IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm

Sharmila G K, Kamalapadu Varsha, Naga Bhavani K, Sandhya R B, G Archana
Department of Computer Science &Engineering
Rao Bahadur Y Mahabaleswarappa Engineering College, Ballari

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. 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 logistic algorithm provide better solution when its work with ensemble machine learning algorithms for diagnosing breast cancer even the variables are reduced. There are 2 types tumours are there. One is Benign Tumour and the other is malignant in which benign Tumour is non-cancerous and the malignant is a cancer Tumour.*

Keywords: Logistic Regression, SVC, Random Forest, Decision Tree, Cat Boost, KNeighbours, MLP Classifier.

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 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.

II. LITERATURE REVIEW

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 This paper demonstrates the use of various supervised machine learning algorithms in classification of breast tissues into less-dense fatty and dense fibroglandular or malignant classes from the measured scattered electric field data obtained through antennas placed around the breast tissue.
H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, 'Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis', Procedia Computer Science, vol. 83, pp. 1064–1069, 2016, doi: 10.1016/j.procs.2016.04.224
In this paper, a performance comparison between different machine learning algorithms: Support Vector Machine

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

(SVM), Decision Tree (C4.5), Naive Bayes (NB) and k Nearest Neighbors (k-NN) on the Wisconsin Breast Cancer (original) datasets is conducted.

L. Latchoumi, T. P., & Parthiban, "Abnormality detection using weighed particle swarm optimization and smooth support vector machine," Biomed. Res., vol. 28, no. 11, pp. 4749–4751, 2017. In this paper, a new hybrid classification approach, which uses Weighted-Particle Swarm Optimization (WPSO) for data clustering in sequence with Smooth Support Vector Machine (SSVM) for classification is proposed.

Fabian Pedregosa and all (2011). "Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 12: 2825–2830 Scikit-learn exposes a wide variety of machine learning algorithms, both supervised and unsupervised, using a consistent, task-oriented interface, thus enabling easy comparison of methods for agiven application.

III. EXISTING SYSTEM

The increasing growth of machine learning, computer techniques divided into traditional methods and machine learning methods. This section describes the related works of classification of Breast Cancer Institute Using Machine Learning Model Detection and how machine learning methods are better than traditional methods. The existing method in this project have a certain flow is used for model development Support Vector Machine (SVM)
are used algorithms in existing system. But it requires large memory and result is not accurate.

IV. PROPOSED SYSTEM

Many machine learning algorithms are available for prediction and diagnosis of breast cancer.Some of the machine learning algorithm are Decision Tree Classifier, K Neighbors Classifier, Random Forest Classifier, Logistic Regression, MLP Classifier, Cat Boost Classifier and GaussianNB. We used proposed Ensemble Voting method and compute best method for diagnosis breast cancer disease.In this stage we have first implement Random Forest Classifier algorithm on these dataset and the implement algorithm individual then we are implement Voting Ensemble algorithm for combine these results and an compute the final accuracy.

4.1 Objectives

- The primary goal of this project is to determine the Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm with Early Classification.
- Based Approach for Fault Classification in whether the result is Cancer are Not Cancer this we used Stochastic Gradient Descent Classification Logistic Regression, SVC, Random Forest, Decision Tree, CatBoost , K Neighbours , MLP Classifier

V. METHODOLOGY

1. Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

2. Support Vector Machine Algorithm:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane.

3. Random Forest

A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms. The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome. A random forest eradicates the limitations of a decision tree algorithm. It reduces the over fitting of datasets and increases precision. It generates predictions without requiring many configurations in packages.

4. Decision Tree

A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal. A decision tree is drawn upside down with its root at the top. In the image on the left, the bold text in black represents a condition/internal node, based on which the tree splits into branches/ edges. The end of the branch that doesn't split anymore is the decision/leaf, in this case, whether the passenger died or survived, represented as red and green text respectively. Although, a real dataset will have a lot more features and this will just be a branch in a much bigger tree, but you can't ignore the simplicity of this algorithm. The feature importance is clear and relations can be viewed easily. This methodology is more commonly known as learning decision tree from data and above tree is called Classification tree as the target is to classify passenger as survived or died. Regression trees are represented in the same manner, just they predict continuous values like price of a house. In general, Decision Tree algorithms are referred to as CART or Classification and Regression Trees.

5. MLP Classifier

Multi-Layer perceptron defines the most complex architecture of artificial neural networks. It is substantially formed from multiple layers of the perceptron. TensorFlow is a very popular deep learning framework released by, and this notebook will guide to build a neural network with this library. If we want to understand what is a Multi-layer perceptron, we have to develop a multi-layer perceptron from scratch using Numpy. MLP networks are used for supervised learning format. A typical learning algorithm for MLP networks is also called back propagation's algorithm. A multilayer perceptron (MLP) is a feed forward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input nodes connected as a directed graph between the input and output layers. MLP uses backpropagation for training the network. MLP is a deep learning method.

VI. PROBLEM STATEMENT

To overcome the problem 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.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

Expected Output

1. System:

1.1 Store Dataset:

The System stores the dataset given by the user.

1.2 Model Training:

The system takes the data from the user and fed that data to the selected model.

1.3 Model Predictions:

The system takes the data given by the user and predict the output based on the given data.

1.4 Data Splitting:

Here the system can split the data into two parts for training and testing.

2. User:

2.1 Load Dataset:

The user can upload the dataset he/she want to work on.

2.2 View Dataset:

The User can view the dataset.

2.3 Select model:

User can apply the model to the dataset for accuracy.

2.4 View results:

User can view the predicted results weather the system is attacked or not.

VII. 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. 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

VIII. FUTURE ENHANCEMENT

Breast cancer diagnosis has undergone significant advancements in recent years, and there is a lot of potential for future development in this field. Here are some possible future scopes for breast cancer diagnosis. This is a non-invasive test that can detect cancer cells or DNA fragments released into the bloodstream by a tumor. Liquid biopsy is still in its early stages, but it has the potential to revolutionize cancer diagnosis and treatment. AI has already shown promise in breast cancer diagnosis by analyzing medical images to identify suspicious areas that may require further testing. With further development, AI could potentially improve the accuracy of breast cancer detection and reduce false positives. New imaging techniques, such as digital breast tomosynthesis (DBT), could provide better visualization of breast tissue and help detect cancers that may be missed by traditional mammography.Researchers are looking for new biomarkers that can detect breast cancer earlier and with greater accuracy. For example, scientists are exploring the use of microRNAs, which are small molecules that regulate gene expression, as potential biomarkers for breast cancer.Advances in genetic testing and molecular profiling may allow for more personalized treatment options based on an individual's specific cancer characteristics. This could lead to more effective and targeted treatments for breast cancer patients.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

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]. 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]. 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.

[6]. Y. Tsehay et al., "Biopsy-guided learning with deep convolutional neural networks for Prostate Cancer detection on multiparametric MRI," 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 642-645.