iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44488

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Breast Cancer Prediction Using Machine Learning

Ramya Challa[2], Reethika Bijjala[3], Dr. Mekala Sreenivas[1]
[1, 2]*B.Tech Students,* [3]*Associate Professor, Dept of IT, Sreenidhi Institute of Science and Technology(A), Hyderabad*

Abstract: ***Breast cancer is one of the diseases which cause number of deaths ever year across the globe, early detection and diagnosis of such type of disease is a challenging task in order to reduce the number of deaths. Now a days various techniques of machine learning and data mining are used for medical diagnosis which has proven there metal by which prediction can be done for the chronic diseases like cancer which can save the life's of the patients suffering from such type of disease. The major concern of this study is to find the prediction accuracy of the classification algorithms like Support Vector Machine, J48, Naïve Bayes and Random Forest and to suggest the best algorithm.***

I. NTRODUCTION

Now a day, breast cancer is one of the burning issue all over the world. It is one of the major health problem for women. Globally the incidence of breast cancer is only second to that of Lung cancer. The disease represents the main cause of cancer death among women. Breast cancer is developed from breast tissue. Signs of breast cancer may include a breast lump, skin dimpling, fluid coming from the nipple, breast shape change, a newly inverted nipple, or a scaly patch of skin.

Breast cancer typically attack post menopausal women. Both genetic and ancestral factor play a role. About 5-10% of breast cancer are hereditary and occur in the patient with mutation BRCA1, BRCA2 genes. Prolong estrogen exposure associated with early menarche, late menopause uses of hormone replacement therapy (HRT) has been associated with increased risk. For extracting medical knowledge, data mining techniques are extremely useful for medical education. Physicians can use data mining applications to identify operative treatments so that patients can get better and more reasonable healthcare services. Matching and mapping strategies become so operative in diagnosis with the help of data mining application.

Data in the healthcare industry is really complex and enormous. Dealing with a large amount of data is really hard. Data mining provide several types of methodologies and techniques to process a large number of data and pulling out useful information for decision making which is a fundamental part of the healthcare sector. Experts believe that data mining techniques in the healthcare industry will reduce the cost to 30% of overall healthcare spending. Electronic Health Records (EHR) is quickly becoming more common among healthcare facilities.

II. DESIGN OF THE SYSTEM

A. Existing System

Many comparative analysisi has beeni madei betweeni Decisioni Treei J48 algorithm and Bayesiani classification to determinei the breast cancer among the women and alsoi used different test option such asi cross validation and percentagei split to givei better result of comparison of Adaboot, SVM, Naives Bayes, Decision Tree, J48 algorithms and also implemented Sequential Minimali Optimization (SMO), Best Firsti Tree and IBKi data mining algorithm to obtaini the classificationi accuracy for the prediction of the breasti cancer.

B. Architectural Design

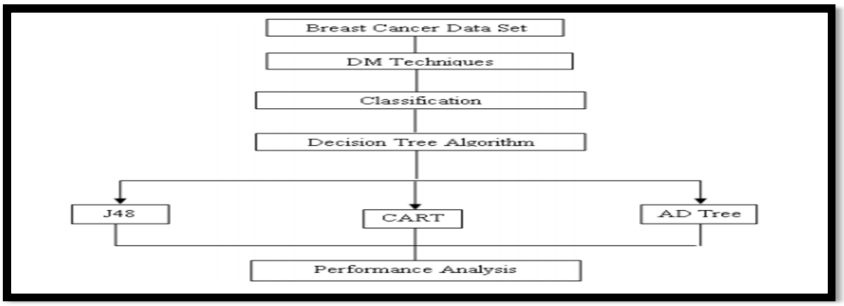


Figure 1.1 Architectural Design

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

III. METHODOLOGY

A. Language Or Technology Used

Python language is used to write the code. Python provides a wide variety of libraries forscientific and computational usage.

