INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

Early Detection of Breast Cancer Using Machine Learning Approaches

[1]Ashish Das, [2]Manmohan Krishna, [3]Soubhik Kundu, [4]Debasish Mondal, [5]Anmol Mishra

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Computer Science & Engineering, [2 3 4 5] Computer Science & Engineering

[1]Durgapur Institute of Advanced Technology and Management, Rajbandh, India
[2 3 4 5] Durgapur Institute of Advanced Technology and Management, Rajbandh, India

1. Abstract

Accurate prediction of breast cancer can help doctors begin treatment earlier and improve patient outcomes. This project develops a machine learning system that analyzes clinical dataset features to classify breast tumors into benign or malignant categories. The workflow begins with preparing the dataset through preprocessing methods such as removing inconsistent values, scaling numerical features, and selecting the most relevant attributes for training [1]. Several supervised learning models were implemented for performance comparison, including Decision Tree, Random Forest, K-Nearest Neighbors, Logistic Regression, Support Vector Classifier, and Linear SVC. Each model was evaluated using multiple performance measures such as accuracy, precision, recall, and F1-score to determine its effectiveness in prediction tasks. Experimental analysis showed that Random Forest produced the most reliable results compared with the other algorithms tested. The developed model provides a data-driven approach that can support healthcare professionals by offering faster preliminary analysis and assisting in diagnostic decisions. This work highlights the growing role of machine learning techniques in medical prediction systems and disease risk assessment.

Index Terms - Breast Cancer Detection, Machine Learning, Random Forest, Classification, Medical Diagnosis.

2. Introduction

Breast cancer is one of the most serious health problems affecting women across the world. It occurs when abnormal cells in the breast grow uncontrollably and form a tumor. Over the years, the number of breast cancer cases has increased significantly, making it one of the leading causes of death among women. According to medical experts, detecting breast cancer at an early stage can greatly improve the chances of successful treatment and long-term survival. When the disease is identified early, doctors can provide timely medical care, reduce complications, and improve the quality of life for patients [2]. Because of this, accurate and fast diagnosis plays a very important role in modern healthcare systems.

Traditionally, breast cancer is diagnosed using techniques such as mammography, biopsy, ultrasound imaging, and clinical examinations performed by medical professionals. Mammography is one of the most common screening methods used to detect unusual changes in breast tissue. A biopsy is another important procedure in which a small sample of tissue is collected and examined to determine whether cancer cells are present. While these methods are effective, they also have several limitations [3]. In many hospitals and healthcare centers, these diagnostic procedures can be expensive and time-consuming. Patients may need to wait for multiple tests and reports before receiving a final diagnosis. In addition, the accuracy of these methods often depends on the experience and expertise of doctors or radiologists. Human interpretation errors, fatigue, or lack of experience may sometimes lead to incorrect or delayed diagnoses.

With the rapid development of technology, artificial intelligence (AI) and machine learning (ML) have started transforming the healthcare industry [4]. Machine learning is a branch of artificial intelligence that enables computer systems to learn from data and make predictions or decisions without being explicitly programmed for every task. In the medical field, machine learning techniques are increasingly being used to analyze large amounts of healthcare data and identify hidden patterns that may not be easily visible to humans. These technologies have the potential to improve the speed, accuracy, and reliability of disease detection systems.

In breast cancer diagnosis, machine learning algorithms can analyse patient data and classify tumors as either benign or malignant based on different medical features. A benign tumor is non-cancerous and generally less harmful, whereas a malignant tumor is cancerous and can spread to other parts of the body if not treated in time [5]. By training machine learning models on medical datasets, the system can learn to recognize patterns associated with different tumor types and provide accurate predictions. This reduces the dependence on manual interpretation and supports doctors in making better clinical decisions.

The main objective of this project is to develop a machine learning-based breast cancer detection system that can accurately classify tumors using diagnostic features from medical datasets. The system uses various attributes such as tumor size, texture, radius, smoothness, and other cellular characteristics to determine whether a tumor is benign or malignant. Different machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) can be used to train and test the model. These algorithms are evaluated based on their prediction accuracy, efficiency, and reliability.

One of the major advantages of using machine learning in breast cancer detection is the ability to process large volumes of medical data quickly and efficiently. Unlike traditional methods, machine learning models can provide faster results and reduce the workload on healthcare professionals. These systems can also minimize human errors and improve diagnostic consistency [6]. In addition, automated prediction systems can be especially helpful in remote or underdeveloped areas where access to experienced medical specialists may be limited.

Another important benefit of this project is its potential to support early diagnosis. Early detection of breast cancer can save lives by allowing treatment to begin before the disease reaches an advanced stage. A machine learning-based system can act as a supportive tool for doctors by providing a second opinion and increasing confidence in the diagnostic process [7]. Although such systems are not intended to replace medical professionals, they can significantly assist in improving healthcare services and patient outcomes.

In conclusion, the use of machine learning and artificial intelligence in breast cancer detection represents an important advancement in modern healthcare. By combining medical knowledge with intelligent algorithms, it is possible to create systems that are faster, more accurate, and more reliable than many traditional diagnostic approaches. This project aims to contribute to the healthcare sector by developing an efficient breast cancer detection model that supports early diagnosis, reduces human error, and improves overall diagnostic accuracy. As technology continues to evolve, machine learning-based healthcare solutions are expected to play an even greater role in saving lives and improving patient care in the future.

3. Problem Statement

Even with continuous improvements in medical research and healthcare systems, diagnosing breast cancer accurately remains a major challenge. In many regions, especially rural and less developed areas, patients often face delays in screening and limited access to skilled radiologists or advanced medical facilities. Conventional diagnostic techniques such as mammography, biopsy, and physical examinations rely greatly on the expertise and judgment of medical professionals. Because of this dependency, the interpretation of test results may vary from one specialist to another, leading to inconsistent outcomes. In addition, factors such as heavy workload, stress, fatigue, and incomplete patient records can increase the possibility of human error during diagnosis.

