Application of Machine Learning Techniques in Predicting Breast Cancer – A Survey

K.Prasuna, K.V.S.N.Rama Rao, CH.M.H.Saibaba

Abstract: ***In counties like India where population is on the rise and living conditions are improving, medical field remained the focal area. In the era of Artificial Intelligence, Machine Learning is getting prominence in diverse areas as it provides an accurate solution to wide range of problems in medical domains. Breast cancer is a disease that is posing a challenge to women health all over the world. The global scenario indicates that breast cancer stands in second place with regard to causing deaths among the females when it comes to the cancer casualities. However, breast cancer is a curable disease if it can be diagnosed early. Most of the deaths in women between the age of 40 and 55 are due to breast cancer. As per the WHO report, approximately 1.2 millions of people are suffering from breast cancer every year all over the world . This paper aims on discussing breast cancer in women and several machine learning techniques proposed by the researchers in diagnosing the disease.***

Index Terms:. Breast cancer, diagnosis, machine learning, prognosis, recurrence, survivability

I. INTRODUCTION

Technology has many facets and one of its wide applicability is found in the field of clinical research and diagnosis. Counties like India where population is on the rise and living conditions are improving, medical health remained the focal area. Adopting the technological developments in the field of medicine has resulted in making the healthcare affordable to larger sections of the society. Medical care for the fairer sex is gaining importance of late, as women of present generation are facing several challenges with regard to their health. Work load on the domestic front coupled with the challenges at the workplace are deterring them from taking adequate care about their health. As a result, a good number of women are subjected to ailments such as cancer. Breast cancer is an epidemic disease that is posing a challenge to women health all over the world. Unfortunately, there is no simple diagnosis to check whether a person is suffering cancer. Lack of awareness and the non-availability of diagnosis is resulted in untimely deaths of millions of patients every year. Technological development in cancer diagnosis is the need of the hour to curtail those deaths and to build a healthy society. It can be done by bridging the gap between the available technology and its interface.

In the era of Artificial Intelligence, Machine Learning is getting prominence in diverse areas. It provides an accurate solution in diagnosing several diseases. ML is the pioneer composing several methods, techniques and tools to solve the medical problems, it has found wide applicability in use for patient management. From disease prediction to knowledge extraction and for other research development activities, the role of Machine Learning has been increasing manifold.Understanding the algorithms coupled with providing mathematical justifications to their properties will help acquire valuable insight in terms of performance and behavior. However, in medical context the need for clarity in the diagnosed result remained a common concern for the Machine Learning applications in the medical context. In this paper, we are discussing the breast cancer in women and the different algorithms proposed by the researchers in diagnosing this disease. Our study helps the researchers to know the different machine learning techniques to predict breast cancer. The subsequent sections of this paper includes as follows: Section II.A discuss about the predictive models for the prognosis and diagnosis of breast cancer. Section II.B discuss about breast cancer survivability predictive models. Section II.C discuss about breast cancer recurrence predictive models.

Revised Manuscript Received on June 05, 2019

K.Prasuna, Department of CSE, Koneru lakshmaiah Education Foundation, Vaddeswaram, Guntur district, A.P, India.

K.V.S.N.Rama Rao, Department of CSE, Koneru lakshmaiah Education Foundation, Vaddeswaram, Guntur district, A.P, India.

CH.M.H.Saibaba, Department of CSE, Koneru lakshmaiah Education Foundation, Vaddeswaram, Guntur district, A.P, India.

II. PREDICTING BREAST CANCER – USAGE OF MACHINE LEARNING TECHNIQUES

Breast Cancer is largely predominant among women and leads to more and more casualties every year. It is second largest type of cancer when it comes to the toll, as the number of women succumbed to Breast Cancer has been increasing every year all over the world. The brighter side is that Breast Cancer is the more curable types of cancers provided if it is diagnosed in the initial stage. Several breast Cancer victims are losing their lives at the age of 40-55 years. World Health Organization's (WHO) recently reported that Breast cancer has been spotted among 1.2 million females across the globe [12]. Treatments such as chemotherapy, hormone therapy and radiation are being used as combinations to increase the survival rate of the victims [10].

