of them. But it is difficult to obtain accurate and reliable results from a single feature set, fusion technology is used to integrate and form a new combined feature set which is used for classification of emotions with Hybrid Emotional Neural Network.

3. Hybrid Emotional Neural Network:

In this paper a model is simulated with emotions in a simple supervised neural network structure as shown in figure 2, and it is used to investigate the effect of the added emotional factors on the learning and decision making capabilities of the neural network. Here two emotions (anxiety and confidence) [2] are considered to believe which have an effect on learning and decision making in humans. At the beginning of learning of a new task the anxiety level is high and the confidence level is low. After time, practice and getting positive feedback, the anxiety level decreases while at the same time Confidence level increases. Once learning is achieved, network experiences to be less anxious and more confident while performing a task. Therefore, anxiety and confidence are two dependent dimensions, where confidence is defined as the negative rate of change of anxiety.

In this section, the modification of this algorithm, by adding two emotional coefficients, provides the proposed HENN. The flow of information within the neural network, which consists of three layers: input layer with 'i' neurons, hidden layer with 'j' neurons, and output layer with 'k' neurons.

Initialization: Set all the weights and biases of the network to small random numbers that are uniformly distributed. Set learning rate α and the momentum factor μ.

$\mathbf{y_{in}}$: input of output nodes.

$\mathbf{z_{in}}$: input of hidden nodes.

m: number of output nodes

n: number of input nodes

p: number of hidden nodes

Feed Forward pass:

Each input unit receives the input signal and transmits the signal to all units in the layer above i.e., hidden nodes. Sigmoid activation function is used to activate each neuron in hidden and output layers. Here, one hidden layer is taken; however the same process can be applied for more than one hidden layer. The output of each hidden-layer neuron is defined as

$$Z_j = \left(\frac{1}{1+\exp\left(-Z_{in\ j}\right)}\right) \tag{1}$$

The input to a hidden-layer neuron Z_{inj} is calculated using the total potential of all input values coming into that neuron. The total potential is the sum of multiplications of input values and their associated weights.

Where Z_{inj} and Z_j are the input and output values of neuron j in the hidden layer, respectively. Input and output configuration of a hidden layer is shown in Fig 2.

$$\mathrm{Z_{inj}} = \mathrm{T_{jc}} + \mathrm{T_b} + \mathrm{T_{jm}} \tag{2}$$

$$\text{Where } T_{jc} = \sum_{i=1}^{n} \mathrm{W_{ji}} \cdot \mathrm{Z_{in}} \tag{3}$$

$$\mathrm{T_b} = \mathrm{W_{jb}} \cdot \mathrm{X_b} \tag{4}$$

Where W_{jb} is the weight associated with bias neuron, and X_b is the input value to the bias neuron, which is always 1.

$$T_{jm} = W_{jm} \cdot X_m \tag{5}$$

Where W_{jm} the weight is associated with emotion neuron m to hidden neuron, and Xm is the input value to the emotion neuron. According to Baumgartner, "Most of the published neuro-imaging papers examining emotional processes have used visual stimuli in order to evoke emotions (positive or negative)." Therefore, the Gabor filter bank is used to extract emotional data which is similar to perception in the human visual system and is given as an input to the emotional neuron

$$X_m = \sum_{i=1}^{40} G_{avg} \tag{6}$$

Where G_{avg} is the average of Gabor features extracted from input image.

The Emotional Back Propagation Parameters

Here the proposed emotional parameters are used together with the existing learning coefficient (η) and momentum rate (α) in order to adjust the neural network weights, based on error minimization.

