ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

Breast cancer detection using machine learning for Early Diagnosis

N.Srinivasan(Guide)A.Haseena[1], D.sanjeevani[2], E.Harika[3],K.Vardhini[4], K.Havila[5]
Chaitanya Bharathi Institute of Technology, Proddatur, A.P, India.

aravahaseena@gmail.com[1],dudekulasanjeevani@gmail.com[2], harikaelluri9877@gmail.com[3], kalamakuntlavardhini@gmail.com[4], kanugahavila103@gmail.com[5]

ABSTARCT

Breast cancer remains a significant global health concern, emphasizing the need for early and accurate diagnosis to improve patient outcomes. This project proposes an intelligent system aimed at automating breast cancer identification using ultrasound imaging and machine learning algorithm. The system begins with the collection of diverse datasets comprising ultrasound images categorizing breast conditions into normal, benign, and malignant cases. Employing robust pre-processing techniques such as data augmentation, cropping, and resizing standardizes the images for subsequent analysis. Transfer learning models-VGG-19, Mobile Net, and Inception Net -are utilized to extract meaningful features from the ultrasound images, enhancing the efficiency of model training and evaluation. The dataset under goes rigorous evaluation using 10foldcrossvalidation, ensuring robust model performance. Machine learning models including Convolutional Neural Network(CNN) is employed for breast cancer classification, fine-tuned to distinguish between benign and malignant tumours.

Keywords: Breast Cancer Detection, Machine Learning, Early Diagnosis, Convolutional Neural Network (CNN), Ultrasound Imaging, Transfer Learning, Tumour Classification.

I. INTRODUCTION

Breast cancer is one of the diseases that causes significant loss of life each year. It is among the most common cancers affecting women worldwide, accounting for approximately 25% of all cancer cases and affecting 2.1 million people in 2015. Early diagnosis significantly increases the likelihood of survival.

Machine learning is a method of training computers with data so that they can make decisions under similar conditions. Its applications can be observed in various domains, including medicine, networking, object identification, and security. Machine learning approaches can be single-model or hybrid, such as those using Convolutional Neural Networks (CNNs).

Breast cancer has been ranked as the most common cancer among Indian women and accounts for 27% of all cancers in women.

About 1 in 28 women are likely to develop breast cancer during their lifetime. In urban areas, the incidence is approximately 1 in 22, compared with 1 in 60 in rural areas. In India, the number of breast cancer cases

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

is rising. According to a 2018 report on breast cancer statistics, there were 1,62,468 newly registered cases and 87,090 reported deaths.

II. EXISTING SYSTEM

The existing system for breast cancer diagnosis using ultrasound imaging relies heavily on manual interpretation by radiologists, which can lead to variability and subjectivity in diagnostic outcomes. Currently, there is a lack of standardized automated systems that can efficiently and accurately classify breast tissue into normal, benign, and malignant categories based on ultrasound images. Most diagnostic processes involve visual inspection and subjective judgment, which may result in errors or delays in diagnosis. Moreover, existing automated systems often face challenges in handling the variability of ultrasound images, including differences in image quality, resolution, and patient demographics. These systems typically lack the robustness to generalize across diverse datasets and may not incorporate advanced machine learning techniques for feature extraction and classification. Therefore, there is a critical need for an advanced automated system, as proposed in this project, that integrates deep learning, transfer learning, and ensemble methods to enhance diagnostic accuracy and efficiency in breast cancer detection using ultrasound imaging.

III. ADVANTAGES OF THE EXISTING SYSTEM

Early Detection and Improved Accuracy: Machine learning algorithms can detect cancerous patterns at an early stage, improving survival rates. They can reduce false positives and false negatives compared with traditional methods.

Faster Diagnosis: Machine learning models can analyze mammograms, ultrasound images, and histopathological images faster than human radiologists. They help doctors make quicker decisions, reducing patient anxiety and waiting times.

Automation and Reduced Human Error: Automated analysis reduces dependence on manual interpretation, minimizing errors due to fatigue or bias and ensuring consistent and repeatable results.