B. UML Diagrams

Any complex system is best understood by making some kind of diagrams or pictures. These diagrams have a better impact on our understanding. If we look around, we will realize that the diagrams are nota new concept but it is used widely in different forms in different industries.

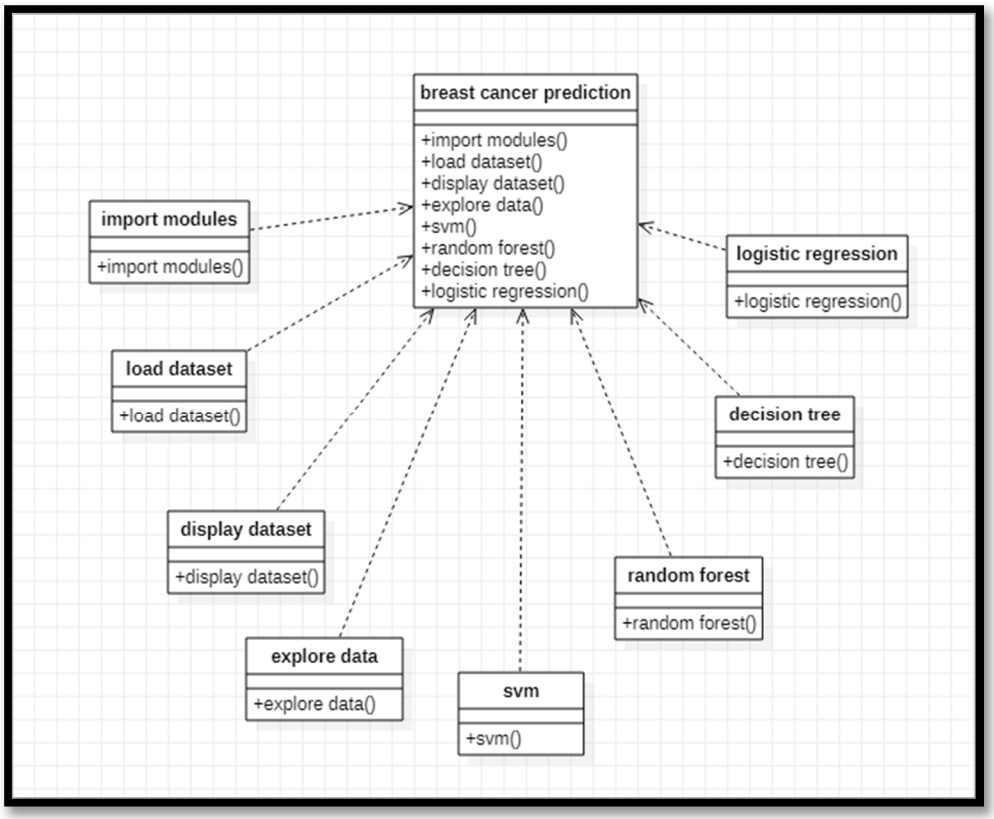
We prepare UML diagrams to understand the system in a better and simple way. A single diagram is not enough to cover all the aspects of the system. UML defines various kinds of diagrams to cover most of the aspects of a system. You can also create your own set of diagrams to meet your requirements. Diagrams are generally made in an incremental and iterative way. There are two broad categories of diagrams and they are again divided into sub categories –