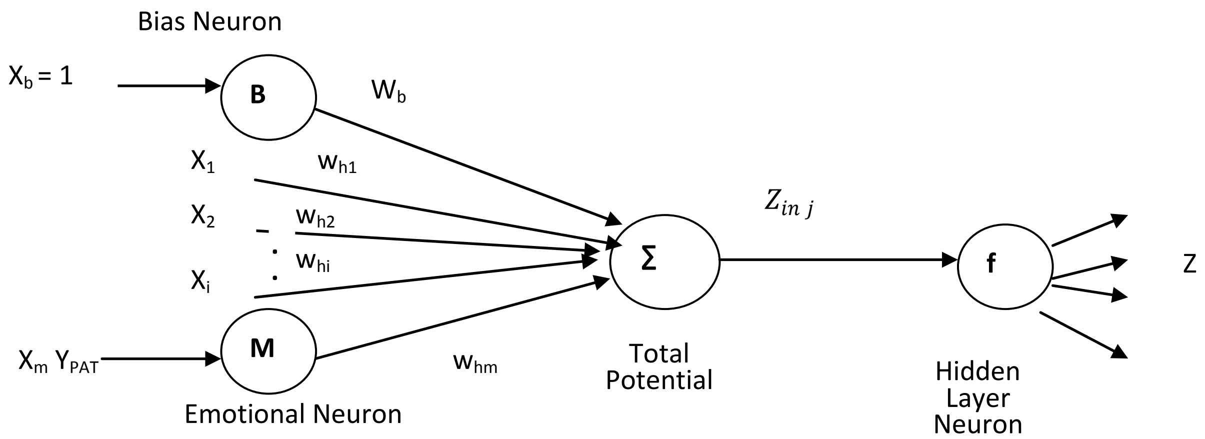

Fig 2: input and output configuration of a hidden-layer neuron

Architecture of the proposed Hybrid Emotional Neural Network classifier is shown in Fig 3. The proposed emotional parameters are the anxiety coefficient (μ) and the confidence coefficient (k). According to Adnan Khashman [2] hypothesis, while learning a new task, the anxiety level is high at the beginning and the confidence level is low. After time, practice, and by getting a positive feedback, the anxiety level decreases while the confidence level increases. Once learning is achieved anxiety is less and confidence is high. The proposed Hybrid emotional neural network uses both emotional coefficients and normal coefficient during the learning and the generalization processes. Both coefficients have values between "0" and "1." In order to model the emotional factor in an artificial neural network, the following rules have been applied.

Rule 1: The anxiety level is dependent on the input patterns, where new patterns cause higher anxiety. During the first iteration (new task learning), the initial anxiety coefficient value is set to "1."

The anxiety level is dependent on the difference (error) between the actual output of the neural network and the desired (target) output. This is a kind of feedback that the emotional neural network uses to measure how successful its learning is. Anxiety decreases with the minimization of the error.

Rule 2: The confidence level increases with the decrease in anxiety level. During the first iteration (new task learning), the initial confidence coefficient value is set to "0." the anxiety coefficient value is defined as

$$\mu = G_{AVPAT} + E \qquad (7)$$

Where G_{AVPAT} is the average value of all presented pattern Gabor features to the neural network in each iteration, which is defined as

