



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Special Issue 2, December 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569

Diagnosis Breast Cancer with Neural Network and Boosting Algorithm with Random under Sampling Techniques

Mrs. K. KowsalyaDevi¹, Dr.M.Deepamalar², P. Ranjani Devi³

Assistant Professor, Paravathy's Arts and Science College, Dindigul, TamilNadu, India¹

Associate Professor, Paravathy's Arts and Science College, Dindigul, TamilNadu, India²

PG Scholar, Paravathy's Arts and Science College, Dindigul, TamilNadu, India³

ABSTRACT: The diagnosed cases of Breast cancer are increasing annually and unfortunately getting converted into a high mortality rate. Cancer, at the early stages, is hard to detect because the malicious cells show similar properties (density) as shown by the non-malicious cells. The mortality ratio could have been minimized if the breast cancer could have been detected in its early stages. But the current systems have not been able to achieve a fully automatic system which is not just capable of detecting the breast cancer but also can detect the stage of it. We proposed adaptive ensemble voting method for diagnosed breast cancer using Wisconsin Breast Cancer database. The system is based on Image processing and machine learning domains. Classification Imbalance problem, a machine learning problem, occurs when instances of one class is much higher than the instances of the other class resulting in an inefficient classification of samples and hence a bad decision support system. Therefore, EUS Boost, ensemble-based classifier is proposed which is efficient and is able to out perform other classifiers as it takes the benefits of both-boosting algorithm with Random Under sampling techniques. Also, comparison of EUS Boost, with other techniques is shown in the paper.

KEYWORDS: Breast Cancer, Neural Network, ADA Boost, Machine Learning Algorithm, WBCD

I. 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 researcher are still performing research for detecting and diagnosing cancer in an early stage. Because the early stage cancer is not a so painful and expensive for complete its treatment and many researchers 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.

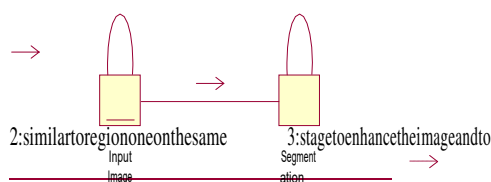
II. RELATED WORK

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), K- Nearest

Neighbour (KNN) and Convolutional Neural Network etc. Many researchers have done research in breast cancer by using 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'd Rasoul used DWT tool for image filtering and BPNN for processing and achieved 93.7% accuracy [1]. The author Clemen Deng used WHAVE algorithm with Wisconsin Breast Cancer Database and achieved 99% accuracy [2]. The author Ashwaq Qasem used marker Controller Watershed algorithm and achieved 95% accuracy [3]. The author Alireza Isareh used signal to noise ratio with SVM and achieved 98.80% accuracy [4]. The author Junaid Ahmad used Adaptive Resonance Theory with UCI database and got 84.21% accuracy [5]. By B.M. Gayathri work on comparative study about Relevance vector machine and achieved 97% accuracy [6]. The author Ms.H.R. Mhaske used KNN and SVN with 150 images database and achieved 80-90% accuracy [7]. The author S. Aruna used Naïve Bayes and SVM with UCI database and achieved 68-79% accuracy [8]. By Yohannes Tsehay developed a weakly supervised computer aided detection system that was used biopsy for learn data [8]. By Sudarshan Nayak used Naïve Bayes and SVM and got 98% accuracy [10]. 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.

PROCESS FLOW DIAGRAM: First, we collected the 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. 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 format. 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 distribution can be compared. In proposed work we have created top 16 feature violin

1:



5: rule based reasoni

plot for comparison.

III. THE EXPERIMENTAL RESULT

4: Computer Aided Diagnosis

→



PROPOSEDMETHODOLOGY: 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 behaviors 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. In this method the dataset is a common requirement for many machine learning estimators.

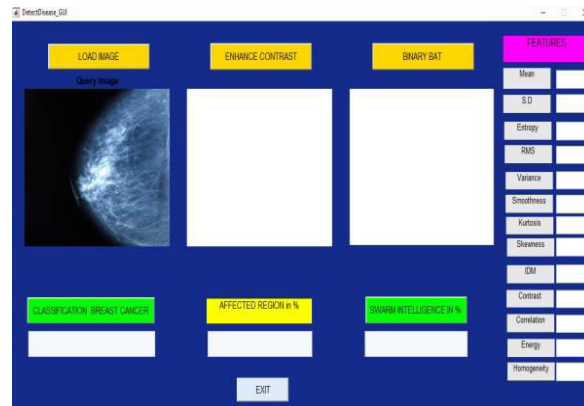


Fig.1 Load Image

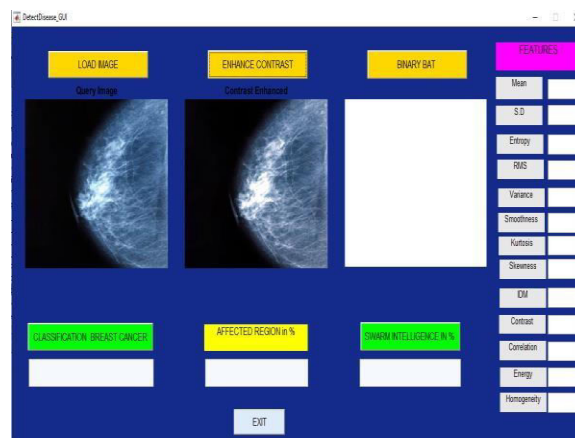


Fig.2 Enhance Contrast

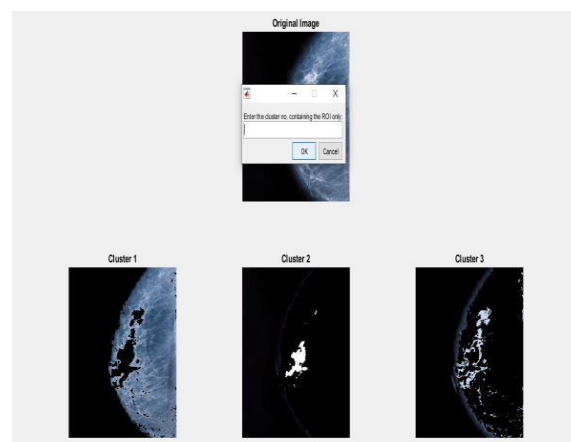


Fig.3 Binary Bat

15th & 16th September 2021

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

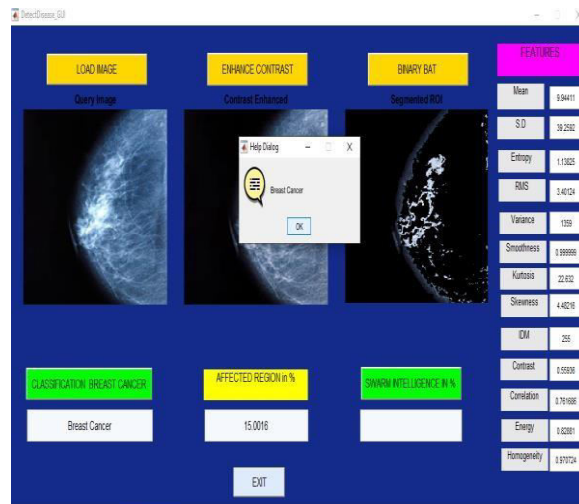


Fig.4 Classification of Breast Cancer

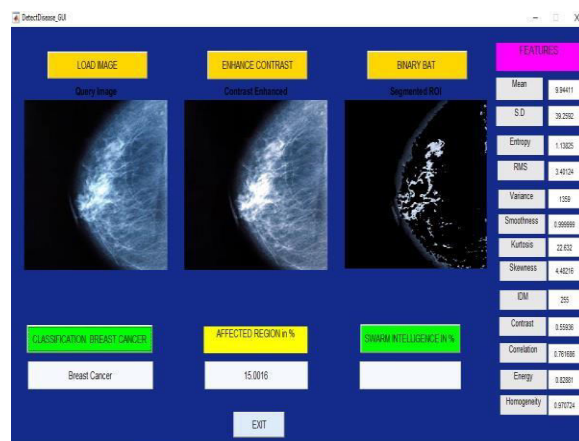


Fig.5 Affected Region

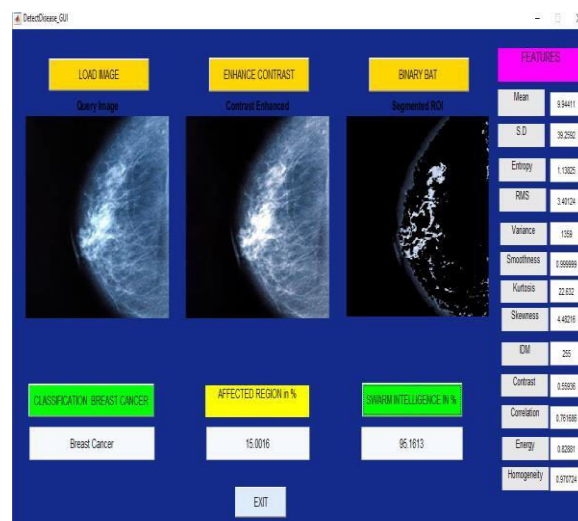


Fig.6 Swarm Intelligence

15th & 16th September 2021

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

IV. RESULT

In this paper we proposed Ensemble Machine Learning algorithm with Logistic and Neural Network for diagnosis and detection of breast cancer. We have used standardization method for pre-processing breast cancer dataset then we have applied Univariate Features Selection algorithm. Univariate Feature Selection algorithm used chi2 method for selection Best 16 Features from UCI dataset. After collect final 16 features from univariate Feature Selection algorithm we implement logistic and neural network algorithm on these 16 features and final applied voting algorithm on result and achieved 98.50% accuracy. Wisconsin Breast Cancer Dataset have contain 699 rows with features categories 30 features. After applied Univariate Feature Selection method top 16 features are decided from final model implementation. Because large features are effect on cost of model implementation. Achieved accuracy is good from individual achieved accuracy from both machine learning algorithm.

V. CONCLUSION AND FUTURE SCOPE

After extracting the features and collecting the feature dataset, the EUS Boost algorithm was applied on a population size of 50 instances. The Geometric Mean (GM) was taken as the evaluation measure, majority was taken as the majority type with true balancing and the total number of iterations carried out were 10. As a result, a Boosted classifier was gained which could classify even the ambiguous instances in the proper classes very efficiently. It helped maintaining the diversity of the base classifiers and hence provided a stability to the classification function. It also benefitted the classification scheme by providing a confidence level of the base classifiers in the weight update.

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.

The difficulty is not only connected with the uneven distribution of the instances in the dataset but also the difficulties present in the nature of the data used. By looking at the above table it can be concluded that the standard approaches like Bagging and Boosting independently are insufficient to deliver a satisfactory classification result. Hence the classification imbalance problem can be solved with EUS method as it provides maximum efficient and satisfactory result as compared to classical models when different segmentation techniques and measures are used.

REFERENCES

- [1] User Interfaces in C#: Windows Forms and Custom Controls by Matthew MacDonald. Applied Microsoft® .NET Framework Programming (Pro- Developer) by Jeffrey Richter.
- [2] Practical. Net2 and C#2: Harness the Platform, the Language, and the Framework by Patrick acchia.
- [3] Operating System Concepts, by Abraham Silberschatz.
- [4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A Berkeley view of cloud computing," University of California, Berkeley, Tech. Rep. USB-EECS-2009-28, Feb 2009.
- [5] "The apache Cassandra project," <http://cassandra.apache.org/>. L. Lamport, "The part-time parliament," ACM Transactions on Computer Systems, vol. 16, pp. 133–169, 1998.
- [6] N. Bonvin, T. G. Papanicolaou, and K. Aberer, "Cost-efficient and differentiated data availability guarantees in data clouds," in Proc. of the ICDE, Long Beach, CA, USA, 2010.
- [7] O. Regev and N. Nisan, "The popcorn market. online markets for computational resources," Decision Support Systems, vol. 28, no. 1-2, pp. 177 – 189, 2000.
- [8] A. Heusinger and T. Wright, "Cougaar: A robust configurable multi agent platform," in Proc. of the IEEE Aerospace Conference, 2005. J. Brunelle, P. Hurst, J. Huth, L. Kang, C. Ng, D. C. Parkes, M. Seltzer, J. Shank, and S. Youssef, "Egg: an extensible and economics-inspired open grid computing platform," in Proc. of the GECON, Singapore, May 2006. J. Norris, K. Coleman, A. Fox, and G. Candea, "Oncall: Defeating spikes with a free-market application cluster," in Proc. of the International Conference on Autonomic Computing, New York, NY, USA, May 2004.
- [9] C. Pautasso, T. Heinis, and G. Alonso, "Autonomic resource provisioning for software business processes," Information and Software Technology, vol. 49, pp. 65– 80, 2007.
- [10] A. Dan, D. Davis, R. Kearney, A. Keller, R. King, D. Kuebler, H. Ludwig, M. Polan, M. Spreitzer, and A. Youssef, "Webservices on demand: Wsla-driven automated management," IBM Syst. J., vol. 43, no. 1, pp. 136–158, 2004.



15th & 16th September 2021

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

[11]14.M. Wang and T. Suda, "The bio- networking architecture: abiologically inspired approach to the design of scalable, adaptive, and survivable/available network applications," in Proc. of the IEEE Symposium on Applications and the Internet, 2001.

[12]N.Laranjeiro and M.Vieira, "Towards fault tolerance in web services compositions," in Proc. of the workshop on engineering fault tolerant systems, New York, NY, USA, 2007.

[13]C. Engelmann, S. L. Scott, C. Leangsuksun, and X. He, "Transparent symmetric active/active replication for service level high availability," in Proc. Of the Acrid, 2007.



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details