Of late, many countries are maintaining electronic patient records for public health care management systems [2]. Data is collected and stored in large information systems known as Electronic Medical Records (EMR) in the modern hospitals [4]. Details like the hospitals, patients, diseases, types of treatment, treatment cost etc. are available in these records. Patient data usually contains missing values, as the patients with same disease are not expected to undergo the same lab examinations. These lab examinations differ from patient to patient according to their age, family history and the percentage of risk in the disease [4]. Patient data may be incomplete and incorrect comprises of several difficulties such as missing parameter values, noise in the data. The datasets are huge and complex to analyze and hence it is difficult to take a decision with regard to the

patient's health. The accurate prediction of disease became the most challenging task to the physicians.

The statistical methods deal with only numeric data whereas Machine Learning deals with multiple types of data such as images, speech, sounds, text, etc. Machine Learning handles both numeric and categorical data of complex datasets to identify the patterns and relationships among the features for diagnosing the outcome of the disease [3,4]. Machine learning techniques are also used to analyze medical datasets and serves as a major platform for intelligent data analysis in medical domain.

Hence, a comprehensive study of patient data will help understand the stages of Breast Cancer in women that undergo the screening test and the study is going to help women from all over the world.

Prediction of breast cancer focus on three tasks such as 1) breast cancer susceptibility prediction (prognosis and diagnosis) 2) breast cancer survivability prediction 3) breast cancer recurrence prediction. The taxonomy of works on prediction of breast cancer is represented in Fig 1.

A. PREDICTIVE MODELS FOR THE FORECAST AND DIAGNOSIS OF BREAST CANCER

Researchers are putting their efforts in finding innovative ways for detection of breast cancer at initial stages. Different machine learning techniques namely SVM, ANN are proposed by the researchers to predict the disease.

Maglogiannis et al [11] proposed SVM classifier in comparison with Bayes classifier and Neural networks models for the prognosis and diagnosis of breast cancer using Wisconsin dataset. Dataset was classified into four classes namely C1, C2, C3, C4 based on the recurrence time which then corresponds to the instances of recurrence between 1 to12 months, 1 – 3 years, 3 - 6 years, more than 6 years respectively. SVM classifier has been applied on each class of recurrence and the results were compared with Bayes classifier and ANN. These models were evaluated basing on precision, sensitivity, specificity and the experimental results proved that SVM classifier attained better accuracy of 96.91%, specificity of 97.67% and sensitivity of 97.84%.

M.F. Akay et al [12] proposed feature selection method combined with SVM as the predictive model for breast cancer analysis. Wisconsin breast cancer dataset is used by authors to evaluate the model. Results show that SVM model has achieved maximum accuracy of 99.51%. Ayer et al [13] analysed the mammographic data available from Wisconsin state repository system and developed a model using ANN to envisage the hazard of breast cancer. The data set consists of 62,219 patient's mammographic findings with 2 classes malignant and benign. The authors developed 3-layer feed forward ANN model with around 1000 hidden nodes. Model is ten-fold cross-validated and the AUC of their ANN is 0.965 which was significantly higher when compared to other radiologists. Vikas et al [16] analysed the breast cancer data available from Wisconsin dataset in predicting breast cancer. The dataset contains 699 examples, 2 classes (malignant and benign) and 9 integer valued attributes. As a part of pre-processing 16 instances were removed and used the dataset with 683 instances. The authors used three classification algorithms sequential minimal optimisation (SMO), IBK and BF tree and compared the results using WEKA tool. Results proved that sequential minimal optimisation(SMO) has the highest prediction accuracy of 96.2%.

