Implementing Machine Learning Classifiers In Breast Cancer Diagnosis

Dr. T.Lalitha[1] , C.Krishnan [2], P.Malathi [3] , V.R.Niveditha [4]

[1]Professor, Computer Science, Jain Deemed to be University, Bengaluru, Karnataka.
[2]Research Scholar, Computer Science, Research and Development Centre, Bharathiar University, Coimbatore, India.
[3]Assistant Professor, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India.
[4]Research Scholar, Department of Computer Science and Engineering, Dr.M.G.R Educational and Research Institute, India.
Email I'd: lalithasrilekha31@gmail.com[1], krishnan.peaceful@gmail.com[2]

Abstract

Cancer is a disorder which is caused by the uncontrollable multiplication of abnormal cells in the part of the body. Cancer cells have the ability to extend to the other portion of the body in a very fast manner with a short span of time. One of the most dreadful types of cancer is breast cancer. As it name conveys, it is the cancer where the abnormal cells originates in the breast and continues to accumulate, forming a mass of cell. In earlier the breast cancer are detected by mammogram. Machine learning is a usage of artificial intelligence which provides system the capacity to study and develop automatically from experience without being explicitly programmed. The usages of machine learning are in the area like diagnosis, image processing, and classification of data, predication, and regression learning association. In this proposed system the machine learning concepts SVM and NN (Neural Network)are applied for the detection of canceron breast. Compared with SVM classifier NN produces more accurate result.

Keywords: Breast Cancer, Machine Learning, Classifiers, Accuracy, Diagnosis, Prediction

I. INTRODUCTION

Cancer is also known in other terms such as malignant tumor and malignant neoplasm. 22% of the cancer is because of the usage of tobacco and 10 % are due to the obesity, insufficient diet and lack of physical activities. Cancer cannot be cured completely but it should be treated in its very earliest stage. Sometimes the untreatable case leads to death. The risk of cancer can be decreased by maintaining a good weight, not smoking and not drinking alcohol, then consumption of more water and green vegetables. Some of the common kinds of cancer disease are sarcoma, carcinoma, lymphoma, leukemia and melanomas. Carcinoma is the type of cancer which is diagnosed commonly. It forms in the skin part, lungs, breast, and pancreas part and in the other glands. Lymphomas are the cancers in lymphocytes. Leukemia is the cancer in the blood sarcoma originates in the bone muscle blood vessels or cartilage. Melanoma originates in the cells which is responsible for the pigment of the skin.

These breast cancers also have certain types. This types of cancer makes difference in the skin cells which lead to the feeling of pain, discomfort in the breast. In cancers it is very important that not to ignore pain or any sign of symptoms. Even though these abnormal cells can extend to any bones on the body, the common sites are ribs, spine, pelvis portion, long bones that is in arms and leg. Indication of breast cancer include, one breast will be larger than the other, the nipple becomes inverted discharge from the nipple, continues pan and swelling around the collar bone. Some of the research point outintroduction to light pollution is also a risk feature for the development of the cancer on breast part.

The usage of machine learning is multiplying in all fields. Machine learning is built with the mathematics and computer science concepts. They are explained well with the linear and matrix algebra. Common machines learning methods are

supervised machine learning technique, unsupervised machine learning approach, semi supervised machine learning concept, and reinforcement machine learning type. Basic concepts of machine learning include decision tree, graphical, neural networks, SVM and sets of rules.

II LITERATURE REVIEW

E. A. Bayrak et al., 2019 uses ANN (Artificial Neural Network) and SVM (Support Vector Machine) to predict the cancer disease. This machine learning concepts are tested with on WBC (Wisconsin Breast Cancer) dataset. The performances of the classifiers are compared by using various matrices like precision value, exactness, ROC value

and recall. Final output shows that SVM technique produces better result than other classifier. The classifiers are implemented with the help of WEKA software tool [1].

Machine learning concepts are mainly used to identify the cancer on the breast. N. Khuriwal et al., 2019 projected a novel ensemble voting technique for identifying cancer using WBC dataset. The main objective of this current research work is to evaluate and make clear about ANN and logistic concepts and it also provides best output even the number of identifiers are also decreased. Final output values are showed better accuracy when integrate ANN technique and logistic concept [2].

In cancer identification domain also machine learning approaches are involved. Machine learning techniques are helps the healthcare professionals to provides a second suggestion and assist them to find an enhanced result. F. Teixeira et al., 2019 uses deep learning methods for diagnosis cancer disease on the breast. The intention this research article is to assess the performance of various predictive system classifiers. Here the authors uses 6m various classifiers for assessing the performance. The experiment results shows that DNN classifier concept produces better result than other traditional methods. All the classifiers are tested with Wisconsin dataset [3-4].

