

Breast Cancer Detection Using Machine Learning Techniques: A Comparative Analysis of Decision Tree, Support Vector Machine, and Artificial Neural Network

Nandani Kumari

B.Tech Scholar, Dept. of Computer Science &
Engineering
RTC Institute of Technology
Ranchi, India

Shiv Kumar

Assistant Professor, Dept. of Computer Science
& Engineering
RTC Institute of Technology
Ranchi, India

Abstract—Breast cancer is among the most prevalent cancers affecting women worldwide and remains a major cause of mortality. Early diagnosis significantly improves treatment success and survival rates. Machine learning techniques have emerged as powerful tools in medical diagnosis for detecting complex disease patterns from healthcare datasets. This study presents a comparative analysis of three supervised machine learning algorithms, namely Decision Tree (DT), Support Vector Machine (SVM), and Artificial Neural Network (ANN), for breast cancer detection using numerical features extracted from breast cancer biopsy data. The Wisconsin Breast Cancer Dataset is utilized for experimentation. Data preprocessing, feature normalization, exploratory data analysis, model training, and performance evaluation were performed. The proposed models achieved classification accuracies of 94.74%, 96.49%, and 98.24% respectively. The ANN model exhibited superior performance, demonstrating its capability to capture highly nonlinear relationships within medical data. The research highlights the potential of machine learning systems in supporting healthcare professionals for early breast cancer diagnosis and reducing misclassification rates.

Keywords—Breast Cancer Detection, Machine Learning, Artificial Neural Network, Support Vector Machine, Decision Tree, Medical Diagnosis, Healthcare Analytics.

I. INTRODUCTION

Breast cancer is one of the most common cancers affecting women globally and presents a significant public health challenge. According to global cancer statistics, breast cancer accounts for a considerable percentage of new cancer cases and cancer-related deaths among women. Early diagnosis is essential because treatment effectiveness decreases significantly as cancer progresses.

Traditional diagnostic procedures involve mammography imaging, biopsy examinations, magnetic resonance imaging (MRI), and pathological evaluation. Although these methods are effective, diagnostic errors and delayed detection remain concerns. The advancement of machine learning has enabled the creation of intelligent diagnostic support systems capable of analyzing large volumes of patient data with high precision.

Machine learning models can identify hidden patterns and correlations among clinical variables that may not be easily recognized through conventional approaches. Consequently, machine learning-based cancer prediction systems are increasingly being adopted in healthcare settings. The objective of this study is to investigate the effectiveness of three supervised machine learning algorithms in breast cancer classification and compare their predictive performance.

II. LITERATURE REVIEW

Several studies have demonstrated the importance of machine learning in cancer diagnosis. Al-Jawad et al. utilized convolutional neural networks (CNNs) for mammographic image analysis and reported substantial improvements in classification performance. Khourdifi and Bahaj implemented Support Vector Machines, Decision Trees, and Random Forest algorithms for breast cancer classification using feature extraction methods and found SVM to provide superior accuracy. Esteva et al. demonstrated dermatologist-level cancer classification using deep neural networks, establishing the effectiveness of AI-based healthcare diagnostics.

Recent studies have emphasized deep learning for image-based diagnosis, feature selection methods for dimensionality reduction, transfer learning approaches, and Explainable Artificial Intelligence (XAI) for healthcare. However, comparative evaluation of classical machine learning algorithms and neural networks using numerical features remains a relevant research area.

III. PROBLEM STATEMENT

Although extensive medical data are available, accurate breast cancer prediction remains challenging because:

- Healthcare datasets contain high-dimensional features.
- Irrelevant features may degrade classification performance.
- Misclassifications can lead to life-threatening consequences.
- Early and accurate diagnosis remains essential for reducing mortality.

Therefore, there is a need to investigate robust machine learning models capable of providing accurate predictions using numerical diagnostic features.

IV. RESEARCH OBJECTIVES

The specific objectives of this research are:

1. To collect breast cancer diagnostic data.
2. To preprocess and transform the dataset.
3. To perform exploratory data analysis.
4. To develop Decision Tree, SVM, and ANN models.
5. To compare model performances.
6. To identify the most effective classifier for breast cancer prediction.

V. DATASET DESCRIPTION

The Wisconsin Breast Cancer Dataset was used for this study. The dataset contains 569 instances and 30 diagnostic features targeting a binary target variable: Malignant (M) or Benign (B). Important features include Radius Mean, Texture Mean, Perimeter Mean, Area Mean, Smoothness Mean, Compactness Mean, Concavity Mean, Symmetry Mean, and Fractal Dimension Mean. The dataset provides Standard Error (SE) and Worst Value measurements for each attribute to capture variability and extreme observations.

