International Journal of Computer Sciences and Engineering
Vol.**11**, Issue **3**, pp.**34-38**, March **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Paper

Predictive Analysis for Decision Making Using Machine Learning in Health Care

Walli Prasadu[1*], Ravishankar Kumar[2], Ravi Kiran Katta[3], Pinki Sagar[4]

[1,2,3,4,5]Dept. of Computer science and Engineering, Faculty of Engineering and Technology, Manav Rachana International of Research and Studies, Faridabad, Haryana, India

Received: 10/Feb/2023; **Accepted**: 17/Mar/2023; **Published**: 31/Mar/2023. **DOI**: https://doi.org/10.26438/ijcse/v11i3.3438

Abstract: Breast Cancer has turned into the normal reason for death among ladies. There are several machine learning algorithms that can be used for breast cancer predictive analysis, such as logistic regression, decision trees, random forests, support vector machines, and neural networks. These algorithms can be trained on large datasets of patient information, including demographic data, medical history, and genetic markers, to identify patterns and make accurate predictions. One of the key benefits of machine learning in breast cancer predictive analysis is the ability to personalize treatment plans based on individual patient characteristics. By analyzing a patient's unique combination of risk factors, doctors can develop tailored treatment plans that are more effective and less invasive. Our point is to group whether the breast cancer is harmless or dangerous and foresee therepeat and non-repeat of threatening cases after a specific period. To accomplish this we have utilized AI strategies, for example, "Support Vector Machine", "Logistic Regression", "KNN and Naive Bayes". We additionally have investigated the precision of expectation of by applying different calculation on the new arrangement of information that has been joined. This paper explores the use of predictive analytics in healthcare decision-making through machine learning. The application of machine learning algorithms inhealthcare can assist in identifying patterns and trends in patient data, which can then be used to predictpotential health risks, recommend treatment options, and optimize healthcare delivery. The paper discusses various predictive analytics techniques such as decision trees, random forests, logistic regression, and neural networks, and their applications in healthcare. Additionally, the paper also highlights the challenges associated with the implementation of predictive analytics in healthcare, such as data quality, privacy concerns, and ethical issues. Finally, the paper concludes by emphasizing the need for healthcare organizations to leverage predictive analytics to improve patient outcomes, reduce costs, and enhance the quality of care.

Keyword: Breast Cancer, Support Vector Machine, K-Nearest Neighbor, Naive Bayes, Logistic Regression, Classification.

1. Introduction

Breast cancer is a type of cancer that develops in the cells of the breast tissue. It is one of the most common forms of cancer affecting women worldwide, and can also occur in men although it is much rarer. Breast cancer occurs when cells in the breast tissue grow uncontrollably and form a tumor that can invade nearby tissue and spread to other parts of the body. There are several types of breast cancer, with the most common being invasive ductal carcinoma which accounts for about 80% of cases. Other types include invasive lobular carcinoma, inflammatory breast cancer, and triple-negative breast cancer. The aim of breast cancer predictive analysis using machine learning is to identify patients who are at high risk of developing breast cancer in the future. By analyzing large datasets of patient information, including demographic data, medical history, and genetic markers, machine learning algorithms can identify patterns and make accurate predictions. One of the key benefits of machine learning in breast cancer predictive analysis is the ability to personalize treatment plans based on individual patient characteristics. This can help doctors and patients make more informed decisions about preventative measures and treatment options. Risk factors for breast cancer include age, family history, certain gene mutations, obesity, alcohol consumption, and exposure to radiation. Early detection of breast cancer is important for successful treatment and the best chances of survival. Screening methods such as mammograms, clinical breast exams, and self-exams can help detect breast cancer early. Treatment options for breast cancer include surgery, radiation therapy, chemotherapy, hormone therapy, and targeted therapy. The choice of treatment depends on the type and stage of the cancer, as well as the patient's overall health and personal preferences. With advancements in medical technology and treatment options, many people withbreast cancer can go on to live long, healthy lives.. The impacted cells and tissues then, at that point, progress through various stages, with going with modifications in the encompassing tissue probably assuming a part in whether the harm prompts a malignant growth. These occasions adding to ensuing tumors might happen precipitously as a result of mistakes in typical cycles, like DNA replication, or possibly through impacts of ecological openings. Almost certainly, numerous such

occasions procardiogenic might in all likelihood never be no doubt preventable on the grounds that, albeit possibly modifiable, they are ramifications of basic biologic cycles, like oxidative harm to DNA from endogenous digestion, or feeling of cell development through typical hormonal cycles.

