International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Artificial Intelligence for early Detection of Breast Cancer and Classification of Mammographic Masses

Sujata Patil, Shweta Madiwalar, V M Aparanji

Abstract—Every single year thousands of women endure painful and invasive surgery to remove breast lesions. Most of the time the mammographic image analysis leads to false positive detection and the majority of this actions reveal the lesions to be benign. Refining present detection and diagnostic tool is a major priority of our work. MATLAB R2015a is been used to develop the algorithm, which aids in detection of breast cancer in its early stage. The algorithm comprises of image processing and applying artificial intelligence where in the system is trained with a set of images so that when the input or the test image is given, the algorithm performs the image processing techniques and then applies the Probabilistic Neural Network (PNN) technique for detection of cancer. The system performance is also been calculated in order to estimate its reliability.

Keywords: Neural Network, Artificial Intelligence, Benign, Malignant, Probabilistic Neural Network (PNN),

I. INTRODUCTION

In Egyptian papyrus(Breasted, 1930) the existence of breast cancer was found from 3000BC. The innermost coating of milk vessels or lobules stock the milk in ducts [2]. It is clonal disease, solitary changed cell, the end outcomes is sequence of somatic (acquired) or germ line mutations is able to express full malignant potential [7]. The non-invasive illness or an invasive breast cancer can happen for elongated period but non metastatic disease.

Lobes and ducts forms the breast. 15 to 20 sections called lobes will be present in each breast. The reduced sections are called as lobules. Lobules end with many tiny bulbs which produce milk .the thin tubes are related by lobes, lobules and bulbs called as ducts .The monochrome liquid called lymph is almost carried by the blood vessels and lymph vessels present in the breast.

To decrease the high amount of redundant breast biopsies, numerous computer-aided analysis systems have been projected in recent years[3] The designed systems support physicians in making the better decision to operate on breast biopsy on apprehensive lesion observed on mammogram or accomplish immediate follow-up inspection instead. Artificial intelligent procedures as neural networks has shown great potential in scientific analysis field [19]. The intelligence showed by machines is called as Artificial Intelligence (AI, also machine intelligence, MI), in distinction to the intellect showed by people and animals. AI research is well-defined learning of "intelligent agents" in the discipline of computer science[4]. In supercomputer discipline devices which notices the atmosphere and reacts takes the best advantage of its chance of effectively achieving its goals. when a machine mimics "cognitive" functions the term "artificial intelligence" is used .that humans associate with other human minds, such as "machine learning" and "problem solving".

In this project we have used MATLAB (R2015a) tool for image processing and detection of cancer in which Gray Level Co-occurrence Matrix (GLCM) is used for feature extraction and probabilistic neural networks (PNN) technique is used for the persistence of detection and cataloguing of breast cancer [6]. Decision making is achieved in two phases: training the classifiers with structures from the catalogue and then testing[4]. The performance is evaluated in terms of entropy, energy, contrast, homogeneity, and correlation. As the testing image is fed to the system, it will automatically compare the image.

with the training set images then we classify image either as benign or malignant. In this way the system helps the physician in classification of the image and for further treatment[17].

II.METHODOLOGY

The flow chart is divided into two equal parts: Training of images, testing of images

1) Training of images

Training images are a bunch of images for which the required outcome is known. You input them to a program that analyse their features and passes the features through a classification routine that determines the appropriate weights to use on the features in order to best achieve the required result [14]. An algorithm is trained by using training data. Certain percentage of the overall dataset is used as training data and remaining is used as test data. To achieve the better classifier performance the training data is needed should very large[5] .Fig.1 shows the flow chart for detection of breast cancer and steps involved .

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.

Dr.Sujata N Patil*, Associate Professor at KLE Dr.M.S.Sheshgiri College and Engg, Belagavi,India.sujata.patil@klescet.ac.in.com,

Prof. Shweta M Madiwalar, Assistant Professor at KLE Dr.M.S.Sheshgiri College and Engg ,Belagavi, India shwetamadiwalar85@gmail.com

Dr. V M Aparanji, Assistant Professor at Siddaganga Institute of Technology, Tumkur, India. vmasit@rediffmail.com

