Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.1, May : 2025

A SYSTEMATIC STUDY OF FEATURE SELECTION AND ML ALGORITHMS FOR BREAST CANCER DETECTION

Siripurapu Sridhar, Faculty in ECE, Lendi Institute of Engineering & Technology
Sateesh Gudla, Faculty in CSE, Lendi Institute of Engineering & Technology
Ch Hema Divya, Student, Lendi Institute of Engineering & Technology
A Gnana Teja, Student, Lendi Institute of Engineering & Technology
Ch Yashwanth, Student, Lendi Institute of Engineering & Technology
A Manohar, Student, Lendi Institute of Engineering & Technology

Abstract:
Breast cancer remains one of the leading causes of mortality among women worldwide. Early and accurate diagnosis is crucial for improving patient outcomes. In this study, we employ three machine learning techniques—Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree (DT)—to classify breast cancer as either benign or malignant. The Wisconsin Diagnosis Breast Cancer (WDBC) dataset is used for model evaluation. The dataset is pre-processed, and feature selection is applied to enhance model performance. A 10-fold cross-validation approach is used to train and validate the models. Performance is measured based on accuracy, precision, recall, F1-score, and confusion matrix analysis. The results indicate that SVM achieves the highest classification accuracy, outperforming LR and DT. However, LR demonstrates competitive performance with high recall, making it valuable in medical applications where false negatives must be minimized. The Decision Tree model, while interpretable, exhibits lower accuracy compared to SVM and LR. This study highlights the effectiveness of machine learning models in breast cancer diagnosis and emphasizes the need for selecting the most suitable model based on performance metrics.

Keywords:
Breast Cancer Diagnosis, Machine Learning, Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), Classification, Feature Selection.

1. Introduction:

Breast cancer is one of the leading causes of mortality among women worldwide, making early and accurate diagnosis essential for improving patient outcomes. Traditional diagnostic methods, such as mammography and biopsy, rely heavily on expert interpretation, which can be time-consuming and prone to human error. With advancements in artificial intelligence and machine learning, automated classification models have emerged as powerful tools for improving diagnostic accuracy and assisting medical professionals in early detection.

In this study, we employ three widely used machine learning models—Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree (DT)—to classify breast cancer as either benign or malignant. These models have demonstrated significant potential in medical applications due to their ability to process complex datasets and identify meaningful patterns. SVM is known for its ability to handle high-dimensional data and deliver high classification accuracy. LR, a statistical model, provides strong interpretability and is effective in binary classification problems. The Decision Tree model, while offering transparency in decision-making, often struggles with overfitting and lower accuracy compared to other approaches.

To enhance model performance, we apply feature selection techniques that help remove irrelevant and redundant data, improving computational efficiency and classification accuracy. The Wisconsin Diagnostic Breast Cancer (WDBC) dataset is used for evaluation, and a 10-fold cross-validation approach ensures robust model training and validation. Performance is assessed using key metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis.

The results indicate that SVM achieves the highest classification accuracy, outperforming both LR and DT. However, LR demonstrates competitive performance with high recall, making it particularly valuable in medical applications where minimizing false negatives is critical. The Decision Tree model, while interpretable, exhibits lower accuracy compared to the other two models. This study highlights the importance of selecting the most suitable machine learning model based on performance metrics and contributes to the advancement of AI-driven breast cancer diagnosis systems.

2. Literature Review:

Reem Alyami et. al. [1] (2017) proposed a breast cancer diagnosis model that integrates Support Vector Machines (SVM) and Artificial Neural Networks (ANN) with feature selection. The model was evaluated using the well-known Wisconsin Diagnosis Breast Cancer (WDBC) dataset for experimentation. Empirical studies have shown that SVM outperformed ANN in classification tasks, achieving an accuracy of 97.14% compared to ANN's 96.71%.

Kemal Polat et. al. [2] (2007) conducted a study on breast cancer diagnosis using the least square support vector machine (LS-SVM) classifier. The performance of LS-SVM was assessed based on classification accuracy, sensitivity and specificity analysis, k-fold cross-validation, and a confusion matrix. The achieved classification accuracy of 98.53% is highly promising when compared to previously reported classification methods.

Edriss Ebrahim Ali et. al. [3] (2016) conducted a review in which experiments were performed using the Wisconsin Diagnosis Breast Cancer database to classify tumors as benign or malignant. Supervised learning algorithms, including Support Vector Machine (SVM) with linear kernels and Neural Network (NN), were compared for this task. The performance analysis revealed that the Support Vector Machine achieved lower accuracy and precision compared to the Neural Network approach.