1) Structural Diagrams

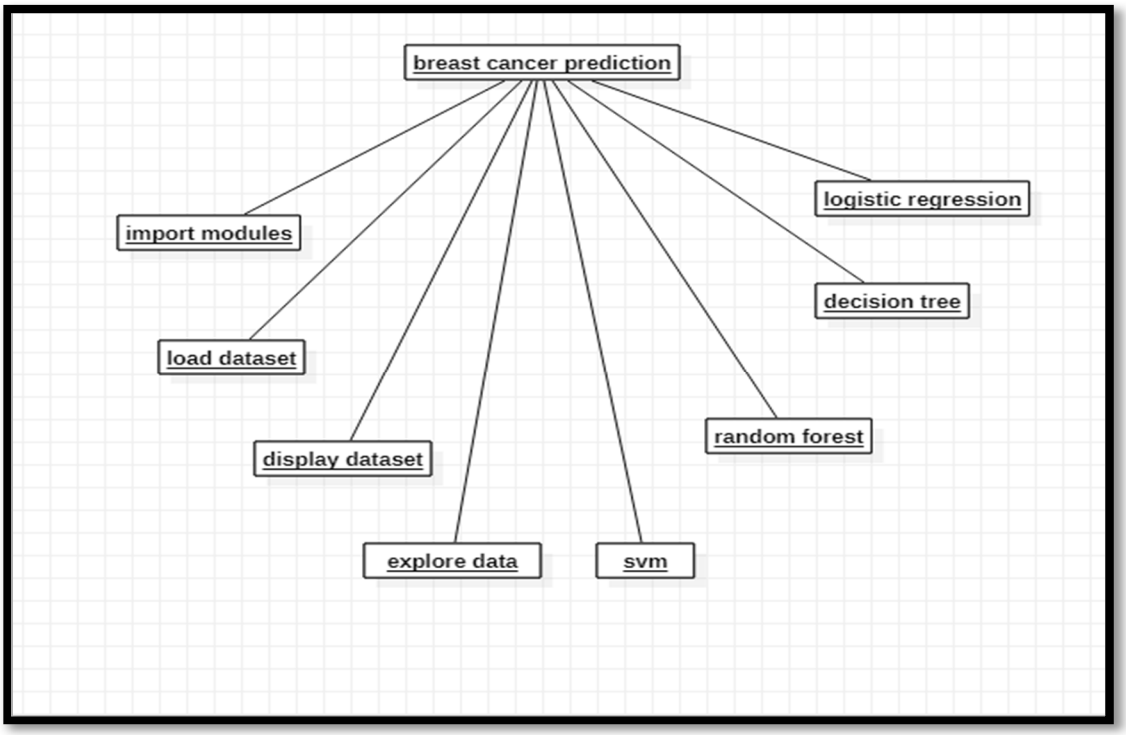
The structural diagrams represent the static aspect of the system. These static aspects represent thoseparts of a diagram, which forms the main structure and are therefore stable.

These static parts are represented by classes, interfaces, objects, components, and nodes. The fourstructural diagrams are −
Class diagram Object diagram
Component diagram Deployment diagram

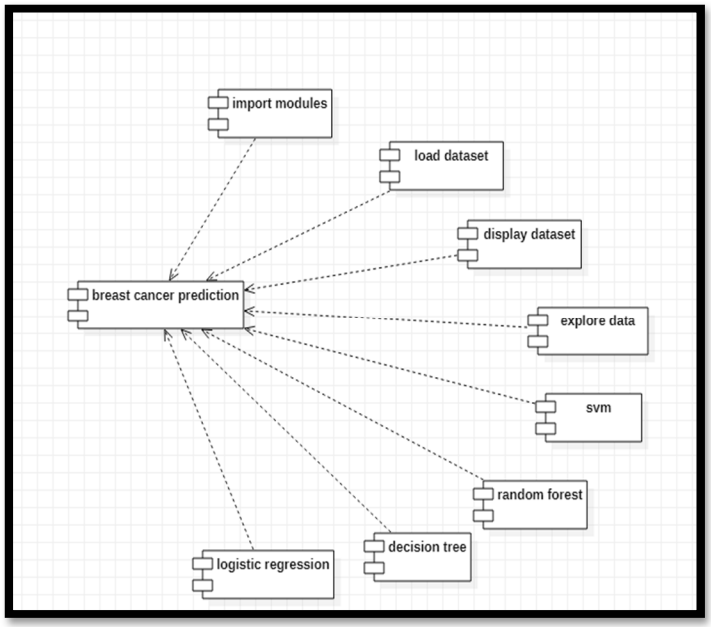
a) Class Diagram: Class diagrams are the most common diagrams used in UML. Class diagram consists of classes, interfaces, associations, and collaboration. Class diagrams basically represent the object-oriented view of a system, which is static in nature. Active class is used in a class diagram to represent the concurrency of the system.