Wrong diagnosis can seriously affect a patient's health and emotional well-being. False positive cases may cause unnecessary fear, extra medical procedures, and additional treatment costs, while false negative cases can delay proper treatment and allow cancer to spread further. Therefore, ensuring accurate and reliable breast cancer detection is extremely important in modern healthcare.

To overcome these limitations, an automated system based on machine learning is proposed. The system is designed to analyze medical data efficiently, improve prediction accuracy, reduce diagnostic mistakes, and support doctors in identifying tumors more effectively and consistently.

4. Objectives of the Project

The primary goal of this project is to develop an intelligent breast cancer detection system using machine learning techniques. The project focuses on collecting and analyzing breast cancer diagnostic datasets to understand important medical patterns and characteristics related to tumor detection. Another objective is to preprocess the collected data by removing unnecessary information, handling missing values, and selecting the most relevant features that improve prediction accuracy.

The project also aims to implement and test different machine learning classification algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbors [8]. These models are evaluated and compared using performance measures like accuracy, precision, recall, and F1-score to determine their effectiveness in predicting breast cancer.

A key objective is to identify the most accurate and reliable model for classifying tumors as benign or malignant. Ultimately, the proposed system is designed to support healthcare professionals by enabling faster, more consistent, and early diagnosis of breast cancer, which can improve treatment outcomes and patient survival rates.

5. Literature Survey

In recent years, machine learning techniques have gained significant attention in the field of medical diagnosis, particularly for the early detection of breast cancer. Researchers and healthcare professionals have explored various computational approaches to improve diagnostic accuracy, reduce human error, and support clinical decision-making [9]. Breast cancer detection has become one of the most widely studied applications of artificial intelligence because early identification of tumors can greatly increase survival rates and improve treatment outcomes.

Initial studies in this area mainly focused on the use of traditional machine learning classification algorithms. Methods such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), and Naïve Bayes were commonly used for predicting whether a tumor was benign or malignant. These models were trained using medical datasets containing tumor characteristics such as radius, texture, perimeter, area, and smoothness. Among these approaches, Support Vector Machines and Logistic Regression showed strong classification performance due to their ability to handle high-dimensional medical data effectively. Many researchers reported prediction accuracies ranging from 85% to 94%, demonstrating the potential of machine learning in breast cancer diagnosis [10].

Although early machine learning models achieved promising results, they also presented several limitations. One common issue was overfitting, where the model performed well on training data but failed to provide accurate predictions on unseen data. This problem reduced the generalization capability of the models and affected their reliability in real-world medical applications. In addition, some studies relied on small or highly specific datasets, limiting the adaptability of the developed systems to diverse patient populations. Researchers also observed that certain algorithms were sensitive to irrelevant or redundant features, which negatively impacted overall performance.

To overcome these challenges, later research introduced advanced techniques such as ensemble learning. Ensemble methods combine the predictions of multiple models to improve accuracy, stability, and robustness [4]. Random Forest became one of the most widely used ensemble algorithms in breast cancer prediction because it constructs multiple decision trees and combines their outputs to produce more reliable results. This approach reduces the risk of overfitting and improves classification performance compared to individual models. Several studies demonstrated that Random Forest could achieve higher accuracy and better consistency than traditional classifiers when applied to medical datasets.

Another important area highlighted in research is feature selection and data preprocessing. Medical datasets often contain irrelevant, missing, or duplicated information that can reduce the efficiency of machine learning algorithms. Researchers found that preprocessing techniques such as normalization, handling missing values, and removing unnecessary attributes significantly improve model performance. Feature selection methods help identify the most important tumor characteristics, allowing algorithms to focus on relevant information and produce more accurate predictions. Effective preprocessing not only improves accuracy but also reduces computational complexity and training time.

Deep learning techniques have also emerged as a growing research area in breast cancer diagnosis. Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) have been applied to medical imaging data such as mammograms and ultrasound images [11]. These methods are capable of automatically extracting complex features from images and identifying hidden patterns associated with cancer. While deep learning approaches often provide high accuracy, they usually require large datasets, powerful computational resources, and longer training times. As a result, traditional machine learning methods continue to remain practical and widely used for structured medical datasets.

Despite the progress made in this field, many existing studies still have certain limitations. Some systems focus only on a single algorithm without comparing multiple models under the same conditions. Others lack practical implementation or user-friendly interfaces that could assist healthcare professionals in real clinical environments. In addition, many research works prioritize accuracy alone while ignoring other important evaluation metrics such as precision, recall, and F1-score, which are essential in medical diagnosis.

This project builds upon previous research by developing a comprehensive machine learning-based breast cancer detection system. The proposed approach includes systematic data preprocessing, feature extraction, feature selection, and comparative evaluation of multiple machine learning algorithms [12]. By analyzing and comparing different models, the project aims to identify the most accurate and reliable classifier for breast cancer prediction. The final system is designed to support early diagnosis, reduce diagnostic errors, and provide practical assistance to medical professionals in improving patient care and treatment outcomes.

6. Methodology

The proposed breast cancer detection system follows a systematic and organized methodology to ensure accurate prediction and efficient model performance. The methodology consists of several important stages, including data collection, preprocessing, feature selection, model training, and performance evaluation. Each stage plays a significant role in improving the effectiveness and reliability of the machine learning-based diagnostic system [3]. The complete process is designed to analyze breast cancer data carefully and classify tumors as benign or malignant with high accuracy.

6.1. Dataset Collection

The first step in the proposed methodology is collecting a suitable breast cancer dataset from a reliable medical source. The dataset contains important diagnostic information related to breast tumors and includes several medical features used for cancer prediction [13]. These features may include tumor radius, texture, smoothness, perimeter, compactness, concavity, symmetry, and other cellular characteristics observed during clinical examinations.