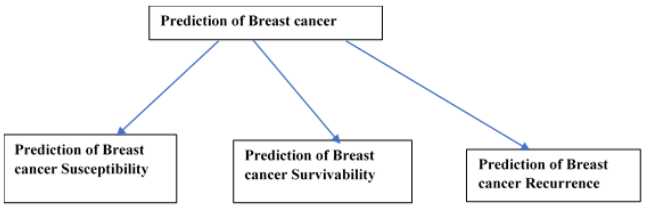

Fig 1: Taxonomy of works on prediction of breast cancer

Abdel et al [17] proposed a computer aided diagnostic (CAD) scheme for the discovery of breast cancer using deep belief networks and back-propagation. The authors used Wisconsin dataset (WBCD). In the related work, the authors applied deep belief network (DBN) to learn the input features in the unsupervised phase. The obtained weighted matrix of DBN was transferred to back-propagation to start the supervised phase. The proposed system delivers an efficient classification model with an accuracy of 99.68%.

Murat karabatak et al [18] proposed an automatic diagnostic system for predicting breast cancer. The authors used Wisconsin breast cancer database and reduced dimensions using association rules(AR) for reducing the dimension of the feature space and neural network(NN) for classification. In the related work, the authors proposed AR+NN model and evaluated using 3-fold cross validation. The classification precision of the proposed model was 95.6%.

Emina et al [20] proposed a model for diagnosing breast cancer using GA feature selection and rotation forest. Authors used two different Wisconsin breast cancer datasets and applied genetic algorithm for attribute selection and rotation forest for classification. In the related work, the authors compared the models logistic regression, decision tree, Bayes net, Random forest, ANN, RBFN, SVM and Rotation Forest. The results of these models were compared with and without GA based feature selection and the model was evaluated by cross validation. GA based feature selection and rotation forest has an accuracy of 99.48%.

H Asri et al [21] developed a model for predicting breast cancer risk and diagnosis using ML algorithms on Wisconsin breast cancer dataset. The authors compared the algorithms such as SVM, C4.5, Naïve Bayes and K- Nearest Neighbours (K-NN). These models were evaluated using the performance measures accuracy, sensitivity, precision and specificity. The results demonstrated that SVM has the highest accuracy of 97.13%. In the above approaches, deep belief network (DBN) achieved highest accuracy in predicting the breast cancer occurrence. Table 1 shows all the studies related to the predictive models for breast cancer susceptibility.

Dataset: The Wisconsin Diagnostic Breast Cancer (WDBC) dataset and Wisconsin Prognostic Breast Cancer (WPBC) dataset are publicly available at http://ftp.ics.uci.edu/pub/machine-learning-databases/breast-cancer-wisconsin/

The features used in the dataset mainly are Clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single ellipthelial cell size, bare nucleoi, bland chromatin, normal nuclei, mitosis.

Table 1: Predictive models for the prognosis and diagnosis of breast cancer

Study	Dataset	features	Instances	evaluation parameters	comparative models
Maglogiannis et al[11]	WPBC from Wisconsin	30 features	198	10-fold cross validation	Bayes, ANN, SVM
M.F. Akay et al [12]	WBCD from Wisconsin	9 features	683	Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, ROC curve and confusion matrix	SVM combined with feature selection
Ayer et al [13]	WCRS from Wisconsin	35 features	62,219	sensitivity, specificity, ROC curve, 10-fold cross validation	ANN
Vikas et al [16]	Wisconsin	9 features	699	10-fold cross validation	SMO,IBK, BF tree
Abdel et al [17]	Wisconsin	9 features	683	Accuracy, sensitivity, specificity	DBN-NN, RIW-BPNN
Murat karabatak et al [18]	Wisconsin	9 features	699	3-fold cross validation	AR+NN,NN
Emina et al [20]	Wisconsin	9 features	683	10-fold cross validation	Logistic regression, Decision trees, Bayes net, Random forest, ANN,RBFN, SVM, Rotation Forest
H Asri et al [21]	Wisconsin	11 features	699	Accuracy, Precision, Sensitivity, Specificity, 10-cross fold validation	SVM, C4.5, Naïve Bayes, K-NN

B. Predictive models for breast cancer survivability

A variety of MLtechniques such as Decision trees, ANN are used to predict the survivability of breast cancer patients. Delen et al [7] proposed a predictive algorithm for breast cancer survivability using SEER(Surveillance Epidemiology End Results) dataset with 433,272 records. After pre-processing the data to remove the redundancies and missing values, the final dataset contains 202,932 records which then categorised into two categories 'survived (93273)' and 'did not survive (109,659)'. The category 'survived' represents a person who survived after 60 months from the date of diagnosis. In this work, [7] used three classification models artificial neural networks, decision trees and logistic regression and were evaluated using 10-fold cross validation. The decision tree model (C5) results in better accuracy of 0.9362 when compared to the other models.