Mr.P.SATHIYANARAYANAN et. al. [4] (2019) conducted a review stating that Breast cancer is classified into two types: Malignant and Benign. Malignant cancer is considered curable, while Benign cancer is often deemed non-curable. Our approach involves using the Decision Tree algorithm, a supervised learning technique, to detect breast cancer. The key advantage of this algorithm is its ability to accurately classify cancer as either Malignant or Benign, achieving an impressive 99% accuracy.

E. Venkatesan et. al. [5] (2015) analyzed that in medical diagnosis, the DM approach facilitates the rapid identification of diseases based on symptoms, with various DM techniques, including classification, being employed to uncover hidden information in medical data. Their research analyses breast cancer data using classification algorithms such as J48, Classification and Regression Trees (CART), Alternating Decision Tree (AD Tree), and Best First Tree (BF Tree).This study specifically aims to compare the performance accuracy of four decision tree algorithms in predicting breast cancer outcomes. In the future, the breast cancer data will be analyzed using the same algorithms, focusing on images. Additionally, the study will explore microcalcifications in breast cancer imagery.

Ronak Sumbaly et. al. [6] (2014) Their research highlights that breast cancer is the second leading cause of cancer-related deaths among women and has become the most prevalent cancer in women. Their review made a decision tree-based data mining technique for the early detection of breast cancer, focusing on distinguishing between benign and malignant breast tumors.

Ziba Khandezamin et. al. [7] (2020) defined that it is now possible to achieve greater accuracy in the direction and early diagnosis of cancer by combining the results of Fine Needle Aspiration (FNA) catology with machine learning techniques. The method involves two steps: initially, logistic regression is employed to eliminate less important features, followed by the use of the Group Method Data Handling (GMDH) neural network in the second step to diagnose benign and malignant sample. The simulation results achieves a precision of 99.4% for the WBCD dataset, 99.6% for the WDBC dataset, and 96.9% for the WPBC dataset.

Hamzah A, Alsayadi et. al. [8] (2022) explained that Early detection of breast cancer can be achieved with greater accuracy, leading to increased life expectancy at a lower cost. This can be accomplished by enhancing the efficiency and precision of breast cancer detection through the analysis of large

datasets using technologies like machine learning and fusion-based decision support systems. The machine learning (ML) algorithms used include decision tree regressor, MLP regressor, SVR, random forest regressor, and K-Neighbors regressor. The models are improved through average ensemble and ensemble methods using the K-Neighbors regressor. The results indicate a reduction in RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI.

Esther. M. Umoren et. al. [9] (2024) aimed to determine determine whether breast cancer is benign or malignant based on a patient's medical history, using random forest and logistic regression machine learning techniques. A 5-fold cross-validation was employed, in dataset 20% used as the test set and the remaining 80% used for training the models. Among the models, the logistic regression performed the best, achieving 98.7% training accuracy, 96.7% validation accuracy, and 97.2% test accuracy. Additionally, the logistic regression model achieved the highest test F1-score of 97.163% and the highest recall score of 96.92%. Upon evaluating the model performance results, it was found that the logistic regression model outperformed the random forest model.

3. Materials & Methods used:

The dataset is selected from the Kaggle.The dataset consists of 569 rows and 32 columns. The inputs are **(1)id (** A unique identifier for each sample). **(2) diagnosis** (The classification of the tumor M = Malignant, B = Benign).**(3) Mean Features (_mean):** These columns contain statistical mean values of different tumor characteristics containing radius_mean, texture_mean, perimeter_mean, area_mean smoothness_mean, compactness_mean, concavity_mean, concave_points_mean, symmetry_mean, fractal_dimension_mean. **(4) Standard Error (_se):** These columns provide standard error measurements for each feature: radius_se, texture_se, perimeter_se, area_se, smoothness_se, compactness_se, concavity_se, concave_points_se, symmetry_se, fractal_dimension_se. **(5) Worst-case Features (_worst):** These columns capture the worst or most extreme values recorded for each characteristic: radius_worst, texture_worst, perimeter_worst, area_worst, smoothness_worst, compactness_worst, concavity_worst, concave_points_worst, symmetry_worst, fractal_dimension_worst.

Here, (1) is not considered and the rest are taken into account. For these attributes they calculate mean, standard error, worst. These data's are fed into the machine. These data's are used for training the machine using Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT) algorithms. By doing so we can get the output as either Malign or Benign. Using the Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT) algorithm the breast cancer is identified.