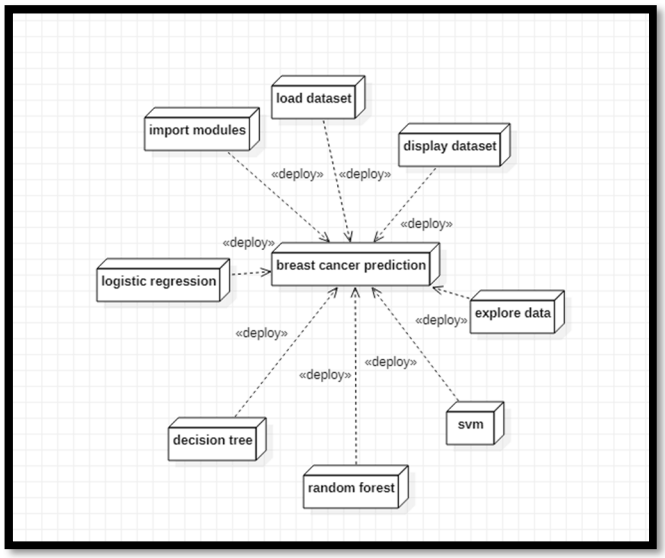
b) *Object Diagram:* Object diagrams can be described as an instance of class diagram. Thus, these diagrams are more close to real-life scenarios where we implement a system. Object diagrams are a set of objects and their relationship is just like class diagrams. They also represent the static view of the system.



c) *Component Diagram:* Component diagrams represent a set of components and their relationships. These components consist of classes, interfaces, or collaborations. Component diagrams represent the implementation view of a system. During the design phase, software artifacts (classes, interfaces, etc.) of a system are arranged in different groups depending upon their relationship.



d) Deployment Diagram: Deployment diagrams are a set of nodes and their relationships. These nodes are physical entities where the components are deployed. Deployment diagrams are used for visualizing the deployment view of a system. This is generally used by the deployment team.



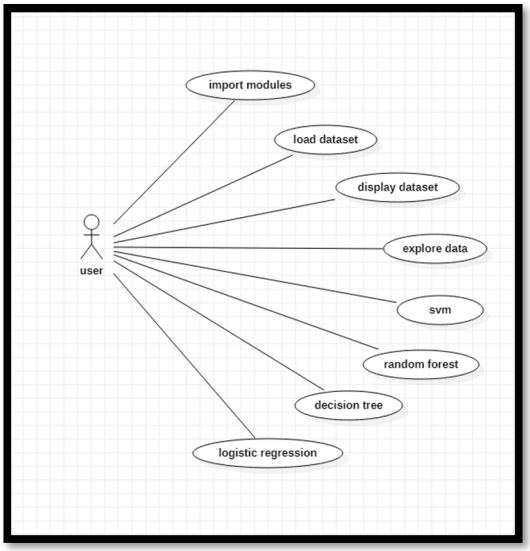
2) Behavioural Diagrams

Any system can have two aspects, static and dynamic. So, a model is considered as complete when both the aspects are fully covered.

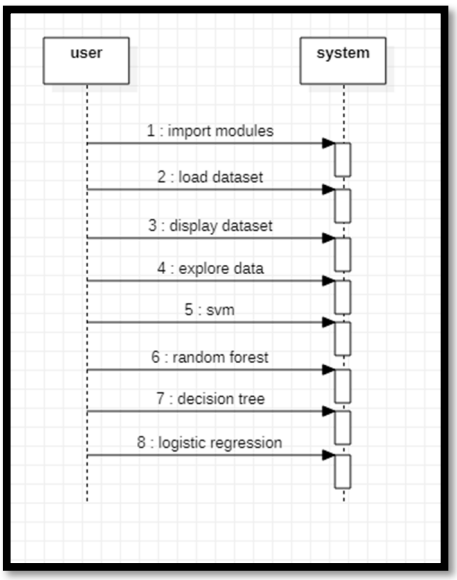
Behavioral diagrams basically capture the dynamic aspect of a system. Dynamic aspect can be further described as the changing/moving parts of a system.

