JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1569 - 1575
https://publishoa.com
ISSN: 1309-3452

Analyze Remnant Impact of Covid -19 On Breast Cancer Using Machine Learning

Devanand

M.Tech.Scholar

Computer Science & Engineering Department, VBSPU, Jaunpur

Email: engineerdevanand@gmail.com

Dileep Kumar Yadav

Assistant Professor

Computer Science & Engineering Department ,VBSPU , Jaunpur

Sanjeev Gangwar

Assistant Professor

Computer Application Department ,VBSPU , Jaunpur

ABSTRACT

Background*-Breast cancer is classified as either malignant or benign. Breast skins and other blood-forming organs create an overabundance of aberrant or immature white blood cells while inhibiting the development of normal cells. Breast cancer is commonly used in machine learning methodologies, whether it's to classify different forms of breast cancer or to determine whether a patient has breast cancer.*

Methods:*Support Vector Equipment, Close Neighbor K, Nave Bayes, Random Forest, Decision Tree, and Logistic Regression are used to assess the impact of breast cancer remnants during Covid-19. The using data mining method to compare the accuracy of the data set with all attributes to the accuracy of the classifier with the specified features.*

Result: *Understanding the influence of test data on diagnostic outcomes, as well as the connection between qualities, is the focus of this research study on Machine Learning Algorithms on Breast Cancer Data. Set the accuracy and training score to 100 percent. who have unending breast cancer and have not enervated the algorithm on the basis of the computation situated among the KNN, SVM, N.B., Random Forest, Logistic Regression, and Decision Tree method utilised in this proceeding? Random Forest has the greatest accuracy of 96.61 percent, while Decision Tree has a 94.73 percent accuracy. Random Forest has a high Training Score of 100 percent, while Decision Tree has a high Training Score of 99.34 percent.*

Conclusion: *The region is being studied using the Information Cultivate system. Random Forest and Decision Tree, Support Vector Machine, K-Nearest Neighbor, Logistic Regression, and Naive Bayes were among the machine learning algorithms used in the research. A Random Forest and Decision Tree algorithm is used to predict if patients with recurrent breast cancer regret infection and whether they will not endure this illness. The new findings suggest that the random forest classifier excels at predicting show outcomes with the highest precision and shortest execution time.*

Keywords: SVM, Random Forest Algorithm, Close NeighborK, Logistic Regression, Decision Tree Algorithm. Naive Bayes.

1. INTRODUCTION

The Wisconsin diagnostic site and Kaggle's breast cancer data collecting The upgrade has been communicated to our computer's data set and facts. Breast cancer is a major health problem that affects a large number of women. This problem was investigated using the data set. According to the doctor, a mammography and an ultrasound are used to analyse the patient with breast cancer[1]. The area is close to a research facility. Sunlight exposure lowers the risk of breast cancer[2].

1.1 Breast Cancer and its symptoms

Breast cancer is usually asymptomatic. While the tumour is still in its infancy. After then, you may utiliseeasy to double-check it. Women who are ill are first and primarily screened for breast cancer. Following that, the body produces a barren physical lump. Breast cancer can spread to the lymph nodes under the armpit and cause a lump or swelling even before the lower breast tumor is large enough to be felt.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1569 - 1575
https://publishoa.com
ISSN: 1309-3452

2.MACHINE LEARNING INTRODUCTION

Machine learning is currently being used to reduce breast cancer's impact. Machine learning was used to look for patterns in the prediction and features of breast cancer. Machine learning is widely used in the fields of research and software development. New data sets related to breast cancer have been added to the web platform. The following are some examples of machine learning components: produced data with the use of machine learning, such as conversing with it and calculating attendance with it[1].

2.1Concepts of Learning

Explain how to internalise learning routes based on machine learning knowledge content. We can learn machine learning in three different classes. Association, Student, and Earth Zone are just a few examples. There are four stages to the machine learning process.

- Learning that is supervised.
- Process of learning without supervision.
- Learning process that is semi-regulated.
- The learning process is based on reinforcement.