ML models used:

1. **Support Vector Machine (SVM):**

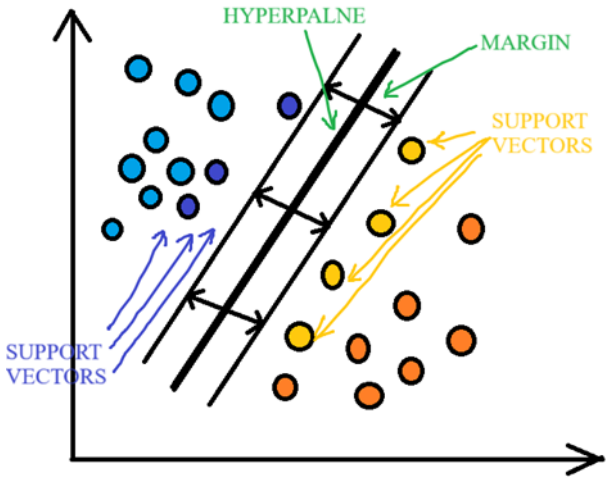


Figure 1: *structure of SVM*

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.1, May : 2025

SVM is a widely used machine learning method that falls under the broader category of kernel methods. Kernel methods utilize a kernel function, enabling them to generate nonlinear decision boundaries using linear techniques and classify data with no obvious distribution. These methods construct a hyperplane by dividing the attribute space, classifying data using support vectors—instances that define each hyperplane.

SVM is particularly useful in biometrics due to its ability to process high-dimensional and diverse data sources, such as gene expressions. It classifies datasets into different classes based on class membership using a hyperplane. If multiple hyperplanes exist, the optimal one is the one that is farthest from the data points. The goal of SVM is to correctly classify new data points into the right class, a process known as linear classification.

SVM and KNN are distinct machine learning algorithms, each based on unique principles. However, in some hybrid models, KNN can be integrated with SVM to enhance classification performance. he SVM model is a supervised learning algorithm primarily used for classification and regression tasks. It is based on the concept of finding an optimal hyperplane that best separates different classes in a high-dimensional space. It effectively handles non-linearly separable data by using kernel functions to map it into a higher-dimensional space. On the other hand, KNN is an instance-based, non-parametric learning algorithm that classifies data points based on the majority class among their k-nearest neighbors in the feature space. A hybrid approach that incorporates SVM into KNN leverages the strengths of both models. While SVM defines decision boundaries, KNN refines classification by considering local neighborhood information. This integration enhances KNN's decision-making process by incorporating a more structured separation of classes while mitigating the limitations of SVM in handling large datasets. It can be particularly effective for high-dimensional datasets, where traditional KNN struggles with the curse of dimensionality.

2. Logistic Regression:

Logistic regression is a supervised machine learning algorithm used for classification tasks, where the goal is to predict the probability that an instance belongs to a particular class. It is a statistical algorithm that analyzes the relationship between two data factors. Primarily used for binary classification, logistic regression employs the sigmoid function, which takes independent variables as input and produces a probability value between 0 and 1.

Logistic Regression is a classification algorithm used in machine learning for binary classification problems, where the output is either 0 or 1, Yes or No, or True or False. Despite its name, it is not a regression algorithm but rather a method for estimating the probability that a given input belongs to a particular class. It achieves this using the sigmoid (logistic) function, which maps any real-valued number into the range (0,1), making it effective for probability-based classification. The decision boundary is determined by a threshold, typically set at 0.5—if the predicted probability is greater than 0.5, the output is classified as 1; otherwise, it is classified as 0.

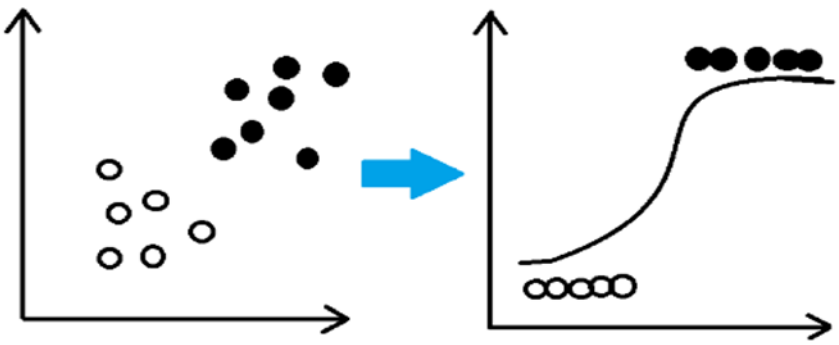


Figure 2: *Logistic Regression*

Types of Logistic Regression:

Binary Logistic Regression: Binary logistic regression is a type of regression analysis used for predicting a binary dependent variable. Its goal is to estimate the probability that an observation belongs to one of two categories based on one or more independent variables.