In the above approach, the patients with survival rate less than 60 months and still

alive or the patients with survival rate less than 60 months are considered but the patients with the cause of their death is not breast cancer are ignored. Bellaachia et al [8] used the SEER database for the years 1973-2002 and included 2 more attributes vital survival recode(VSR) and cause of death(COD) with survival time recode(STR) in predicting the survivability of breast cancer patients. In the related work, the authors compared three classification algorithms Naïve Bayes, back-propagated neural network and C4.5 decision tree algorithms on SEER breast cancer dataset. The results proved that C4.5 decision tree algorithms have good performance with an accuracy of 86.7%.

In the above approaches the decision tree algorithms proved as superlative classifiers for SEER data. The decision tree models are sharp at each node of the decision tree boundaries and a small change in the value of attribute result in misclassification [9]. Hybrid fuzzy decision tree classification is more robust and balanced than independently applied crisp classification [9]. Khan et al [9] proposed a fuzzy model for predicting breast cancer survivability. The authors used the SEER database for the years 1973-2002 consisting 162,500 records with 17 variables after pre-processing. Sigmoid, linear, and other membership functions were. Sigmoid and linear functions produced better results. Park et al [10] used SEER breast cancer survivability dataset which is a major database in US for cancer statistics which has 162500 records with 16 predictor features and one target class variable. The target variable is 'Survivability' which is binary with value +1 if the patient had survived and -1 if patient had not survived more than 5 years after the diagnosis. The authors compared the three models ANN, SVM and best performance was achieved by SSL model with AUC of 0.76. In the above studies, decision tree proved as the best classifier. Table 2 shows all the studies related to the predictive models for breast cancer survivability.

Table 2: **Predictive models for breast cancer survivability**

Study	Dataset	Instances initially	Instances after preprocessing	evaluation parameters	comparative models
Delen et al[7]	SEER dataset for the years 1973-2000	433,272	202,932	10-fold cross validation	ANN, Decision trees (C5), Logistic Regression
Bellaachia et al[8]	SEER dataset for the years 1973-2002	482,052	151,886	10-fold cross validation	Naïve bayes, Back propogated neural networks, C4.5
Khan et al [9]	SEER dataset for the years 1973-2003	433,272 with 86 variables	162,500	Kernel functions	Decision trees, fuzzy decision trees
Park et al [10]	SEER dataset for the years 1973-2003	433,272 with 86 variables	162,500	k-fold cross validation	ANN, SVM, SSL

Dataset: Authors used SEER(Surveillance Epidemiology End Results) database. After data preparation, the final dataset consists of 17 variables (16 predictor variables and 1 dependent variable). The features used in the dataset are Tumor size, Number of nodes, Number of primaries, Age at diagnosis, Number of positive nodes, Marital status, Race, Behaviour code, Grade, Extension of tumor, Node involvement, Histological type ICD, Primary site, Site specific surgery, Radiation, Stage of cancer.

