

Human Facial Expression Recognition using MANFIS Model

V. Gomathi, Dr. K. Ramar, and A. Santhiyaku Jeevakumar

Abstract—Facial expression analysis plays a significant role for human computer interaction. Automatic analysis of human facial expression is still a challenging problem with many applications. In this paper, we propose neuro-fuzzy based automatic facial expression recognition system to recognize the human facial expressions like happy, fear, sad, angry, disgust and surprise. Initially facial image is segmented into three regions from which the uniform Local Binary Pattern (LBP) texture features distributions are extracted and represented as a histogram descriptor. The facial expressions are recognized using Multiple Adaptive Neuro Fuzzy Inference System (MANFIS). The proposed system designed and tested with JAFFE face database. The proposed model reports 94.29% of classification accuracy.

Keywords—Adaptive neuro-fuzzy inference system, Facial expression, Local binary pattern, Uniform Histogram

I. INTRODUCTION

FACIAL image analysis is an active research in the field of machine vision. Psychological studies have suggested that facial motion is fundamental to the recognition of facial expression. Fully automatic and real-time facial expression system helped for many applications such as human computer interaction, behavioral research, video conferencing and various vision systems. Facial expression analysis is the part of facial image analysis. Mehrabian [2] indicated that the verbal part of a message contributes only for 7% to the effect of the message as a whole, the vocal part contributes for 38%, while facial expression of the speaker contributes for 55% to the effect of the spoken message. This implies that the facial expressions form the major modality in human communication.

Most approaches to automatic facial expression analysis attempt to recognize a principal set of prototypic emotional facial expressions, i.e., fear, sadness, disgust, anger, surprise, and happiness. From the survey, it was revealed that most of the facial expression recognition systems were based on the Facial Action Coding System (FACS) [7], [14]. It is a system designed for human observers to describe changes in the facial

expression in terms of visually observable activations of facial muscles.

The changes in the facial expression are described with FACS in terms of 44 different action units (AUs), each of which is anatomically related to the contraction of either a specific facial muscle or a set of facial muscles. A problematic issue is that AUs can occur in more than 7000 different combinations, with wide variations due to age, size and ethnicity.

M.Pantic and Rothkrantz [14] presented detail survey about facial expression recognition. Most of the research on facial expression recognition has been based on static images. The facial images are extracted and classified using Template based methods, Feature based methods, Neural Networks and Rule based methods. Recently it has been carried out using dynamic facial image sequences. The psychological experiments by Bassili [3] have suggested that facial expressions are more accurately recognized from a dynamic image than from a single static image. Kotsia et al. [4] used facial wire frame model and a Support Vector Machine (SVM) performs classification. Zhang et al. [5],[6] suggested IR illumination camera for facial feature detection and tracking and recognized the facial expression using Dynamic Bayesian networks (DBNs). Tian et al. [7] proposed multi state face component model and neural network to recognize the facial expressions. The System has achieved average recognition rates of 96.4 percent for upper face AUs and 96.7 percent for lower face AUs.

Yeasin *et al.* [9] designed a discrete hidden Markov models (DHMMs) to recognize the facial expressions. Anderson et al. [10] used the multichannel gradient model (MCGM) to determine facial optical flow. The motion signatures produced are then classified using Support Vector Machines. Cohen et al. [11] used Naive-Bayes classifiers and hidden Markov models (HMMs) for automatically segmenting and recognizing human facial expression from video sequences. Pantic *et al.* [12] used face-profile-contour tracking and rule-based reasoning to recognize 20 AUs occurring alone or in a combination in nearly left-profile-view face image sequences and they achieved 84.9% accuracy rate.

Recently Ahonen *et al.* [13] proposed a novel facial representation scheme for static images based on Local Binary Pattern (LBP) features. In this approach, the face image is divided into several regions from which the LBP features are extracted and concatenated into an enhanced feature vector. This approach is proving to be a growing success. It has been adopted by many research groups, and has been successfully used for facial expression recognition.