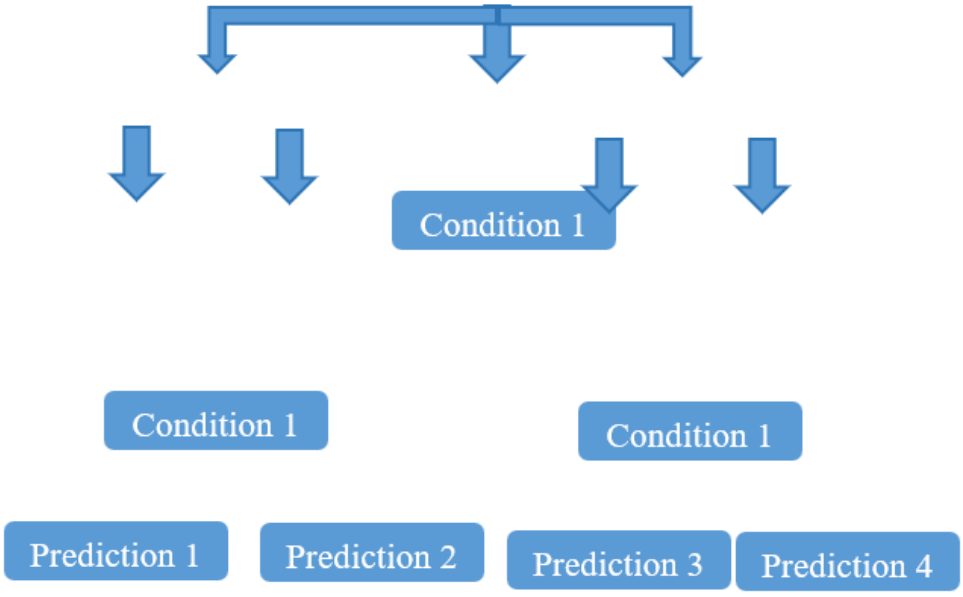
Multinomial Logistic Regression: Multinomial logistic regression is a statistical technique used to model the probabilities of three or more categorical outcomes that lack a natural order. Also known as nominal logistic regression, it employs a linear combination of independent variables to analyze correlations with outcome likelihoods and predict outcomes based on specific input conditions.

Multinomial logistic regression determines which factors significantly impact a categorical outcome. For example, when predicting transportation mode choice, the model can assess the effects of variables like distance, income, and environmental preference. Unlike linear regression, which predicts outcome values, multinomial logistic regression focuses on probabilities. It examines how changes in predictors influence the odds of different categories occurring.

Ordinal Logistic Regression: A statistical method for modelling and analysing ordinal categorical outcomes is ordinal logistic regression, commonly referred to as ordered logistic regression. Ordinal results are categorical variables having a built-in order, but the gaps between the categories are not all the same. An example of an ordinal scale is a Likert scale with responses ranging from "Strongly Disagree" to "Strongly Agree."

3. Decision Tree:

Figure 3: *Decision tree for proposed model*



A **Decision Tree** is a supervised machine learning algorithm used for classification and regression tasks. It is structured as a tree, where each node represents a decision based on a feature, branches represent the outcome of a decision, and leaves represent the final prediction or class label. The dataset is split based on feature values to create decision nodes. Each node represents a condition on a feature that splits the data further. The leaf nodes represent the final decision or class labels. The tree is built recursively by selecting the best feature split at each step. And the decision tree has:

J48 Algorithm: The J48 algorithm is an implementation of the C4.5 decision tree algorithm in the WEKA machine learning library and is used for classification problems. It is a supervised learning algorithm that builds a decision tree by recursively partitioning the data into smaller subsets based on attribute values. At each node, the algorithm selects the attribute that best splits the data using measures such as information gain or gain ratio. The tree construction begins with a root node and continues by adding child nodes until a stopping criterion is met, such as when all instances in a node belong to the same class. Once built, the decision tree classifies new instances by following a path from the root

node to a leaf node. Due to its simplicity, efficiency, and effectiveness, J48 is a widely used algorithm for classification tasks.

