Vol. 1, 18
Proceedings of SPIT-IEEE Colloquium and International Conference, Mumbai, India

Breast Cancer Malignancy Identification Using Jordan Method

Sudhir D. Sawarkar, [1] Rajesh D. Kadu [2] and Amol P. Pande [3]

***Abstract*--An artificial neural network (ANN) is an information-processing paradigm inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information. The key element of the ANN paradigm is the novel structure of the information processing system. Learning in ANN typically occurs by example through training, or exposure to a set of input/output data where the training algorithm iteratively adjusts the connection weights. The SRN was originally developed to learn time-varying patterns or temporal sequences, specifically, character string. For breast cancer malignancy identification using the neural network various databases available on the Internet as well as gathered information from hospitals is used.**

***KeyWords*--Jordan ANN, breast cancer, efficiency, Specificity, receiver operating characteristics**

I. Introduction

IN developed countries such as the USA and other Western countries, it is the most commonly diagnosed cancer in women, causing about 16% of all deaths due to cancer. Worldwide, 308,000 women died of breast cancer in 1985, with annual totals expected were around 340,000 deaths in 1990 and 420,000 by 2000. The most recent statistics by Asian Project Report 2002 clearly shows the severity of the disease.[1] Currently, breast imaging for the detection and characterization of suspicious breast lesions relies upon mammography and ultrasound. Mammography is the modality of choice for early detection of breast cancer. Although mam[1]mography is very sensitive at cancer detection, it results in many false positives. As few as 20% of currently biopsy cases actually reveal cancer. The remainders are all benign cases, which underwent a potentially unnecessary surgical procedure.

Preventing benign biopsies is the most important way to improve the efficacy of mammography screening, especially as screening becomes more widespread. This clearly demonstrates a need for efficient breast cancer detection and diagnosis techniques. Some of the works done in this direction include linear programming approach Mangasarian,1995), machine-learning approach (Wolberg, 1994). [2,3,4]

II. DATA

This work grew out of the desire by Dr. Wolberg [2,3,4] to accurately diagnose breast masses based solely on a Fine Needle Aspiration (FNA). The feature extraction process is performed as follows:

An FNA is taken from the breast mass. This material is then mounted on a microscope slide and stained to highlight the cellular nuclei. A portion of the slide in which the cells are well differentiated is then scanned using a digital camera and a frame-grabber board and identified nine visually assessed characteristics of an FNA sample, which he considered relevant to diagnosis. The resulting data set is well-known as the Wisconsin Breast Cancer Data.

(Total attributes 9.; Number of instances- 699; Missing attributes-16; Benign- 458; Malignant-241.)

The reported sensitivity (i.e., ability to correctly diagnose cancer when the disease is present) of mammography varies from 68% to 79%, of FNA with visual interpretation from 65% to 98%, and of surgical biopsy close to 100%. Therefore, mammography lacks sensitivity, FNA sensitivity varies widely, and surgical biopsy, although accurate, is invasive, time consuming, and costly. The goal of the diagnostic aspect of this paper is to develop a relatively objective system.

III. JORDEN NEURAL NETWORK

The simple recurrent network, often referred to as the Elman network [5], is a single hidden-layer feedforward neural network. However, it has feedback connections from the outputs of the hidden layer neurons to the input of the network. This network is similar to an architecture proposed by Jordan [6]. The SRN was originally developed to learn time-varying patterns or temporal sequences, specifically, character strings. The Jordan and Elman network is shown in figure 1. The function of the context units is to replicate the hidden layer output signals at the previous time step. The algorithm used for Jordan network is shown in figure 2.

[1] Dr. Sudhir D. Sawarkar (sudhir_sawarkar@yahoo.com),
[2] Rajesh D. Kadu (raj69kadu@yahoo.com) and
[3] Amol P. Pande (amol_ap@yahool.com) are with Department of Computer Engineering, Datta Meghe College of Engineering, Airoli. Navi Mumbai, Mumbai University(INDIA) Pin: 400708