VI. PROPOSED METHODOLOGY

A. Data Collection

The data were obtained from publicly available cancer repositories and healthcare datasets.

B. Data Preprocessing

The preprocessing stage involved missing value handling, data cleaning, outlier analysis, feature scaling, and normalization.

C. Exploratory Data Analysis

The dataset was analyzed using histograms, scatter plots, and correlation heatmaps. Heatmaps provide relationships among diagnostic variables and help identify highly correlated features.

D. Feature Engineering

The feature engineering stage included feature normalization, dimensionality analysis, and feature importance evaluation.

E. Training and Testing

The dataset was divided into a Training Set (80%) and a Testing Set (20%). Cross-validation techniques were applied to improve generalization.

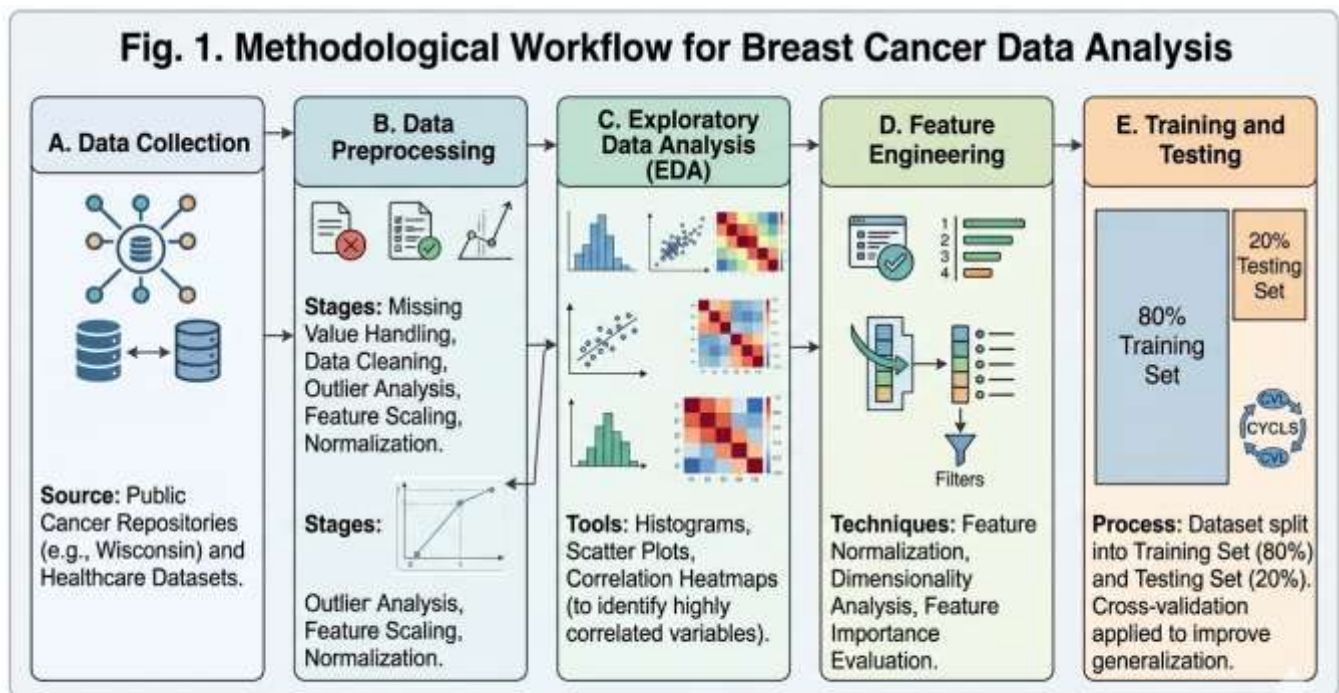


Fig. 1. Methodological Workflow for Breast Cancer Data Analysis

VII. MACHINE LEARNING MODELS

A. Decision Tree Classifier

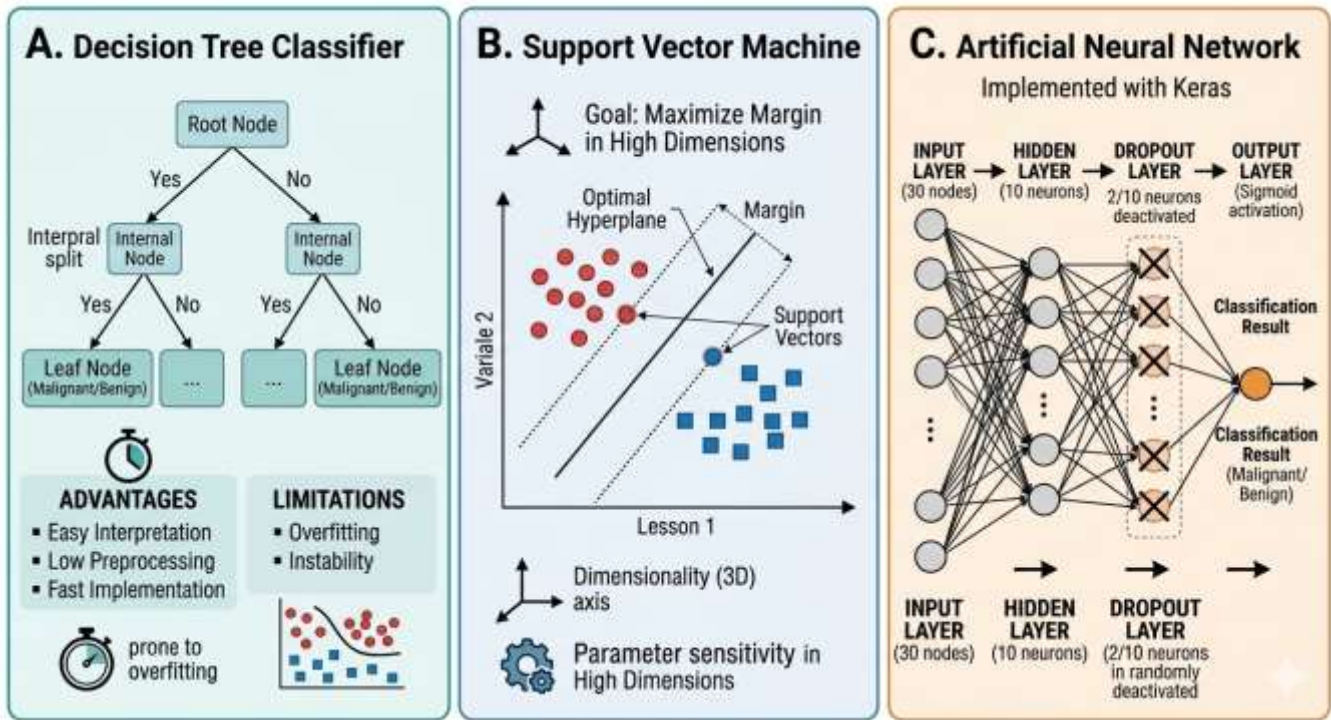
Decision Tree is a supervised learning algorithm that partitions the dataset into subsets using feature-based decision rules. Its advantages include easy interpretation, low preprocessing requirements, and fast implementation, though it suffers from limitations like overfitting and instability.

B. Support Vector Machine

Support Vector Machine (SVM) attempts to identify an optimal hyperplane separating the classes. It is highly effective in high dimensions and offers good generalization, though it can be computationally expensive and sensitive to parameter selection.

C. Artificial Neural Network

The ANN model was implemented using the Keras framework and developed for binary classification. The architecture consists of an Input Layer (30 nodes), a Hidden Layer (10 neurons), a Dropout Layer, and an Output Layer (Sigmoid activation).



VIII. EXPERIMENTAL RESULTS AND ANALYSIS

A. Class Distribution Analysis

The distribution of malignant and benign cases was analyzed before training. Malignant and benign sample frequencies help in understanding dataset balance.

B. Correlation Analysis

Correlation heatmaps revealed strong relationships between perimeter, radius, and area features.

C. Confusion Matrix Analysis

Decision Tree: Accuracy: 94.74%, Precision: 93.75%, Recall: 93.75%, F1-Score: 93.75%.

Support Vector Machine: Accuracy: 96.49%, Precision: 97.83%, Recall: 93.75%, F1-Score: 95.74%.