Machine Learning is performed through Learning that is supervised and Process of learning without supervision Semi Regulated and Reinforcement Learning Process

2.1.1 SupervisedLearning Supervised learning is an approach in machine learning. Real-time application of supervised learning is used to peel, schedule tests, and check information. For classification and regression, supervised learning is applied. Supervised learning is used in machine learning to assign and adjust the values of valid data. Supervised learning is used to cure weak tumours in men and women who are related to health. Supervised learning is a model of women's health that is connected to, among other things, unlimited credit. The supervised technique is used in machine learning. The purpose of this test is to develop a framework for determining the kind of fever. Which ones are used is determined by the patient. Display temperature, which is the name of the force of migration, as an example. This section offers data on a wide range of topics.

2.1.2Unsupervised Learning

Unsupervised learning enables for the detection of disparities. Unsupervised learning is a unit of comparison in learning. There is no way to handle it because there is no data. Unsupervised learning is a linguistic process in this study work, and the search for signs, names, or structures, and distinctions is done from the top side.

2.1.3 Semi-Regulated Learning

Some learning tests are being marked. Semi-regulated learning is the designated stage in the learning process. It can prepare unlabeled data for a model of the quantity of named data needed for a try-out score. The most recent learning processing in the state is a semi-regulated learning permit. The cost of obtaining the dataset's absolute name. To a sign of a tiny subset, we are Progressive pragmatic[1].

2.1.4 Reinforcement Learning

The goal is to give the learning data with the framework and dynamic condition of the specific goal. The frameworks are exhibition dependency into input reaction responses in a good-looking manner[3].

3. ALGORITHMS FOR MACHINE LEARNING

3.1 ALGORITHM OF THE K-NEAREST NEIGHBOR

K Nearest Neighbor is a type of algorithm of supervised learning. It can allow for both classifications as well as prediction of the regressive part in the issues. Its work involves the assistance of the following advances-

$$d=\sqrt{((x2-x1)^2+(y2-y1)^2)}$$

calculating the distance between two spots.

3.2 VECTOR MACHINE SUPPORT

The support vector machine is a machine that generates support vectors learning algorithm is incredible. We're adaptable, and we're the ones in charge of the machine learning algorithm. It can use both categories and divisive to address the problem in a linked manner. A way is the support vector machine. In contrast to continuous and categorical variables, the SVM provides the popular outcomes of the volume they are working with.

- Support Vectors - This is a designation for the subject at hand, and the support vector component is well recognised. With the help of the characterised into the data point, the isolating line is designated.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1569 - 1575
https://publishoa.com
ISSN: 1309-3452

- Hyper Plane - It can locate an outline and choose a plane or space. Whose is separated from more than one object of a different class?
- Margin may be defined as the space between two lines where data points are indicated in different classifications. The tendency of determining the otherwise optimal path from a line with the use of support vectors. A large edge of a decent edge is referred to as a horrible edge, whereas a tiny edge is referred to be a terrible edge[5].

3.3LOGISTIC REGRESSION

Because the data may not always match a straight line, linear regression may not always fit. However, the straight line esteems that are more than 1 and less than 0 may be more conspicuous.
Simple logistic regression
Y = mx + c
Output = 0 or 1, hypothesis =>Z =WX+B hθ(x) = sigmoid (Z)
If 'Z' reaches infinity, Y (predicted) becomes me, whereas if 'Z' reaches negative infinity, it becomes zero.
The supervised learning of the classification algorithm used to predict and likelihood of target task is the kind of regression. Nature's dichotomous objective is dichotomous, which implies there are just two viable classifications[1].

