A Survey on Machine Learning Techniques to Detect Breast Cancer

Ms. Pooja Panapana
dept. of Information Technology
GMR Institute of Technology
Rajam, India
pooja.p@gmrit.edu.in

Supraja Bingi
dept.of Information Technology
GMR Institute of Technology
Rajam, India
19341a1219@gmrit.edu.in

Rohith Kumar Bhairi
dept.of Information Technology
GMR Institute of Technology
Rajam, India
19341a1216@gmrit.edu.in

Sai Kumar Dhavvala
dept. of Information Technology
GMR Institute of Technology
Rajam, India
19341a1233@gmrit.edu.in

Venkata Manikanta Kandula
dept.of Information Technology
GMR Institute of Technology
Rajam, India
19341a1253@gmrit.edu.in

2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 978-1-6654-5630-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAAIC56838.2023.10140277

***Abstract*— One of the diseases that most commonly affect women worldwide is breast cancer. In 2015, 8.8 million people lost their lives to cancer. Currently, in terms of incidence, lung cancer is second only to breast cancer among women worldwide. Cancer patients' lives can be saved through early detection and diagnosis.Therefore, it's critical to identify the tumor as soon as feasible. This survey covered several studies that were utilized to pinpoint this breast cancer as well as the methods they employed. This research study also discusses about the accuracyof different strategies and the datasets that were used and finally gives an overview of machine learning methods for recognizing and categorizing breast cancer. Research has been conducted on a variety of techniques, including Support Vector Machines (SVMs), Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Neural Network, DecisionSupport System, K-Fold Cross Validation, etc. Finally, all the techniques are compared based on their merits andshortcomings. This research study aims to gather and assess different breast cancer detection methods.**

Keywords— ***Breast Cancer, feasible, Support vector machine, K-Fold cross validation, Decision SupportSystem.***

I. INTRODUCTION

Even with major advancements in patient care and screening, breast cancer is remains the second-most common disease affecting women and the second-leading cause of cancer deaths in the United States. The World Health Organization reports that breast cancer is the most prevalent type of cancer among women worldwide (WHO). 627,000 women will pass away from breast cancer in 2018, predicts the WHO. The most recent cancer data from 2020 show that breast cancer has surpassed lung cancer to take the fifth spot as the leading cause of cancer mortality globally. The most common disease impacting women today is breast cancer. A good prognosis for breast cancer is often attained with early detection and treatment. Breast tissue may undergo abnormal cell division, resulting in growth. A variety of screening methods are available to find breast cancer. Early breast cancer identification is crucial to advantageous for prompt diagnosis and treatment because long-term survival depends on the prognosis. Because they reduce the risk of mortality, early cancer diagnosis, treatment, and detection all considerably increase the patient's chance of survival. A number of aspects, including effective performance on healthcare-related datasets involving pictures, x-rays, blood samples, etc., were presented with machine learning algorithms. This essay contrasts various techniques for locating breast cancer.

II. LITERATURE SURVEY

Radiologists must distinguish between normal and pathological cell growth using a computer-aided detection (CAD) system. There are two sections to this research. The first part presents a fast review of the major image modalities by gathering data on ultrasound, histography, and mammography from a range of research sources and evaluating a number of publications. Many examples are provided in the second part. In the future, by adding more data sources and picture modalities, the usage of bio-imaging-based machine learning algorithms for breast cancer detection can be improved even further. For better detection, mammography can be used in conjunction with MRI, ultrasound, and PET scans. Furthermore, the use of multi-parametric imaging (for instance, the use of biomarkers, genetic tests, and proteomics data) may result in a more accurate and individualized diagnosis and course of action for breast cancer. Moreover, interactive visualizations and real-time machine learning algorithms can be created to give the doctors immediate feedback and help them interpret the results better. Finally, study in this area can be expanded to include other illnesses and tumors, like lung, colorectal, and prostate cancer. [1]

This research intends to analyze the efficacy of machine learning Objective as a model-development tool for downstream analysis, knowledge discovery and to find unique pain features that might help detect whom sufferers could be vulnerable to NP after BC surgery in order to develop an efficient prognosis model that can be used without patient-doctor interaction shortly after surgery. In the future, More research into the causes of neuropathic pain following breast cancer surgery can be done using machine learning. In order to find patterns related to neuropathic pain, machine learning can be used to evaluate enormous volumes of data, including demographics, lifestyle, clinical and laboratory data, and medications. Moreover, machine learning can be utilized to discover novel risk variables that may not have been discovered earlier as well as to comprehend how different factors interact with one another. Moreover, customized prediction models that take into account unique patient features can be created using machine learning. This could aid in determining which patients are most vulnerable to neuropathic pain and assist in the creation of specialized treatment regimens. [2]