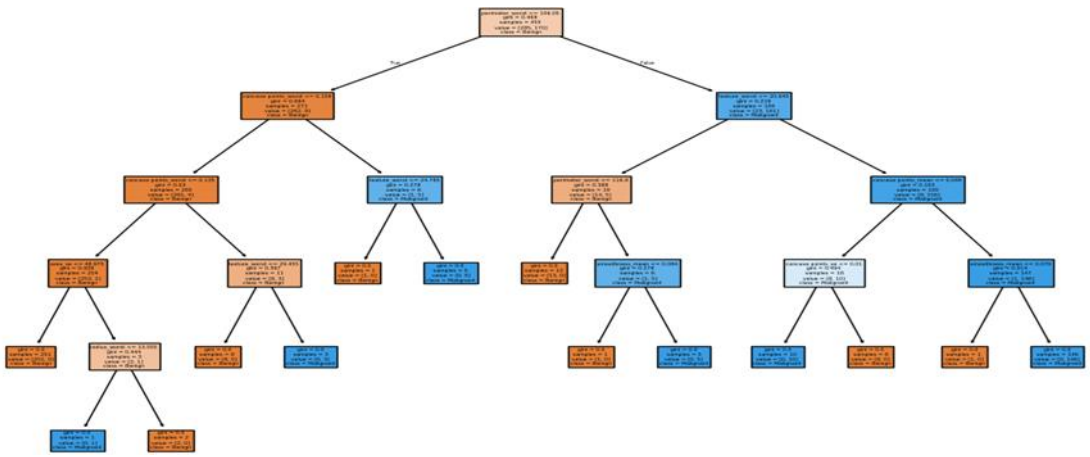
CART Algorithm: The CART (Classification and Regression Trees) algorithm is a widely used decision tree method for both classification and regression tasks. Introduced by Breiman, Friedman, Olshen, and Stone in 1984, it builds a binary decision tree by recursively partitioning the data based on attribute values. At each node, the algorithm selects the attribute that best splits the data using measures such as Gini impurity or squared error. The tree construction starts with a root node and continues by adding child nodes until a stopping criterion is met, such as when all instances in a node belong to the same class or the tree reaches a maximum depth. Once built, the decision tree can classify new instances or make predictions for regression problems. Known for its simplicity, interpretability, and effectiveness, CART remains a popular choice in machine learning.

BF Tree (B+ Tree): A B+ tree, also known as a BF tree, is a self-balancing search tree data structure widely used in databases and file systems. It is a variation of the B-tree that maintains sorted data and supports search, insert, and delete operations in logarithmic time. The structure of a B+ tree consists of a root node, internal nodes, and leaf nodes. Each node stores key-value pairs, where the key acts as a unique identifier, and the value represents the associated data. The tree grows by recursively splitting nodes when they become too full and merges nodes when they become too empty. One of its key features is that all leaf nodes remain at the same level, making it highly efficient for range queries and disk I/O operations. Due to its high performance, reliability, and scalability, the B+ tree is extensively used in database management systems.

AD Tree (Attribute-Dependence Tree): An Attribute-Dependence (AD) tree is a probabilistic graphical model that represents conditional independence relationships between attributes in a dataset. It is a type of Bayesian network that utilizes a tree structure to illustrate dependence relationships. In an AD tree, each node corresponds to an attribute, while edges signify conditional dependence relationships. The tree is built by recursively partitioning attributes into smaller subsets based on these relationships. AD trees are valuable for tasks such as feature selection, dimensionality reduction, and probabilistic inference. They help model complex attribute relationships and identify the most informative attributes in a dataset. With applications in data mining, machine learning, and artificial intelligence, AD trees play a crucial role in understanding and analyzing structured data.

4. EXPERIMENTAL EVALUATIONS AND RESULTS:

The results of the ML models are generated from the Anaconda navigator, pycharm, jupyter notebook sources.



The confusion matrix is a fundamental tool in evaluating the performance of classification models in machine learning, it provides a comprehensive breakdown of the model's predictions compared to the actual (true) values. The confusion matrix shows that the Logistic Regression model performs well, achieving a good balance between precision and recall.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.1, May : 2025