3.4 NAIVE BAYES

The naïve bays rule can apply for a method for going from P(X/Y), it is known by the preparation of the dataset to discover P(X/Y).
What occurs if Y has multiple classes? We process the likelihood of each class of Y and let the most elevated success.
P(X/Y) =P (X∩Y)/P(Y)
[P (Evidence / outcome) (Known from training data)]
P(Y/X) = P (X∩Y)/P(X)
[P (Outcome / Evidence) (to be predicted for test data)]
Nave Bays calculations are an arranging approach based on Bays' hypothesis and the assumption that each of the indicators is independent of the others. The governing aim in naive Bayesian depiction is to identify the back probabilities, such as the probability of a name given some watched feature, (L | features).

3.5 RANDOM FOREST

In another situation of the numerous classifications issue, supervised learning is a random forest that allows classifications and regression. This research proposes the creation of a forest based on the comparison of a single tree and several trees using a progressive robust forest. It also uses a random forest technique to use a decision tree on data samples and generate a forest from each one, resulting in the selection of a suitable solution at a cost of a vote. A single selected tree is the best of all things.
It's gone since it's below the term that's over-fitting by average.
Random Forest Algorithm

- Step 1- The first stage is selecting random samples from a dataset that has been supplied..
- Step 2- The forest will be built backwards from any more expected choice tree in the second stage of the computation, which will create the choice tree in each case..
- Step3 – For each predicted outcome, the next stage in the process will be to use the progression and cast the ballot..
- Step 4- Finally, choose the most expensive ballot forecast outcome, as well as the advanced prediction result..

3.6 DECISION TREE

A decision tree is a certain way to communicate with data. That is provided to the dataset's knowledge set, the hard work of classification and labelling methods, and the appearance for the most basic rules Until they reach a certain level of resemblance, the observations are not the same.. It may be utilised of the layer and they are used in a layered giveaway process, where each layer separates the details into one to many groups, with the data frequently falling under an equivalent group that is quite identical to every other group[4].

4. PATIENT DATA SET

The entire dataset regarding breast cancer has been gathered from the Kaggle and UCI repositories in 570 instances with 32 characteristics. The diagnostic' qualities are specified as quantifiable, with zero indicating that the patient does not have breast cancer (B= Benign), and one indicating that the patient does have breast cancer (M = Malignant). The value attribute in the breast cancer dataset is described in Table I. The dataset contains 570 records, 357 of which have no breast cancer (B = Benign) and 212 of which have breast cancer[3].

5. PROPOSED TECHNIQUE

As asserted by machine learning algorithms for breast cancer disease divination, the goal of the approach is to build the

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1569 - 1575
https://publishoa.com
ISSN: 1309-3452

best technique possible. According to access, the machine learning algorithms have been discussed, and this is the execution of the metre approach that has been evaluated[2].

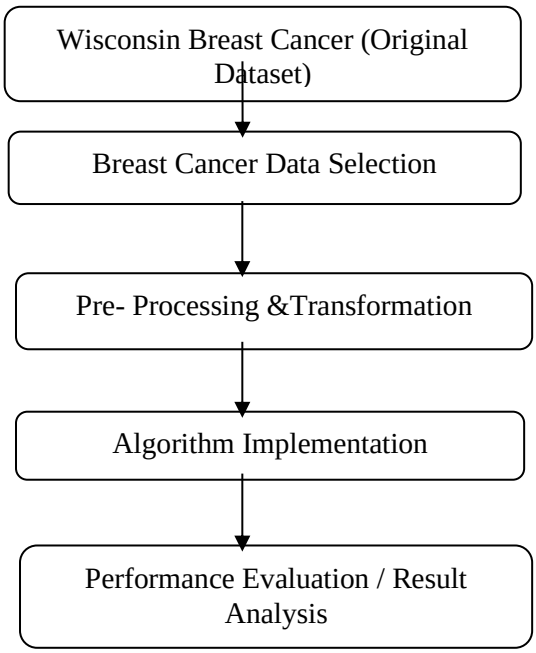

Fig.1 Breast Cancer Proposed Technique Methodology

5.1 PERFORMANCE MEASURES FOR CLASSIFICATION

It is indicated to the confusion matrix is some condition follows as.

