International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

BREAST CANCER DETECTION WITH MACHINE LEARNING

[1]A Sai Pujitha, [2]Badam Lakshmi Sravani, [3]Grandhisaisudheer, [4]Shaik Usha Begum,[5] Kanamarlapudi Avinash
[1,2,3,4,,5]Assistant professors, Department of CSE in Narasaraopet Institute Of Technology

ABSTRACT
According to Breast Cancer Institute (BCI), Breast Cancer is one of the most dangerous type of diseases that is very effective for women in the world. As per clinical expert detecting this cancer in its first stage helps in saving lives. As per cancer.net offers individualized guides for more than 120 types of cancer and related hereditary syndromes. For detecting breast cancer mostly machine learning techniques are used. In this paper we proposed adaptive ensemble voting method for diagnosed breast cancer using Wisconsin Breast Cancer database. The aim of this work is to compare and explain how ANN and logistic algorithm provide better solution when its work with ensemble machine learning algorithms for diagnosing breast cancer even the variables are reduced. In this paper we used the Wisconsin Diagnosis Breast Cancer dataset. When compared to related work from the literature. It is shown that the ANN approach with logistic algorithm is achieved 98.50% accuracy from another machine learning algorithm.

1. INTRODUCTION

The most dangerous disease in the world is cancer in which breast cancer is the dangerous for women. Many women die every year because of breast cancer. Detecting the breast cancer manually takes a lot of time and it is difficult for the physician to classification. So the detecting the cancer through various automatic diagnostic techniques is very necessary. There are various method and algorithm are available for detecting breast cancer such as Support Vector Machine, Naïve Bayes, KNN and Convolution Neural Network is the latest algorithm in deep learning that is also used for classification. CNN and deep learning algorithm mainly used for images classification and object detection. In this paper we use UCI open database for training and testing purpose in which two classes of Tumor are available, one is Benign Tumor and the other is malignant in which benign Tumor is non-cancerous and the malignant is a cancer Tumor. Many reasecher are still performing research for detecting and diagnosing cancer in an early stage. Because the early stage cancer is not a so panful and expensive for complete its treatment and many researcher are still trying to developing a proper diagnosis system for detection the Tumor as early as possible. So the treatment can be started earlier and the rate for resolution may increase. This work main aim is comparatively study of various machines learning algorithm with Artificial Neural Network. The rest of the paper is organized as follows: Section 2 presents the literature review of the proposed work. Section 3 includes the architecture of the proposed work .Section 4 describes the methodology which is used for the proposed work. Section 5 describes feature selection process for the proposed work. In section 6 we discuss the model implementation of proposed work. Section 7 discusses the results and Section 8 concludes the proposed work.

2.LITERATURE SURVEY

M. R. Al-Hadidi, A. Alarabeyyat and M. Alhanahnah, "Breast Cancer Detection Using K-Nearest Neighbor Machine Learning Algorithm," 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, 2016, pp. 35-39. Breast cancer detection images are the standard clinical practice for the diagnosis of breast

cancer. Digital Mammogram has emerged as the most popular screening technique for early detection of Breast Cancer and other abnormalities. In this paper we introduce a Computer-Aided Detection (CAD) system to perform automatic diagnosing of malignant/non-malignant breast tissues using Polar complex Exponential Transform (PCET) moments as texture descriptors. The input Region of Interest (ROI) is extracted through histogram ROI selection and further pre-processing stages are carried out. The calculated PCET moments are used for feature extraction. A new classifier Adaptive Differential Evolution Wavelet Neural Network (ADEWNN) is used to improve the classification accuracy of the CAD system.

A. Qasem et al., "Breast cancer mass localization based on machine learning," 2014 IEEE 10th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, 2014, pp. 31-36.