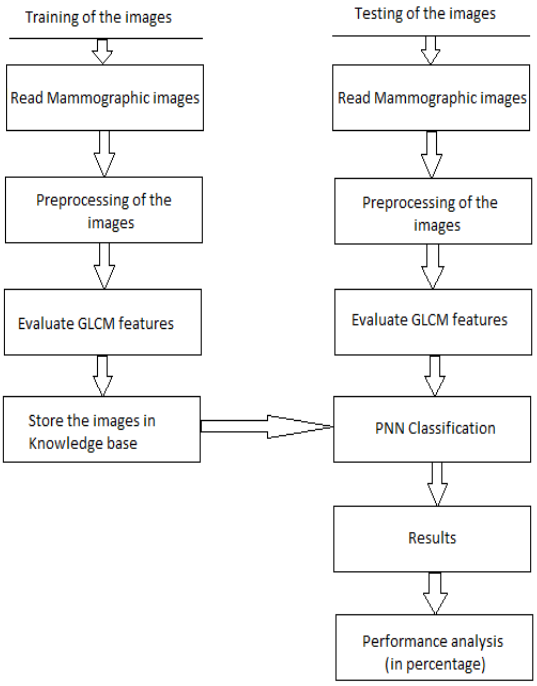

Figure .1: Flow chart for detection of breast cancer with Artificial intelligence

A.Read Mammographic images

An input image from the scanning units such as Mammogram, Ultrasound, CT, MRI etc. is taken and is given to the system for analysis.

B.Pre-processing of the images

When operating with the lowest level of abstraction pre-processing operations are used both output and output are intensity images [1][15]. The image function values (brightness) by matrix is characterized by the intensity of the image matrix. The main goal of pre-processing operations is to improve the image data that suppresses distortions or improves some features of the image which play an important role in processing, the pre-processing methods are classified as geometric transformations of images (e.g. rotation, scaling, and translation) [19]. In this project we have first cropped the image, adjusted the image pixel values, removed the unwanted background and extracted the required features of the image using GLCM evaluation.

C.GLCM Evaluation

over the image the distribution of co-occurrence pixel values (grayscale values) at given offset is defined by co-occurrence distribution of matrix. How often different combinations of pixels brightness values occur in an image is given by GLCM tabulation..The spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM),is considered by examining texture using statistical method.

The GLCM functions describe the texture of an image.In an image the GLCM is created by calculating how commonly the pairs of pixel with the same exact value in spatial relationship[8]. The several statics can be derived once we create GLCMs.the information about the texture of an image is provided by statistics the obtained by GLCM[5].

In an image the GLCM is matrix which contains the number of gray level is equal to the number of columns and rows .the matrix elements gives the relative frequency of two pixel separated by distance of ΔX and ΔY.the pixel occurs within given neighbor hood with intensity of i,and j. Lot of momentary data storing is possible by using large number of intensity levels G.i.e. G × G matrix for each grouping of (ΔX, ΔY) or (d, ɵ). the GLCM's is very composite to the size of the texture trials on which they are projected due to their huge dimensionality[16][17]. The amount of gray levels is frequently reduced [9]. An example showed with the four different gray levels can be used to explain the GLCM matrix. The bigger offset is possible if the window is large . one pixel offset is used.(a location pixel and its immediate neighbour). The number of times the combination (0,0) occurs the value will be stored in the topmost left cell. Once all the required features are extracted the next step is to store the extracted values in the knowledge base so that they can be used in the further process [7]. For the testing set similar procedure is carried on and then PNN technique is been applied to the images testing.

D. Entropy

The entropy gives the information about the loss of message in transmitted signal and also measures image information .Entropy gives the details of amount of information needed for the image compression .

Statistics	Description
Entropy	When all the elements of matrix are equal then entropy attains maximum value .The condition of the image can be measured .
Contrast	In gray level co-occurrence matrix contrast value gives the local variation of the matrix.
Correlation	For the specified pixel pair the correlation gives the measurement of the joint probability occurrence .
Energy	The energy is also known as angular second moment or uniformity which gives the sum of squared essentials in the GLCM..
Homogeneity	immediacy distribution of elements in GLCM to the GLCM diagonal.

Table 1: GLCM feature description

E. Contrast

In brief method it is also called as CON. The additional term of contrast is 'Sum of Square Variance'. The design of the intensity contrast connecting pixel and its neighbour over the whole image. 0 is the contrast value for the constant image. Weight increases exponentially as continues from the diagonal in contrast measurement[19] .