2. Literature Review

Mandeep Rana et. al[1] proposed SVM, KNN, Strategic Relapse and navie bayes methods on analysis (WDBC) and prescient (WBPC) dataset acquired from UCI vault and made end in view of various upsides of precision. As per them KNN strategy has given the best outcome and SVM is solid procedures in prescient examination and thus they presumed that SVM involving Gaussian bit is the most appropriate strategies for repeat and non-repeat expectation of Brest cancer. B M Gayathri et. Al [2] proposed SVM and RVM methods in light of the information of breast cancer. In this paper they made sense of about crafted by breast cancer finding brain organization. They likewise found that Importance Vector Machine (RVM) is reasonable for breast cancer conclusion in present as well as in future. Siham A. Mohammed et al[3] proposed guileless Bayes, Successive insignificant improvement (SMO) and a choice tree based on J48 calculation on the WBC dataset got from UCI machine. learning respiratory. In this paper they made sense of that how for manage lopsidedness information that have missing qualities utilizing resampling methods to improve the grouping precision of identifying breast cancer. Wendie A. Berg1 [4] Five experienced mammographers, not explicitly prepared in BI-RADS, utilized the dictionary to depict and survey 103 screening mammograms, including 30 (29%) showing malignant growth, and a subset of 86 mammograms with symptomatic assessment, including 23 (27%) showing disease. A subset of 13 screening mammograms (two with dangerous discoveries, 11 with demonstrative assessment) were rereviewed by every onlooker 2 months after the fact. Kappa insights were determined as proportions of understanding past possibility. Nithya et al. [5] accept that the fundamental issue of breast cancer is tied in with ordering the brest cancer. PC Helped Determination (computer aided design) has been utilized for the discovery and portrayal of breast cancer. Their primary thought was further developing bosom disease expectation by utilizing information mining techniques. Stowing, multiboot, irregular subspace to the grouping execution of innocent Bayes, support vector machine-consecutive negligible streamlining (SVM-SMO), and multi-facet perceptron were applied. Delen et al. [6] concentrated on the expectation of bosom disease information with 202,932 patient records. The dataset was isolated into two unique gatherings as made due (93,273) and not made due (109,659), then, at that point, credulous Bayes, brain organization, and c4.5 choice tree calculations were applied. The accomplished outcomes showed that the c4.5 choice tree would do well to execution than different procedures.

3. Existing Methodology

Our methodology includes use of Machine learning techniques such as; Support Vector Machine, K-Nearest Neighbor, Logistic Regression and Naïve Bay.

3.1 Support Vector Machine

Support vector machine is an extremely impressive and modern AI calculation particularly with regards to prescient investigation. We have examined and executed SVM utilizing two parts: straight and Gaussian. With regards to ordering divisible dataset we favor direct portion (as figure 1) though for non-straight dataset arrangement we decide on bit determination, for example, Gaussian (as displayed in figure 2), polynomial.

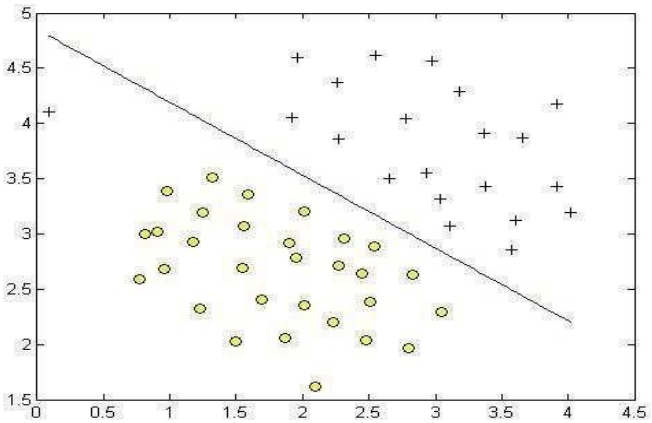

Fig.1. linear Support Vector Machine hyperplane construction

SVM focusses in determining the hyperplane such that it divides the region into two classes. Hyperplane equation

$$(y) = w * X' + b$$

where y is the predicted class label, X is the input feature vector, w is the weight vector, b is the bias, and X' denotes the transpose of X. The weight vector w and bias b are determined during the training process by optimizing the objective function that maximizes the margin between the hyperplane and the closest data points, which are known as support vectors. The hyperplane is chosen such that it maximizes the margin, which is the distance between the hyperplane and the support vectors. This approach makes SVMs a powerful tool for binary classification tasks, where the goal is to find the optimal hyperplane that separates the two classes with maximum margin.