Because of their variable expression in healthy and malignant tissues, several small non-coding microRNAs (miRNAs) are attractive candidates for cancer biomarkers. The research used machine learning and miRNA expression data to establish the importance of these miRNAs. The collection of relevant miRNAs was prioritized using Information Gain (IG), Chi-Squared (CHI2), Least Absolute Shrinkage, and Selection Operation (LASSO). In the future, MiRNA indicators for breast cancer may be further validated using machine learning methods. For instance, a machine learning strategy might be applied to categorize miRNAs according to the patterns of their expression in various forms of breast cancer or to find miRNA biomarkers that can foretell tumor recurrence. Also, based on the expression of miRNAs, therapeutic targets for breast cancer could be identified using machine learning algorithms. Researchers could find more efficient treatments for breast cancer by identifying miRNAs linked to drug response. Finally, miRNAs that may be employed as prognostic markers for predicting the outcomes of breast cancer patients could be found using machine learning. [3]

This research aims to provide an ultrasound-based method for detecting and diagnosing breast tumors. Ultrasound images are viewed and sectioned using six metric capacity unit procedures to determine the condition or type of tumor. The fractal approach is used to extract features from photographs. In addition, classification methods such as Images are identified using decision trees, support vector machines, k-nearest neighbor algorithms, and Nave Bayes. Following that, breast cancer is directly classified using the convolutional neural network (CNN) architecture utilizing only ultrasound images. In the future, the development of more complex machine learning algorithms to boost the precision of the diagnostic results may be part of the project's future scope. Deep learning techniques could also be applied to increase the system's accuracy even more. The system might also be expanded to incorporate more medical imaging modalities, including MRI or CT scans, to detect cancers other than breast, colon, or lung. The technique might also be utilized to create a web- or mobile-based platform that would help doctors find and cure breast cancer in isolated areas. The method might also be used to create individualized treatment plans for each patient based on factors like age, ethnicity, and family history. [4]

This study developed an automated, machine learning-based approach for identifying breast cancer quickly and reducing mortality. In addition to obtaining the function technique of analysis using linear discrimination, two machine learning (ML) methodologies are suggested for the Wisconsin Breast Cancer Dataset. Wisconsin breast cancer: a 562 instance and 39 attribute dataset patients gathered from Kaggle was used for RF and SVM analysis. In the future, the creation of more effective and precise models that may be utilized for early breast cancer diagnosis could be one of the future possibilities for A Linear Discriminant Analysis and Classification Model for Breast Cancer Diagnosis. Further study could also look into the possibility of using additional data sources, including imaging data, for better sensitivity and accuracy. Research might also be done to see if the model could benefit from adopting machine learning techniques like deep learning. The possibility of using the model to generate predictions and guide treatment choices, such as directing tailored treatment programmes, might also be studied. [5]

This study aims to assess the need for predictive data extraction employing a decision support system, such as frequently collected information on the demographic, clinical, and biochemical characteristics of breast cancer patients (DSS) based on machine learning (ML) and random optimization (RO). A DSS model created in a practice set of 318 people using a testing set of 136 people realized accuracy of 86% and a C-index of 0.84 for progression-free survival. Recently, the promise of a multiple kernel learning (MKL) based semi-explainable DSS was demonstrated. It provides the opportunity to assess the taught model. In the future, more precise and individualized prognosis predictions for breast cancer patients could be made using machine learning techniques. This can entail combining many types of data, including genomic data, imaging data, and lifestyle data. To make more accurate predictions and better guide therapeutic decisions, machine learning algorithms could be utilized to create predictive models that take into account the unique traits of each patient. Also, high-risk patients might be identified using these predictive models, and the appropriate therapies could be given. Moreover, machine learning could be utilized to find new therapeutic targets and comprehend the basic principles that underlie breast cancer. [6]

In this work, machine learning techniques were utilized to build models that may be used to find and display useful indicators of breast cancer survival rates. One dependent variable, the patients' survival status, and 23 predictor variables were included in the dataset (alive or dead). Models for prediction have been created utilizing decision trees, random forests, neural networks, extreme boost, logistic regression, and support vector machines to identify the key prognostic factors affecting breast cancer survival rates. In the future, Machine learning algorithms can be used to estimate breast cancer patients' chances of survival more thoroughly and correctly. For instance, when analyzing vast volumes of medical imaging data, researchers can use deep learning models like convolutional neural networks (CNNs) to discover precise information about the position and size of a tumor, which can then be used to more precisely forecast a patient's chance of life. Moreover, more sophisticated machine learning algorithms can be utilized to examine lifestyle, linked data, and patient medical records to provide predictions that are more accurate. Moreover, biomarkers linked to breast cancer survival can be found using machine learning approaches, which could aid physicians in making better treatment choices. [7]