F. Correlation

The linear dependency of grey levels is measured by its correlation of neighbouring pixels [11].The tracing and image recording approaches is employed by optical method in digital image correlation. for precise 3D and 2D dimensions of variations in images [7]. It is used in the measurement of distortion, movement, straining and optical movement, extensively useful in many parts of engineering and science .

In measuring the motion of an optical mouse it is most commonly used.

G. **Energy**

For doing the work the energy is used .The texture is used to compute the order of an image . Energy gives the sum of square elements in GLCM[13]. Entropy gives total different from energy when the space is proficiently arranged .high value of Energy achieved . Energy is used as the square root of angular Second Moment of the texture character.0 to 1 range is used ..The value is 1 for the constant image [9].

H. **Homogeneity**

In short term it is also called as HOM.it allows us to calculate the value that defines the tightness distribution of the elements in GLCM diagonal and GLCM.the range is 0 to 1.the diagonals GLCM has value 1.

I. Store image in knowledge base

The database of the software stores the Trained images obtained It includes various feature of the images, which further act as the source / base for testing on which the conclusion are made.

2) Testing of images

The training images are used to determine the best weights and procedures to follow, you then pass through a set of "test" images, for which you also know the outcome. The images are extracted and the procedures and weights determined above are applied to make a prediction of how the image should classify [1]. Then the predicted classification is compared to the actual known classification to determine how well the procedures and weights do on analysing data that was not originally used to determine the weights. If you got a high score on the training images but a low score on the test images then the implication is that the routine "over-trained" and has become too specific to the training images and not able to predict for images whose results are not known. Step of reading mammographic images, pre-processing and GLCM feature extraction are same as the training set process.

3) Probabilistic neural network (PNN Technique)

It is widely used in classification and pattern recognition problems this algorithm is know as feedforward neural network[10]. In this method the parental probability distribution function of each group is approximated by parcen window and non-parametric function[18]. Using PDF the class probality of new input data is estimated for each class and then Bayes rule is applied to allocate the class with highest posterior probability to new input data By this method, the probability of mis-classification is minimized[20].

If the spread is value is nearby zero then the network behaves like nearest neighbor classifier. If the value of spread is larger then the designed network is taken into account by many nearby design vectors[16].

Example

With a set of inputs `S` and class indices `T1` the classification problem is defined S = [1 2 3 4 5 6 7];
T1 = [1 2 3 2 2 3 1];
The target vectors are obtained by class indices to design PNN.
T1 = ind2vec (T1)
net = newpnn (S, T1);
Y = sim (net, S)
Yc = vec2ind(Y)

A. Advantages of PNN over other neural network techniques

Requires less time to train the images. Even with few training sets it can give more accurate results. After applying the PNN technique the algorithm automatically give the result and classifies the given test images as benign
or malignant for cancerous breast images [6].

III.PERFORMANCE ANALYSIS

The performance analysis is carried out for testing the performance of the algorithm [14]. This is necessary as it gives the reliability of the algorithm in terms of percentage. In our system the performance analysis is carried out by testing of the known images, the system will predict whether the image is malignant or benign. This prediction is compared with the already saved database in the system [7]. Based on similarity the system will give the performance analysis in terms of percentage. The accuracy of the system is calculated by the following equation

Accuracy= Sum of true positive + sum of true negative

Sum of total population

Test images	Actual image test	System prediction	Analysis
1	Malignant	Malignant	True positive
2	Malignant	Malignant	True positive
3	Malignant	Malignant	True positive
4	Malignant	Malignant	True positive
5	Malignant	Benign	False negative
6	Malignant	Malignant	True positive
7	Malignant	Malignant	True positive
8	Malignant	Malignant	True positive
9	Malignant	Malignant	True positive
10	Malignant	Malignant	True positive
11	Malignant	Malignant	True positive
12	Malignant	Benign	False negative
13	Malignant	Benign	False negative
14	Malignant	Malignant	True positive
15	Malignant	Malignant	True positive
16	Malignant	Benign	False negative
17	Malignant	Benign	False negative
18	Malignant	Benign	False negative

19	Malignant	Benign	False negative
20	Malignant	Malignant	True positive
21	Benign	Benign	True negative
22	Benign	Benign	True negative
23	Benign	Benign	True negative
24	Benign	Benign	True Negative
25	Benign	Benign	True negative
26	Benign	Benign	True negative
27	Benign	Benign	True negative
28	Benign	Benign	True negative
29	Benign	Benign	True negative
30	Benign	Benign	True negative

Table 2.Performace analysis table

A. Receiver operating characteristics (ROC) Curve

ROC analysis is commonly used to evaluate medical tests and, in order to get an understanding of its use. It is important to distinguish between disorder and diagnosis. A patient either has or has not a specific disorder during the period of testing [15].The ROC curve is created at various threshold settings by plotting the false positive rate (FPR)against true positive rate (TPR).The sensitivity is also know by true positive rate .The analysis of the results falls above the grey line as shown in below figure.3 then the system is said to be reliable and can be implemented for various applications else the system is not reliable.