$$G_{AVPAT} = \frac{\sum_{P=1}^{N_p} G_{AVG}}{N} \qquad (8)$$

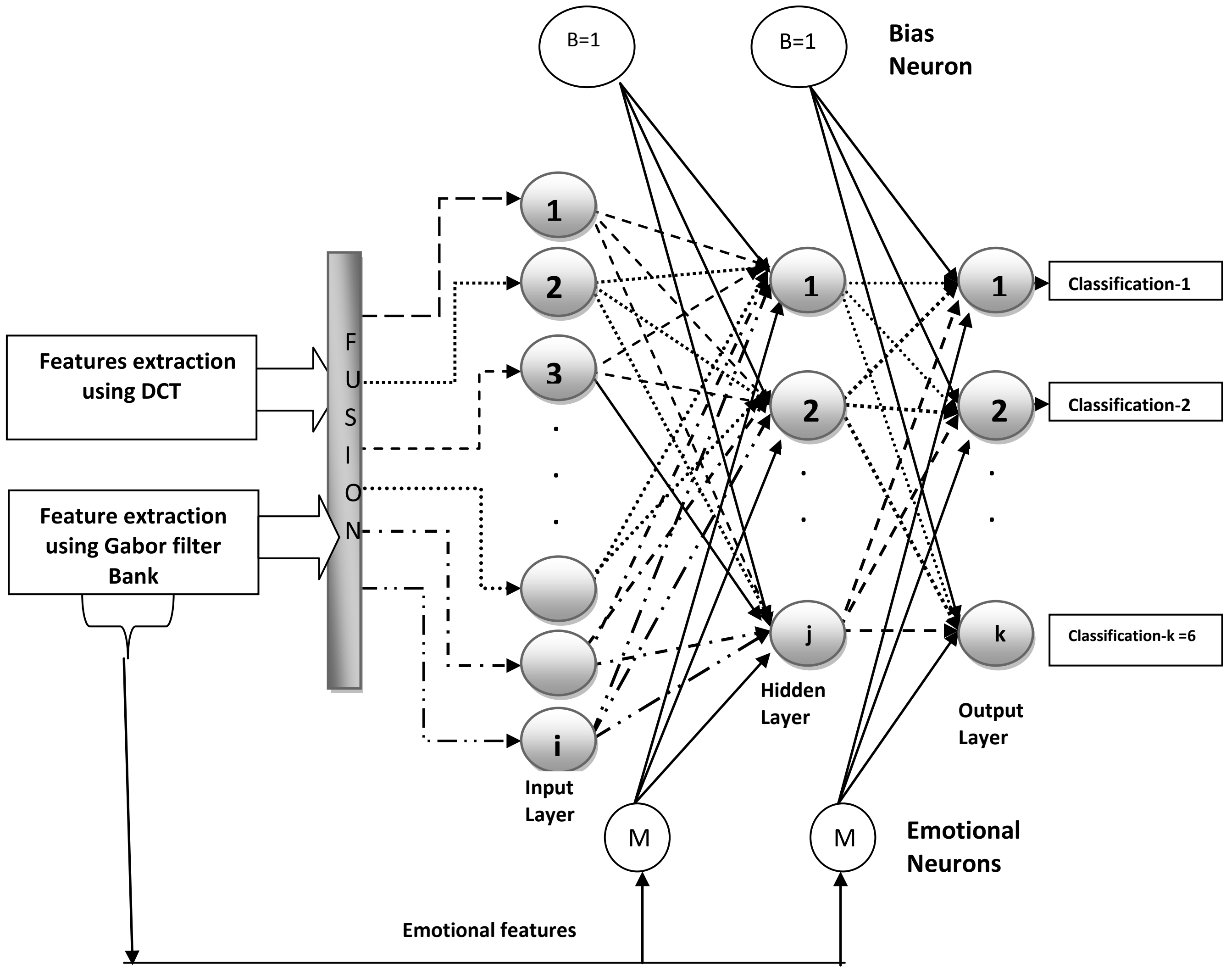

Fig 3: Architecture of the Hybrid Emotional Neural Network

Where p is pattern index from first to last pattern N_p. G_{AVG} is average value of pattern p. The error feedback E is defined as

$$E = \sum_{j=1}^{N_j} \left(T_j - Y_j\right)^2 \Big/ N_p . N_j \qquad (9)$$

where j is the output neuron index from first to the last neuron N_j and N_p is the total number of patterns. Y_j and T_j are the actual and target output value for neuron j, respectively. The confidence coefficient (k) value is defined as

$$k = \mu_0 - \mu_i \qquad (10)$$

Where (μ_0) is the anxiety coefficient value after the first iteration (exposure to new patterns) and (μ_i) is the anxiety coefficient value in at subsequent iteration i.

The Error BP Calculations

The conventional error signal BP algorithm aims at updating the weights between the various neural network layers after each iteration or epoch. Similarly, the modified back propagation algorithm propagates back an error signal to update its weights at the end of each iteration. Central to the concept of training a neural network is the definition of network error. Rumelhart [11] defined an error E_p is the difference between the output value and the target value.

$$E_p = \sum_{j=1}^{Nj} (Tj - Ypj)^2 \qquad (11)$$

The first derivative of this function $F'(X_j)$ is an important element in error BP. For output-layer neurons, a quantity called the error signal is represented by Δj,

$$\Delta j = f'\left(Y_{in\ j}\right).\left(T_j - Y_j\right) \qquad (12)$$

$$\Delta j = Y_j\left(1 - Y_j\right).\left(T_j - Y_j\right) \qquad (13)$$

Hidden-to-Output Weight Updating

$$W_{kj}\,(new) = W_{kj}\,(old) + \Delta\, Wc_{kj} + \Delta\, Wm_{kj} \qquad (14)$$