BIRADS is a Breast Imaging, Reporting and Data System. A tool to standardize mammogram reports and minimizes ambiguity during mammogram image evaluation. Classification of BIRADS is one of the most challenging tasks to radiologist. An apt treatment can be administered to the patient by the oncologist upon acquiring sufficient information at BIRADS stage. This study aspired to build a model, which classifies BIRADS using mammograms images and reports. Through the implementation of type-2 fuzzy logic as classifier, an automatically generated rules will be applied to the model. To evaluate the proposed model, accuracy, specificity and sensitivity of the modal will be calculated and compared vis-à-vis rules given by the experts. The study encompasses a number of steps beginning with collection of the data from Radiology Department, Hospital of National University of Malaysia (UKM). The data was initially processed to remove noise and gaps. Then, an algorithm developed by selecting type-2 fuzzy logic using Mamdani model. Three types of membership functions were employed in the study. Among the rules that used by the model were obtained from experts as well as generated automatically by the system using rough set theory. Finally, the model was tested and trained to get the best result. The study shows that triangular membership function based on rough set rules obtains 89% whereas expert rules achieve 78% of accuracy rates. The sensitivity using expert rules is 98.24% whereas rough set rules obtained 93.94%.

3. SYSTEM ANALYSIS

Existing system

Many machine learning algorithms are available for prediction and diagnosis of breast cancer. Some of the machine learning algorithm are Artificial Neural Network (ANN), Naïve Bayes, Support Vector Machine (SVM), KNearest Neighbour (KNN) and Convolutional Neural Network etc. Many researcher have done research in breast cancer by sing various dataset such as using Mammogram images as dataset, SEER dataset, Wisconsin Dataset and also dataset from various hospitals. By using these dataset authors extract and select various features and complete your research. These are some important research review. The Author Moh'dRasoul used DWT tool for image filtering and BPNN for processing and achieved 93.7% accuracy. The author Clemen Deng used WHAVE algorithm with Wisconsin Breast Cancer Database and achieved 99% accuracy . The author Ashwaq Qasem used marker Controller Watershed algorithm and achieved 95% accuracy. The author AlirezaIsareh used signal to noise ratio with SVM and achieved 98.80% accuracy. The author Junaid Ahmad used Adaptive Resonance Theory with UCI database and got 84.21% accuracy. By B.M. Gayathri work on comparative study about Relevance vector machine and achieved 97% accuracy.The

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

author Ms.H.R. Mhaske used KNN and SVN with 150 images database and achieved 80-90% accuracy. The author S. Aruna used Naïve Bayes and SVM with UCI database and achieved 68-79% accuracy. By Yohannes Tsehay developed a weakly supervied computer aided detection system that was used biopsy for learn data. By Sudarshan Nayak used Naïve Bayes and SVM and got 98% accuracy. The Ryota Shimizu used Deep learning and neural network and achieved 90% accuracy. Many other authors also done research for detection and diagnosis breast cancer using various machine learning algorithm.

Disadvantages:

- Less accuracy score
- Low performance
- Unable to predict the resolution

Proposed system

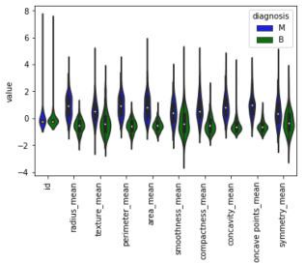
In this section, we introduce the proposed First collected Wisconsin Breast Cancer Database from UCI website then pre-processed the dataset and select 16 important features. For feature selection we used Recursive feature Elimination Algorithm using Chi2 method and get 16 top features. After that applied ANN and Logistic algorithm individually and compute the accuracy. Finally, we used proposed Ensemble Voting method and compute best method for diagnosis breast cancer disease. In this paper we have using ensemble method for diagnosis breast cancer with neural network and logistic algorithm. All process consist main three parts: Pre-processing data, features selection and voting models. In this work we have used BCI dataset having 569 rows and 30 column of dataset. In experiment part we have first evaluated the features from default dataset. For features selection we have used Univariate Features selection method and Recursive Features Selection method with Cross Validation Method.

A. Pre Processing Data Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real world data is often incomplete, inconsistent, and lacking in certain behaviours and likely to contain many errors. Data pre-processing is a proven method of resolving such issues. Data pre-processing prepares raw data for further processing. For pre-processing we have used standardization method to pre-process the UCI dataset. B. Standardization Method: In this method the dataset is a common requirement for many machine learning estimators. In this paper we have created different data visualization for data pre-processing. First we have count malignant and benign from all dataset and plot in graph. In second stage we have created a Violin plot for dataset compression. It shows the distribution of quantitative data across several levels of one categorical variable such that those distributions can be compared. Then we draw a scatterplot with non-overlapping points. This gives a better representation of the distribution of values. we have created a relationship scatterplot between dataset features for more understanding. This shown the relationship scatterplot of between dataset features.