Dr. William H. Walberg's breast cancer tumor data from the University of Wisconsin Hospital was utilized to make predictions about the various types of breast cancers. For data visualization and machine learning on this dataset. We used R and Python to develop these machine learning algorithms and visualizations. They seek to evaluate data presentation and machine learning techniques for diagnosing and finding breast cancer. In the future, more complex models for breast cancer detection and diagnosis can be created using data visualization and machine learning techniques. Diagnostic precision can be increased by utilizing machine learning techniques like Deep Learning and Reinforcement Learning. The detection of breast cancer can also be aided by the use of cutting-edge techniques like Natural Language Processing and Computer Vision to extract meaningful data from medical images. In order to comprehend the data and show results in a way that is more understandable, new techniques for data visualization, such as 3D visualization, can also be used. Finally, new tools and applications that may be utilized to increase diagnosis accuracy and shorten the time it takes to diagnose breast cancer can be created using future research in data visualization and machine learning. [8]

They want to do machine learning research on forecasting patient outcomes using not greater than a picture of a as input, a tumor sample. By grouping samples into groups with low or high digital risk scores, the prediction is made (DRS). After being trained using example photos of 868 patients, the output classifier is analyzed and put up against image specialist categorization in a test set of 431 instances. The DRS categorization in this investigation led to a hazard ratio for breast cancer-specific survival of 2.10 (95% CI 1.33-3.32, p=0.001). Finally, they suggest. In the future, using scans of tumor tissue, researchers may be able to forecast breast cancer outcomes more precisely using more advanced machine learning methods, such as deep learning. Compared to conventional machine learning algorithms, deep learning algorithms are more capable of identifying patterns in data. Furthermore, in order to increase prediction accuracy even more, researchers may be able to apply more recent methods like reinforcement learning or generative adversarial networks. Furthermore, to further increase the precision of their predictions, researchers may be able to leverage bigger datasets, such as those produced by contemporary digital pathology. In order to increase the precision of their predictions, researchers may also be able to merge images of tumor tissue with other patient data sources, like a patient's medical history. [9]

They used four different prediction models and data exploration approaches to enhance the identification of breast cancer (DET). Wisconsin Diagnostic Breast Cancer (WDBC) and Coimbra Breast Cancer Dataset are two instances of pertinent data sets (BCCD). Parameters for radius, texture, area, perimeter, smoothness, and other features are available in datasets. Using standard performance metrics including confusion matrices and K-fold cross-validation techniques, each classifier's effectiveness and training duration were evaluated. In the future, utilising information from medical imaging, such as MRI and mammography, to build machine learning models for more precise breast cancer screening. creating algorithms to recognise and categorise cancers in medical imaging data. Machine learning algorithms can be made more accurate by including genetic and epigenetic data. constructing methods for integrating machine learning models into clinical judgement. systematically locating high-risk individuals and communities in order to offer early diagnosis and specialized care. utilizing deep learning and NLP techniques to automate the examination of patient data. To predict treatment outcomes, patient-specific models are built. constructing models to predict the course of an illness and whether it will recur. developing classification and recognition models for metastatic malignancies. employing machine learning techniques, forecast drug resistance and treatment effectiveness. [10]

An overview of existing noise reduction approaches and how they are used in machine learning algorithms. Analyze the efficiency of appearing sound avoidance approaches in minimizing mistakes in machine learning algorithms. The influence of noise reduction guidelines for accuracy of machine learning methods for identifying breast cancer tumors. Many noise reduction methods have their computational complexity graded. They address the possible advantages of using algorithms for machine learning that reduce noise to identify breast cancer tumors. In the future, Machine learning algorithms for breast cancer tumor diagnosis can benefit from applying noise-reduction techniques to increase precision and accuracy. This will enhance diagnosis precision and lessen false positive and false negative results. The data can be cleaned up using noise reduction techniques like Independent Component Analysis (ICA) and Principal Component Analysis (PCA), which can then be used to train machine learning models. The accuracy of the models can be further increased by combining these noise reduction approaches with already-existing feature selection and extraction algorithms. [11]

This article demonstrates how to employ machine learning to leverage next-generation sequencing data to predict breast cancer. They explain why it is critical to analyze breast cancer data and why an ideal machine learning approach is required to effectively anticipate the illness. In this section, they give an overview of the state of breast cancer research today and the challenges associated with diagnosing the disease. It is examined to what extent current machine learning techniques are employed to analyze data on breast cancer. They also highlight the limits of present ways to interpreting next-generation sequencing data. The suggested machine learning approach and its implementation are described in this section. They go through data pre-processing, feature selection, and assessment measures. In the future, using next generation sequencing data, machine learning techniques can be utilized to create models that are more precise in predicting the existence of breast cancer. These models could be used to find gene expression patterns linked to the disease as well as mutations in particular genes linked to breast cancer. Moreover, novel biomarkers or signatures linked to breast cancer may be found using machine learning approaches, which may aid in better understanding the condition and forecasting its course. Finally, distinct subtypes of breast cancer that are more likely to respond to particular treatments could be found using machine learning algorithms. [12]