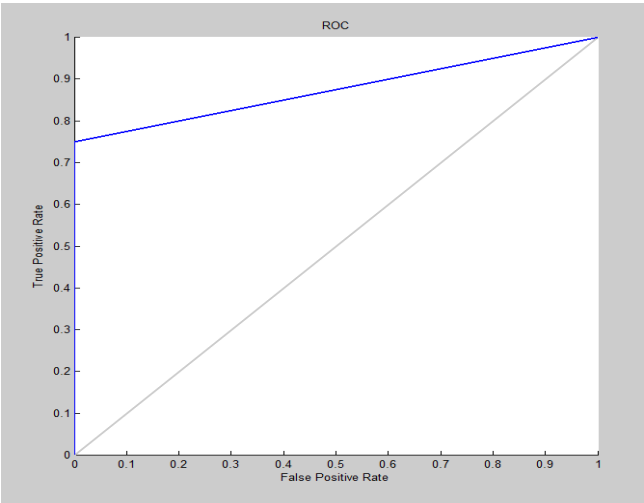

Figure .3: ROC performance analysis curve

IV.RESULTS

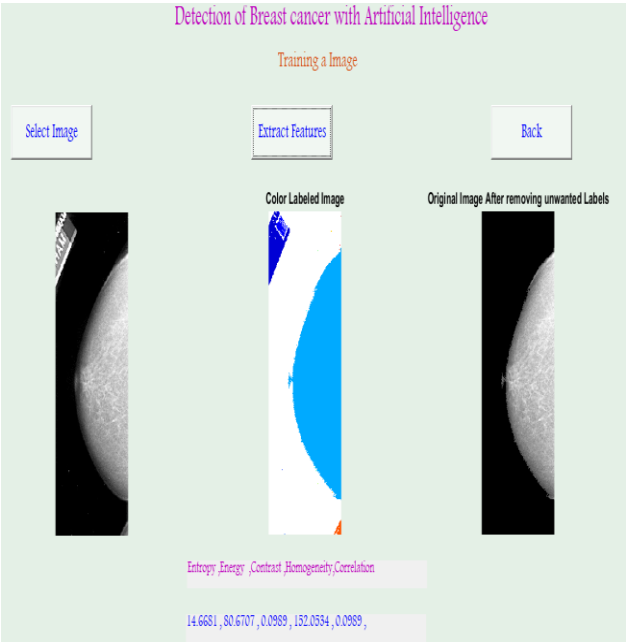

Figure.4: Output window for training of an image

The training of image process is carried to extract the features like Energy, Entropy, Contrast, Homogeneity, Correlation , stored in the database. The process of training includes selecting the image, extracting features and obtaining a processed image as shown in the below window. The training of images can be carried out on multiple number f images [7] [1].

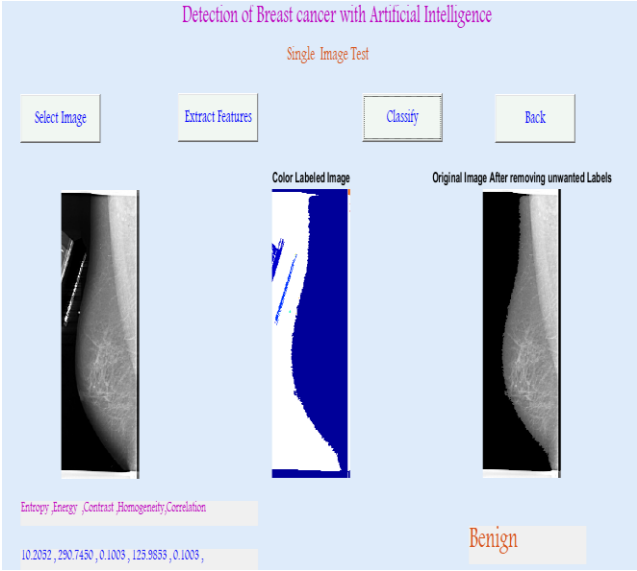

Figure .5 Output windows for classification of malignanc

The single test image test is carried out for testing of any random image. The process includes selecting of the image, extracting features, and classifying them either as **'benign' or 'Malignant'**.