UML has the following five types of behavioral diagrams −

a) Use Case Diagram: Use case diagrams are a set of use cases, actors, and their relationships. They represent the use case view of a system. A use case represents a particular functionality of a system. Hence, use case diagram is used to describe the relationships among the functionalities and their internal/external controllers. These controllers are known as actors.

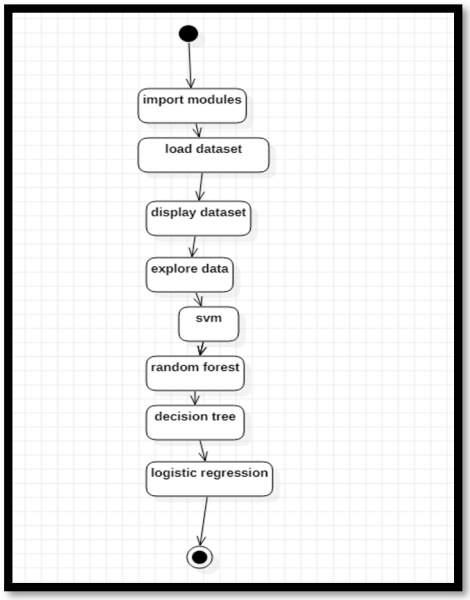


b) Sequence Diagram: A sequence diagram is an interaction diagram. From the name, it is clear that the diagram deals with some sequences, which are the sequence of messages flowing from one object to another.

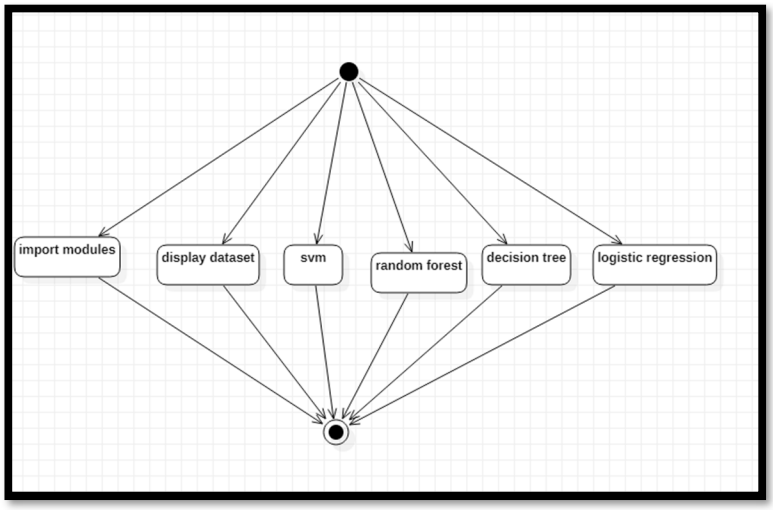


c) Collaboration Diagram: Collaboration diagram is another form of interaction diagram. It represents the structural organization of a system and the messages sent/received. Structural organization consists of objects and links. The purpose of collaboration diagram is similar to sequence diagram. However, the specific purpose of collaboration diagram is to visualize the organization of objects and their interaction.

d) State Chart Diagram: Any real-time system is expected to be reacted by some kind of internal/external events. These events are responsible for state change of the system. Statechart diagram is used to represent the event driven state change of a system. It basically describes the state change of a class, interface, etc. State chart diagram is used to visualize the reaction of a system by internal/external factors.



e) Activity Diagram: Activity diagram describes the flow of control in a system. It consists of activities and links. The flow can be sequential, concurrent, or branched. Activities are nothing but the functions of a system. Numbers of activity diagrams are prepared to capture the entire flow in a system. Activity diagrams are used to visualize the flow of controls in a system. This is prepared to have an idea of how the system will work when executed.