A high-quality dataset is essential for building an efficient machine learning model because the accuracy of prediction largely depends on the quality and reliability of the input data. The collected dataset contains two main classes: benign tumors, which are non-cancerous, and malignant tumors, which are cancerous and dangerous if not treated at an early stage. The dataset serves as the foundation for training and testing the machine learning algorithms used in the project.

6.2. Data Preprocessing

After collecting the dataset, the next important step is data preprocessing. Raw medical data often contains missing values, duplicated records, irrelevant attributes, and inconsistencies that can negatively affect the performance of machine learning algorithms. Therefore, preprocessing is necessary to clean and prepare the dataset before model training [1].

In this stage, missing values are identified and handled appropriately to maintain the quality of the dataset. Unnecessary or irrelevant columns that do not contribute to prediction are removed to reduce complexity and improve efficiency. Data normalization and scaling techniques are also applied because machine learning algorithms perform better when numerical values are within a similar range. Standard scaling is commonly used to transform the data so that features have a mean value close to zero and a standard deviation close to one [14].

Preprocessing improves the reliability of the dataset and helps machine learning models learn patterns more effectively. It also reduces computational complexity and prevents biased predictions caused by inconsistent data values.

6.3. Feature Extraction and Selection

Feature extraction and selection are essential stages in the development of an efficient breast cancer detection system. Medical datasets often contain a large number of features, but not all of them contribute equally to the prediction process. Some features may be irrelevant or redundant, which can reduce model accuracy and increase training time.

To address this issue, statistical techniques are used to identify and select the most important features from the dataset [4]. In this project, SelectKBest with the ANOVA F-test method is applied for feature selection. This method evaluates the relationship between input features and the target variable to determine which attributes are most significant for classification.

By selecting only the most relevant features, the system becomes more efficient and easier to interpret. Feature selection also helps reduce overfitting, improve prediction accuracy, and enhance the overall performance of machine learning models. As a result, the algorithms can focus on the most meaningful tumor characteristics during training and testing.

6.4. Data Balancing

Class imbalance is a common problem in medical datasets, including breast cancer datasets. In many cases, one class may contain significantly more samples than the other. For example, benign cases may outnumber malignant cases, leading the machine learning model to become biased toward the majority class.

To overcome this issue, data balancing techniques are applied before training the models. Balancing ensures that both classes are represented fairly in the dataset, allowing the algorithms to learn equally from benign and malignant samples [15]. Techniques such as oversampling or under sampling may be used to achieve balanced class distribution.

Balanced data improves the reliability of predictions and helps reduce false negative and false positive results. This is especially important in healthcare applications, where incorrect classification can seriously affect patient treatment and diagnosis.

6.5. Data Splitting

Once preprocessing and balancing are completed, the dataset is divided into two separate parts: training data and testing data. In this project, 70% of the dataset is used for training, while the remaining 30% is reserved for testing and evaluation.

The training dataset is used to teach the machine learning models how to recognize patterns and relationships between input features and tumor classification. During this phase, the algorithms learn from historical data and adjust their internal parameters to improve prediction accuracy.

The testing dataset is used to evaluate how well the trained models perform on unseen data. This step is important because it measures the generalization capability of the models and ensures that they can make accurate predictions on new patient data rather than only memorizing the training samples.

6.6. Model Training

In the model training stage, multiple machine learning classification algorithms are implemented and trained using the prepared dataset. Different algorithms are selected because each model has unique strengths and performance characteristics.

The classifiers used in this project include Logistic Regression, Decision Tree, K-Nearest Neighbours (KNN), Linear Support Vector Classifier (Linear SVC), Support Vector Machine (SVC), and Random Forest. Logistic Regression is widely used for binary classification problems due to its simplicity and effectiveness. Decision Tree provides easy interpretation and visualization of decisions. KNN classifies data based on similarity with neighboring samples. SVC and Linear SVC are effective for high-dimensional datasets, while Random Forest improves prediction reliability through ensemble learning techniques.

Training multiple algorithms allows comparative analysis and helps identify the most suitable model for breast cancer detection [16].

6.7. Model Evaluation

The final stage of the methodology is model evaluation. After training, each classifier is tested using the testing dataset to measure its predictive performance. Several standard evaluation metrics are used to analyze model effectiveness, including accuracy, precision, recall, and F1-score.

Accuracy measures the percentage of correctly classified samples. Precision evaluates how many predicted positive cases are actually correct, while recall measures the ability of the model to identify actual positive cases. The F1-score provides a balance between precision and recall and is especially useful for medical diagnosis tasks [17].

By comparing these evaluation metrics, the best-performing classifier can be identified. The selected model is expected to provide reliable and accurate breast cancer prediction, assisting healthcare professionals in early diagnosis and improving patient treatment outcomes.