Recent studies have improved the diagnostic capability of healthcare systems by utilizing a variety of methodologies from Data Science, AI, ML, and DL. For the early categorization of heart valve problems utilizing Doppler heart sound data, one study used LDA and ANFIS. Another study increased accuracy by 3.09% by recommending PSO-SVM for feature selection in the early classification of heart disease. SVM was suggested in one study for the early classification of cervical cancer, and SMOTE-RF and feature reduction approaches were utilized in another study to diagnose cervical cancer. A quicker regions-CNN neural network model was employed in a study to automate cell detection and cervical cytology diagnosis. In the future, to further enhance the effectiveness of feature selection approaches for breast cancer diagnosis, researchers can employ machine learning-based optimization strategies. For instance, more sophisticated optimization techniques like genetic algorithms, ant colony optimization, and particle swarm optimization might be employed to improve the feature selection procedure. Moreover, researchers might include other feature selection techniques, including support vector machines and evolutionary algorithms, in the optimization procedure. The effectiveness of feature selection procedures can also be enhanced by using more advanced data analysis techniques, such as deep learning and natural language processing. Finally, to help physicians better comprehend and evaluate the findings of feature selection, researchers should investigate the use of interactive visualization approaches like dynamic networks. [13]

The goal of this project is to develop a pathologic complete response (PCR) prediction model for patients with breast cancer receiving neoadjuvant chemotherapy (NAC). Following the collection of 287 people with stage II-III breast cancer, 14 candidate genes were selected. The genes were examined using a quantitative polymerase chain reaction using TaqMan probes. The Naive Bayes method was shown to have the highest predictive value when compared to other algorithms. The odds ratio between the 17-gene prediction model and PCR was 8.914 (95% confidence interval: 4.430–17.934), which was very significant. The patients are divided into two groups such as sensitive and insensitive based on the prediction model, with PCR rates of 42.3% and 7.6%, respectively (P 0.001). The model had a 62.0% specificity and an 84.5% sensitivity. In the future, the creation of more complex models to boost forecast accuracy is a potential future application for the Naive Bayes algorithm for neoadjuvant chemotherapy prediction. The employment of more sophisticated algorithms, such as support vector machines (SVMs) or deep learning techniques, may be part of this to better capture the intricate connections between the features and the result. The most crucial features that are connected to the outcome could also be found using data mining approaches like feature selection and feature extraction. Also, the growth of bigger datasets with more specific data will enable more accurate modelling of the response to neoadjuvant chemotherapy. [14]

This study proposes a unique technique to aid in the early detection of breast cancer using thermography. After gathering thermal images, pre-processing is supposedly done to enhance contrast, get rid of flaws, and extract information including statistical, geometrical, intensity, and textural aspects. Using cubic SVM and a combination of different features, the retrieved features are then used for automatic classification using machine learning techniques, with a top accuracy of 93.3% being reached. In the future, The automatic detection of breast cancer using thermal pictures can be made more accurate by study employing curvelet transform and machine learning. This may entail additional research into the application of various feature extraction methods, various machine learning algorithms, and other image processing methods, such as image segmentation and registration. The curvelet transform and machine learning algorithms' parameters can also be studied in order to increase accuracy. Moreover, research can be conducted to create a method for automatically diagnosing breast cancer from thermal pictures that is more effective and precise. The automated detection of breast cancer using various types of medical imaging, such as x-ray, MRI, and ultrasound pictures, can also be studied. [15]

III. VARIOUS METHODS ARE USED IN BREAST CANCER

A. *Support Vector Machine*

supervised machine learning models called support vector machines (SVMs) that look at information from classification and regression analysis. They also contain related learning methods. The fundamental SVM uses a set of input data to make predictions about which of two potential classes a new input will belong to. The method divides the examples of each class into as many distinct segments as feasible by mapping the data as points in space.

Steps in SVM:

1. The preparation of the data is the initial stage in employing an SVM. The data must first be gathered before being cleaned, normalized, and scaled as necessary. Additionally, data should be divided into training and testing sets.

2. The type of SVM to employ must be chosen as the next step. The kind of data, the quantity of characteristics, and the level of accuracy required all play a role in this.
3. The next step is to choose the parameters after choosing the model. The kernel type, regularization parameter, and cost parameter can all be chosen in this manner.
4. After choosing the parameters, the SVM may be trained using the training set of data. In order to do this, an optimization technique must be used to identify the ideal hyperplane that divides the two classes.
5. The test data can be used to assess the model once it has been trained.
6. To enhance the model, the parameters can be adjusted, and/or more data can be included. This might improve the model's precision.

SVMs are especially effective in classifying complicated but small- to medium-sized datasets. Although they are strong and precise, they also require a lot of memory, which might be a problem for larger datasets. They are also highly susceptible to overfitting, making correct model tuning crucial.