IV. RESULTS AND ANALYSIS

Algorithm	Accuracy
Logistic Regression Method	0.982456
Decision tree Classifier Method	0.900585
Random Forest Classifier Method	0.959064
Support Vector Classifier Method	0.976608

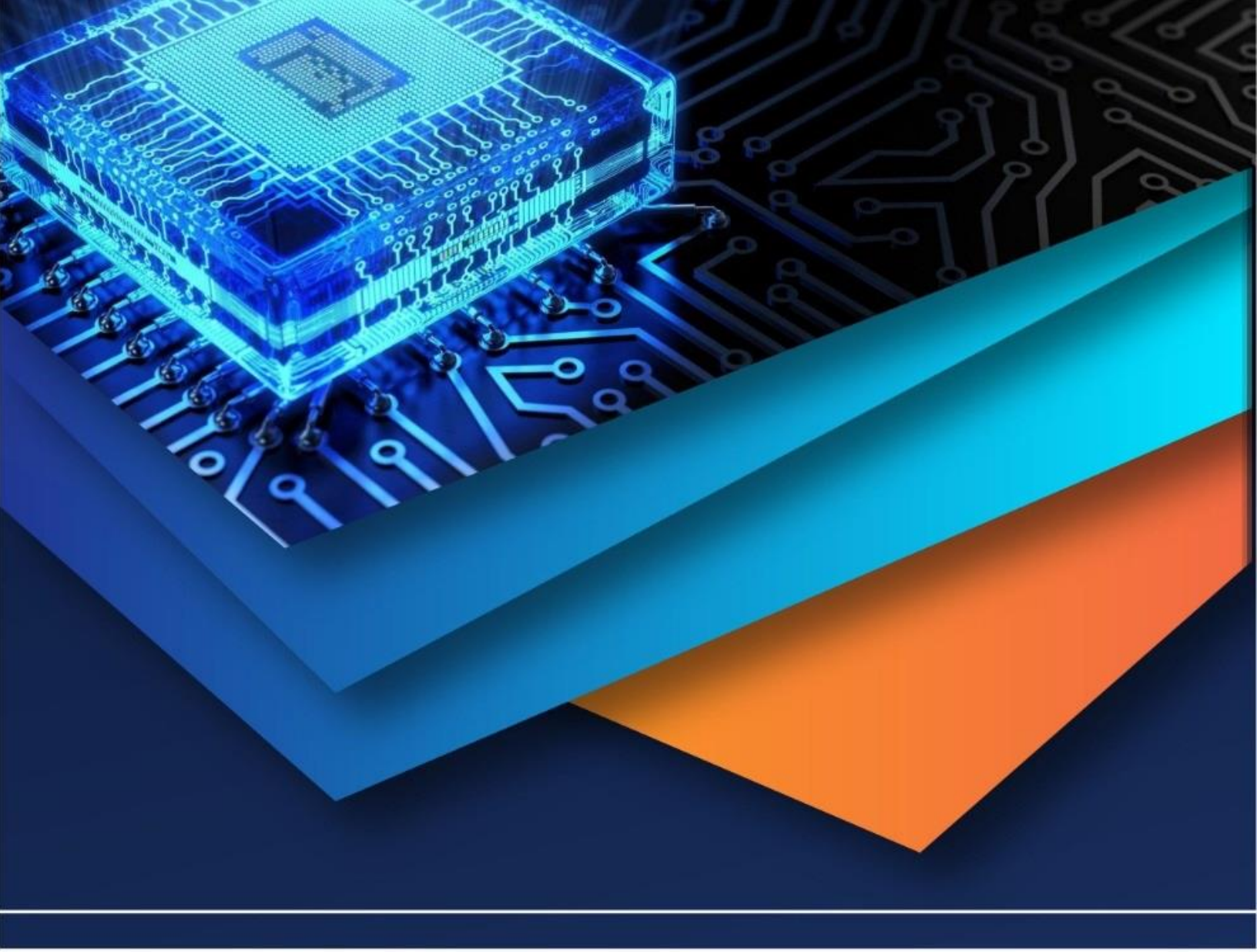
V. CONCLUSION AND FUTURE SCOPE

Breast cancer has been predicted and considered for some classifiers such as Naïve Bayes, Random Forest, Logistic Regression, Multilayer Perceptron, K-nearest neighbors classifier. WEKA data mining tool has been used and compared the presentation of these classifier algorithms. It is observed the performance results of K-nearest neighbors classifier algorithm. It provides the highest correctly classified instances of 97.9021%. The second most accurate classifier is Multilayer Perceptron with correctly classified instances of 96.5035 %. This paper mainly visualized 286 instances with 10 attributes to predict and analyse breast cancer dataset has been showed. It discusses the performance of different classification algorithm on the basis of distribution plot. This paper also observed Kappa statistic, Mean absolute error, Fmeasure, MCC, ROC Area, Relative absolute error, FP rate, TP rate, Root mean squared error, Precision Recall and. It has also been comprised that K-nearest neighbors classifier has highest percentage (99.9%) of ROC Area and Multilayer Perceptron has second highest percentage (98.0%) of ROC Area.

In future, the other classificationi algorithms arei utilized for thei analysis of thei same mammogram images to predict their performances. . In future will work to increase this accuracy up to 99% or more.

REFERENCES

[1] H. Elouedi, W. Meliani, and Z. Elouedi, "A hybrid approach based on decision trees and clustering for breast cancer classification," in 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2014.
[2] F.Kharbat, H.Ghalayini, "New algorithm for Building Ontology from Existing Rules: A Case Study," in International Conference on Information Management and Engineering, 2009, pp. 12-16.
[3] P. Hamsagayathri and P. Sampath, "Performance analysis of breast cancer classification using decision tree classifier," International Journal Of Current Pharmaceutical Research (IRCPR), vol. 9, 2017.