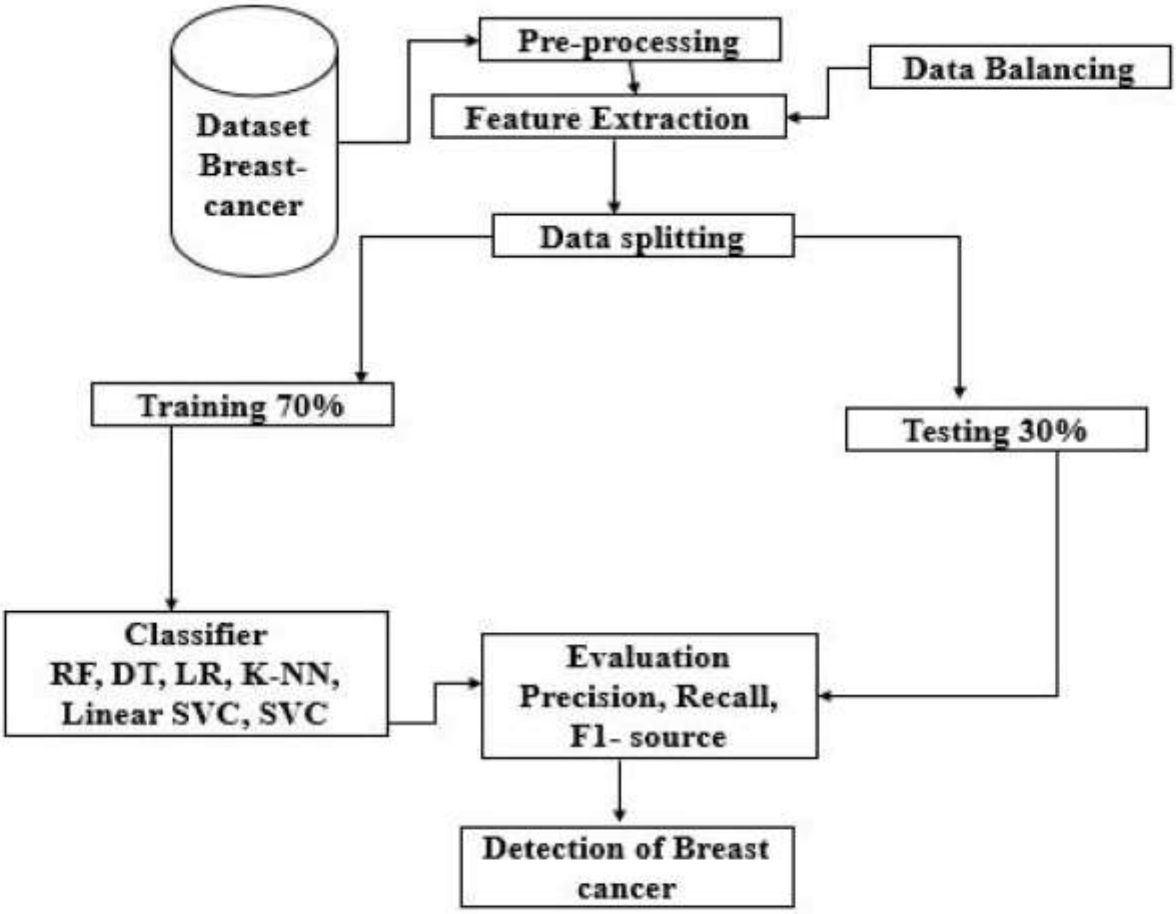

Fig. 1. Proposed methodology

The system architecture illustrates the complete workflow of the proposed breast cancer detection system. It begins with data input and preprocessing, followed by feature extraction and data balancing. The processed data is then split into training and testing datasets. Various classifiers are trained using the training data, and their performance is evaluated using standard metrics. The final output is the accurate detection of breast cancer as benign or malignant.

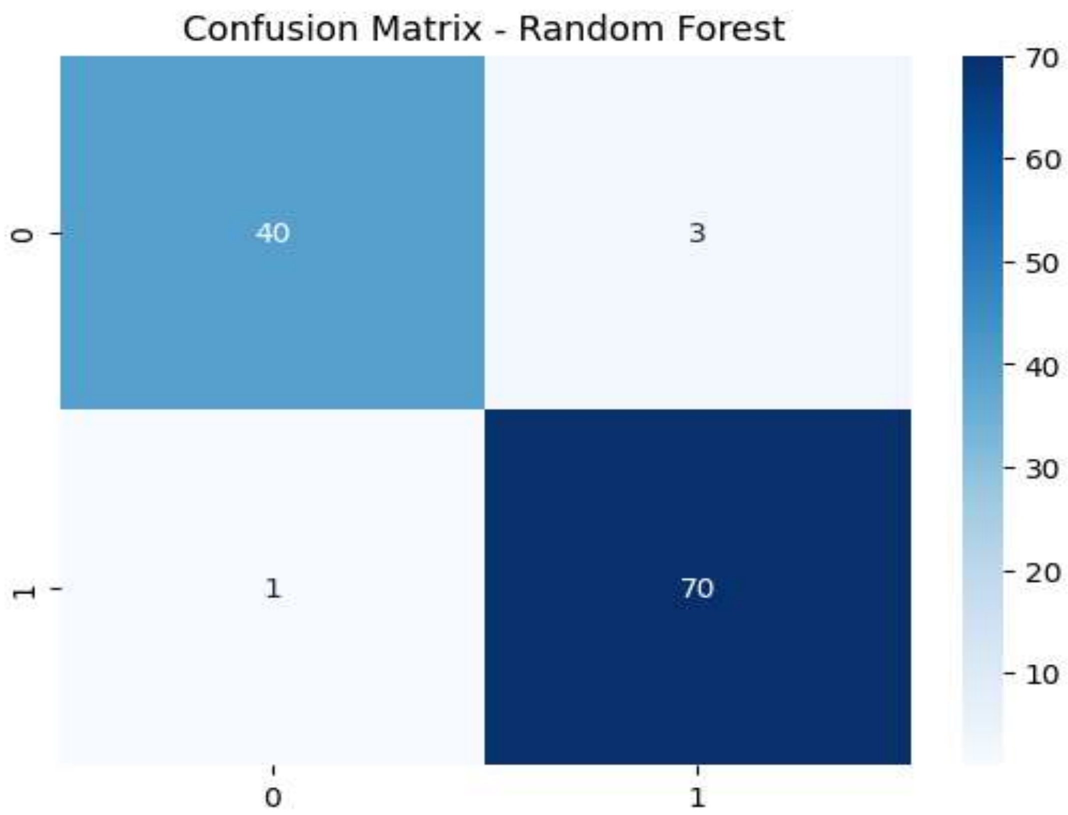

Fig. 2. Confusion matrix

7. Results and Discussion

The trained machine learning models were tested after dataset preparation, feature filtering, and classifier development were completed. Performance comparison was carried out to understand how effectively each model could classify breast tumors into benign and malignant categories. The implemented algorithms included Logistic Regression, Decision Tree, K-Nearest Neighbors, Linear SVC, Support Vector Classifier, and Random Forest. Model quality was assessed using evaluation measures such as accuracy, precision, recall, and F1-score.

