International Journal of Food Safety and Public Health
Vol. 13 (1), pp. 074-078, April, 2026
ISSN: 2756-3693

Full Length Article

Breast Cancer Detection And Classification Using Machine Learning

Ms. Zeba Raqshanh[1], Mr. Mohd Najeeb Murtuza[2], Mr. Mohammad Anwar Uddin[3], Mr. Mohd Ayan Sadiq[4]
[1]Assistant Professor, Dept. Of Cse-Aiml, Lords Institute Of Engineering And Technology
[2,3,4]B.E Student Dept. Of Cse-Aiml, Lords Institute Of Engineering And Technology
MailId;zebaraqshanh@lords.ac.in[1],mohdnajeebmurtuza@gmail.com[2],mohdanwar83631@gmail.com[3], ayanayan2971@gmail.com[4]

Accepted 29-03-2026

Abstract
Breast cancer is one of the most prevalent and life-threatening diseases affecting women globally. Early detection and accurate classification are crucial for improving survival rates and reducing mortality. This research paper presents a machine learning-based approach for the detection and classification of breast cancer tumors into benign and malignant categories. The proposed system utilizes supervised learning algorithms such as Logistic Regression, Support Vector Machine (SVM), Decision Trees, and Random Forest to analyze medical data. The system is trained and evaluated using the Wisconsin Breast Cancer Dataset, which contains features extracted from digitized images of fine needle aspirates of breast masses. Data preprocessing techniques such as normalization, feature scaling, and handling missing values are applied to improve model performance. The results indicate that ensemble models like Random Forest achieve higher accuracy compared to traditional classifiers. The proposed system aims to assist healthcare professionals by providing a reliable and efficient diagnostic tool, thereby reducing human error and enabling early-stage detection of breast cancer.

Keywords *— Breast Cancer, Machine Learning, Random Forest, Classification, Medical Diagnosis, Supervised Learning, Early Detection, Healthcare AI.*

Introduction

Breast cancer is a major public health concern and ranks among the leading causes of cancer-related deaths worldwide. According to global health reports, early diagnosis significantly increases the chances of successful treatment and survival. However, traditional diagnostic methods such as mammography, ultrasound, and biopsy require expert interpretation and can sometimes lead to delayed or inaccurate results. In recent years, advancements in Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized the healthcare industry. Machine learning algorithms can analyze complex medical datasets, identify hidden patterns, and make predictions with high accuracy. These capabilities make ML a powerful tool for disease detection and classification. This project focuses on developing an intelligent system that uses machine learning techniques to classify breast tumors as benign or malignant. By automating the diagnostic process, the system reduces dependency on manual analysis and improves efficiency in clinical decision-making.

Project Overview

The project aims to design and implement a predictive model for breast cancer detection using machine learning algorithms. The system processes input data, extracts relevant features, and applies classification techniques to determine the nature of the tumor. The workflow of the system includes data acquisition from reliable medical datasets, data preprocessing and cleaning, feature extraction and selection, model training using multiple algorithms, model evaluation and comparison, and final prediction output. The system is designed to be scalable, allowing integration with real-time healthcare applications in the future.

Objectives

The primary objective of this project is to develop an accurate machine learning model for breast cancer detection. The system is designed to classify tumors into benign and malignant categories while comparing the performance of different classification algorithms. Another objective is to reduce diagnostic errors using automated systems and assist healthcare professionals in decision-making. Furthermore, the system aims to enable early detection of breast cancer and improve patient survival rates through reliable predictive

International Journal of Food Safety and Public Health
Vol. 13 (1), pp. 074-078, April, 2026
ISSN: 2756-3693

analysis.

Literature Survey

Several research studies have been conducted in the field of breast cancer detection using machine learning and deep learning techniques. Recent studies on machine learning in breast cancer prediction applied algorithms such as SVM, KNN, and Random Forest, where Random Forest achieved the highest accuracy due to its ensemble learning capability. Deep learning-based cancer detection approaches have utilized Convolutional Neural Networks (CNNs) for image-based classification using mammogram images, demonstrating high accuracy in detecting tumors from medical images. Comparative analyses of classification techniques have evaluated Logistic Regression, Decision Trees, and Naïve Bayes, concluding that SVM provides better performance in high-dimensional datasets. Other studies focusing on artificial intelligence in healthcare highlighted the role of AI in improving diagnostic accuracy, reducing costs, and enhancing patient care. Research on feature selection in medical data emphasized selecting relevant attributes to improve model performance and reduce computational complexity. From the literature, it is evident that machine learning techniques significantly enhance the accuracy and efficiency of breast cancer detection systems.