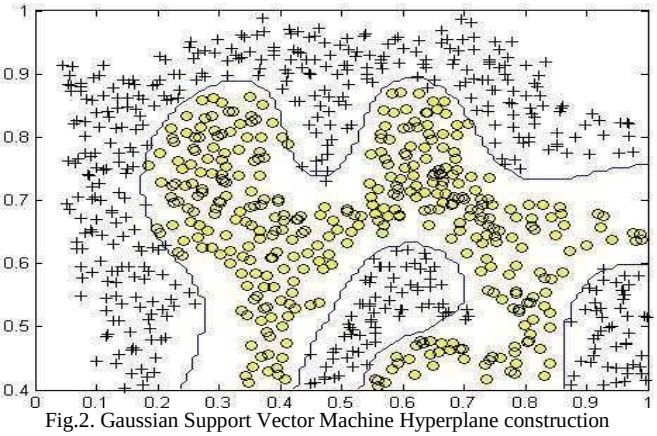

Fig.2. Gaussian Support Vector Machine Hyperplane construction

3.2 Logistic Regression

Logistic Regression is a statistical method used to model the relationship between a binary dependent variable (also known as the response variable) and one or more independent variables (also known as predictor variables or covariates). It is a type of generalized linear model (GLM) that is commonly used for classification problems.

Logistic Regression is a supervised learning algorithm used for classification tasks in machine learning. It is a linear model that uses a sigmoid function to predict the probability of an event occurring, given the input features. The output of logistic regression is a binary classification, i.e., either 0 or 1. In logistic regression, the goal is to find the optimal set of parameters that minimize the error between the predicted values and the true values. The optimization is typically done using gradient descent, a popular optimization algorithm. The input features are transformed using a linear function, which produces a weighted sum of the features. The weighted sum is then passed through a sigmoid function, which transforms the output to a value between 0 and 1. This value represents the predicted probability of the event occurring. The predicted probability is then converted to a binary classification based on a threshold. Typically, a threshold of 0.5 is used, where values less than 0.5 are classified as 0, and values greater than or equal to 0.5 are classified as 1.

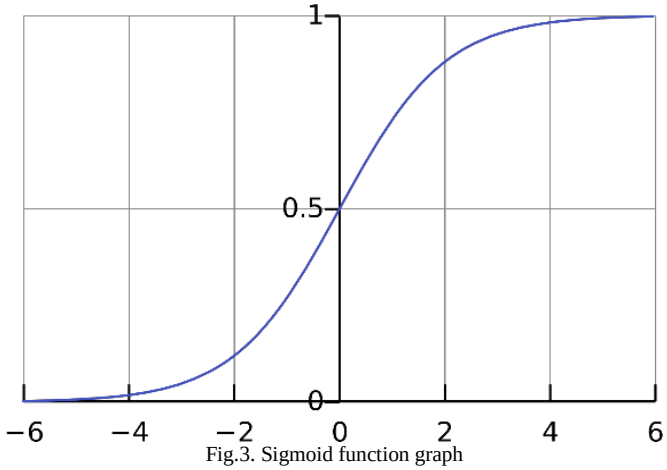

Fig.3. Sigmoid function graph

3.3 : k-Nearest Neighbor

In k-NN, the goal is to predict the class of a new data point by looking at the k nearest neighbors in the training dataset. The value of k is a hyperparameter that is chosen by the user. The algorithm works as follows: It is a non- parametric method that does not make any assumptions about the underlying data distribution. The distance between the new data point and all the points in the training dataset is calculated using a distance metric, such as Euclidean distance or Manhattan distance. The k nearest neighbors of the new data point are selected based on the calculated distances. For classification tasks, the class of the new data point is determined by majority voting among the k nearest neighbors. That is, the class that appears the most among the k neighbors is assigned to the new data point. For regression tasks, the value of the new data point is determined by taking the average of the values of the k nearest neighbors. The k-NN algorithm has some advantages and disadvantages. One of the main advantages is thatit is simple to implement and can work well with small datasets. It can also handle multi-class classification tasks. However, it can be computationally expensive for large datasets, and the choice of k can affect the performance of the algorithm. Additionally, k-NN is sensitive to the distance metric used, and the curse of dimensionality can affect the algorithm's performance in high-dimensional spaces.