Experimental observations showed noticeable differences in predictive performance across the classifiers. Among all tested approaches, Random Forest generated the most stable and accurate results on the breast cancer dataset. The model recorded an accuracy close to 96%, outperforming the remaining algorithms used in the experiment. This indicates that the classifier was able to identify most tumor samples correctly during testing.

A major reason for this strong performance is the ensemble-based design of Random Forest. Instead of depending on a single learning structure, the algorithm builds multiple decision trees from different portions of the training data. Predictions from these trees are then aggregated to produce the final output. This strategy improves robustness and lowers the probability of overfitting.

The findings suggest that ensemble methods are highly suitable for structured medical datasets, where stable prediction and reduced classification error are important for reliable diagnosis support.

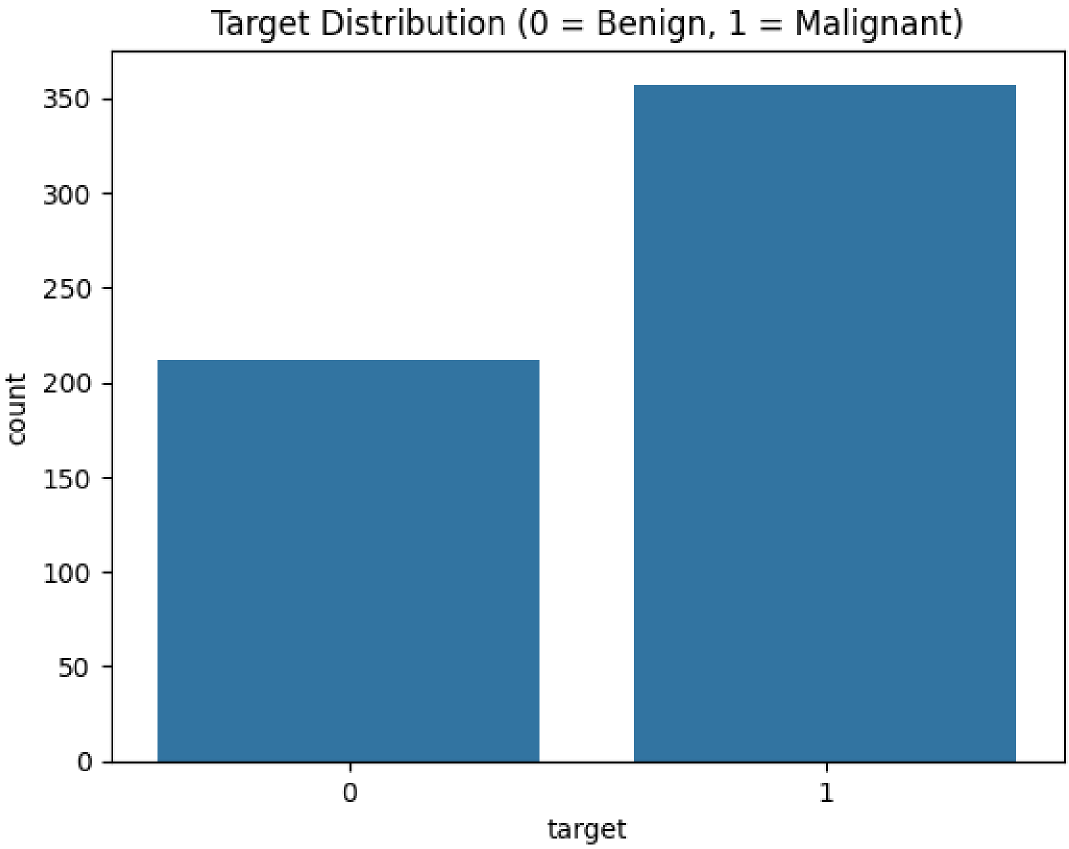

Fig. 3. Breast cancer target and count

Overfitting occurs when a model performs extremely well on training data but fails to generalize properly on unseen data. The Random Forest classifier reduces this issue by averaging predictions from multiple trees rather than depending on a single model. As a result, it provides better generalization and more stable performance when applied to new data samples. This makes Random Forest highly suitable for medical diagnosis applications where reliability and consistency are extremely important [19].