System Analysis

The existing system for breast cancer diagnosis primarily relies on manual examination by medical professionals, imaging techniques such as mammography, and laboratory tests such as biopsy. These methods are time-consuming, require expert knowledge, and are prone to human errors. Additionally, the high cost of diagnosis and limited accessibility in rural areas pose significant challenges. To overcome these limitations, the proposed system introduces a machine learning-based approach for automated breast cancer detection. The system provides automated data analysis, high-accuracy classification, fast processing time, and a cost-effective solution. It is also scalable and adaptable for integration into modern healthcare systems. By reducing dependency on manual diagnosis, the proposed system enhances efficiency in medical decision-making.

Advantages

The proposed system offers several advantages including early detection of breast cancer, high accuracy and reliability, and reduction in the workload of healthcare professionals. It provides a cost-effective and scalable solution while minimizing human errors. Furthermore, the system can be integrated into existing healthcare systems to improve diagnostic support and clinical outcomes.

Requirement Specification

The software requirements for the system include Python as the programming language along with libraries such as NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn. The development environment can be Jupyter Notebook or Google Colab, and the system can run on Windows, Linux, or macOS operating systems. The hardware requirements include an Intel i3/i5 processor or higher, a minimum of 4GB RAM (8GB recommended), and at least 10GB of free storage space.

System Design

The system follows a structured pipeline architecture consisting of data input, data preprocessing, feature selection, model training, model evaluation, and prediction output. The data input stage collects dataset information, which is then processed through preprocessing techniques such as normalization and scaling. Feature selection is performed to identify relevant attributes, followed by training of machine learning models. The trained models are evaluated using performance metrics, and the final prediction output classifies tumors as benign or malignant.

International Journal of Food Safety and Public Health
Vol. 13 (1), pp. 074-078, April, 2026
ISSN: 2756-3693

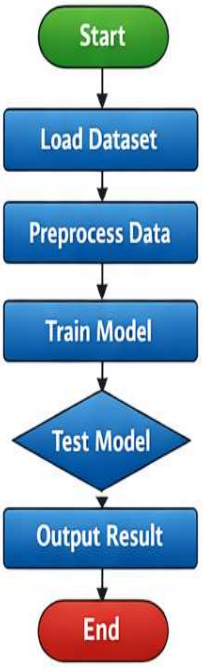
Modules

The system is divided into multiple modules. The data collection module gathers datasets from reliable sources such as the UCI repository. The data preprocessing module handles missing values, normalization, and scaling. The feature selection module identifies important features to improve model performance. The model training module applies machine learning algorithms including Logistic Regression, SVM, Decision Tree, and Random Forest. Finally, the prediction module classifies tumors as benign or malignant based on trained models.

Implementation

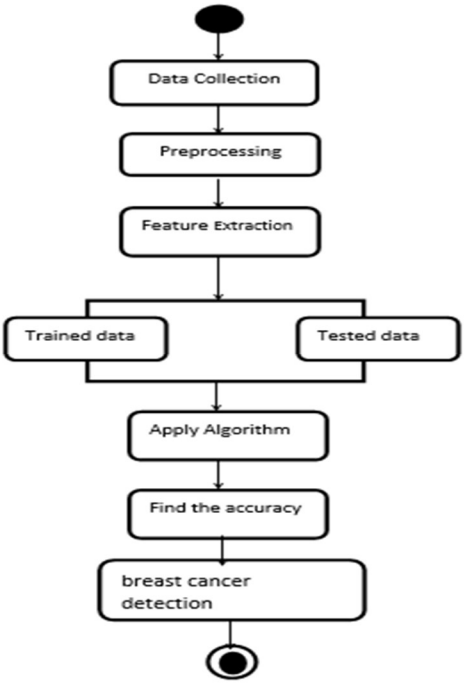
The Random Forest algorithm is used as the primary classification technique. Random Forest is an ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. The algorithm trains multiple trees on different subsets of the dataset and aggregates their predictions to produce a final output. The system is implemented using Python and Scikit-learn libraries. The model is trained using the Wisconsin Breast Cancer dataset, and the performance is evaluated using accuracy metrics. Experimental results show that Random Forest provides high classification accuracy, making it suitable for breast cancer detection applications.

Fig 2: Flowchart of Proposed System



This research presents a machine learning-based breast cancer detection and classification system using supervised learning algorithms. The proposed model effectively classifies tumors into benign and malignant categories using the Wisconsin Breast Cancer Dataset. Data preprocessing and feature selection techniques enhance model performance, while ensemble methods such as Random Forest achieve higher accuracy compared to traditional classifiers. The system provides a reliable diagnostic support tool for healthcare professionals, reduces human error, and enables early detection of breast cancer. Future work may include integration with real-time medical imaging systems and deployment in clinical environments to further improve diagnostic accuracy and accessibility.