C. Predicting breast cancer recurrence

Kim W et al [14] analyzed the breast cancer dataset available from Korean tertiary hospital and developed a SVM model for breast cancer recurrence. The authors used the data set collected from Korean tertiary hospital of 1541 patients who underwent breast cancer surgery during the years 1994 and 2002. As a part of pre-processing, they removed 808 patient data and used the data of 733 patients. In the related work, the authors compared SVM, Artificial neural network and Cox-proportional hazard regression model with each other. The SVM model proved to be better when compared with other models with an accuracy of 84.6% and AUC of 0.85. Pritom et al [15] proposed a ML model for breast cancer recurrence using different data mining techniques. The authors used Wisconsin dataset from UCI repository which consists of 35 attributes and 198 instances. The output attribute consists of two nominal values 'R' and "N" representing recursive cancer and non-recursive cancer respectively. For recurrence breast cancer prediction, the authors investigated three algorithms SVM, Naïve bayes and C4.5 and were evaluated by cross-validation. To select best features and removing redundant/irrelevant features, authors used Ranker algorithm.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The results proved that SVM has the highest accuracy of 75.75% in predicting the recurrence of breast cancer. Eshlaghy et al [19] developed a prediction model for breast cancer recurrence using three machine learning techniques. The authors used the dataset of Iranian centre for breast cancer(ICBC) for the years 1997 to 2008 which consists of 1189 instances and 23 features. In the related work, they implemented three machine learning techniques C4.5, SVM and ANN and evaluated the prediction models using 10-fold cross validation. SVM model has the highest accuracy of 0.957 and least error rate in predicting the breast cancer recurrence. Table 3 shows all the studies related to the predictive models for breast cancer recurrence.

Table 3: Predictive models for breast cancer recurrence

Study	Dataset	features	Instances	evaluation parameters	comparative models
Kim W et al [14]	Korean tertiary hospital	Age, histological grade, local invasion of tumor, no.of tumors, tumor size, lympovascular invasion, estrogen receptor, no.of metastatic lymph nodes	808	Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, AUC	SVM, ANN, Cox-proportional hazard regression
Pritom et al [15]	Wisconsin	35 attributes	198	10-fold cross validation	SVM, Naïve bayes, C4.5
Eshlaghy et al [19]	ICBC	22 variables	1189	10-fold cross validation	Decision trees, SVM, ANN

I. DISCUSSION

Machine Learning techniques are used for analyzing medical datasets and serves as a major platform for intelligent data analysis in the medical domain. In our study, we focus on three different issues in predicting breast cancer in women namely 1) prediction of breast cancer susceptibility (prognosis and diagnosis) 2) prediction of breast cancer survivability 3) Prediction of breast cancer recurrence. Table 4 represents the predictive models for breast cancer in women.

Table 4: Predictive models for breast cancer in women

Study	Year	Clinical End point	Best Method	Accuracy
Delen et al [7]	2005	Survivability	C4.5	0.9362
Bellaachia et al [8]	2006	Survivability	C4.5	86.7%
Khan et al [9]	2008	Survivability	Fuzzy Decision trees	85%
Park et al [10]	2013	Survivability	SSL	0.76
Maglogiannis et al [11]	2009	Prognosis	SVM	96.91%
M.F.Akay et al [12]	2009	Prognosis	SVM	99.51%
Ayer et al [13]	2010	Prognosis	ANN	AUC=0.965
Vikas et al [16]	2017	Prognosis	SMO	96.2%

Kim et al [14]	2012	Recurrence	SVM	84.6%
Pritom et al [15]	2016	Recurrence	SVM	75.75%
Abdel et al [17]	2016	Diagnosis	DBN-NN	99.68%
Murat karabatak et al [18]	2009	Diagnosis	AR+NN	95.6%
Eshlaghy et al [19]	2013	Recurrence	SVM	0.957
Emina et al [20]	2015	Diagnosis	GA+Rotation Forest	99.48%
H Asri et al [21]	2016	Diagnosis	SVM	97.13%

In our study, it has been observed that most of the researchers used the datasets of Wisconsin and SEER. Several classification techniques such as decision trees, regression, SVM, ANN etc. are proposed in predicting breast cancer in women. Deep belief network (DBN) has achieved highest accuracy of 99.68 % in predicting breast cancer susceptibility. When it comes to the breast cancer survivability prediction, decision trees achieved highest accuracy of 0.9362. The study further revealed that SVM model with 0.957 accuracy in predicting breast cancer recurrence. Hence it is observed that Deep belief network (DBN) model are proved to be efficient when compared to other models in predicting breast cancer in women.