The HENN contains emotional neurons, and therefore, an extra set of weights must be accounted for. Here the change in the emotional weights is calculated as follows:

$$\Delta\, Wm_{kj} = \mu\ .\ \Delta_k\ .\ G_{avg} + \kappa\ .\ [\ \delta\ W_{kj}\,(old)] \qquad (15)$$

where μ is anxiety coefficient, Δ_k is the error signal for the output layer

Input-to-Hidden Weight Updating

$$\Delta j \;=\; f'\,(Z_{in}).\sum_{k=1}^{Nj} W_{kj}\ .\Delta_k \qquad (16)$$

$$\Delta_j \;=\; Z_{in}\ .(1 - Z_{in})\sum_{k=1}^{Nj} W_{kj}\ .\Delta_k \qquad (17)$$

The weight adjustments for the connections feeding the hidden layer from the input layer are now calculated in a similar manner to those feeding the output layer. These adjustments are calculated using the following:

$$W_{ji}\,(new) = W_{ji}\,(old) + \Delta\, Wc_{ji} + \Delta\, Wm_{ji} \qquad (18)$$

where $\Delta\ Wc_{ji}$ is the change in the conventional weights (including the bias weights) feeding the hidden layer, and $\Delta\ Wm_{ji}$ is the change in the emotional weights feeding the hidden layer $\Delta\, Wc_{ji} = \eta.\ \Delta_j\ .\ Z_{in} + \alpha\ [\delta\ W_{ji}\ .(old)]$ (19)

$$\Delta\, Wm_{ji} = \mu\ .\ \Delta_j\ .\ Y_{PAT} + k.\ [\delta\ W_{ji}\,(old)] \qquad (20)$$

4. EXPERIMENTAL RESULTS

In this paper, the static images of Cohn-Kanade [3] are used. Raw faces are preprocessed to overcome illumination effects. Each face image represents six basic emotions. To validate proposed algorithms against variations of illumination and rotation, three different databases are generated with rotation angles of 2^o, 5^o and 10^o from standard benchmark database such as Cohn-Kanade facial expression database apart from original. Test set 1 have a rotational noise of -2^0. Similarly Test set 2 and Test set 3 have rotational noise of -5^0 and -10^0 respectively. From the total of 600 face images 180 images (3 for each expression) are used for training and the remaining 420 images were classified with recognition rates of 98.3333%, 96.6666% and 95.4761% on the Test Sets 1, 2 and 3 respectively.

Expression	Surprise	Sad	Anger	Fear	Disgust	Happy
Surprise	70	0	0	0	0	0
Sad	0	68	0	0	2	0
Anger	0	0	68	1	1	0
Fear	1	0	0	69	0	0
Disgust	0	0	0	0	69	1
Happy	0	0	0	0	0	70
Overall	98.5714					

Table 1: Confusion matrix for Test set 1 with HENN using Wavelet Fusion.

Expression	Surprise	Sad	Anger	Fear	Disgust	Happy
Surprise	69	0	0	0	1	0
Sad	1	67	0	0	2	0
Anger	0	0	68	1	1	0
Fear	1	2	0	67	0	0
Disgust	1	1	0	0	67	1
Happy	0	0	2	0	0	68
Overall	96.6666					

Table 2: Confusion matrix for Test set 2 with HENN using Wavelet Fusion.

Expression	Surprise	Sad	Anger	Fear	Disgust	Happy
Surprise	68	1	0	0	1	0
Sad	0	65	1	2	2	0
Anger	0	2	66	1	1	0
Fear	1	2	0	67	0	0
Disgust	0	0	2	0	66	2
Happy	1	0	0	0	0	69
Overall	95.4761					

Table 3: Confusion matrix for Test set 3 with HENN using Wavelet Fusion.