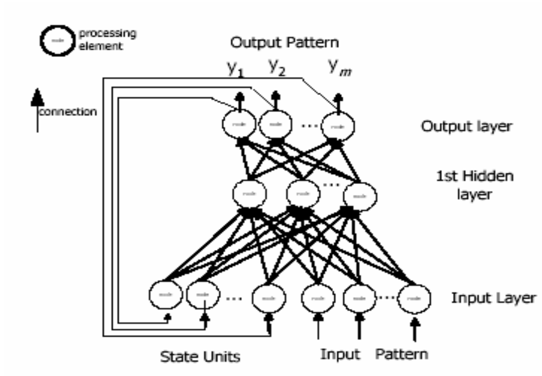

Figure 1 Configuration of Jordan network

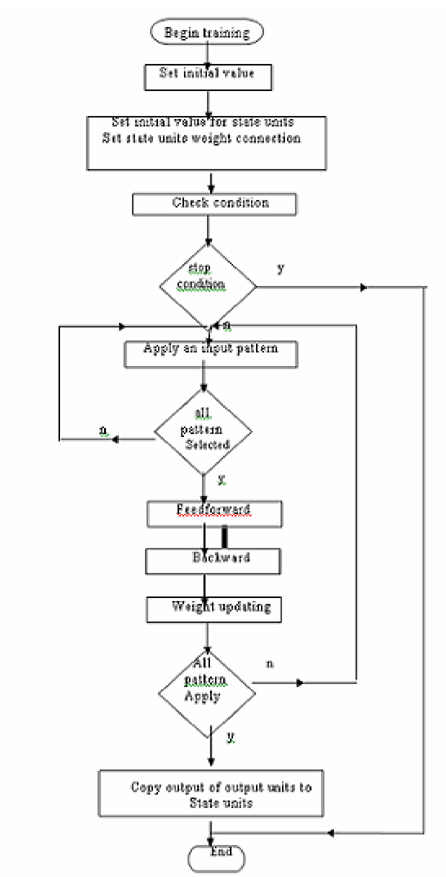

Figure 2. The algorithm for Jordan neural network

IV. RESULT ANALYSIS

A. Pred value

The accuracy or efficiency of the diagnosis and prognosis of breast cancer by ANN is evaluated by using the magnitude of relative error which is calculated using formula:

MRE=Abs((Actual detection–Desired detection) /Actual detection)…………………………………………………..4.1

Measure of Average efficiency is calculated using:

Pred(p) = if (Mre < 0.25,1,0) …………………………...4.2

Pred(p) = K/N [7]

N total no of Historical Data

K is number of cases output with MRE less than or equal to p.

Pred (0.25) gives % of input that were predicted with an MRE is less than 0.25. Efficiency calculated always in %.

B. Specificity, sensitivity, receiver operating characteristics and area under curve

Classification efficiency has been widely used as the main criterion for comparing the classification quality of classifiers. Initially, if the class distribution is skewed rather than constant and relatively balanced in the real world, then the evaluation based on accuracy breaks down. [8,9,10] later, classification accuracy assumes equal misclassification costs (for false positive and false negative errors), which is problematic because for real-world problems one type of classification error is much more expensive than the another, e.g. classifying a healthy

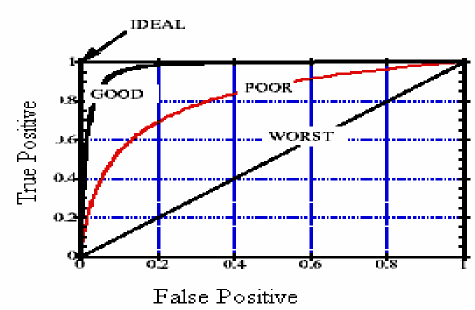

Figure 3. ROC curves

woman to have breast cancer and classifying a breast cancer patient to be healthy will have different misclassification cost, since the latter may cost the patient's life.

True Positive (TP): the number of correct positive predictions.

True negative (TN): the number of correct negative predictions.

False Positive (FP): the number of incorrectly positive predictions.

False negative (FN): the number of incorrect negative predictions.

True Positive (TP) rate: TP/(TP+FN), true positive examples over total positive examples.

False Positive (TP) rate: FP/(FP+TN), false positive examples over total negative examples.