IV. CONCLUSION:

In this survey, we discussed different machine learning techniques proposed by the researchers in predicting breast cancer. Also, we outlined the predictions of the authors in terms of prognosis/diagnosis, recurrence and survivability. Several studies that have been proposed in the past few years focused on developing predictive models for breast cancer by means of supervised learning. Based on the analysis of the results, SVM classifier combined with feature selection provides highest accuracy in predicting the disease outcome.

REFERENCES

1. Walters, V. (1993). Stress, anxiety and depression: women's accounts of their health problems. *Social science & medicine*, *36*(4), 393-402.
2. Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient machine learning for big data: A review. Big Data Research, 2(3), 87-93.
3. Hand, D. J. (1999). Statistics and data mining: intersecting disciplines. ACM SIGKDD Explorations Newsletter, 1(1), 16-19.
4. Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J. F., & Hua, L. (2012). Data mining in healthcare and biomedicine: a survey of the literature. Journal of medical systems, 36(4), 2431-2448.
5. Tomar, D., & Agarwal, S. (2015). Hybrid feature selection based weighted least squares twin support vector machine approach for diagnosing breast cancer, hepatitis, and diabetes. Advances in Artificial Neural Systems, 2015, 1.
6. Jemal, A., Center, M. M., DeSantis, C., & Ward, E. M. (2010). Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiology and Prevention Biomarkers, 1055-9965.
7. Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. Artificial intelligence in medicine, 34(2), 113-127.
8. Bellaachia, A., & Guven, E. (2006). Predicting breast cancer survivability using data mining techniques. Age, 58(13), 10-110.
9. Khan, M. U., Choi, J. P., Shin, H., & Kim, M. (2008, August). Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 5148-5151). IEEE.
10. Park, K., Ali, A., Kim, D., An, Y., Kim, M., & Shin, H. (2013). Robust predictive model for evaluating breast cancer survivability. Engineering Applications of Artificial Intelligence, 26(9), 2194-2205.
11. Maglogiannis, I., Zafiropoulos, E., & Anagnostopoulos, I. (2009). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. Applied intelligence, 30(1), 24-36.
12. Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 36(2), 3240-3247.
13. Ayer, T., Alagoz, O., Chhatwal, J., Shavlik, J. W., Kahn Jr, C. E., & Burnside, E. S. (2010). Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration. Cancer, 116(14), 3310-3321.
14. Kim, W., Kim, K. S., Lee, J. E., Noh, D. Y., Kim, S. W., Jung, Y. S., ... & Park, R. W. (2012). Development of novel breast cancer recurrence prediction model using support vector machine. Journal of breast cancer, 15(2), 230-238.
15. Pritom, A. I., Munshi, M. A. R., Sabab, S. A., & Shihab, S. (2016, December). Predicting breast cancer recurrence using effective classification and feature selection technique. In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 310-314). IEEE.
16. Chaurasia, V., & Pal, S. (2017). A novel approach for breast cancer detection using data mining techniques.
17. Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. Expert Systems with Applications, 46, 139-144.
18. Karabatak, M., & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. Expert systems with Applications, 36(2), 3465-3469.
19. Ahmad, L. G., Eshlaghy, A. T., Poorebrahimi, A., Ebrahimi, M., & Razavi, A. R. (2013). Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inform, 4(124), 3.
20. Aličković, E., & Subasi, A. (2017). Breast cancer diagnosis using GA feature selection and Rotation Forest. Neural Computing and Applications, 28(4), 753-763.
21. Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science, 83, 1064-1069.
22. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.

AUTHORS PROFILE

K.Prasuna is working as an Assistant Professor in Koneru Lakshmaiah Education Foundation in Dept.of CSE. Had 8+ years of experience in academics and industry. Her research interests include machine learning, wireless sensor networks.

Dr.K.V.S.N.Rama Rao is working as a Professor in Koneru Lakshmaiah Education Foundation in Dept.of CSE. Had 20+ years of experience in academics and industry. International research experience at Australian university. His research interests include cyber security, machine learning and bioacoustics.

CH.M.H.Saibaba is working as an Assistant Professor in Koneru Lakshmaiah Education Foundation in Dept.of CSE. Had 16+ years of experience in academics and industry. His research interests include Data sciences.