Proposed Method	Fusion Technique	Test Set 1	Test Set 2	Test Set 3
Gabor+ DCT +PCA+HENN	Maximum	95	92.8571	91.6666
	Average	95.9142	93.8095	92.3809
	PCA	97.1428	94.7619	94.2857
	Wavelet	98.5714	96.6666	95.4761

Table 4: Correct Classification rates for HENN with different fusion techniques.

Table 4 describes that the proposed method tested with different fusion techniques and wavelet fusion gives highest classification rate than Maximum, Average and PCA fusions. It is shown in the figure 5.

The Proposed method (Gabor+ DCT+ PCA+ Wavelet Fusion+ HENN) has highest classification rates than the previous existing neural networks like Feed Forward Neural Network (FFNN)[2] and Radial Basis Function (RBF)[6].And comparative study of these networks with wavelet fusion as a fusion technique is shown in the Table 5. And it is represented in the figure 6.

Method	Test Set 1	Test Set 2	Test Set 3
Gabor+ DCT +PCA+FFNN	90.7142	86.9047	83.5714
Gabor+ DCT +PCA+RBF	97.3809	89.2857	86.1904
Gabor+ DCT +PCA+HENN	98.5714	96.6666	95.4761

Table 5: Classification Rates of different Neural Networks on Test Sets 1, 2 and 3

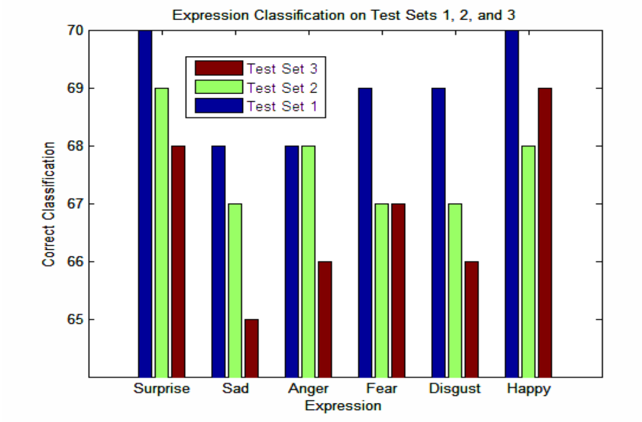

Fig 4: Expression Classification with Proposed method on Test Sets 1, 2 and 3.