B. Logistic Regression

For classification tasks, logistic regression is the one of the machine learning Technique. It is a kind of linear regression that models a binary dependent variable using a logistic function. The model uses one or more independent factors to estimate the likelihood of a binary result.

Steps in logistic regression:
1. Gathering data and investigating it
2. Data preprocessing (e.g., normalizing, imputing missing values)
3. Separating the data into test and training sets
4. Using the training data to fit the model.
5. Assessing the model based on test data
6. Modifying the model's parameters (if necessary)
7. Making forecasts based on fresh data
8. Assessing and analyzing the model's findings.

The likelihood of an event occurring, such as whether or not a consumer would purchase a product, can be predicted using logistic regression. It can also be used to forecast a customer's propensity to fall into a specific sector or class. In the end, it is employed to derive conclusions from data and generate future projection.

C. K-Nearest Neighbors

K-Nearest Neighbors is a classification and regression method (KNN). This supervised machine learning approach starts with a data point, then examines its "K" nearest neighbors (the data points closest to it), and finally selects a value or class label based on the majority of the "K" neighbors. Because it makes no assumptions about the underlying data, KNN is an algorithm for slow, non-parametric learning.

Steps in KNN:
1. Select K's value (number of neighbors to consider).
2. Measure the separation between each training data point and the fresh data point.
3. Choose the K nearest data points.
4. Based on the vast majority of the K-nearest neighbors, assign a class name or value to the new data point.

In order to forecast the class label or value of the new data point, it merely calculates the distances between the given data points and their neighbors.

D. Random Forest

To handle classification and regression issues, supervised learning techniques like random forest are applied. This method uses an ensemble, which implies that different decision trees are combined to form one model. Compared to a single decision tree, this model is more accurate and has greater predictive potential. The random forest approach produces a number of decision trees using a random selection of the training data. The decision boundary for each of these trees will vastly different, and the model bases its forecast on the mean of all the projections from all the trees.

Steps in Random Forest:
1. Draw samples at random from the dataset.
2. For each sample, construct a decision tree.
3. Assume outcomes for each tree.
4. Calculate the influence of each prediction.
5. To determine the final prediction, combine the projections from all the trees.

This makes it more accurate than a single decision tree and robust to overfitting. Additionally, it lowers variance and produces outcomes that are more trustworthy. Random forest can handle missing values as well as category and numerical data.

E. Decesion Tree

Decision trees extract from the data a set of simple "if-then" rules for forecasting the value of the target variable. The tree's nodes stand in for "tests" on various each branch indicating the outcomes of those testing. The classification of the instance is in the leaves of the tree.

Steps in Decision Tree:

1. To divide the population into two halves, start with selecting the best attribute from either the entire population or sample.

2. Divide the population in the root node into subsets based on the specified quality.

3. Repeat step 3 for each subgroup, choosing the most appropriate characteristic to divide them in half.

4. For the subsets that cannot be separated further, create leaf nodes.

5. Pruning will lessen overfitting.

6. Use the decision tree to make predictions based on unknown facts.

One of the best and most used machine learning methods is the decision tree algorithm. It is appropriate for large datasets and may be applied to both classification and regression problems.

F. Least Square

A technique for determining which function fits a collected data points Least squares is the best. In order to find an equation that well characterizes the data points, regression analysis in this sense is performed. The height of the line sum of squares above the data points is decreased using the least squares approach to find the line that most closely fits a particular data set. It is used to identify links between variables and generate predictions about the data.

Steps in Least Square:

1. First gather the data points.

2. Averaging the y-values and x-values is required.

3. By resolving a series of equations, determine the slope and intercept of the best-fit line.

4. Draw the line of best fit on the scatterplot of the data points.

5. The sum of these squared distances is used to determine the vertical separation between the data points and the line of best fit.

6. To assess the quality of the fit, calculate the coefficient of determination (R2).

G. Neural Network

Artificial intelligence systems called neural networks assemble the necessary data points to work. Giving each of a set of inputs a weight and then using those weights to calculate each input's influence on the outcome is how this is done. Following that, the weights are modified based on the output, enabling the neural network to gradually learn and alter its actions. This technique is repeated until the desired outcome is attained.

Steps in Neural Network:

1. Identifying the issue that needs to be solved is the initial stage in any neural network research. This include gathering information pertinent to the issue, comprehending the desired result, and selecting the model architecture.
2. After that, the data must be preprocessed to ensure that it is appropriate for the neural network. This entails transforming the data into removal of any outliers, training, validation, and test sets.
3. The model architecture must be created once the data is ready. Choosing quantity of layers, the kind to activate functions, and the learning rate are all included in this.
4. The model is then trained using the training set. This includes changing the model's weights in response to the data, which can be accomplished using a gradient descent technique or back propagation.
5. To ensure that the model is operating as anticipated, it must be validated after training. This can be done by testing it on the comparing the findings to the validation set.
6. Once the model exhibits the anticipated behavior, it can.