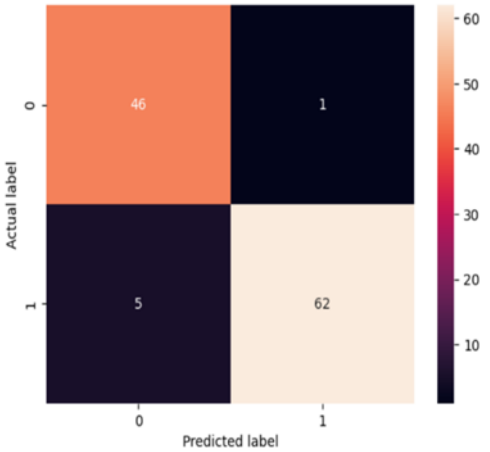


Figure 4. *Confusion matrix for LR*

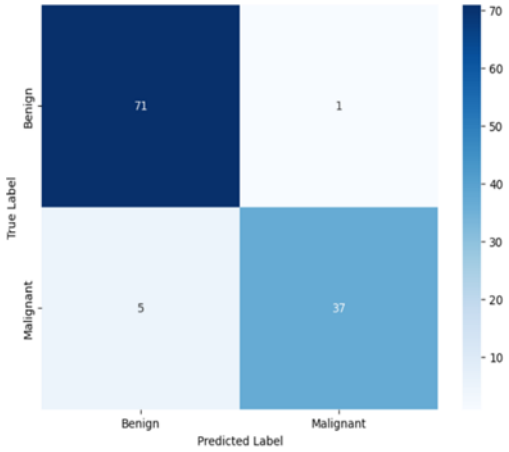


Figure 5. *Confusion matrix for DT*

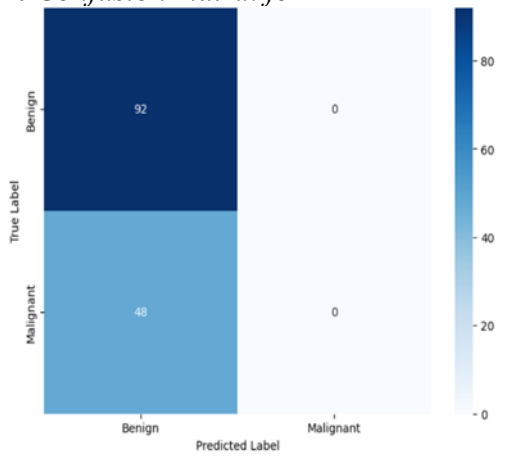


Figure 6. *Confusion matrix for SVM*

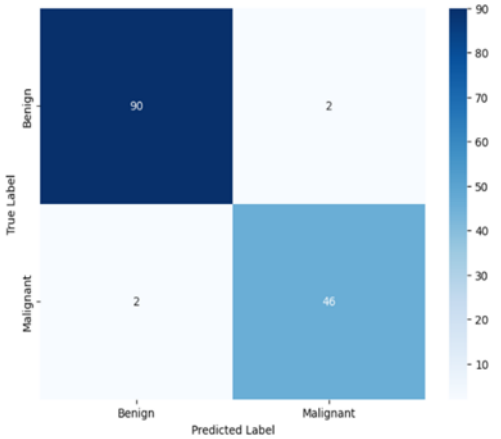


Figure 7. *Confusion matrix for KNN*

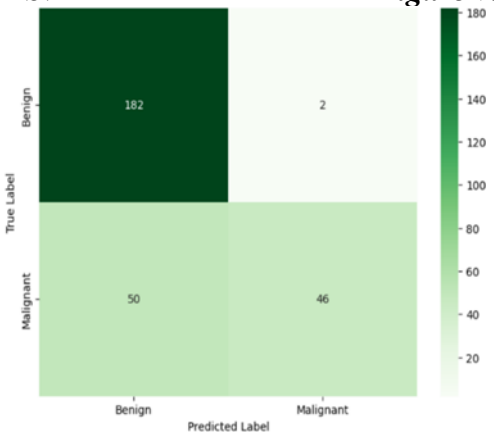


Figure 8. *Combined confusion matrix for SVM-KNN model*

For multiple classes, the confusion matrix extends beyond 2x2 and shows predictions across multiple categories. **The tables shows Benign and Malignant results for selected models:**

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.1, May : 2025

Table 1: *True/Predicted values for LR model*

True/Predicted (LR)	Benign	Malignant
Benign	46	1
Malignant	5	62

Table 2: *True/Predicted values for DT model*

True/Predicted (DT)	Benign	Malignant
Benign	71	1
Malignant	5	37

Table 3: *True/Predicted values for SVM model*

True/Predicted (SVM)	Benign	Malignant
Benign	92	0
Malignant	48	0

Table 4: *True/Predicted values for KNN model*

True/Predicted (KNN)	Benign	Malignant
Benign	90	2
Malignant	2	46

Table 5: *True/Predicted values for SVM + KNN model*

True/Predicted (SVM-KNN)	Benign	Malignant
Benign	182	2
Malignant	50	46

Malignant tumors are cancerous, meaning they grow uncontrollably, invade surrounding tissues, and can spread (metastasize) to other parts of the body. They pose a serious health risk and often require aggressive treatments like surgery, chemotherapy, or radiation. Benign tumors are non-cancerous, grow slowly, and do not spread to other tissues. When they do, surgery is usually the only treatment required.

SVM is not a suitable choice as it completely fails to identify malignant cases. KNN performs the best, achieving the lowest false positive and false negative rates. The combination of KNN and SVM does not lead to significant performance improvement. If KNN is not combined with SVM, then the Decision Tree is a reasonable model for malignant detection.

Table 6: *The Results on Accuracy, Precision, Recall and F1-Score*

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
DT	93.85%	94.87%	88.09%	91.35%
LR	94.73%	98.41%	92.53%	95.38%

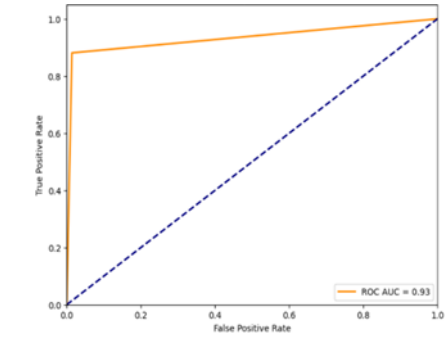
Tabel 7: *The Results on Accuracy, Precision, Recall and F1-Score*