An alternative criterion is sensitivity and specificity analysis, [11] with which sensitivity measures the ability of a test to be positive when condition is actually present, or numbers of the positive test examples are recognized.

$$sensitivity = \frac{truePositiveExamples}{totalPositiveExamples} = \frac{TP}{TP + FN} \quad \text{………4.3}$$

and specificity measures the ability of a test to be negative when the condition is actually not present, or how many of the negative test examples are excluded:

$$specificity = \frac{trueNegativeExamples}{totalNegativeExamples} = \frac{TN}{FP + TN} \quad \text{………4.4}$$

ROC stands for receiver operating characteristic, which is increasingly practiced in the machine learning research field. ROC graphs have long been used in signal detection theory to depict tradeoffs between hit rate and false error rate. ROC analysis has been extended for use in visualizing and analyzing the behavior of diagnostic systems, and is used for visualization in medicine etc. [9,10,11,12,13,14]

An ROC curve is a two-dimensional depiction of classifier performance, on which *TP* (short for *true positive rate*) is plotted on the *Y* axis and *FP* (short for *false positive rate*) is plotted on the *X* axis. The (*FP*, *TP*) pairs (each pair comes from a confusion matrix) vary together as a threshold on a classifier's continuous output is varied between its extremes, and the resulting curve is called the ROC curve. A confusion matrix can be illustrated as following, in which *TP rate* and *FP rate* are values between 0 and 1. The point (0,0) means predicting all cases to be negative, (0,1) represents a perfect classifier, classifying all positive cases and negative cases correctly, (1,0) is the worst classifier, incorrect for all classifications, and (1,1) means predicting all cases to be positive. The closer to the point (0,1) the better the classifier.

C. ROC and AUC

Fig. 3 shows ROC curves. Therefore, although an ROC curve is a valuable visualization technique, ROC analysis does a poor job of aiding the choice of classifiers. [9] AUC (area under the ROC curve) has been recently used as an alternative measure for machine learning algorithms. AUC is calculated using Trapezoidal rule which is shown in figure 4.

There many advantages of AUC, such as its independency to the decision sensitivity in analysis of variance (a collection of statistical models and their associated procedures which compare means by splitting the overall observed variance into different parts) tests, its independency to the decision threshold, and its invariance to *a priori* class probability (recognized in advance as equally probable) distribution etc.

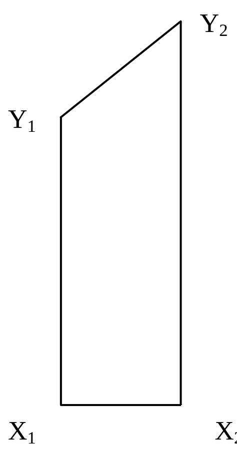

Figure 4 Trapezoid area calculation

Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1. However, because random guessing produces the diagonal line between (0, 0) and (1, 1), which has an area of 0.5, no realistic classifier should have an AUC less than 0.5. [15,16,17,18,19,20,21]

It has feedback connections from the outputs of the hidden layer neurons to the input of the network. This network is similar to an architecture proposed by Jordan. The modified back-propagation algorithm proposed by Jordan is used. [6]
Experimental results are mainly categorized between two phases namely training phase and testing phase.

This experiment is aimed at finding the efficiency using Jordan neural network. For this total 480 experiments are performed by selecting different sets of training and testing samples out of 683 samples of data set .

Activation functions selected,
TanhAxon,
SigmoidAxon,
Linear TanhAxon,
Linear SigmoidAxon,
SoftMaxAxon,
BiasAxon,
LinearAxon, and
Axon
Momentum = 0.7, Step size: Hidden layer =0.1, output layer =0.1
Number of Hidden layer = 1, 2, 3
Number of neurons in hidden layer = 4
Number of epochs = 5000

The artificial neural network is trained with proper training parameters. Training returns new values of weights and biases. The training errors are plotted with respect to epochs. The example with activation function TanhAxon and hidden layer 2 training error versus number of epoch is plotted and as shown in figure 5. For 20 readings with activation function TanhAxon and hidden layer 2, efficiency versus number of testing samples is shown in figure 6. For the same testing error versus number of training samples is shown in figure 7.

The percentage efficiency calculated for 20 experiments each, using different activation functions, varying number of hidden layers (1, 2, 3) and varying number of training and testing samples. The neural network is trained for 5000epoch.

Result Analysis: Figure 8 shows keeping hidden layer as 3 for BiasAxon 97.05% efficiency is obtained which is good as compare to other activation functions.
Figure 9 shows overall training error i.e. Minimum MSE is good for TanhAxon.
Figure 10 shows overall training error i.e. Final MSE is good for TanhAxon.
Figure 11 shows overall testing error (MSE) is good for TanhAxon.
From above observations, it is seen that the Jordan neural network is trained for transfer function BiasAxon with hidden layer 3. It is also observed that for TanhAxon function and Linear TanhAxon function, the percentage efficiency increases as hidden layers increases. For Sigmoid function, the efficiency is nearly Zero for hidden layer 2 and 3 indicating no pattern gets detected.