H. K-Fold Cross Validation

A prominent method for assessing a machine learning algorithm's performance is K-fold cross validation. Randomly, k equal-sized subsamples are selected from a dataset. In order to build a model, training data from the other k-1 subsamples are used for each subsample, which is then tested on the held-out subsample. Each subsample is kept out and utilized as a validation dataset once during the course of this process, which is repeated k times. The results are then averaged to determine the final model performance metric. By lowering the variance in the estimation, this approach offers a more precise assessment of model performance than a single split of the data.

Steps in K-fold cross validation:

1. There should be k equal subsamples taken from the dataset. 2. Reiterate this process for each subsample:

a. Training data are drawn from the remaining k-1 subsamples, while the held out subsample serves as a validation dataset.

b. With the training data, develop a model.

c. Analyze the model using the validation dataset.

3. Determine the performance metric's average over all k iterations.

4. To estimate model performance, use the average performance metric.

I. Decision Support System

A decision support system (DSS) is a device that runs on computers and helps decision makers resolve challenging problems with a lack of clear solutions. Typically, it is employed to aid and enhance the decision-making process. Systems for supporting decisions can offer a variety of assistance, from simple to complex, from providing data and analysis to making suggestions and making predictions.

Steps in DSS:

1. Defining the issue that is being addressed is the first step in developing a DSS. Obtaining information about the issue, identifying parties and their interests, and comprehending the decision context are all part of this process.

2. Once the problem has been identified, the next stage is to gather and examine the pertinent data. This include locating data sources, compiling data in a consumable format, and conducting data analysis to spot patterns and connections.

3. The third phase is to develop a model or simulation to aid in the exploration of potential solutions after the data has been gathered and processed. In this step, the problem and its related factors are modelled mathematically or computationally, and simulations are done to evaluate the possible efficacy of various solutions.

4. Providing decision support is the DSS procedure's fourth stage. In order for the decision-makers to make an informed choice, this stage entails giving them the data, analysis, models, and simulations they require. The decision assistance tool can also offer recommendations and pointers regarding the optimal course of action.

J. Multiple Kernel Learning

A machine learning technique called multiple kernel learning (MKL) mixes numerous kernels into a single model to increase prediction accuracy. MKL uses numerous kernels to capture various facets of the data, which can enhance the model's performance. The weights for each kernel in MKL are determined via the optimization procedure. The kernel weights are changed until the best prediction is made as a result of the optimal solution.

Steps in MKL:

1. The first stage is to pre-process the data, which include cleaning, normalizing, and converting the data.

2. Choosing the right kernels for the job and adjusting the settings for the optimum performance.

3. Using the chosen kernels and parameters, train the model.

4. evaluating the model's performance using metrics like recall, accuracy, and precision.

5. Getting the best performance out of the model by tweaking the weights of the kernels and parameters.

6. Making the model available for application use.

Natural language processing, computer vision, and bioinformatics are just a few of the domains where MKL has been used as a current research topic.

TABLE-I. COMPARISON OF DIFFERENT METHODS

Author Name	**Methodsand tools used**	**Merits**	**Demerits**
Safdar, S., Rizwan, M., Gadekallu et al.	1.Support Vector Machine (SVM) 2.Logistic Regression 3. K-Nearest Neighbor (KNN)	The system can diagnose breast cancer more rapidly and accurately, saving time and money.	It is a time- and resource-consuming, computationally complex process. It is susceptible to mistakes like false positives and is only as reliable as the quality of the photographs used.
Juwara, L., Arora, N., Gornitsky et al.	6 ML algorithms least square, ridge, elastic network, random forest, gradient boosting, neural network	Machine Learning can help physicians better target intervention and improve patient outcomes by identifying which patients are most likely to have post breast cancer surgery, neuropathic discomfort.	Machine Learning approaches can be used to find potential indicators of post-breast cancer surgery neuropathic pain, however this method has the drawback of not fully capture the complexity of the underlying mechanism
Rehman, O., Zhuang et al.	1.Random Forest 2.Support Vector Machine (SVM)	By utilizing machine learning algorithms to assess miRNAs as breast cancer biomarkers, it is feasible to discover new, reliable, additionally to prognostic indicators for the early detection of breast cancer.	It might not be able to depict the intricate nature of the breast cancer's underlying molecular processes. The selected machine learning algorithm and parameter parameters may skew the findings.
Pourasad, Y., Zarouri et al.	1.K-Nearest Neighbors (KNN) 2.Support Vector Machine (SVM) 3.Decision Tree 4.Naive Bayes	The proposed architecture offers a rapid and accurate method for identifying breast cancer and determining its specific location from ultrasound images and machine learning. It enables a quicker and more precise detection method than traditional techniques, which reduces the time and cost associated with making a breast cancer diagnosis.	It took a lot of time and money to build this intricate construction. The location of breast cancer may not be correctly identified due to the inherent complexity of ultrasound imaging.
Adebiyi, M. O., Arowolo et al.	1.Random Forest 2.Support Vector Machine (SVM)	By selecting important predictors from the data and computing their discriminant weights, linear discriminant analysis offers an effective and efficient method for categorizing breast cancer diagnoses.	Its accuracy is constrained because it does not account for the existence of other cancer types. Due to its propensity for overfitting, careful parameter adjustment is necessary.
Ferroni, P., Zanzotto et al.	1.Support Vector Machine (SVM) 2.Random Optimization (RO) 3. Multiple Kernel Learning 4. Decision Support System	The prognosis of breast cancer can be predicted accurately, automatically, and effectively using machine learning approaches. It can aid in finding minute patterns in data and offer a more accurate prognosis than conventional techniques.	Due of the complexity and heterogeneity of breast cancer, it is challenging to develop a trustworthy machine learning model. Some businesses may find the expense of procuring and processing data for machine learning to be exorbitant.