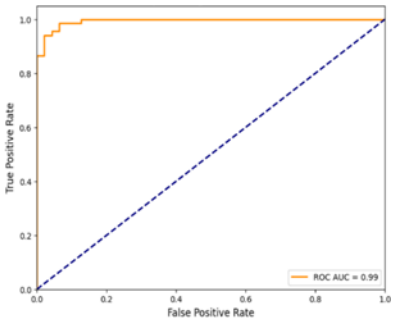
SVM	**Precision**	**Recall**	**F1-score**	**Support**
2	0.63	1.00	0.70	88
4	0.00	0.00	0.00	52
Accuracy			0.63	140
Macro avg	0.31	0.50	0.39	140
Weighted avg	0.40	0.63	0.49	140

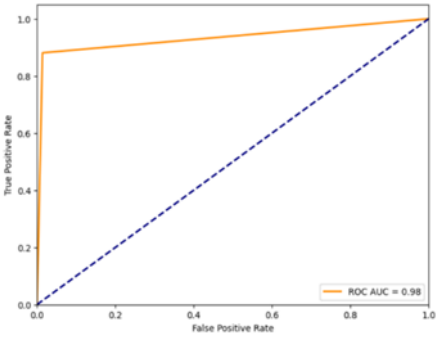
Figure 9: *ROC for DT*

Figure 10: *ROC for LR*

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.1, May : 2025







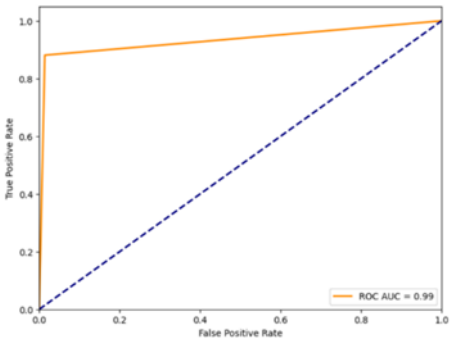


Figure 11: *ROC for SVM*

Figure 12: *ROC for KNN*

The ROC (Receiver Operating Characteristic) curve depicts the balance between the True Positive Rate (TPR) and the False Positive Rate (FPR) for a classifier. The Area Under the Curve (AUC) serves as a measure of the model's overall performance.

The LR model (AUC = 0.99) demonstrates the best performance, as it achieves the highest AUC, indicating an optimal balance between sensitivity and specificity. The SVM model (AUC = 0.98) ranks second, while the DT model (AUC = 0.93) has the lowest performance. If the SVM model is combined with KNN, the Hybrid Model (SVM + KNN) becomes the best choice, as it leverages the strengths of both models. However, if they are not combined, the LR model is the better option for overall performance.

5. CONCLUSION:

This paper compared three models to classify Wisconsin Diagnosis Breast Cancer dataset. **Support Vector Machine (SVM)** models for classifying the Wisconsin Diagnostic Breast Cancer dataset. The results demonstrated that SVM achieved a high classification accuracy of highlighting its effectiveness in distinguishing malignant and benign cases. This study demonstrates the application of **Decision Trees** in modeling breast cancer diagnosis for both local and systematic treatment. Additionally, it explores other techniques that can be utilized for classification. Experimental results validate the effectiveness of the proposed Decision Tree model in accurately diagnosing breast cancer. The study further highlights the potential of Decision Trees as a reliable technique for breast cancer diagnosis, emphasizing their interpretability and efficiency in medical decision-making. This study focused on using **Logistic Regression** as a machine learning technique to diagnose breast cancer based on patients' medical records. By evaluating the algorithm's performance on the Wisconsin breast cancer dataset, valuable insights were gained into its effectiveness for breast cancer diagnosis. The results demonstrate the potential of Logistic Regression as a reliable and interpretable model for identifying malignancies, contributing to early detection and improved patient outcomes. Further exploration into optimizing the model and comparing it with other techniques could enhance its diagnostic accuracy and applicability in clinical settings.

For small and linearly separable data, **Logistic Regression** is useful for better performance. If interpretability is crucial, **Decision Tree** is useful and preferred as a great choice. And for non-linear and complex data, **SVM (with RBF kernel)** is correct for occurring more accuracy percentage. As

compared to SVM, DT, LR the Decision Tree(DT) is better for detection and optimizing the breast cancer diagnosis. Since it provides a good balance between accuracy and robustness.