Cancer on the Breast is common among females and it is expand from the tissues on the breast. Timely identification and proper handling is important to improve the survival rate of the affected people. B. Bektaş et.al. 2016 uses various machine learning classifiers to recognize and identify the cancer on the breast. With the help of identifiers choosing technique the authors identifies the genes in the affected portion [5].

Machine learning concepts are also used in medical domain to detect the cancer cells. Due to breast cancer also the death rate is increased. It is one of familiar type of cancer and the major reason of female death count in the whole world. The affected cells are categorized as Benign type or Malignant type. Various machine learning approaches are used to predict the cancer on the breast. A.Bharat et al, 2018 uses SVM classifier on WBC dataset to diagnosis the cancer [6-7].

Uncontrollable growth of breast cells is known as cancer on breast.   It is one of the important issues in the healthcare domain. Cancer on the breast is the important reason of increasing fatal rate of females. Naveen, R et al., 2019 diagnosis the cancer with high level of accuracy value using machine learning concepts. Here the authors use the dataset Coimbra from UCI to develop better models. The important concepts used in this proposed system are scaling features, cross validation and models with the technique of bagging. The experiment result shows that KNN and decision tree approaches produces better result [8].

M. Nemissi et al., 2018 says that cancer disease on the breast is the major fatal disease among female people. Beginning stage identification of this cancer can increase the survival rate of the affected people. Here the authors propose neural network concepts to predict the cancer. This system consists of hidden layer and machine learning approaches are used to train the proposed system.   The important aim of the suggested system is to apply various activation methods on the unseen neurons and optimized using the genetic approach. Performance of the system can be tested with WDBC (Wisconsin Diagnostic Breast Cancer) data set. The retrieved output is compared with other traditional methods [9].

J. A. Bhat et al., 2015 developed new tools which can assist the healthcare professionals in hospitals to identify cancers on breast. It is the most important kind of cancer in all over the country and it is identified at the final stages. Here the authors are presented the output of their project report BCDM implemented using MATLAB tool. The approach designed for this cancer research work depends upon adaptive resonance method. The target of this research work is to execute a concept on cloud systems and early identification of cancer disease from the concern people using internet based communication from any part of the country [10-14].

III. PROPOSED SYSTEM

Cancer disease is one of the common diseases in the whole world.   From that breast cancer disease is the common cancer for females.Based on BCI (Breast Cancer Institute) report, in the whole world breast cancer disease is most risky disease for female people. Healthcare professional said that identifying cancer in the earlier stage increases the life of affected people. The following fig 1 shows the flow diagram of proposed work. In this system mammogram images are given as input. The original image contains various unwanted data. Preprocessing techniques are used remove noisy and unwanted data from the input image. After preprocessing the images are segmented by using various image processing methods. In this work images are segmented by cropping ROI manually or threshold region based techniques. The preprocessed images are consists of huge number of features.   The required features are extracted and selected with the help of various feature extraction techniques. After the process of feature selection SVM and NN classifiers are used to classify the images. The classifiers are used to classify the breast cancer type Malignant or Benign. Finally assess the performance of the classifiers using accuracy and precision matrices.
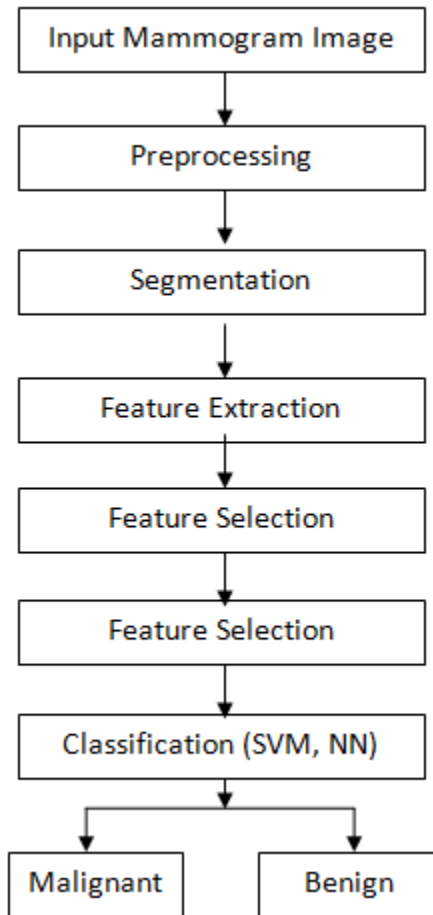
Figure 1 Proposed System Flow Diagram