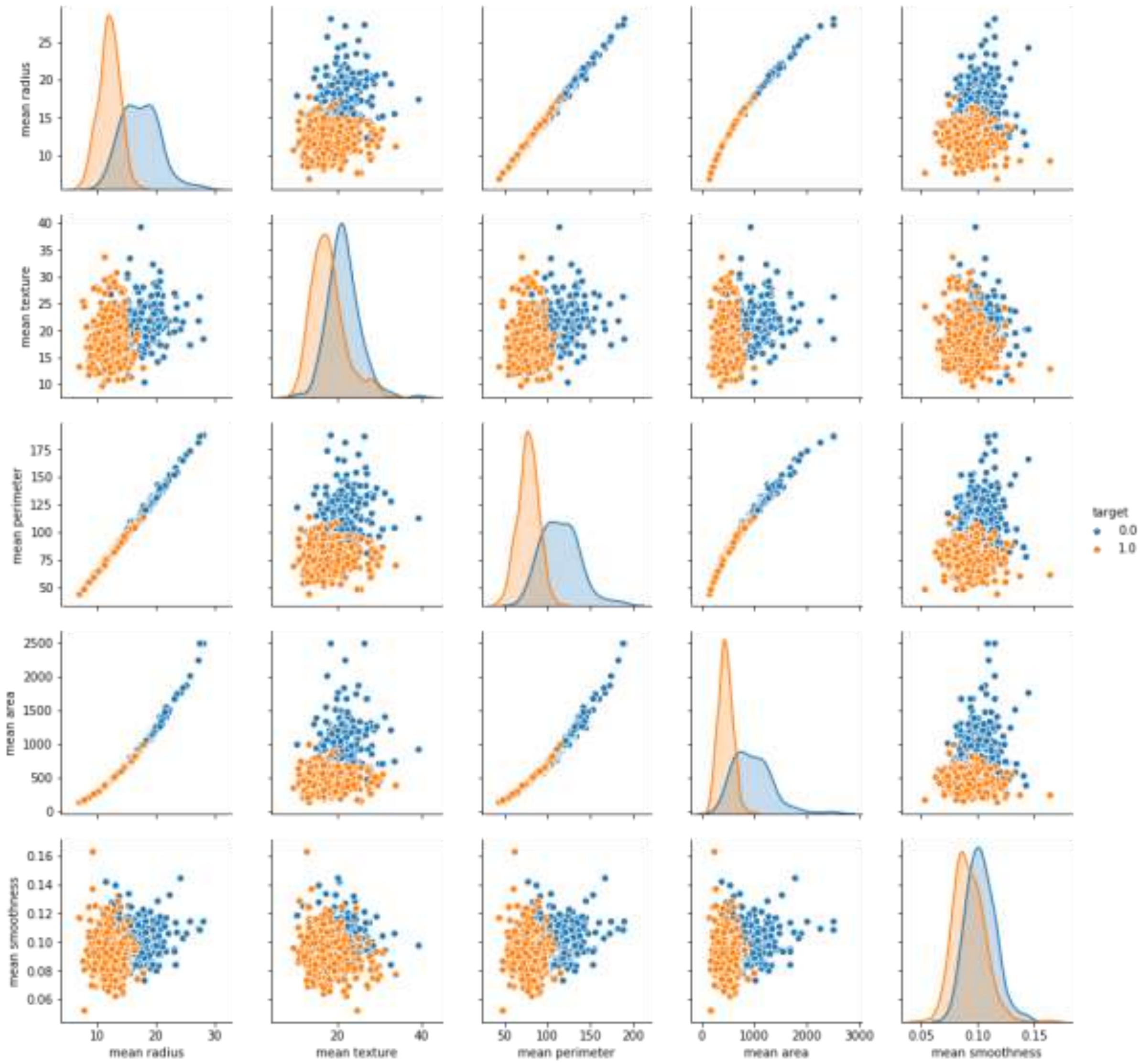

Fig. 4. Breast cancer diagnosis graph

Other machine learning algorithms used in the project also produced satisfactory results, although their performance was slightly lower than Random Forest. Logistic Regression showed good classification capability due to its effectiveness in binary classification problems. Support Vector Machine and Linear SVC also achieved strong accuracy because they are capable of handling high-dimensional datasets efficiently. Decision Tree models provided interpretable results and easy visualization of classification rules, while KNN demonstrated acceptable performance based on neighborhood similarity [1]. However, some of these algorithms were more sensitive to data imbalance, noise, or feature complexity, which affected their overall prediction accuracy.

The evaluation metrics further confirmed the reliability of the Random Forest classifier. High precision values indicated that the model produced fewer false positive predictions, reducing the chances of unnecessary medical concern or treatment for patients. Similarly, high recall values demonstrated the model's strong ability to correctly identify malignant tumor cases, which is especially important in healthcare applications because missed cancer diagnoses can delay treatment and increase health risks. The balanced F1-score also reflected the overall stability and effectiveness of the model.