Ganggayah, M. D., Taib et al.	1. Decision Tree 2. Random Forest 3.Neural Network 4.Logistic Regression 5.Support Vector Machine	Machine learning can help medical professionals better understand the many complicated elements that affect breast cancer survival, leading to more individualized and efficient treatments.	The reliability of the data is not always present, making it challenging to assess the accuracy of predictions. The method can be time consuming and expensive to implement because it calls for specific skills.
Ak, M. F.	1.Logistic Regression 2.K-Nearest Neighbors (KNN) 3. Support Vector Machine (SVM) 4. Decision Tree 5. Random Forest 6. Naïve Bayes	While using machine learning techniques, create more precise and predictive models for recognizing and diagnosing cancer, data visualization techniques make it easier to spot patterns and trends in breast cancer data.	Applications for data visualization may be constrained by the calibre and volume of available data, which could result in inaccurate diagnosis.
Turkki, R., Byckhov, D et al.	1. support vector machine 2. Random Forest	Machine learning may be able to provide very precise predictions regarding the progression of breast cancer by examining images of tumor tissue, enabling more specialized and precise treatment regimens. This can lessen the possibility of a false positive and raise the standard of patient care.	Data that is too well-fit and incomplete. Getting correct data is challenging since tumour tissue pictures are complicated.
Rasool, A., Bunterngchit, C., et al.	1. Support Vector Machine 2. Logestic Regression 3.K-Fold Cross Validation	In order to accurately treat patients, medical in order to detect breast cancer, doctors mainly rely on prediction models based on machine learning since they boost diagnostic accuracy and speed up diagnosis timeframes.	Predictive models have the potential to produce erroneous data and false diagnoses. Large volumes of data are needed for machine learning algorithms, yet in some circumstances, the data may not be available.
Ahuja, A., Al-Zogbi, L., & Krieger, A et al.	1. Decision Tree 2. Logestic Regression 3.Support Vector Machine 4.Artificial Neural Network	By lowering the amount of noise in the data and allowing the algorithms to concentrate on the crucial features, noise-reduction approaches to breast cancer machine learning algorithms tumour diagnosis improve algorithm performance.	Machine learning algorithms may be less effective because noise-reduction strategies may filter out crucial information. It may also make the method more complex, lengthening computing times and raising expenses.
Kurian, B., & Jyothi, V. L.	1.Logistic Regression 2.Linear Discriminant Analysis 3. K-Nearest Neighbors (KNN) 4. Naïve Bayes 5. Support Vector Machine 6. Random Forest	From next-generation sequence data, such as whole-genome, transcriptome, and methylome sequencing, machine learning approaches can be utilized to assess and predict breast Cancer	Large volumes of data from next-generation sequencing must be analyzed, which is a difficult and time-consuming process. Creating the best machine learning strategy for this type of data is challenging because it has a variety of properties and is frequently noisy.

Sharma, A., & Mishra, P. K.	1.Naive Bayes 2.Logistic Regression 3. K-Nearest Neighbors 4. Support Vector Machine 5. Random Forest 6. Artificial Neural Network	The most pertinent aspects for an accurate and trustworthy cancer diagnosis can be found using improved feature selection techniques for utilizing machine learning in the diagnosis of breast cancer.	Large datasets are needed for the machine learning-based optimum feature selection algorithms, which can be challenging to acquire. Due to the lengthy feature selection process, it is also computationally expensive.
Yang, L., Fu, B., Li, Y et al.	1. Random Forest 2. Support Vector Machine 3. K-Nearest Neighbors	Due to their capacity to produce precise forecasts with minimal computing expense and data needs, naive Bayes algorithms are helpful in predicting how breast cancers will respond to neoadjuvant treatment. The algorithm is very strong in handling both continuous and discrete information, giving it a great option for forecasting breast cancer patients' responses to neoadjuvant treatment.	They are also restricted in their capacity to manage complicated data sets, which might not adequately reflect the true underlying connection between the breast tumors and the response to neoadjuvant chemotherapy.
Karthiga, R., & Narasimhan, K	1. Support Vector Machine	It reduces the time and expense of diagnosis by automating the diagnostic procedure. Due to the usage of machine learning and the Curvelet transform, diagnosis accuracy has increased.	The curvelet transform is a costly computing method. Large datasets are necessary for accurate findings from machine learning algorithms, which are subject to bias.