SVM classifier is commonly used in binary classification with linear data. The mail goal of SVM is to search the suitable hyperplane. It means the margin among the two separate types of classes and it is also used to divide the data into two various classes.$\{x_1, x_2, \dots \dots x_n\}$Is the real value data $y_i \in \{-1, 1\}$ is the data label and ω is represents the weighted vector value. The hyperplane is represented as the following equation 1.

f(x)=w$x_i + b = 0$ ---------------------- (1)

Optimal value of margin is get from distance of hyperplane and vector. Classification is represented as

$min \frac{1}{2} \|w\|^2$ ----------------------------- (2)

According to T. Nadira, et.al., 2018 The following figure 2 shows the pictorial representation of SVM classifier[16].
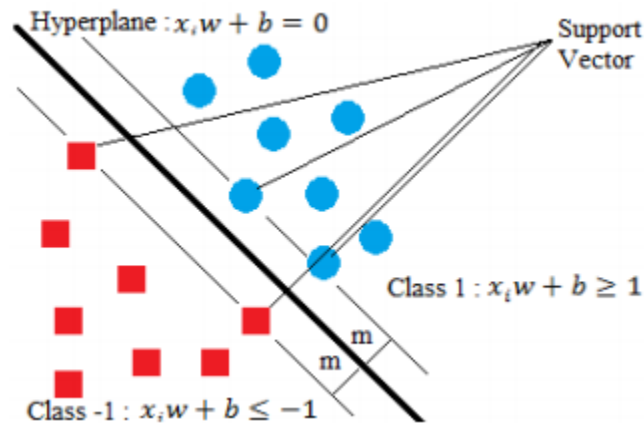

Figure 2 SVM Classifier

NN (Neural Network) approach is also collects the mammogram dataset and predict the chance of cancer on breast of concern people. Every neuron is associated with local type of neurons and weight value is also shared with other neurons. It is one of the effective methods to classify the images and generate image data with related features. The following figure 2 shows the common form of NN.
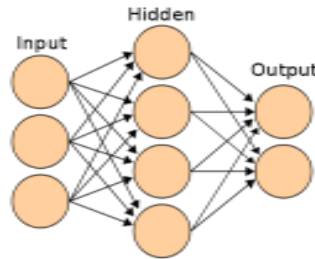

Figure 2 Common Form of NN

IV. RESULT AND DISCUSSION

Prediction and diagnosis of this disease in earlier stage is significant for better life. Various computing techniques are used to diagnosis the cancer. Machine learning concepts are offering various tools to diagnosis cancer in breast. Here SVM and NN classifiers are used to classify the mammogram images. The following figure 4 demonstrates the output of the proposed

system. The classifiers are implemented and tested by using Wisconsin dataset. Finally the performance level of the current work compared in terms of accuracy and precision value.
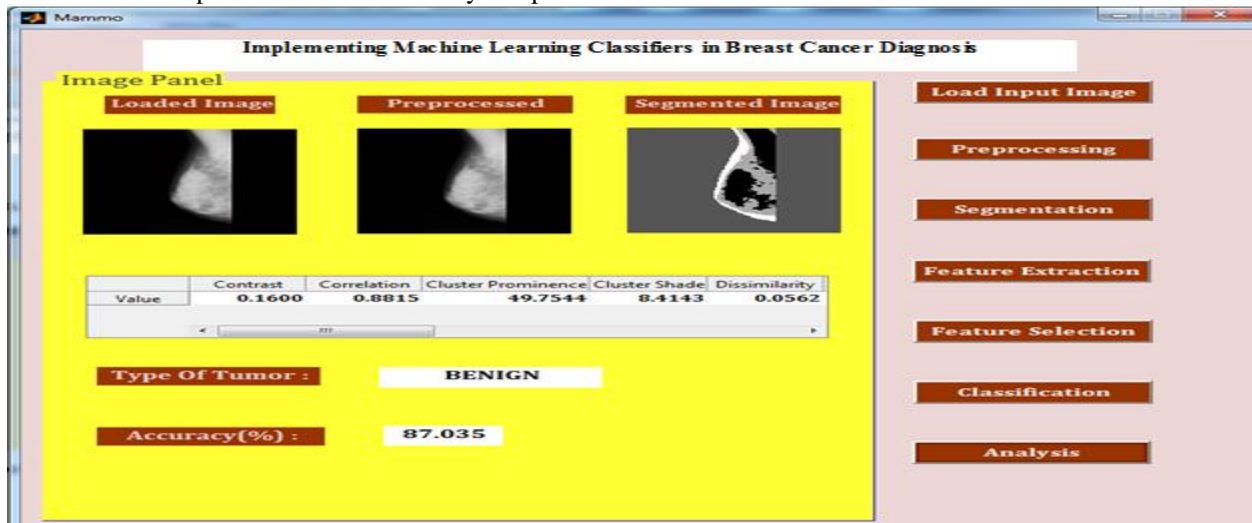

Figure 4 Output of Proposed Work

The current system is developed using MATLAB software. Compared with SVM technique, NN approach produces better result.