3.4 Naive Bayes

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem, which states that the probability of a hypothesis (or class) given some observed evidence (or features) is proportional to the probability of the evidence given the hypothesis, multiplied by the prior probability of the hypothesis. Naive Bayes calculates the probability of each class label for a given input data point, based on the observed values of the features and the prior probabilities of the class labels. The class label with the highest probability is then assigned as the predicted label for that data point. In AI, credulous Bayes classifiers are a group of straightforward probabilistic classifiers in light of applying Bayes' hypothesis with solid (guileless) freedom suspicions between the highlights. Gullible Bayes is a straightforward strategy for building classifiers: models that relegate class marks to issue occurrences, addressed as vectors of component values, where the class names are drawn from some limited set. It's anything but a solitary calculation for preparing such classifiers, yet at the same a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. Breast cancer is a major health concern worldwide, and early detection and diagnosis are critical for successful treatment. Machine learning algorithms like Naive Bayes can analyze patient data and identify patterns that can help healthcare providers make more accurate diagnoses and treatment decisions.

In a Naive Bayes model for breast cancer prediction, the algorithm would be trained on a dataset of patient information, including factors like age, family history, breast density, and other risk factors. The algorithm would then use this data to develop a probability-based model that can predict whether a patient is likely to develop breast cancer or not.

3.5 Existing analysis

Breast cancer is a Exploratory Data Analysis (EDA): EDA is the process of analyzing and summarizing data inorder to understand its underlying patterns, distributions, and relationships. EDA is often used as a first step in MLto gain insights into the data and to identify any issues that need to be addressed before modeling. Feature Engineering: Feature engineering is the process of selecting, transforming, and creating features from the raw data that will be used as input to an ML model. Feature engineering is a critical step in ML, as the performance of the model often depends on the quality of the features. Model Selection and Evaluation: Model

selection involves selecting an appropriate ML algorithm and tuning its parameters to obtain the best performance on a given task. Model evaluation involves testing the performance of the model on a set of held-out data to estimate its generalization ability. One of the significant advantages of using machine learning algorithms is that they can handle large volumes of data and extract hidden patterns and relationships that may not be apparent to human experts. This ability to analyze complex data sets can improve the accuracy of breast cancer diagnosis and prediction. Several machine learning algorithms have been used in breast cancer predictive analysis, including decision trees, support vector machines (SVMs), artificial neural networks (ANNs), and random forests. These algorithms have shown promising results in accurately predicting breast cancer outcomes. In conclusion, breast cancer predictive analysis using machine learning is a promising area of research that can potentially improve the accuracy of breast cancer diagnosis and prediction. However, further research is needed to optimize the performance of these algorithms and to develop a robust and accurate predictive model that can be used in clinical practice.

Table 1 Result for Diagnosis of breast Cancer: -

MACHINE-LEARNING TECHNIQUES- USED	PARAMETERS	
	TRAINIG-ACCURACY	TEST-ACCURACY
SVM- LINEAR	63.68	80.56
SVM-RBF-static 'C' parameter	100	64.03
SVM-RBF-Dynamic 'C' parameter	89	94.05
Logistic Regression generalized	92.56	91
Logistic Regression regularized	93.54	92.08
KNN-Euclidian	100	95.68
K-NN-Manhattan	100	94.96
Naive bayes - normal	100	92.1
Naive bayes - Kernel	100	92.1

Table 2 Result for Prediction of breast:-

MACHINE-LEARNING TECHNIQUES-USED	PARAMETERS TRAINING-ACCURACY	TEST-ACCURACY
SVM- LINEAR	74	68
SVM-RBF-static 'C' parameter	100	68
SVM-RBF-Dynamic 'C' parameter	100	64
Logistic Regression generalized	82	70
Logistic Regression regularized	80	70
KNN Euclidian	100	70
k-NN-Manhattan	100	72
Naïve bayes - normal	81.33	65
Naïve bayes - Kernel	81.33	68
KNN Manhattan	100	94.96