The typical ROC curve for Jordan neural network is shown in figure 11. From ROC curve of Jordan neural network the area under curve (AUC) is calculated using trapezoidal rule. The area under curve is found to be 1.

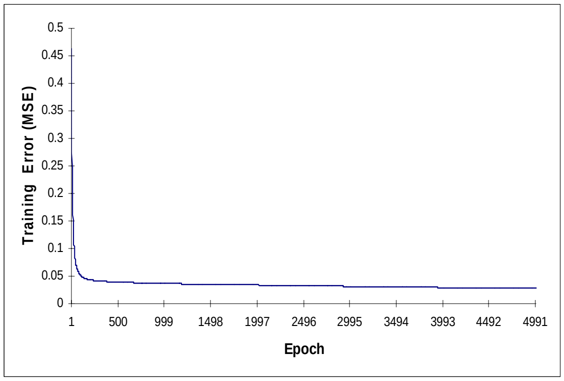

Figure 5. Training error versus number of epochs

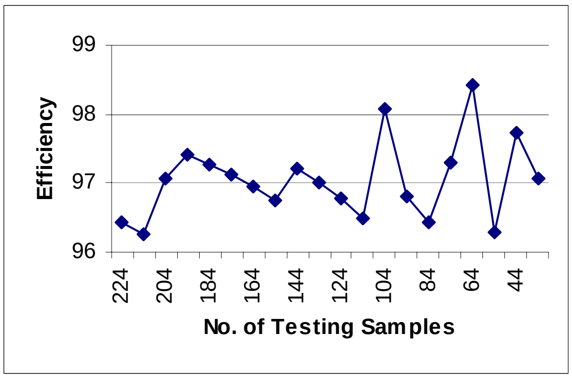

Figure 6. Efficiency versus number of testing samples

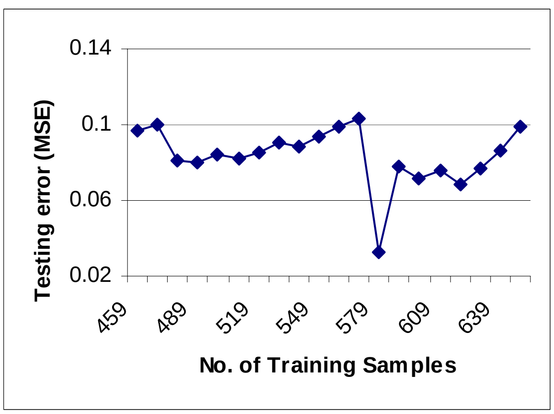

Figure 7. Testing error versus number of training samples

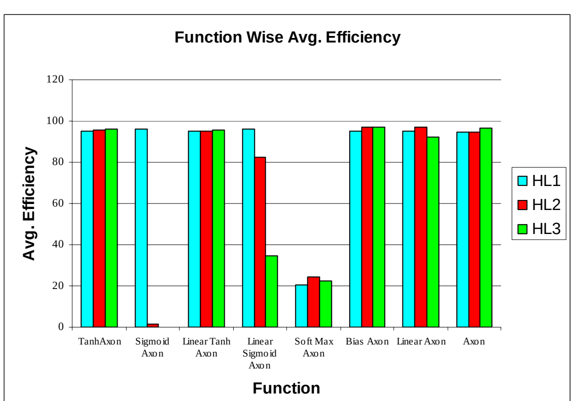

Figure 8. Activation function versus average efficiency in % (JOR NN)

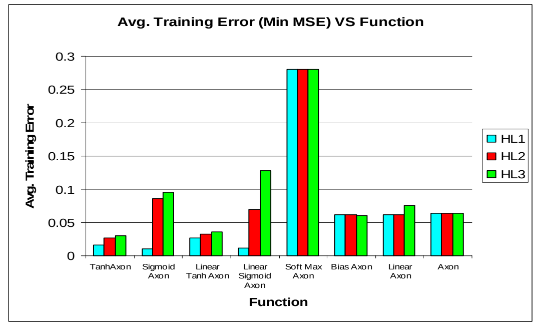

Figure 9. Activation function versus average training error (JOR NN)