V. CONCLUSION

Cancer on the breast is one of the fatal kinds of syndrome.. Mostly female people are affected by this cancer disease. Early detection of cancer diseases are increases the survival period of the affected of the people. Machine learning classifiers are assists healthcare professionals to diagnosis breast cancer in earlier stage. This proposed system uses SVM and NN classifiers to predict the disease. Mammogram images are the input of the current prediction system. Unwanted data are removed by using preprocessing concepts and images are segmented by using cropping or threshold technique. Finally the images are classified by using SVM and NN classifier. The NN classifier produces better result in terms of accuracy level and precision value. This system implemented by using MATLAB software tool.

REFERENCES

1. [1] E. A. Bayrak, P. Kırcı& T. Ensari(2019), "Comparison of Machine Learning Methods for Breast Cancer Diagnosis,"*IEEE Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT),* pp. 1-3.
2. [2]N. Khuriwal & N. Mishra (2018), "Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm,"*IEEMA Engineer Infinite Conference (eTechNxT*), pp. 1-5.
3. [3]F. Teixeira, J. L. Z. Montenegro, C. A. Costa & Rosa Righi(2019)"An Analysis of Machine Learning Classifiers in Breast Cancer Diagnosis*," XLV Latin American Computing Conference (CLEI)*, pp. 1-10.
4. [4]S.Umadevi, S. Nirmala SugirthaRajini, A. Punitha and VijiVinod(2020), "Performance Evaluation Of Machine Learning Algorithms In DimensionalityReduction", *International Journal of Advanced Science and Technology,*Vol. 29, No. 9s, pp. 3845-3853.
5. [5]B. Bektaş & S. Babur(2016), "Machine learning based performance development for diagnosis of breast cancer,"*Medical Technologies National Congress (TIPTEKNO*), pp. 1-4.
6. [6]A. Bharat, N. Pooja& R. A. Reddy(2018), "Using Machine Learning algorithms for breast cancer risk prediction and diagnosis*,"IEEE3rd International Conference on Circuits, Control, Communication and Computing (I4C),* pp. 1-4.
7. [7]Naveen, R. K. Sharma & A. Ramachandran Nair(2019), "Efficient Breast Cancer Prediction Using Ensemble Machine Learning Models,"*4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT),* pp. 100-104.
8. [8]M. NEMISSI, H. SALAH & H. SERIDI(2018), "Breast cancer diagnosis using an enhanced Extreme Learning Machine based-Neural Network,"*International Conference on Signal, Image, Vision and their Applications (SIVA)*, pp. 1-4.
9. [9]J. A. Bhat, V. George& B. Malik(2015), "Cloud Computing with Machine Learning Could Help Us in the Early Diagnosis of Breast Cancer,"*Second International Conference on Advances in Computing and Communication Engineering*, pp. 644-648.
10. [10]B. Bavani, S. Nirmala SugirthaRajini, M.S. Josephine, & V. Prasannakumari (2019), "Heart Disease Prediction System based on Decision Tree Classifier", *Jour of Adv Research in Dynamical & Control Systems*, Vol. 11, 10-Special Issue.
11. [11] Shalini A, Jayasuruthi L, Vinoth Kumar V, "Voice Recognition Robot Control using Android Device" *Journal of Computational and Theoretical Nanoscience,* 15(6-7), pp. 2197-2201.
12. [12]M. S. Yarabarla, L. K. Ravi& A. Sivasangari (2019), "Breast Cancer Prediction via Machine Learning,"*3rd International Conference on Trends in Electronics and Informatics (ICOEI),* pp. 121-124. .
13. [13] S. Sharma, A. Aggarwal &T. Choudhury (2018), "Breast Cancer Detection Using Machine Learning Algorithms,"*International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS),*pp.114-118.
14. [14] T. Nadira and Z. Rustam(2018), "Classification of cancer data using support vector machines with features selection method based on global artificial bee colony",