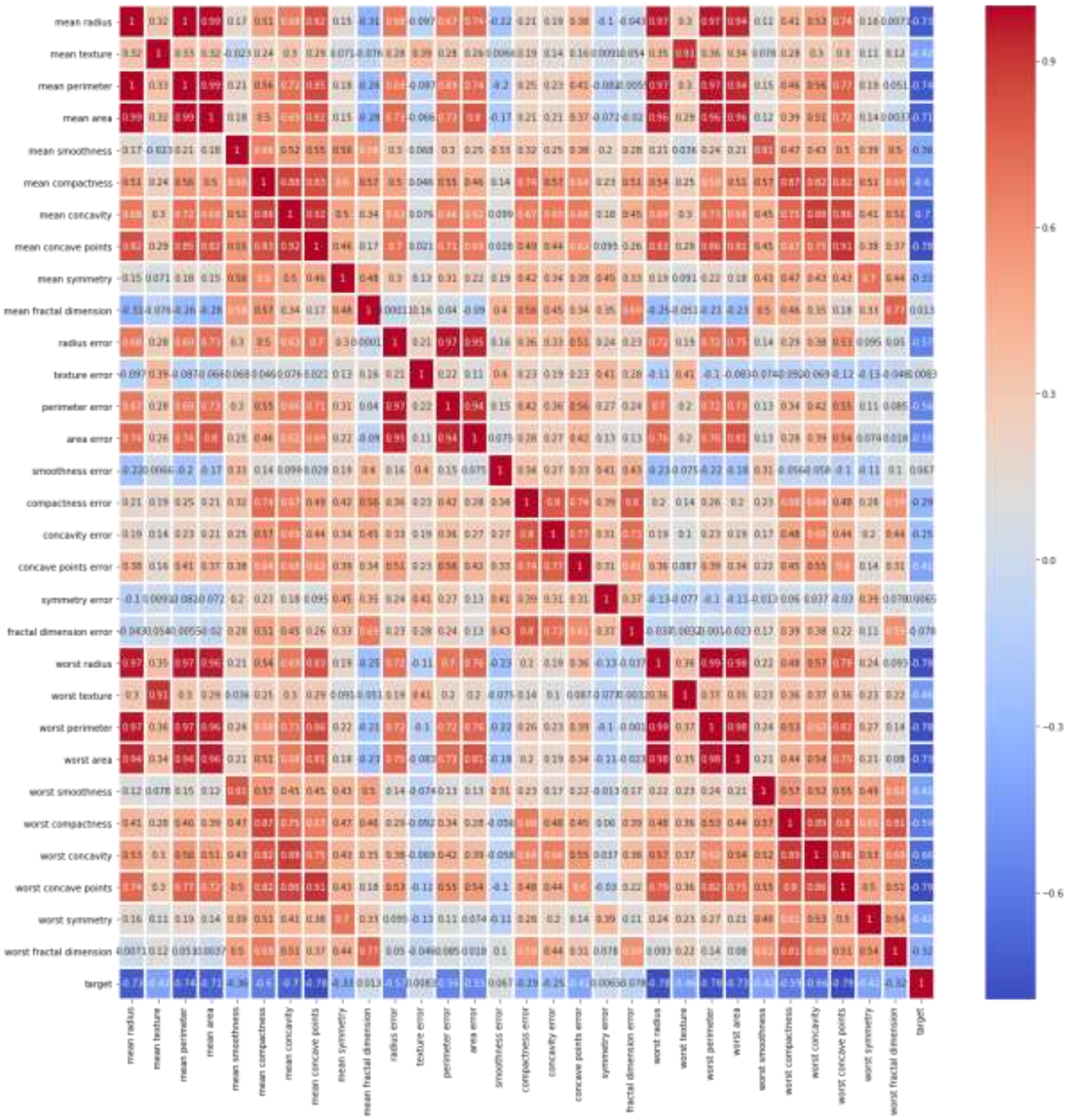

Fig. 5. Correlation between variables

The results of this project clearly show that machine learning techniques can play a significant role in breast cancer diagnosis and prediction. By analyzing medical datasets and identifying hidden relationships among diagnostic features, machine learning models can assist healthcare professionals in making faster and more accurate decisions. Automated systems based on machine learning can also reduce dependence on manual interpretation and minimize errors caused by fatigue or subjective analysis.

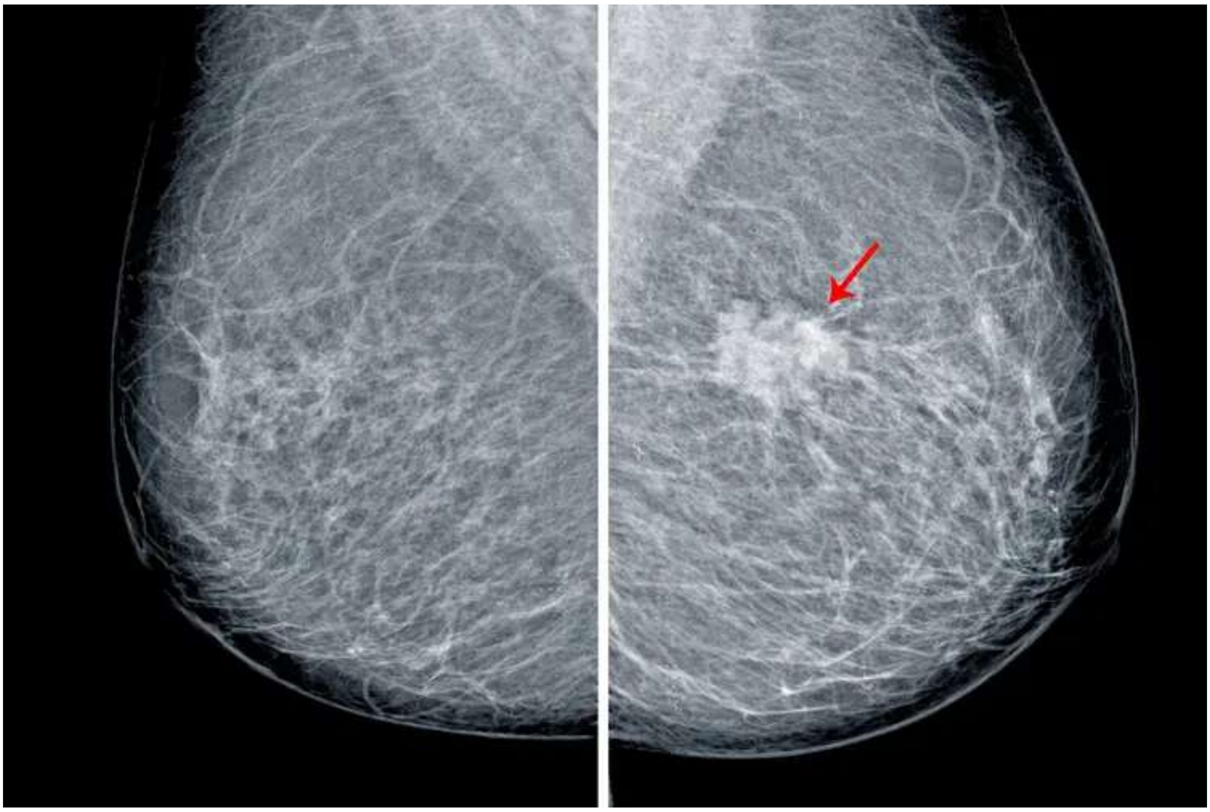

Fig. 6. Breast cancer looks on mammography.

Another important advantage observed during the project is the ability of machine learning systems to process large volumes of medical data quickly and efficiently [20]. Traditional diagnostic methods may require significant time and expert involvement, whereas machine learning models can generate predictions within seconds after training. This can help improve clinical workflow and support early diagnosis, particularly in healthcare centers with limited medical resources.