[4] R. Sumbaly, N. Vishnusri, and S. Jeyalatha, "Diagnosis of breast cancer using decision tree data mining technique," International Journal of Computer Applications, vol. 98, no. 10, p. 975 8887, 2014.
[5] C. P. Utomo, A. Kardina, and R. Yuliwulandari, "Breast cancer diagnosis using artificial neural networks with extreme learning techniques," International Journal of Advanced Research in Artificial Intelligence (IJARAI), vol. 3, no. 7, pp. 10–14, 2014.
[6] S.Jhajharia, H.K.Varshney, S.Verma, and R.Kumar, "A neural network based breast cancer prognosis model with PCA processed features," in International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016.
[7] Kathija and S.Nisha,"Breast cancer Data Classification Using SVN and Naive Bayes Techniques," International Journal of Innovative Research in Computer and Communication Engineering(IJIRCCE), vol. 4, pp. 21 167–21 175, 2016.
[8] D.Lavanya and D. Rani, "Evaluation of Decision Tree Classifiers on Tumor Datasets," International Journal of Emerging Trends Technology in Computer Science (IJETTCS), vol. 2, pp. 418–423, 2013.
[9] T. M. Mohamed, "Efficient breast cancer detection using sequential feature selection technique," in 7th International Conference on Intelligent Computing and Information Systems (ICICIS), 2015.
[10] J.-W.Liu, Y.H.Chen, and C.H.Cheng, "Owa based information fusion method with PCA preprocessing for data classification," in International Conference on Machine Learning and Cybernetics, 2012, pp. 3322– 3327.
[11] T. R. Patil and S. S. Sherekar, "Performance analysis of naive bayes and j48 classification algorithm for data classification," International Journal Of Computer Science And Applications, vol. 6, no. 2, 2013.
[12] I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques. MorganKaufmann Publishers Inc., San Francisco,CA, USA, 2nd edition, 2005.
[13] A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software:. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/downloading.html



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY