JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm

Ms. Tejaswi Yamala [*1], Mr.CH.Dinesh [*2]

[1]MCA Student, Department of Master of Computer Applications,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.
[2]Assistant Professor, Department of Information Technology,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.

Abstract:

Breast Cancer, identified by the Breast Cancer Institute (BCI) as one of the most perilous diseases affecting women globally, underscores the critical need for effective detection. Early diagnosis, particularly in the initial stages, significantly contributes to saving lives. Platforms such as cancer.net provide tailored guides covering a spectrum of over 120 cancer types and related hereditary syndromes.In the realm of breast cancer detection, machine learning techniques take center stage. This work introduces an innovative approach – the adaptive ensemble voting method – for diagnosing breast cancer, leveraging the Wisconsin Breast Cancer database. The primary objective is to scrutinize and elucidate how logistic algorithms can offer superior solutions when collaborating with ensemble machine learning algorithms for breast cancer diagnosis, even in scenarios where variables are reduced.

Keywords: Breast Cancer, Adaptive Ensemble,Cancer Database, Syndromes

1. INTRODUCTION

Breast cancer stands out as one of the most perilous diseases globally, particularly affecting women, leading to numerous fatalities each year. Manual detection is time-consuming, posing challenges for physicians in accurate classification. Thus, the adoption of automatic diagnostic techniques becomes imperative. Various methods and algorithms, including Support Vector Machine, Naïve Bayes, KNN, and the cutting-edge Convolutional Neural Network (CNN) in deep learning, have emerged for breast cancer detection.

CNN and deep learning algorithms excel in image classification and object detection. This paper leverages the UCI open database for training and testing, featuring two tumor classes: Benign and Malignant. Benign tumors are non-cancerous, while malignant tumors pose a cancer threat. Researchers continue to explore early-stage cancer detection due to its less painful nature and lower treatment costs. Developing an effective diagnosis system for early tumor detection is crucial to initiating prompt treatment and improving resolution rates.

2. LITERATURE SURVEY

1. S. Nayak and D. Gope, "Comparison of Supervised Learning Algorithms for RF-based Breast Cancer Detection," 2017 Computing and Electromagnetics International Workshop (CEM), Barcelona, 2017, pp.

Three-dimensional imaging based on radiofrequency, exploiting dielectric property contrasts in tissues, offers a non-invasive and non-ionizing method for breast cancer detection. This paper explores the application of various supervised machine learning algorithms for classifying breast tissues into less-dense fatty and dense fibro glandular or malignant classes. The classification is based on measured scattered electric field data obtained through antennas around the breast tissue. The paper includes a performance comparison of these algorithms, with potential quantitative non-linear optimization for enhanced tissue profile reconstruction.

2. L. Latchoumi, T. P., & Parthiban, "Abnormality Detection Using Weighted Particle Swarm Optimization and Smooth Support Vector Machine," Biomed. Res., vol. 28, no. 11, pp. 4749–4751, 2017.

This paper proposes a hybrid classification approach employing Weighted-Particle Swarm Optimization (WPSO) for data clustering in conjunction with Smooth Support Vector Machine (SSVM) for classification. The performance of WPSO clustering is compared with K-means and fuzzy methods using inter-cluster, intra-cluster, and validity index metrics. The proposed WPSO-SSVM classification methodology demonstrates accuracy rates of 83.76% for liver disorder, 98.42% for WBCD, and 95.21% for mammographic mass data, surpassing existing literature.

3. Fabian Pedregosa and all (2011). "Scikit-learn: Machine Learning in Python". Journal of Machine Learning Research. 12: 2825–2830.

Scikit-learn, a machine learning library in Python, provides a versatile collection of both supervised and unsupervised algorithms. It offers a consistent, task-oriented interface, facilitating easy comparison of methods for specific applications. Integrated into applications beyond traditional statistical data analysis, it becomes a valuable tool for various use cases, including medical imaging. The implemented algorithms, in a high-level language, serve as adaptable building blocks for case-specific approaches. Future work includes advancements in online learning to scale for large datasets.

3. EXISTING SYSTEM

The rapid growth of machine learning has led to a division between traditional methods and modern machine learning techniques. In the context of Breast Cancer Institute using Machine Learning Model Detection, this section provides an overview of related works and highlights the superiority of machine learning methods over traditional approaches. The existing method in this project involves a specific flow used for model development, employing Support Vector Machine (SVM) algorithms. However, this approach faces challenges such as high memory requirements and less accurate results.

Limitations of the Existing System:

1. Accuracy low
2. Requires more time
3. Difficult to handle

4. PROPOSED SYSTEM

Numerous machine learning algorithms are at one's disposal for predicting and diagnosing breast cancer. Some of these algorithms encompass Decision Tree Classifier, K Neighbors Classifier, Random Forest Classifier, Logistic Regression, MLP Classifier, and Cat Boost Classifier. In our approach, we employ the proposed Ensemble Voting method to determine the best method for diagnosing breast cancer. In this stage, we initially implement the

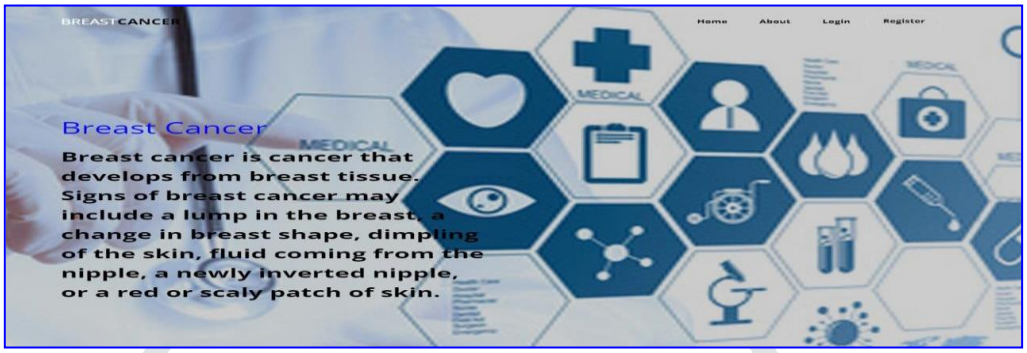
Random Forest Classifier algorithm on the datasets individually. Subsequently, we implement the Voting Ensemble algorithm to combine these results and compute the final accuracy.

Advantages of the Proposed System:

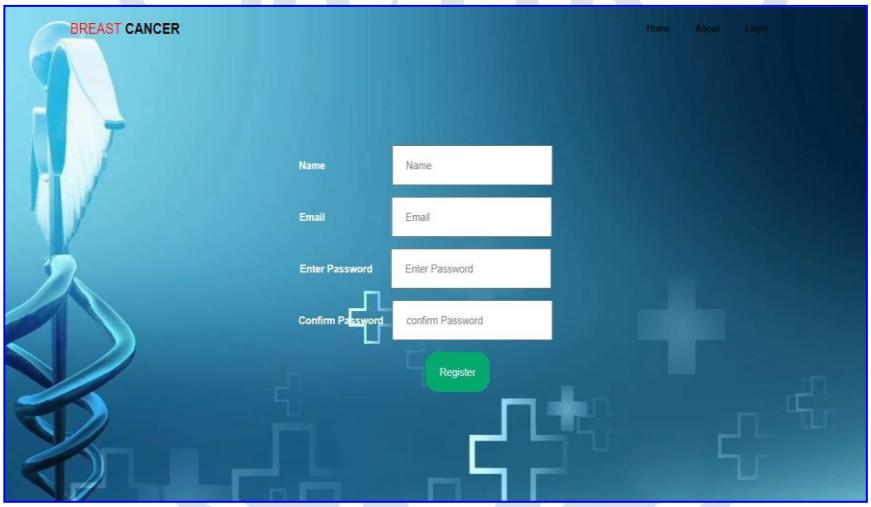
1. Requireslesstime
2. GoodAccuracy
3. Easytohandle

5. EXPERIMENTAL RESULTS

Home Page:



Login:



6. CONCLUSION

This work proposes an ensemble machine learning method for diagnosing breast cancer, as demonstrated in the table and graph, showcasing an impressive accuracy of 98.50%. The study utilizes only 16 features for cancer diagnosis, with plans to extend the analysis to include all features in the UCI dataset for improved accuracy in the future. The research affirms that neural networks are effective for analyzing human vital data and facilitating pre-diagnosis without specialized medical knowledge.In this paper, an Ensemble Machine Learning algorithm is introduced, combining Logistic Regression and Neural Network for the diagnosis and detection of breast cancer. The dataset undergoes standardization as a pre-processing step, followed by the application of the Univariate Features Selection algorithm, utilizing the chi2 method to select the best 16 features from the UCI dataset. After obtaining the final 16 features from the Univariate Feature Selection algorithm, logistic and neural network algorithms are implemented on these features. The ensemble method, employing a voting algorithm on the results, achieves an impressive accuracy of 98.50%.The Wisconsin Breast Cancer Dataset, comprising 699 rows and 30 feature categories,

undergoes feature selection to optimize the model's implementation cost. The achieved accuracy is noteworthy, surpassing individual accuracy results from both machine learning algorithms..

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.