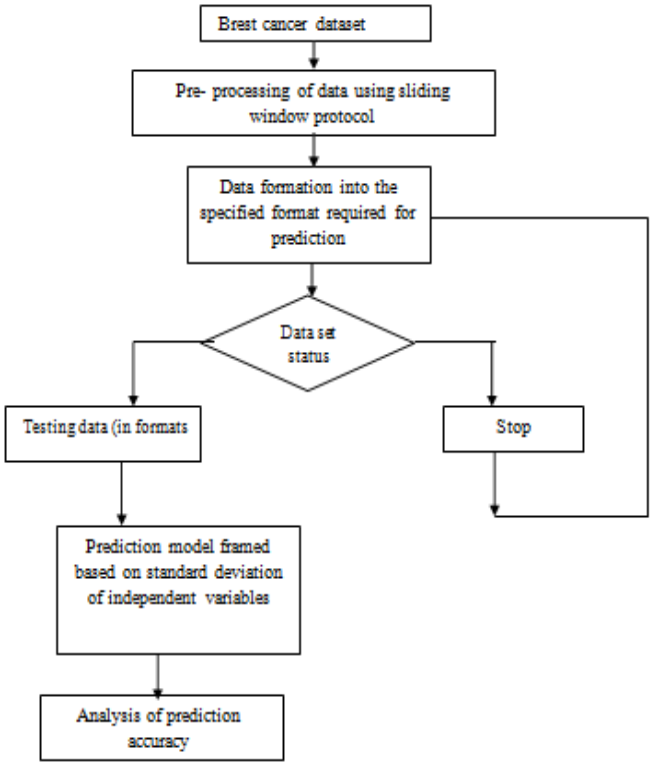

Fig.4. Proposed frame work of prediction model

4. Future Scope

Predictive analysis using machine learning in healthcare has tremendous potential to revolutionize the way healthcare decisions are made. Some potential future scope for predictive analytics in healthcare decision- making using machine learning include:

Personalized Medicine: With the help of predictive analytics, doctors can analyze a patient's medical history and current health condition to provide personalized treatment plans. Machine learning algorithms can be used to analyze large sets of patient data to identify patterns, and suggest tailored treatment options. Early Disease Detection: Predictive analytics can be used to identify early signs of diseases and disorders. Machine learning models can analyze patient data to detect patterns and provide alerts to healthcare providers for early intervention. Resource Optimization: Predictive analytics can be used to optimize resource allocation in healthcare organizations. By analyzing data from electronic health records and other sources, machine learning models can help healthcare providers make better decisions on resource allocation, reducing costs and improving patient outcomes.

References

[1] P. A. Francis, *et al*., "Adjuvant ovarian suppression in premenopausal breast cancer," *New England Journal of Medicine*, Vol.**372**, no.**5**, pp.**436-446, 2015.** [Online]. Doi: 10.1056/NEJMoa1412379

[2] C. E. De Santi's, *et al*., "Breast cancer statistics, 2015: Convergence of incidence rates between black and white women," *CA: a cancer*

journal for clinicians, Vol.**66**, no.1, pp.**31-42, 2016.** [Online]. doi: 10.3322/caac.21320

[3] C. E. De Santi's, *et al*., "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA: a cancer journal for clinicians*, Vol.**67**, no.**6**, pp.**439-448, 2017**. [Online]. doi: 10.3322/caac.21412

[4] N. K. Nikolova, "Microwave imaging for breast cancer," *IEEE microwave magazine*, Vol.**12**, no.**7**, pp.**78-94, 2011.** [Online]. doi: 10.1109/MMM.2011.942702

[5] Xie, Yao, *et al*., "Multistatic adaptive microwave imaging for early breast cancer detection," *IEEE Transactions on Biomedical Engineering*, Vol.**53**, no.8, pp.**1647-1657, 2006.** [Online]. doi: 10.1109/TBME.2006.87805

[6] Rana, pooja chandorkar, Alishiba Dsouza and Nikahat kazi, *Breast cancer diagonosis and recurrence prediction using machine learning techniques.* Vol.4 Issue: 04 | Apr-**2015**.

[7] B. M Gayathri, C. P. Sumathi, and T. Santhanam- Breast cancer diagonosis using machine learning algorithm. Vol.**4**, No.**3**, May **2013.**

[8] Rashmi Aggarwal – Predictive analysis of breast cancer using machine techniques. Vol.**15**, No.**3, 2019.**