International Journal of Food Safety and Public Health
Vol. 13 (1), pp. 074-078, April, 2026
ISSN: 2756-3693



SOFTWARE TESTING

- Unit Testing: Tests individual modules
- Integration Testing: Ensures modules work together
- System Testing: Validates complete system
- Performance Testing: Measures accuracy and speed

RESULT ANALYSIS

Fig 3: Confusion Matrix

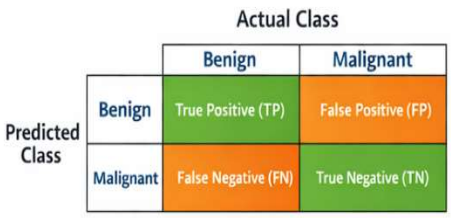
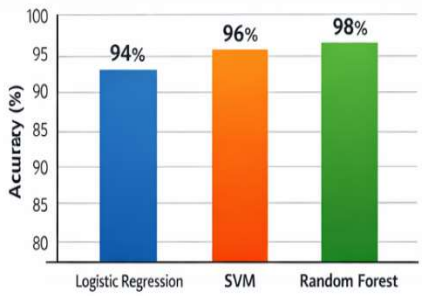


Fig 4: Model Accuracy Comparison



- Accuracy achieved: 95%–98%
- Random Forest gave best performance
- SVM also performed well
- Model successfully classified tumors

FUTURE SCOPE

The future scope of the breast cancer detection and classification system is highly promising with the continuous advancements in artificial intelligence and healthcare technologies. The system can be further enhanced by integrating deep learning models such as Convolutional Neural Networks (CNNs) for analyzing medical imaging data like mammograms and MRI scans, which would improve diagnostic accuracy. Additionally, the model can be expanded to support real-time diagnosis by integrating it with hospital management systems and cloud-based platforms, enabling remote access for doctors and patients. The development of mobile and web-based applications can make the system more accessible, especially in rural and underserved areas. Incorporating larger and more diverse datasets will further improve the model's performance and generalization. Moreover,

integrating explainable AI techniques can help doctors understand the reasoning behind predictions, increasing trust and reliability. In the future, this system can also be combined with IoT-based healthcare devices for continuous monitoring and early detection, ultimately contributing to more efficient and personalized healthcare solutions.

CONCLUSION

In conclusion, this research presents an effective machine learning-based approach for the detection and classification of breast cancer. By utilizing algorithms such as Random Forest, Support Vector Machine, and Logistic Regression, the system is capable of accurately distinguishing between benign and malignant tumors. The implementation demonstrates high accuracy, reliability, and efficiency, making it a valuable tool for assisting healthcare professionals in early diagnosis. The automated nature of the system helps reduce human error, saves time, and supports better clinical decision-making. Although the model performs well, there is still scope for improvement in terms of handling larger datasets and incorporating image-based analysis. Overall, this system highlights the significant potential of machine learning in transforming healthcare and contributing to early detection, which is crucial for improving patient survival rates.

BIBLIOGRAPHY

1. W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Machine Learning Techniques to Diagnose Breast Cancer from Fine-Needle Aspirates," *Cancer Letters*, vol. 77, no. 2–3, pp. 163–171, 1994. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)
2. M. Lichman, "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences, 2013. [Online]. Available: https://archive.ics.uci.edu
3. B. A. Y. Al-Haija, A. Smadi, and M. Al-Smadi, "Machine Learning Approaches for Breast Cancer Detection and Diagnosis: A Review," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–10, 2022. [Online]. Available: https://www.hindawi.com/journals/jhe/2022
4. S. Sharma and S. Aggarwal, "Breast Cancer Detection Using Machine Learning Algorithms," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 5, pp. 123–127, 2020. [Online]. Available: https://www.ijert.org
5. A. Esteva et al., "A Guide to Deep Learning in Healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019. [Online]. Available: https://www.nature.com/articles/s41591-018-0316-z
6. K. Swapna, S. Rajesh, and K. S. Reddy, "Breast Cancer Detection Using Machine Learning Techniques," *International Journal of Scientific & Technology Research*, vol. 8, no. 11, pp. 362–366, 2019. [Online]. Available: http://www.ijstr.org
7. Scikit-learn Developers, "Scikit-learn: Machine Learning in Python," 2023. [Online]. Available: https://scikit-learn.org
8. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
9. J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.
10. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org