V. Gomathi, is with the National Engineering College, Kovilpatti, Tamil Nadu, INDIA (Mobile: 99948 46309; e-mail: vgese@nec.edu.in).

Dr. K.Ramar, is with the National Engineering College, Kovilpatti, Tamil Nadu, INDIA (Mobile: 94426 24114; e-mail: drkrcse@nec.edu.in).

A. Santhiyaku Jeevakumar, was with the National Engineering College, Kovilpatti, Tamil Nadu, INDIA

Zhao *et al.* [1] used facial dynamic texture information with co-occurrences of Local Binary Pattern on the Three Orthogonal Planes (LBP-TOP) and Volume Local Binary patterns (VLBP), for combining motion and appearance. He achieved 95.19% recognition accuracy using VLBP and 94.38% recognition accuracy using LBP-TOP. In his earlier work [15], he tested with the two-dimensional (2-D) discrete cosine transform (DCT) over the entire face image as a feature detector and used a constructive one-hidden-layer feed-forward neural network as a facial expression classifier but achieved less accuracy only.

From the detailed survey, it was revealed that the FACS involves more complexity due to facial feature detection and extraction procedures. Only limited contributions were emerged in recognizing expressions using dynamic textures. Hence, the proposed facial expression recognition system aimed to use LBP texture feature information for building MANFIS based expression model for the frontal view face image sequences.

The rest of the paper has organized to explain the proposed methodology in section II, data collection in section III, experimental results and analysis in section IV, comparison in section V and concludes the work in section VI.

II. METHODOLOGY

The proposed methodology is based on the facial texture information of permanent facial feature components such as eye, eyebrow, nose and mouth to recognize the facial expression. Initially frontal view image is segmented into three regions; upper region *E*(eye and eyebrow), middle region *N*(Nose) and lower region *M*(mouth). The uniform rotation-invariant LBP (RIULBP) texture feature distributions are extracted from these three regions and represented as a weighted 7 dimensional histogram descriptor.

$$h = [h_1, h_2, h_3, \dots, h_7] \quad (1)$$

where, h_i – Frequency of occurrence of i^{th} RIULBP histogram.

As per human physiology, since each expression has distinct variation from a neutral state, the measure of distortion from neutral could be a better discriminator for grouping the class of expression. This variation in histogram descriptor is to be fed as an input feature vector for the proposed Adaptive Neuro-Fuzzy based Inference system model. The proposed neuro-fuzzy model is capable of capturing the subjective nature of facial expressions in terms of defining an 'Expression term set' with linguistic variables like closer to neutral, small variations from neutral, medium variations from neutral, large variations from neutral.

A. Local Binary Patterns (LBP)

The LBP operator labels each of the pixels in a Gray-level image by considering each pixel as center and their 3 x 3 neighborhood pixel variations put together as an 8-bit binary code.

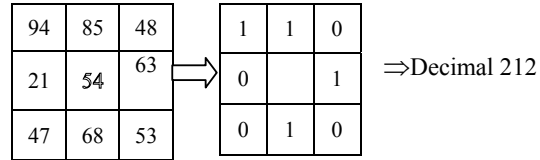


Fig. 1. The LBP Operator

So the original dimension of LBP histogram has 256 bins. See Fig. 1 for an illustration of the LBP operator. The LBP texture T is defined in a local $(P+1)$ neighborhood, as:

$$T = t(g_c, g_0, g_1, \dots, g_{P-1}) \quad (2)$$

where g_c corresponds to the gray value of the center pixel. $g_p (p = 0, \dots, P-1)$ correspond to the gray values of P equally spaced pixels on a circle of radius R ($R > 0$) that form a circularly symmetric set of neighbors. To achieve invariance with respect to any monotonic transformation of the gray scale, only the signs of the differences are considered as shown in Eqs. (3) & (4):

$$T \approx t(s(g_0 - g_c), \dots, s(g_{P-1} - g_c)) \quad (3)$$

where,

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Ojala *et al.* [18], found that the vast majority of the LBP patterns in a local neighborhood are so called "uniform patterns". An LBP is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when it is considered circular. For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform LBPs, whereas the patterns 11001001 (4 transitions) and 01010011 (6 transitions) are not. In the computation of the LBP histogram, uniform patterns are clustered so that the histogram for every uniform pattern has reduced the dimension from 256 to 58 bins.

B. Rotation invariant Uniform LBP (RIULBP)

The $LBP_{P,R}$ operator produces 2^P different output values, corresponding to the 2^P different binary patterns that can be formed by the P pixels in the neighbor set. When the image is rotated, the gray values g_p will correspondingly move along the perimeter of the circle around g_0 . Since g_0 is always assigned to be the gray value of element (0, R) to the right of g_c rotating a particular binary pattern naturally results in a different $LBP_{P,R}$ value. This does not apply to patterns comprising of only 0s (or 1s) which remain constant at all rotation angles. To remove the effect of rotation, i.e., to assign a unique identifier to each rotation invariant local binary pattern uses Eq. 5:

$$LBP_{P,R}^i = \min \{ROR(LBP_{P,R}, i) | i = 0, 1, \dots, P-1\} \quad (5)$$

where $ROR(x, i)$ performs a circular bit-wise right shift on the P -bit number x i times. In terms of image pixels (4), simply corresponds to rotating the neighbor set clockwise so many times that a maximal number of the most significant bits, starting from g_{P-1} is 0. $LBP_{P,R}^{ri}$ quantifies the occurrence statistics of individual rotation invariant patterns corresponding to certain micro features in the image such as

edges, spots and flat areas, over the whole image. For example, for the original LBP code $(00000011)_2$, its codes after rotating clockwise 45, 90, 135, 180, 225, 270, 315, 360 degrees are $(00000011)_2$, $(00000110)_2$, $(00001100)_2$, $(00011000)_2$, $(00110000)_2$, $(01100000)_2$, $(11000000)_2$ and $(10000001)_2$ respectively. Thus RIULBP further reduces the bins from 58 to 7 bins.

C. Aggregation of LBP Feature Descriptor

In this phase, the frontal view image is divided into three regions, namely E, N & M . The LBP operators are applied for each of the three regions. From these regions rotation-invariant uniform texture features are extracted and the histogram is computed independently for each of the three regions such as, E_h, N_h and M_h . These three 7-bin histograms are aggregated based on weighted averaging as,

$$E_h * w_E + N_h * w_N + M_h * w_M \quad (6)$$

where, w_E, w_N and w_M are the weights empirically chosen as 0.65, 0.25 and 0.1

D. Neuro-Fuzzy Approach

In this work, a Neuro-Fuzzy approach is proposed to model the facial expression recognition. To recognize the six facial expressions multiple outputs are required. Due to multiple responses of the target system, MANFIS model is adopted, which is an extension of the well-known ANFIS model with symmetric architectures and sharing the input layer. This is based on the first order Sugeno fuzzy model, which always drives crisp outputs and computationally intensive. The internal functional blocks are illustrated in Fig. 2.

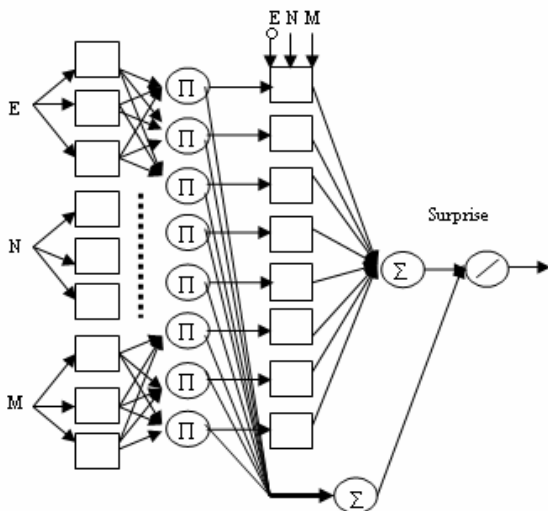


Fig. 2. ANFIS for Surprise Expression

E. Multiple Adaptive Neuro-Fuzzy Inference System

To recognize the six facial expressions, the system describe six output MANFIS structure. In MANFIS, each ANFIS shares the input alone as shown in Fig. 3. The MANFIS for this problem has six output nodes corresponding to the aggregation of three facial region of the frontal view image. To recognize the facial expressions, MANFIS is trained by the

data samples (input variables) collected from JAFEE Database. Each ANFIS has three 5 - dimensional input nodes.

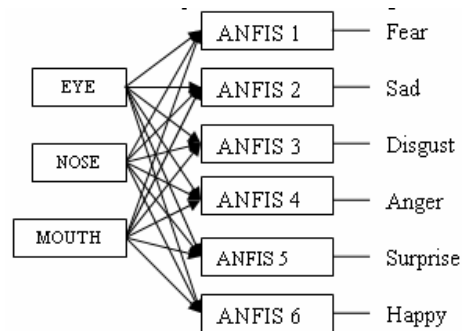


Fig. 3. MANFIS Structure

The first layer is the input layer. In facial image analysis, only dominant uniform patterns (i.e. RIULBP) are selected which are grouped based on their rotation invariant property. These values are considered as the input for the MANFIS Architecture. Premise layer has three membership functions are Small, Medium, Large for each input. The Fig. 4 describes the fuzzy sets for the Eye region's RIULBP histogram inputs.

The second layer is the fuzzy-rule base layer. Depends upon the contributions of input parameters this will derive fuzzy rules with firing strengths as outputs.

The third layer is the functional consequent layer. This layer takes the same input as Layer1 and the links of the nodes in this layer are weighed by firing strengths. Hence the proposed system generates six expression labels with fuzzy-rules, to recognize facial expression as a linear polynomial function.

The fourth layer is a summation layer or fuzzy aggregation layer. This layer will provide the decisive composition for an expression in terms of the appropriate consequent layer's output nodes. Each node in this layer acts as a voting mechanism for predicting the most probable label of the facial expression.

The fifth layer is the output layer. This output layer recognizes six facial prototypic expressions such as, fear, sadness, disgust, anger, surprise, and happiness by discriminating the most probable expression label with higher score. Hence act as final output layer.

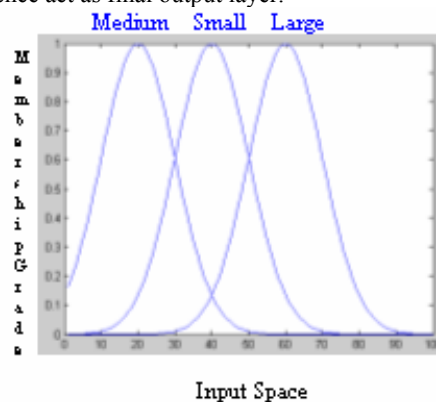


Fig. 4. Fuzzy sets for Eye Region with three MFs

F. Fuzzy Membership Functions and partitioning

The Triangular membership functions (MFs) defined for each input variables. Each fuzzy set has three MFs to express its three variations from neutral. The shape of the MFs chosen based on the objective.

In fuzzy modeling, it is important to determine a reasonable number of membership functions to maintain appropriate linguistic meanings. The proposed neuro-fuzzy facial expression model defines linguistic variables that are characterized by small variations, medium variations and large variations. In this MANFIS input space is partitioned using grid partitioning. Depends upon the input value membership function and fuzzy rule will derive and firing strengths (w_i) as outputs.

III. DATA COLLECTION

JAFEE stands for The Japanese Female Facial Expression (JAFEE) Database. The database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by Japanese female models. Sixty Japanese subjects have rated each image on 6 emotion adjectives. The database was planned and assembled by Miyuki Kamachi, Michael Lyons, and Jiro Gyoba with the help of Reiko Kubota as a research assistant. The photos were taken at the Psychology Department in Kyushu University. Few samples are shown in Fig. 5.



Fig. 5. Anger, Disgust, Fear, Happy, Sad and Surprise Facial Expressions

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, constructed MANFIS model using JAFEE Face Database and considered frontal view facial images. Initially all the images are resized into 120*120 pixels. The facial image is divided into three regions based on the permanent facial feature. The LBP is applied for each of the three regions. Now, the histogram is computed independently for each of the three regions as a uniform texture feature. Based on the LBP uniform rotation invariant property, the 58 uniform bins are reduced into 7 bins. The number of rotation-invariant uniform LBP texture feature occurrence is measured for each of the bin. It was observed that, each bin has discriminative texture histogram value for different expressions. The histogram bin values have excellent discriminate variations for each of the expressions and uniform variations for the test set of 213 images in the JAFEE Database.

Training Phase: In this work, supervised learning is used to train the ANFIS network. The training samples are taken from the JAFEE database. This work has considered 125 training samples for all expressions. After getting the samples, hybrid learning is used to train the network. It reduces the error signals.

Testing Phase: This proposed system tested with JAFEE database. It was taken totally 5 sample images for each of the facial expressions. So, totally 35 testing sample are taken.

The Fig. 6, shows the GUI for the extracted LBP regions and its classification. In this figure, the sample image exhibits anger expression. The near white regions are more uniform LBP, whereas the near black regions are non-uniform. It clearly exhibits that, the muscular changes due to anger are captured by the uniform LBP regions. The score of each expression is listed, while the score for anger was highest.

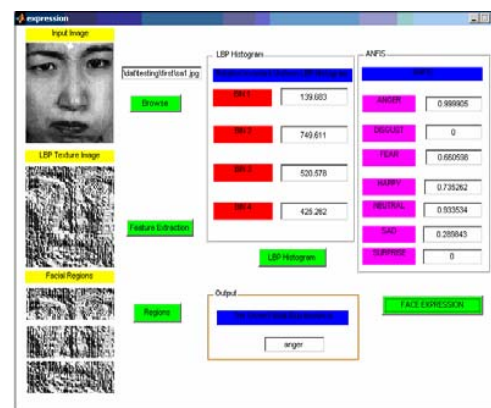


Fig. 6. GUI for recognizing the Facial Expression

V. COMPARISON

The LBP based neuro fuzzy approach was compared with Zhao *et al.*[1] approach of Volume Local Binary patterns (VLBP) and Local Binary Pattern on the Three Orthogonal Planes (LBP-TOP). They achieved 95.19% recognition accuracy using VLBP and 94.38% recognition accuracy using LBP-TOP. And also with Liao *et al.*[9] approach of multi-state face component model and neural network model, which has achieved average recognition rates of 94.56% for upper face AUs and 94.72% for lower face AUs. In the proposed Neuro-Fuzzy based approach, was achieved more recognition accuracy. The comparative analysis of our result was shown in Table I, which confirms the robustness of the proposed approach.

TABLE I
COMPARATIVE ANALYSIS OF RESULT

Approaches	Resolution	Feature Dimension	Recognition Accuracy
Zhao [1]	400*300	2^{3P+2}	95.19%
Liao [9]	64*64	16	94.56%
Our Approach	120*120	12(4*3)	94.29%

VI. CONCLUSION

In this paper, a novel and efficient facial expression recognition system is proposed. The LBP is best operator to extract the facial feature and MANFIS is used to recognize the

facial expression. The proposed neuro-fuzzy approach implementation is reliable for the real-time applications. We achieved 100% accuracy for training sets and 94.29% accuracy for test sets. This work is useful for real-world problems such as human emotion analysis, human-computer interaction, surveillance and online-conferencing and for entertainments.