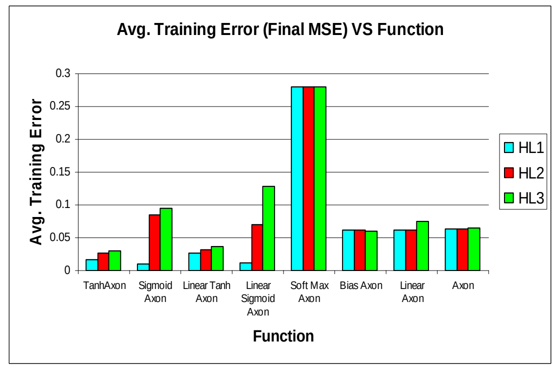

Figure 10. Activation function versus average final training error (JOR NN)

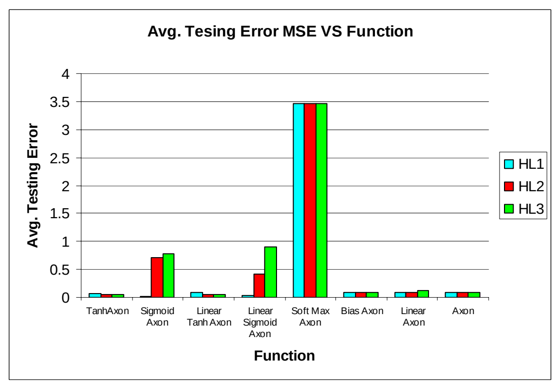

Figure 11 Activation function versus average testing error (JOR NN)

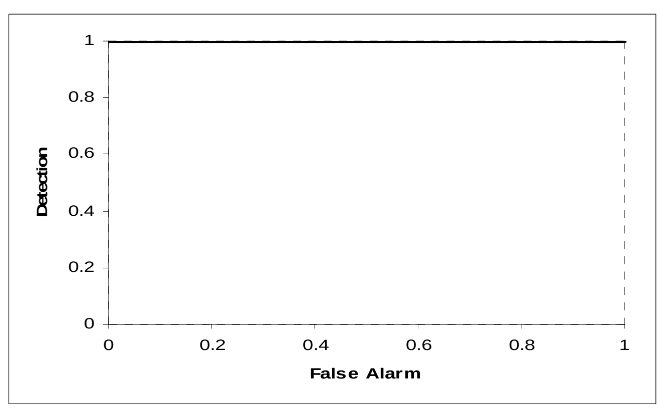

Figure 12. ROC curve for JOR NN model on testing dataset

V. CONCLUSION

Neural network aided breast cancer diagnosis gives promising results. It can be a good supplement to the conventional clinical diagnosis system. This study shows the decision taking ability of artificial neural network model.

For diagnosis the efficiency of Jordan neural network shows that it can support the doctors or physicians to consider it as a second opinion of the learning machine to prevent biopsy. In addition, this neural network based clinical Decision Support System avoid unnecessary excision and expenses.

Using Jordan neural network, the breast cancer diagnosis is comparably accurate than the human being. The efficiency of the Jordan neural network is nearly 97.05%.

The optimal classifier activation function of neurons for the Jordan neural network is seen to be Bias Axon. With this activation function the estimated classifiers are able to discriminate between the malignant and the benign patients. The proposed classifiers are seen to be a good supplementary aid for the physicians to make precise clinical decisions pertaining to diagnosis.

VI. FUTURE SCOPE

It may be possible to improve efficiency of module that has currently been implemented. Some of the areas in which future work can be done to make the present scheme more effective are:

The neural network can be realized in hardware with the programmable weights and biases so that the real time capability of the scheme can be effectively analyzed.

The network has to be trained with more input patterns, so that the generalization ability of network will be enhanced.

The present study uses Euclidean norm (L2 norms) to compute the MSE between the output of the neural network model and the desired one. Possibility of other Lp norms (p>2) for the calculation of the MSE may be investigated to suggest an optimal cost function.

VII. REFERENCES

[1] 2003: Imag (NBCC) National Breast Cancer Center websitewww.nbcc.org.au/pages/info/resource/nbccpubs-/bc21-94/intro.htm

[2] O.L. Mangasarian, W.N. Street and W.H. Wolberg. Breast cancer diagnosis and prognosis via linear programming. Operations Research, 43(4), pages 570-577, July-August 1995.