- True positives refer to a positive breast cancer record that the classifier correctly identified[6].
- True negative records are those that were correctly classified as negative breast cancer records by the classifier[7].
- False positives are records of negative breast cancer that were mistakenly classified as positive.
- False-negative is a term used to describe positive breast cancer records that have been wrongly labelled as negative.
- A matrix to help distinguish between positive and negative records[9].

- **Confusion Matrix Components**

		Yes	**No**	
Actual Class	Yes	Positives that are true (TP)	Negatives that aren't true (FN)	P
	No	Positives that aren't true (FP)	Negatives that are true (TN)	N
		P Complement	N Complement	P + N

Table.1 Confusion Matrix of Breast Cancer Data-set.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1569 - 1575
https://publishoa.com
ISSN: 1309-3452

Table2. Results of Prediction, Recall, F1-Score While Using Various algorithms with Breast cancer Dataset

Name of the Algorithms	Averages	Precision	Recall	F1- Score	Support
K-NN	Macro. Average.	0.74	0.65	0.64	114
	Weighted Average	0.73	0.70	0.67	114
Logistic Regression	Macro Average	0.29	0.50	0.37	114
	Weighted Average	0.35	0.59	0.44	114
Decision Tree	Macro Average	0.94	0.95	0.95	114
	Weighted Average	0.95	0.95	0.95	114
Random Forest	Macro Average	0.96	0.96	0.96	114
	Weighted Average	0.96	0.96	0.96	114
Support Vector Machine	Macro Average	0.29	0.50	0.37	114
	Weighted Average	0.35	0.59	0.44	114
Naïve Bayes	Macro Average	0.63	0.51	0.41	114
	Weighted Average	0.62	0.60	0.47	114

Measure	**Formula**
Accuracy, Recognition Rate	TP+TN …… P+N
Error, Misclassification Rate	FP+FN. ……. P+N
Sensitivity, True Positive Rate Recall	TP……. P
Specificity, True Negative Rate	TN……. N
Precision	TP……. TP+FP
F, F1, F-Score, Harmonic Mean of Precision and Recall	2*Precision*recall…… Precision + Recall

Table3. Problem Solving With Help of the Measurement and Formula of the Breast Cancer Prediction

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1569 - 1575
https://publishoa.com
ISSN: 1309-3452

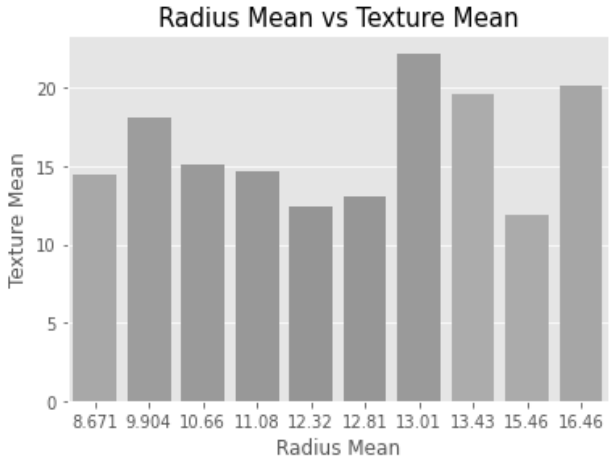

Fig.2 Radius Mean And Texture Mean help BCDS.

6. EXPERIMENTAL RESULT

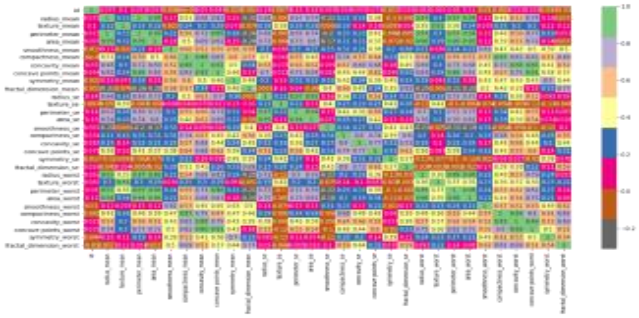

Fig. 3 Impact of the test data on the diagnostic result to observe and correlation between attributes