ACKNOWLEDGMENT

The Authors thank All India Council for Technical Education (AICTE), New Delhi, INDIA for funding this work under Research Promotion Scheme. (Ref. No. : 8023 / BOR / RPS/103/2006-07). Also they acknowledge the Management, Principal of National Engineering College for their continual encouragements in all aspects.

REFERENCES

- [1] Guoying Zhao and Matti Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions", *IEEE Trans. Syst., Man, Cybern.B, Cybern.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.
- [2] A. Mehrabian, "Communication without words", *Psych. Today*, vol. 2, no. 4, pp. 53–56, 1968.
- [3] J.N. Bassili, "Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Area of the Face", *J. Personality and Social Psychology*, vol. 37, pp. 2049–2059, 1979.
- [4] Irene Kotsia, Ioannis Pitas, "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines", *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 172–187, Jan. 2007.
- [5] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 699–714, May 2005.
- [6] Spiros Ioannou, George Caridakis, Kostas Karpouzis, and Stefanos Kollias, "Robust Feature Detection for Facial Expression Recognition" *EURASIP Journal on Image and Video Processing*, vol. 2007, pp. 1–22, May. 2007.
- [7] Y. L. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 97–115, Feb. 2001.
- [8] Mohammed Yeasin, Baptiste Bulot, and Rajeev Sharma, "Recognition of Facial Expressions and Measurement of Levels of Interest From Video", *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 500–508, Jun. 2006.
- [9] Yan Tong, Wenhui Liao, and Qiang Ji, "Facial Action Unit Recognition by Exploiting Their Dynamic and Semantic Relationships", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1–17, Oct. 2007.
- [10] Keith Anderson and Peter W. McOwan, "A Real-Time Automated System for the Recognition of Human Facial Expressions", *IEEE Trans. Syst., Man, Cybern.B, Cybern.*, vol. 36, no. 1, pp. 96–105, Feb. 2006.
- [11] I. Cohen, N. Sebe, S. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: temporal and static modelling", *Comput. Vis. Image Understand.*, vol. 91, pp. 160–187, 2003.
- [12] M. Pantic and Ioannis Patras, "Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments From Face Profile Image Sequences", *IEEE Trans. Syst., Man, Cybern.B, Cybern.*, vol. 36, no. 2, pp. 433–449, Apr. 2006.
- [13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
- [14] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, Dec. 2000.
- [15] Guoying Zhao and Matti Pietikainen, "Facial Expression Recognition Using Constructive Feed forward Neural Networks", *IEEE Trans. Syst., Man, Cybern.B, Cybern.*, vol. 34, no. 3, pp. 1588–1595, Jun. 2004.
- [16] J.-S. Roger Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference Systems", *IEEE Trans. Syst., Man, Cybernatics*, vol. 23, no. 03, May, pp. 665–685, 1993.
- [17] J.-S. Roger Jang C.-T. Sun and E. Mizutani, "Chapter 19: ANFIS Applications" *Neuro-Fuzzy and Soft Computing: a computational approach to learning and machine intelligence*, Prentice Hall, 1997, pp. 503–534.
- [18] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, July 2002.

V. Gomathi obtained her B.E. Computer Science and Engineering from Thiagarajar College of Engineering, Madurai, INDIA. She has done M.Tech Computer Science Engineering at Indian Institute of Technology Madras (IITM), Chennai, INDIA. Currently pursuing PhD in the field of Video analysis.

She is working as Assistant Professor in Department of Computer Science and Engineering, National Engineering College, Kovilpatti, Tamil Nadu, INDIA. Her research interests are Video surveillance, Facial Expression recognition and Computer Vision.

Prof. V.Gomathi is a Life Member of ISTE and IE (I) chapters. She is the recipient of Prof. H.N. Mahabala Endowment Prize for her Best M.Tech thesis Award. Also recipient of IE(I) Young Engineers Award.