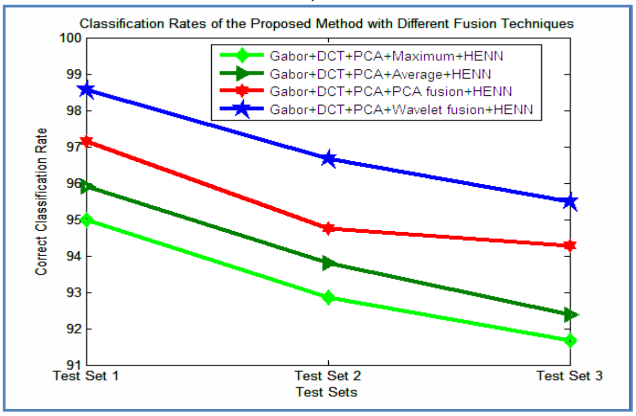

Fig 5: Classification rate of proposed method with different fusion techniques

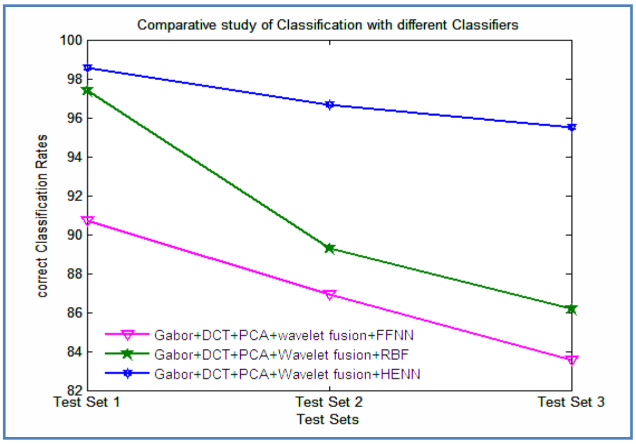

Fig 6: Comparative study of Classification rates with different classifiers

5. CONCLUSION

Hybrid Emotional Neural Network (HENN) for classification of emotions from Facial expressions was presented. Six different facial emotions were classified with the proposed algorithm. The classification results were compared with standard neural networks such as Feedforward Neural Network and Radial Basis Neural Network. The results are compared between different fusion techniques with the proposed architecture. The correct classification rate was significantly high in proposed method comparative with other existing methods. The proposed method also tested with different fusion techniques for the classification of emotions. The percentage of the correct classifications varied across different databases from 75% to 99%. Further studies are needed to develop better training algorithm to speed up the training time and also optimisation methods for feature selection that would give higher recognition rates between similar facial expressions under influence of noise in input images. Further work involves taking many effective measures to reduce the training time and improve the recognition accuracy.

6. REFERENCES

[1] HAYKIN , S., "Neural Networks: A Comprehensive Foundation," Prentice Hall, Upper Saddle River, NJ, 1999.

[2] Adnan Khashman, "A Modified Back propagation Learning Algorithm With Added Emotional Coefficients", IEEE Transactions on Neural Networks, vol 19, No 11, pp 1896-1909, 2008.

[3] Kanade, T., Cohn, J. F., & Tian, Y., "Comprehensive database for facial expression analysis". Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), pp 46-53, 2000.

[4] K.V. Krishna Kishore, G.P.S.Varma, "Efficient Facial Emotion Classification with Wavelet Fusion of Multi Features" ,International Journal of Computer Science and Network Security, Vol. 11, No. 8, pp. 160-165.2011.

[5] Prof. G. U. Kharat, Dr. S.V.Dudul" Neural Network Classifier for Human Emotion Recognition from Facial Expressions Using Discrete Cosine Transform", IEEE, pp 653-658, 2008.

[6] Chien-Cheng Lee and Cheng-Yuan Shih, "Gabor Feature Selection and Improved Radial Basis Function Networks for Facial Expression Recognition", proc., by IEEE, 2009.

[7] Linlin Shen, Li Bai, "A review on Gabor wavelets for face recognition", proc., Pattern Anal Applic vol- 9, pp 273–292, 2006.

[8] Matthem Turk and Alex Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience Volume 3, Number 1, 1991.

[9] Andrew J. Calder , A. Mike Burton , Paul Miller , Andrew W. Young , Shigeru kamatsu, A principal component analysis of facial expressions", Vision Research vol-41, pp 1179–1208, 2001.

[10] Venkatrama Phani Kumar, K.V.Krishna Kishore, K Hemantha Kumar, "Hybrid Face Recognition with Locally Discriminating Projection", Proceedings of International conference on Signal Acquisition and Processing, pp 327-331, 2010.

[11] D. Rumelhart, G.E.Hinton, and R.J.Williams,"Learning internal representations by error propagation", in Parallel Distributed Processing, MIT Press, 1986, vol 1.

7. AUTHORS PROFILE

K.V.Krishna Kishore received his M.Tech degree in Computer Science and Engineering from Andhra University. He is currently working as Professor and Head of Computer Science and Engineering Dept., in Vignan's University. He got more than 19 years of teaching and research experience. His current research interests include Digital image processing, Biometrics, Face recognition, Facial emotional classifications, Neural Networks, Pattern Recognition and Medical Imaging. He has published more than 15 journal and conference papers in these areas. He has received best teacher award for four times in Vignan's Engineering College. He has Chaired National Youth festival for three times.

G.P.S. Varma received his B.Tech, M.Tech degrees in Computer Science and Engineering from Mangalore University and REC Warangal respectively. He got Ph.D Degree from Andhra University, Visakhapatnam. He is currently working as Professor and Head of Information Technology Department, S.R.K.R. Engineering College, Bhimavaram, He got more than 17 years of teaching and research experience. His current research interests include Software Engineering, Data ware Housing and Mining, Digital image processing, Neural Networks, Pattern Recognition and Bio Informatics. He has published more than 49 International and National Journal and conference papers in these areas. He has authored 8 books in the field of Computer Science and Engineering. He Chaired National and International Conferences for Nine times.