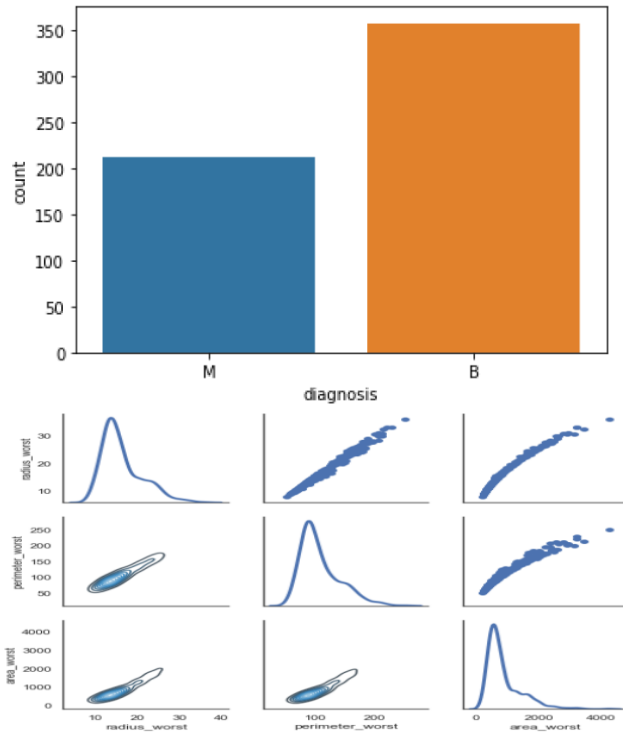
Advantages:

- Good accuracy score and performance
- Predict the higher resolution

4. RESULTS:



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



5. CONCLUSION

This work is the proposed an ensemble machine learning method for diagnosis breast cancer, in which we can see in the table and graph that proposed method is showing with the 98.50% accuracy. In this paper we used only 16 features for diagnosis of cancer. In future we will try on all features of UCI and to achieve best accuracy. Our work proved that neural network is also effective for human vital data analyzation and we can do pre-diagnosis without any special medical knowledge.

REFERENCES

[1] M. R. Al-Hadidi, A. Alarabeyyat and M. Alhanahnah, "Breast Cancer Detection Using K-Nearest Neighbor Machine Learning Algorithm," 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, 2016, pp. 35-39.

[2] C. Deng and M. Perkowski, "A Novel Weighted Hierarchical Adaptive Voting Ensemble Machine Learning Method for Breast Cancer Detection," 2015 IEEE International Symposium on Multiple-Valued Logic, Waterloo, ON, 2015, pp. 115-120.

[3] A. Qasem et al., "Breast cancer mass localization based on machine learning," 2014 IEEE 10th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, 2014, pp. 31-36.

[4] A. Osareh and B. Shadgar, "Machine learning techniques to diagnose breast cancer," 2010 5th International Symposium on Health Informatics and Bioinformatics, Antalya, 2010, pp. 114-120.

[5] J. A. Bhat, V. George and B. Malik, "Cloud Computing with Machine Learning Could Help Us in the Early Diagnosis of Breast Cancer," 2015 Second International Conference on Advances in Computing and Communication Engineering, Dehradun, 2015, pp. 644- 648.

[6] B. M. Gayathri and C. P. Sumathi, "Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-5.

[7] H. R. Mhaske and D. A. Phalke, "Melanoma skin cancer detection and classification based on supervised and unsupervised learning," 2013 International conference on Circuits, Controls and Communications (CCUBE), Bengaluru, 2013, pp. 1-5.

[8] S. Aruna, S. P. Rajagopalan and L. V. Nandakishore, "An algorithm proposed for Semi- Supervised learning in cancer detection," International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2011), Chennai, 2011, pp. 860-864.