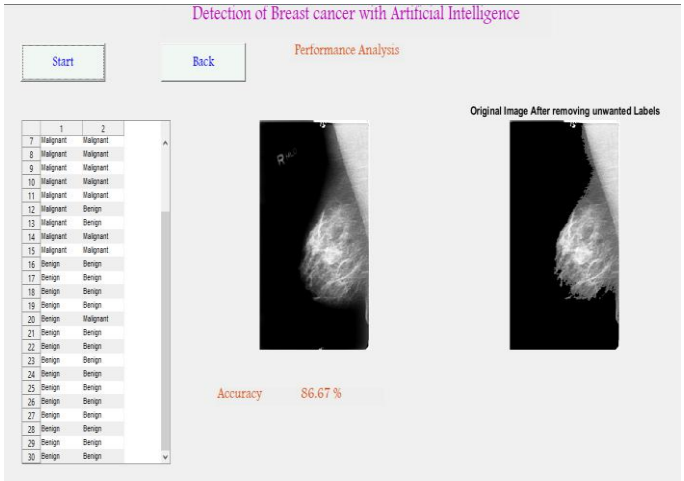

Figure 6. Output window for classification of Benign tumor

This test is carried out to check the mortality rate of the system. This process is done by comparing with actual values and the predicted values of the system. Based on the performance check the Accuracy [14] of the system .

V CONCLUSION

In this project we have developed the algorithm for the early detection of breast cancer using AI (PNN) technique. This technique reduces the number of iterations performed in the detection of the tumor. It helps physician in early detection of cancer with more accuracy and saves much of his time.an Algorithm is developed to check the performance of the system, which predicts how reliable the system is in terms of percentage. In this project the accuracy obtained is around **86.67%**. This accuracy can be increased with the increase in the training set of images.

ACKNOWLEDGMENT

The project was taken as a part of the study from the Hospital (where they do it subjectively) for the early detection of breast cancer. We thank the team for their valuable guidance. The authors are thankful to the KLE Dr. Prabhakar Kore Hospital and Medical Research Center for all the support with the images and help us to understand the analysis part.