[3] W.H. Wolberg, W.N. Street, and O.L. Mangasarian. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. Analytical and Quantitative Cytology and Histology, Vol. 17 No. 2, pages 77-87, April 1995.

[4] W.H. Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning techniques to diagnose breast cancer from fine-needle aspirates. Cancer Letters 77 (1994) 163-171.

[5] J. L. Elman, "Finding Structure in Time," Cognitive Science, vol. 14, 1990, pp. 179-211.

[6] M. I. Jordan, "Serial Order: A Parallel Distributed Processing Approach," Institute for Cognitive Science Report 8604,San Diego: University of alifornia, 1986.

[7] Ali Idri, Taghi M Khoshgoftaar, "Can Neural Network be easily interpreated in software cost estimation", 2002, world congress computational intelligence, Honolulu, Hawaii.

[8] J. R. Quinlan and R. L. Rivest, "Inferring Decision Trees Using the Minimum Description Length Principle", Information and Computation, Vol. 80, pp. 227-248, 1989.

[9] J. R. Quinlan and R. L. Rivest, "Inferring Decision Trees Using the Minimum Description Length Principle", Information and Computation, Vol. 80, pp. 227-248, 1989.

[10] Provost, F. and Fawcett, T., "Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distribution" Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97), pp.43-48, 1997.

[11] Provost, F., Fawcett, T. and Kohavi, R., "The case against accuracy stimation for comparing induction algorithms", Proceedings of the Fifteenth International Conference on Machine Learning (IMLC-98), Madison, WI, 1998.

[12] Cios, K. J. and Moore, G. W., "Uniqueness of Medical Data Mining", Artificial Intelligence in Medical Journal, Vol.26, No.1-2, pp1-24, 2002.

[13] Egan, J. P., "Signal Detection Theory and ROC Analysis", Series in Cognition and Perception, New York: Academic Press, 1975.

[14] Swets, J. "Measuring the Accuracy of Diagnostic Systems", Science, Vol. 240, pp.1285-1293, 1988.

[15] Beck, J.R. and Schultz, E.K., "The Use of ROC Curves in Test Performance Evaluation", Arch Pathol lab Med, Vol.110, pp13-20,1986.

[16] Bradley, A.P., "The Use of the Area under ROC Curve for Multiple Class Classification Problems", Pattern Recognition, Vol.30, pp1145-1159, 1997.

[17] Hand, D. J. and Till R. J., "A Simple Generalization of the Area under the ROC Curve for Multiple Class Classification Problems", Machine Learning, Vol. 45, pp. 171-186, 2001.

[18] Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., "Classification and Regression Trees", Wadsworth International Group, 1984.

[19] Bishop, C, "Neural Networks for Pattern Recognition", Oxford: University Press, 1995.

[20] Drummond, C. and Holte, R. C., "Exploiting the Cost (In)sensitivity of Decision Tree Splitting Criteria", Proceedings of the 17th International Conference on Machine Learning (ICML'2000), pp. 239-246.2000.

[21] Fawcett, T., "ROC Graphs: Note and Practical considerations for Data Mining Researchers", HP Labs Technical Report (HPL-2003-4), 2003.

VIII BIOGRAPHIES

S. D. Sawarkar – received his BE(Electronics) and ME(Electronics) from Amaravati University, India in 1988 and 1995 respectively. He received his Ph.D degree in 2007 from Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra India. He has published total 14 research papers in national / international journals /conferences. He has guided M Tech ME dissertation. Dr. S. D. Sawarkar is currently working as a Professor in Computer Engineeringdepartment in Datta Meghe college of Engineering , Mumbai

R. D. Kadu – Received his BE(Computer) from Amravati university in 1991.He has received his M.tech.(Computer) degree from Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra India in 2003.

He has published research papers in national conferences. Curently he is working as Assistant Professor in Computer Engg Department, Datta Meghe College of Engineering Airoli, Navi Mumbai, India.

A. P. Pande-- Received his BE(Computer) from Amravati university in 1991.He has received his M.tech.(Computer) degree from Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra India in 2005.
He has published research papers in national conferences. Curently he is working as Assistant Professor in Computer Engg Department, Datta Meghe College of