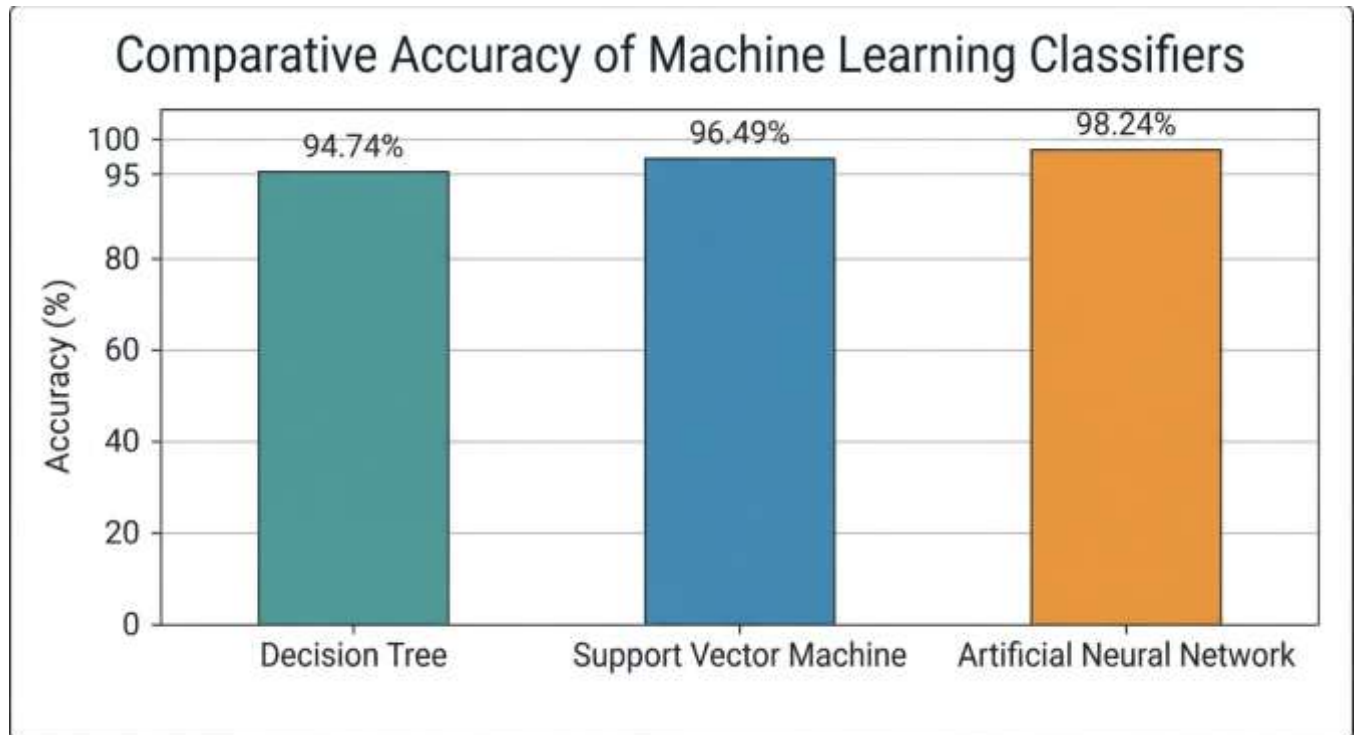
Artificial Neural Network: Accuracy: 98.24%.

IX. COMPARATIVE PERFORMANCE EVALUATION

TABLE I. CLASSIFIER ACCURACY COMPARISON

Classifier	Accuracy (%)
Decision Tree	94.74
Support Vector Machine	96.49
Artificial Neural Network	98.24

The ANN model achieved the highest predictive accuracy, indicating its capability to capture nonlinear relationships among diagnostic features. SVM also performed effectively and outperformed the Decision Tree.



X. DISCUSSION

The comparative study reveals that ANN provides superior classification performance, while SVM demonstrates strong robustness for medical datasets. The Decision Tree offers clear interpretability but lower accuracy. Overall, feature preprocessing significantly improves prediction quality. These findings strongly support the integration of intelligent diagnostic systems into healthcare environments for early cancer screening and risk assessment to assist clinicians in reducing diagnostic errors.

XI. CONCLUSION

This research investigated the effectiveness of Decision Tree, Support Vector Machine, and Artificial Neural Network models for breast cancer classification. Experimental analysis demonstrated that ANN achieved the highest accuracy of 98.24%, followed by SVM and Decision Tree. The study confirms that machine learning techniques can play an essential role in supporting healthcare professionals through accurate, reliable, and early cancer diagnosis. The proposed approach has the potential to enhance clinical decision-making and improve patient outcomes.

XII. FUTURE SCOPE

Future research directions include deep CNN-based medical image analysis, transfer learning applications, and Explainable AI (XAI) frameworks. Additionally, integration with Electronic Health Records (EHR), mobile healthcare applications, real-time cloud-based diagnostic systems, and federated healthcare machine learning remain highly promising paths.

REFERENCES

- [1] A. R. Chaudhury et al., “Diagnosis of Invasive Ductal Carcinoma Using Image Processing Techniques,” Proc. International Conference on Medical Imaging, 2011.
- [2] F. K. Ahmad and N. Yusoff, “Breast Cancer Classification Using Random Forest,” Journal of Clinical Analytics, 2013.
- [3] A. Dey, “Machine Learning Algorithms: A Review,” Computer Science Review, vol. 4, pp. 23–31, 2016.
- [4] H. Sharma and S. Kumar, “Survey on Decision Tree Algorithms,” International Journal of Computer Applications, 2016.
- [5] A. Esteva et al., “Dermatologist-level Classification of Skin Cancer,” Nature, vol. 542, pp. 115–118, 2017.
- [6] Y. Khouardifi and M. Bahaj, “Machine Learning Algorithms for Breast Cancer Prediction,” IEEE Insurance & Healthcare Systems, 2018.