Understanding the influence of test data on diagnostic outcomes, as well as the connection between qualities, is the focus of this research study on Machine Learning Algorithms on Breast Cancer Data. Set the accuracy and training score to 100 percent. who have unending breast cancer and have not enervated the algorithm on the basis of the computation situated among the KNN, SVM, N.B., Random Forest, Logistic Regression, and Decision Tree method utilised in this proceeding? Random Forest has the greatest accuracy of 96.61 percent, while Decision Tree has a 94.73 percent accuracy. Random Forest has a high Training Score of 100 percent, while Decision Tree has a high Training Score of 99.34 percent[10].

Name of Algorithms	Training Score in %	Accuracy in %
K-Nearest Neighbor	80.21%	70.17%
Support Vector Machine	63.73%	58.77%
Naïve Bays	63.29%	56.29%
Random Forest	100.00%	96.49%
Logistic Regression	63.73%	58.77%
Decision Tree	99.34%	94.73%

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1569 - 1575
https://publishoa.com
ISSN: 1309-3452

Table.4 Find out of the Training Score and Accuracy Help of Machine Learning Algorithm.

7. CONCLUSION

The region is investigated using the Information Cultivate system. Machine learning techniques used in the study included Random Forest and Decision Tree, Support Vector Machine, K-Nearest Neighbor, Logistic Regression, and Naive Bayes. A Random Forest and Decision Tree approach is used to predict which patients with recurrent breast cancer regret infection and which do not. According to the most recent data, the random forest classifier surpasses other models in predicting the best show outcome in terms of precision and execution time.

8. FUTURE SCOPE

The Machine Learning Algorithm's accuracy and training score for predicting breast cancer. To increase accuracy and training score, the study will use additional machine learning algorithms such as CNN Algorithm and K- Means Cluster algorithms, regression analysis algorithm, and Cluster Analysis algorithm.

9. REFERENCES

[1] Nikita Rane, Jean Sunny, RuchaKanade, Prof. Sulochana Devi, Breast Cancer Classification And Prediction Using Machine Learning vol.09 Issue 02, February-2020.
[2] Md. Milon Islam, Hasib Iqbal, Md. RezwanulHaque and Md. Kamrul Hasan, Prediction Of Breast Cancer Using Support Vector Machine And K-Nearest Neighbors, 2017 IEEE Region 10 Humanitarian Technology Conference R-10-HTC.
[3] Dr. Stephanie J Seidler, Dr. DE. Huber, Overview of Diagnosis and Treatment of Breast Cancer in Young Women. Gynecology 2020.
[4] MeriemAmrane, SalihaOukid, IkramGagaoua, TolgaEnsari, Breast Cancer Classification Using Machine Learning78-1-5386-5135-3/18 IEEE.
[5] Chuck Easttom, SudipThapa, Justin Lawson, A Comparative Study Of Machine Learning Algorithms For Use In Breast Cancer Studies, 978-1-7281-3783-4/20 IEEE.
[6] Dr. MuktéviSrivenkatesh, Prediction of Breast Cancer Disease Using Machine Learning, International Journal of Innovative Technology and Exploring Engineering, Volume-9 Issue-4, February 2020.
[7] KNN With Different Distance Measures And Classification Rules Using Machine Learning, European Journal of Molecular and Clinical Medicine volume07, Issue 03, 2020.
[8] "Breast Cancer Wisconsin (Diagnostic) Data Set", Sep 2019, [online] Available:
[9] UCI, "Breast Cancer Wisconsin (Diagnostic) Data Set," Kaggle, 25-Sep 2016. [online] Available: https://www.kaggle.com/uciml/breast -cancer Wisconsin - data.
[10] "U.S Breast Cancer Statics", breastcancer.org, Sep 2019,[online] Available: https://www.breastcancer.org/symptoms/understand_bc/statistics/.