6. REFERENCES:

1. Reem Alyami, Jinan Alhajjaj, Batool Alnajrani, Ilham Elaalami, Abdullah Alqahtani, Nahier Aldhafferi, Taoreed O. Owolabi b and Sunday O. Olatunji "Investigating the effect of Correlation based Feature Selection on breast cancer diagnosis using Artificial Neural Network and Support Vector Machines" 978-1-4673-8765-1/17/$31.00 ©2017 IEEE
2. Kemal Polat , Salih Günes "Breast cancer diagnosis using least square support vector machine" doi:10.1016/j.dsp.2006.10.008, Digital Signal Processing 17 (2007) 694–701.
3. Edriss Ebrahim Ali , Wu Zhi Feng "Breast Cancer Classification using Support Vector Machine and Neural Network Ebrahim" DOI: https://doi.org/10.21275/v5i3.nov161719 www.ijsr.net
4. Mr.P.SATHIYA NARAYANAN, Ms. PAVITHRA.S, Ms. SAI SARANYA M, Ms. MAKESWARI.M "Identification of Breast Cancer Using The Decision Tree Algorithm" @IEEE 978-1-7281-1524-5
5. E. Venkatesan and T. Velmurugan "Performance Analysis of Decision Tree Algorithms for Breast Cancer Classification" Indian Journal of Science and Technology, Vol 8(29), IPL0625, November 2015
6. Ronak Sumbaly, N. Vishnusri, S. Jeyalatha "Diagnosis of Breast Cancer using Decision Tree Data Mining Technique" International Journal of Computer Applications DOI: https://doi.org/10.5120/17219-7456
7. Ziba Khandezamin , Marjan Naderan , Mohammad Javad Rashti "Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier" Journal of Biomedical Informatics 111 (2020) 103591
8. Model Hamzah A. Alsayadi , Abdelaziz A. Abdelhamid , El-Sayed M. El-Kenawy , Abdelhameed Ibrahim , Marwa M. Eid "Ensemble of Machine Learning Fusion Models for Breast Cancer Detection Based on the Regression" DOI: 10.54216/FPA.090202 ; Vol. 09, No. 02. PP. 19-26, 2022
9. Esther. M. Umoren, Philip Michael Asuquo, Stephen, Bliss Utibe-Abasi "Breast Cancer Detection Using Logistic Regression And Random Forest Machine Learning Techniques" https://www.jmest.org/wp-content/uploads/JMESTN42354405.pdf
10. Fondón I., A., Sarmiento A., García A. I. Silvestre M. and Eloy C. (2018). "Automatic classification of tissue malignancy for breast carcinoma diagnosis," Computers in Biology and Medicine, vol. 96, pp. 41–51
11. S. V. G. Reddy, K. T. Reddy, V. V. Kumari, and K. V. Varma, "An SVM Based Approach to Breast Cancer Classification using RBF and Polynomial Kernel Functions with Varying Arguments," vol. 5, no. 4, pp. 5901–5904, 2014.
12. Iranpour M, Almassi S and Analoui M, "Breast Cancer Detection from fna using SVM and RBF Classifier", In 1st Joint Congress on Fuzzy and Intelligent Systems, 2007.
13. M.H. Asri, H.A Moatassime, "Utilizing Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis". Procedia Comput Sci, Vol. 83, pp. 1064– 1073, 2016.
14. Saravana Kumar K, Arthanariee AM. "Evaluate the multiple breast cancer factors and calculate the risk by software tool breast cancer risk evaluator." Indian Journal of Science and Technology. 2015 Apr;8(S7):686–91.
15. H. Blockeel and J. Struyf."Efficient algorithms for decision tree cross-validation." Proceedings of the Eighteenth International Conference on Machine Learning (C. Brodley and A. Danyluk, eds.), Morgan Kaufmann, 2001, pp. 11-18
16. Kaustubh Chakradeo, Sanyog Vyawahare, and Pranav Pawar. "Breast cancer recurrence prediction using machine learning." In 2019 IEEE Conference on Information and Communication Technology, pp. 1-7. IEEE, 2019.

17. Mochen Li, Gaurav Nanda, Santosh Chhajedss, and Raji Sundararajan. Machine Learning-Based Decision Support System for Early Detection of Breast Cancer. Indian Journal of Pharmaceutical Education and Research 54, no. 3 (2020): S705-S715.
18. Anusha Bharat, N. Pooja, and R. Anishka Reddy. Using machine learning algorithms for breast cancer risk prediction and diagnosis. In 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), pp. 1-4. IEEE, 2018.
19. Aloraini A. Different machine learning algorithms for breast cancer diagnosis. International Journal of Artificial Intelligence and Applications. 2012 Nov; 3(6):21–30.
20. Al Jarullah, A.A., "Decision tree discovery for the diagnosis of type II diabetes," Innovations in Information Technology (IIT), 2011 International Conference on , vol.,no.,pp.303,307, 25-27 April 2011
21. Mehmet Fatih Akay. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 36(2):3240-3247, 2009.
22. M. F. Akay, “Support vector machines combined with feature selection for breast cancer diagnosis,” Expert Syst. Appl., vol. 36, no. 2, pp. 3240–3247, 2009.