IV. CONCLUSION

This paper given that it might be tough disease prognosis in its early phases, data classification and analysis in the early stages is a popular study topic. Our research leads us to the conclusion that mammography and histopathological records are crucial for correctly identifying and diagnosing breast cancer. According to our research, penalized regression models and tree-based models appear to perform better than other prediction models, with formed machine learning models generally agreeing. We employ popular ML methods including K-Nearest Neighbors, Logistic Regression, Decision Trees, Artificial Neural Networks, Support Vector Machines, Random Forests, etc. in our study. We selected these techniques because they have the best histories of disease diagnosis in the healthcare sector.

REFERENCES

[1] Safdar, S., Rizwan, M., Gadekallu, T. R., Javed, A. R., Rahmani, M. K. I., Jawad, K., & Bhatia, S. (2022). Bio-imaging-based machine learning algorithm for breast cancer detection. Diagnostics,12(5), 1134.

[2] Juwara, L., Arora, N., Gornitsky, M., Saha-Chaudhuri, P., & Velly, A. M. (2020). Identifying predictive factors for neuropathic pain after breast cancer surgery using machine learning. International journal of medical informatics, 141, 104170.

[3] Rehman, O., Zhuang, H., Muhamed Ali, A., Ibrahim, A., & Li, Z. (2019). Validation of miRNAs as breast cancer biomarkers witha machine learning approach. Cancers, 11(3), 431.

[4] Pourasad, Y., Zarouri, E., Salemizadeh Parizi, M., & Salih Mohammed, A. (2021). Presentation of novel architecture for diagnosis and identifying breast cancer location based on ultrasoundimages using machine learning. Diagnostics, 11(10), 1870.

[5] Adebiyi, M. O., Arowolo, M. O., Mshelia, M. D., & Olugbara, O. O. (2022). A Linear Discriminant Analysis and Classification Model for Breast Cancer Diagnosis. Applied Sciences, 12(22),11455.

[6] Ferroni, P., Zanzotto, F. M., Riondino, S., Scarpato, N., Guadagni, F., & Roselli, M. (2019). Breast cancer prognosis using a machine learning approach. Cancers, 11(3), 328.

[7] Ganggayah, M. D., Taib, N. A., Har, Y. C., Lio, P., & Dhillon, S. K. (2019). Predicting factors for survival of breast cancer patientsusing machine learning techniques. BMC medical informatics and decision making, 19, 1-17.

[8] Ak, M. F. (2020, April). A comparative analysis of breast cancerdetection and diagnosis using data visualization and machine learning applications. In Healthcare (Vol. 8, No. 2, p. 111). MDPI.

[9] Turkki, R., Byckhov, D., Lundin, M., Isola, J., Nordling, S., Kovanen, P. E., ... & Linder, N. (2019). Breast cancer outcome prediction with tumour tissue images and machine learning. Breastcancer research and treatment, 177, 41-52.

[10] Rasool, A., Bunterngchit,). Improved machinelearning-based predictive models for breast cancer diagnosis. C., Tiejian, L., Islam, M. R., Qu, Q., & Jiang, Q. (2022International journal of environmental research and public health, 19(6), 3211.

[11] Ahuja, A., Al-Zogbi, L., & Krieger, A. (2021).Application of noise-reduction techniques to machine learning algorithms for breast cancertumor identification. Computers in Biology andMedicine, 135, 104576.

[12] Kurian, B., & Jyothi, V. L. (2021). Breast cancer prediction using an optimal machine learning technique for next generation sequences. Concurrent Engineering, 29(1), 49-57.

[13] Sharma, A., & Mishra, P. K. (2021). Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. International Journal of Information Technology, 1-12.

[14] Yang, L., Fu, B., Li, Y., Liu, Y., Huang, W., Feng, S., ... & Bu, H. (2020). Prediction model of the response to neoadjuvant chemotherapy in breastcancers by a Naive Bayes algorithm. Computermethods and programs in biomedicine, 192, 105458.

[15] Karthiga, R., & Narasimhan, K. (2021). Medical imaging technique using curvelet transform and machine learning for the automated diagnosis of breast cancer from thermal image.Pattern Analysis and Applications, 24(3), 981-991.