	Model	Scores
0	Random Forest Classifier	96.491228
3	Decision Tree	93.859649
1	Logistic Regression	92.982456
2	KNeighbour Clasifier	92.105263
5	Linear SVC	89.473684
4	SVC	87.719298

Fig. 7. Classifiers accuracy result

Overall, the project demonstrates that the Random Forest classifier is highly effective for breast cancer detection and classification. Its high predictive accuracy, strong generalization capability, and robustness make it a suitable choice for developing intelligent healthcare applications [13]. The findings also confirm that machine learning-based diagnostic systems have the potential to improve medical decision-making, support early cancer detection, and enhance patient care. With further improvements and integration into real-world healthcare environments, such systems can contribute significantly to reducing breast cancer-related mortality and improving treatment outcomes in the future.

8. Applications

- Early detection of breast cancer
- Medical decision support systems
- Hospital diagnostic automation
- Healthcare data analysis and research

9. Hardware and Software Requirements

Hardware Requirements:

- Intel i3 or higher processor
- Minimum 4 GB RAM
- 500 GB Hard Disk

Software Requirements:

- Windows 10 or higher
- Python
- Jupyter Notebook / VS Code
- NumPy, Pandas, Matplotlib, Seaborn
- Scikit-learn
- Flask

10. Conclusion and Future work

This project examined the use of machine learning techniques for predicting breast cancer from structured medical data. Different classification models were trained and evaluated after applying preprocessing, feature scaling, and feature selection methods to the dataset. Performance was analyzed using accuracy, precision, recall, and F1-score to compare the effectiveness of each algorithm.

Among all tested models, Random Forest produced the strongest overall performance and delivered the most consistent predictions on the testing dataset. Its ensemble-based structure improved classification reliability and reduced the effect of overfitting when compared with several individual classifiers.

The findings of this work show that machine learning can be applied as a practical support tool in medical prediction tasks. While the developed model is not intended to replace professional diagnosis, it can assist healthcare practitioners by providing faster preliminary classification and additional analytical support.

Future improvements can include training the system on larger and more diverse clinical datasets, integrating real hospital records, and combining tabular medical information with imaging data such as mammograms. The project can also be extended into a user-friendly application for demonstration or healthcare support environments.

11. References

[1] W. Yue, Z. Wang, H. Chen, A. Payne and X. Liu, "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis," *Designs,* vol. 2, no. 2, p. 13, 2018.

[2] A. R. Vaka, B. Soni and S. Reddy, "Breast Cancer Detection by Leveraging Machine Learning," *ICT Express,* vol. 6, no. 4, pp. 320-324, 2020.

[3] A. Asri, H. Mousannif, H. Al Moatassime and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," *Procedia Computer Science,* vol. 83, no. 1, pp. 1064-1069, 2016.

[4] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis and D. I. Fotiadis, "Machine Learning Applications in Cancer Prognosis and Prediction," *Computational and Structural Biotechnology Journal,* vol. 13, no. 1, pp. 8-17, 2015.

[5] H. M. Abdel-Zaher and A. M. Eldeib, "Breast Cancer Classification Using Deep Belief Networks," *Expert Systems with Applications,* vol. 46, no. 1, pp. 139-144, 2016.

[6] M. Fatima and M. Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic," *Journal of Intelligent Learning Systems and Applications,* vol. 9, no. 1, pp. 1-16, 2017.

[7] R. D. Cruz and A. Wishart, "Applications of Machine Learning in Cancer Prediction," *Artificial Intelligence in Medicine,* vol. 65, no. 2, pp. 123-136, 2019.

[8] L. Wang and Y. Zhang, "Medical Data Mining Techniques for Disease Prediction," *Healthcare Analytics Journal,* vol. 4, no. 1, pp. 88-96, 2021.

[9] S. Sharma, A. Aggarwal and T. Choudhury, "Breast Cancer Detection Using Machine Learning Algorithms," *International Journal of Computational Science,* vol. 5, no. 3, pp. 114-118, 2018.

[10] S. A. Mohammed, S. Darrab, S. A. Noaman and G. Saake, "Analysis of Breast Cancer Detection Using Different Machine Learning Techniques," *Data Mining and Big Data,* vol. 12, no. 2, pp. 108-117, 2020.

[11] T. M. Mitchell, Machine Learning, New York: McGraw-Hill, 1997.

[12] C. M. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2006.

[13] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, California: O'Reilly Media, 2019.

[14] K. P. Murphy, Machine Learning: A Probabilistic Perspective, Cambridge: MIT Press, 2012.

[15] E. Alpaydin, Introduction to Machine Learning, Cambridge: MIT Press, 2020.

[16] K. Bhuvaneswari and P. Aruna, "Breast Cancer Prediction Using Data Mining Techniques," *International Journal of Computer Applications,* vol. 97, no. 13, pp. 1-5, 2014.

[17] D. Lavanya and K. U. Rani, "Analysis of Feature Selection with Classification Algorithms for Breast Cancer Detection," *International Journal of Computer Science and Information Technologies,* vol. 3, no. 5, pp. 756-763, 2012.

[18] S. Sharma, A. Aggarwal and T. Choudhury, "Breast Cancer Detection Using Machine Learning Algorithms," in *International Conference on Computational Techniques, Electronics and Mechanical Systems*, 2018.

[19] R. Kumar and S. Gupta, "Classification of Breast Cancer Dataset Using Supervised Learning Techniques," in *International Conference on Data Science and Engineering*, 2019.

[20] N. Patel and V. Shah, "Comparative Study of Machine Learning Classifiers for Breast Cancer Prediction," in *IEEE International Conference on Advances in Computing,Communication and Control*, 2021.