REFERENCES

1. Sujata N Patil, Uday V Wali, M K Swamy, "Deep Learning Techniques for Automatic Classification and Analysis of Human in Vitro Fertilized (IVF) embryos.", February 2018, Volume 5, Issue 2, Journal of Emerging Technologies and Innovative Research (JETIR).
2. Florin gurunescu, Marina gurunescu, smarandagurunescu and Elia El-Darzi "a stastical evaluation of neural computing approaches to predict recurrent events in breast cancer", 4th IEEE International Conference on Intelligent Systems, 2008, pp 38-43.
3. Nuryanti Mohd Salleh , HarsaAmylia Mat Sakim and Nor Hayati Othman "Neural Networks to Evaluate Morphological Features for Breast Cells Classification" IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.9, September 2008, pp 51-58.
4. Mohammad Samedi, Rabab Kreidieh Ward, Jacqueline MorganParkes and BrankoPalcic "Image Feature Extraction in the Last Screening Mammograms Prior to Detection of Breast Cancer" IEEE journal of selected topics in signal processing, vol. 3, no. 1, February 2009, pp 46-52.
5. Yuanjiao MA, Ziwu WANG, Jeffrey Lian LU, Gang WANG, Peng LI Tianxin MA, YinfuXIE ,Zhijie ZHENG "Extracting Microcalcification Clusters on Mammograms for Early Breast Cancer Detection" Proceedings of the 2006 IEEE International Conference on Information Acquisition August 20 - 23, 2006, Weihai, Shandong, China ,pp499-504.
6. Breast Cancer Diagnosis using Artificial Neural Network Models R.R.Janghel1 , Anupam Shukla2 , Ritu Tiwari3 , Rahul Kala4 1,2,3,4 Indian Institute of Information Technology and Management, Gwalior, India.
7. Sujata N Patil, Uday V Wali, M K Swamy, "Application of vessel enhancement filtering for automated classification of human In-Vitro fertilized (IVF) images" 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 27978-1-5090-4697-3/ 16/ $31.00 ©2016 IEEE .
8. Claudio Manna, Loris Nanni, Alessandra Lumini , Article in reproductive Biomedicine online. October 2012 – Artificial intelligence techniques for embryo and oocyte classification.
9. Dr.R.H.Havaldar, Prof.S.S.Ittannavar "Comparative study of Mammogram Enhancement Techniques for Early Detection of Breast Cancer" International Journal of Technology and Science(online)2350-1111,volume IX, issue1, 2016pp.5-8.
10. Hanung Adi Nugroho, Faisal N, Indah Soesanti "Analysis of Computer Aided Diagnosis on Digital Mammogram Images "2014 International Conference on Computer, Control, Informatics and Its Application Bandung, Indonesia.
11. Umesh ,B Ramachandra "Association rule mining based predicting breast cancer recurrence on SEER breast cancer data" International Conference on Emerging Research in Electronics, Computer Science and Technology – 2015 Mandya, India
12. Sagar Metri and Asha T. "Patch Based Wiener filter for Image Denoising" International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018.
13. Sagar Metri and Raghuvamsa GH "Performance improvement in WiMAX networks with femtocells" International Conference on Computing, Electronics and Electrical Technologies, ICCEET 2012.
14. Sujata N Patil, Uday V Wali , M K Swamy, "Selection of Single Potential Embryo to Improve the Success Rate of Implantation in IVF Procedure using Machine Learning Techniques" 2019 International Conference on Communication and Signal Processing (ICCSP), Tamilnadu. 0881-0886 IEEE.
15. Sujata N Patil, Uday V Wali, M K Swamy, "Performance Analysis And Classification of Human In Vitro Fertilized (Ivf) Embryos Using Vesselness Filters and Hough Transform Algorithm", Int J Recent Sci Res. 9(1), pp. 23475-23479.
16. Rastghalam, R., & Pourghassem, H. (2016). "Breast cancer detection using MRF-based probable texture feature and decision-level fusion-based classification using HMM on thermography images". Pattern Recognition, 51, 176-186.
17. Sree, S. V., Ng, E. Y. K., Acharya, R. U., & Faust, O. (2011). "Breast imaging: a survey". World journal of clinical oncology, 2(4), 171.
18. Gorge, P., Serbs, A., & Uçan, O. N. (2015). "Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines". Expert Systems, 32(1), 155-164.
19. Butteroil, T., Lotz, T., Kashif, A., & Chase, J. G. (2014). "Reconstructing 3-D skin surface motion for the DIET breast cancer screening system". IEEE transactions on medical imaging, 33(5), 1109-1118.
20. Hossain, M. D., & Mohan, A. S. (2017). "Cancer Detection in Highly Dense Breasts Using Coherently Focused Time-Reversal Microwave Imaging". IEEE Transactions on Computational Imaging, 3(4), 928-939.

AUTHOR PROFILE

Dr. Sujata N Patil received her PhD from the KLE Academy of Higher Education and Research (KLE University) University, Belagavi, India. ANN, CNN, Machine Learning, Deep Learning Artificial Intelligence, and Neutral Networks applications. She is currently Associate professor in the Department of Electronics and Communication Engineering, KLE's Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi, Karnataka, India. She has published over 5 academic papers in conferences and 2 papers in journal. **Dr. Sujata Patil** is a member of IEEE, ISTE.

Prof. Shweta M Madiwalar .M.Tech working as assistant Professor in the Department of Electronics and Communication Engineering, KLE Dr.M.S,Shesshgir college of Engg and Technology, Belagavi. India Her research interests include cryptography, machine learning, deep learning artificial intelligence, neutral networks applications She has published over 3 academic papers. Prof. Shweta. is a member of ISTE.

Dr. V M Aparanji received Ph.D from VTU, Belagavi. Her research interests include ANN, CNN, Machine Learning, Deep Learning Artificial Intelligence, and Neutral Networks applications. She is currently Assistant Professor in the Department of Electronics and Communication Engineering, at Siddaganga Institute of Technology, Tumkur, and Karnataka, India. She has published over 5 academic papers in conferences and 2 papers in journal. **Dr. V M Aparanji** is a member of ISTE.