Cost-Effective Solution: Improved initial diagnostic accuracy can lower the cost of repeated tests and reduce the need for expensive biopsies in non-cancerous cases.

Personalized Treatment Plans: AI models can predict cancer progression and help doctors customize treatment plans. They enable precision medicine by analyzing individual patient data.

IV. DISADVANTAGES OF THE EXISTING SYSTEM

Data-Related Issues: Medical datasets often have class imbalance, with fewer malignant cases than benign ones, which may lead to biased predictions. Patient data is highly sensitive, and improper handling can lead to ethical and legal issues. High-quality annotated medical images or histopathological data are scarce, affecting model accuracy.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

Model Limitations: Many machine learning models perform well on training data but fail to generalize effectively to new cases. Deep learning models, especially CNNs, often function as "black boxes," making their decisions difficult for doctors to interpret. Poor-quality mammograms or histopathological images can mislead the model and lead to incorrect classifications.

Ethical and Legal Concerns: If models are trained on biased datasets (e.g., more data from a particular ethnicity or age group), they may not perform well on diverse populations.

V. PROPOSED SYSTEM

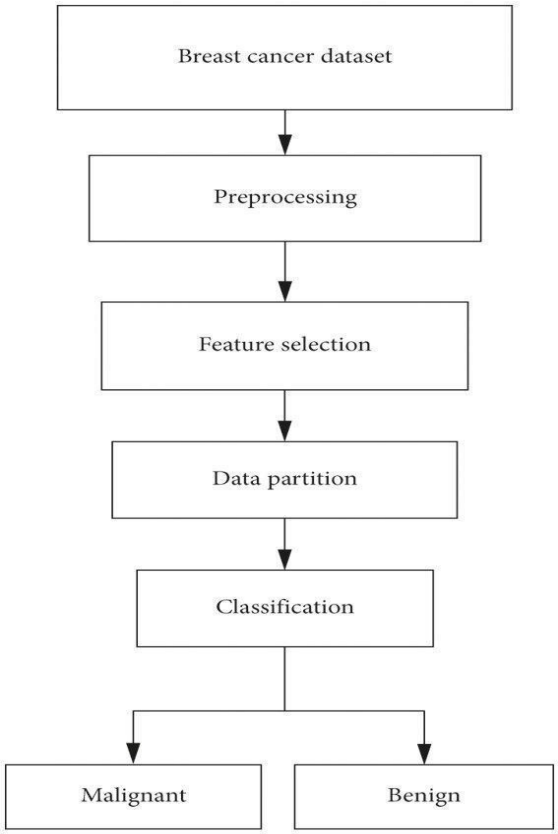
Breast Cancer Dataset: The process starts with a dataset containing medical records or biopsy data of breast cancer
patients.
Preprocessing: This step involves cleaning and preparing the dataset by handling missing values, normalizing data, and possibly performing transformations to enhance model performance.
Feature Selection: Relevant features, such as tumour size, shape, and texture, are chosen to improve accuracy and reduce computational complexity.
Data Partition: The dataset is split into training and testing sets to train the model and evaluate its performance.
Classification: A machine learning algorithm (e.g., a Convolutional Neural Network) is applied to classify the data into two categories.
Malignant: Indicates cancerous tumours that require medical intervention.
Benign: Indicates non-cancerous tumours that are less harmful.

VI. ADVANTAGES OF THE PROPOSED SYSTEM

Systematic Data Processing; Improved Accuracy; Efficient Data Handling; Effective Model Training; Better Decision-Making;
Reduced Computational Complexity; Scalability.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

VII. DATA FLOW DIAGRAM



MODULES

User Management Module

Image Acquisition and Preprocessing Module; Feature Extraction Module

Classification and Detection Module

Result Analysis and Report Generation Module; Patient Record and History Management Module; Integration and Deployment Module.

VIII. CONCLUSION

Breast cancer detection using machine learning has achieved an accuracy of up to 90.4%. The use of machine learning makes the output effective and faster while reducing complexity. A combination of classifiers and algorithms, such as decision tree, random algorithm, and logistic regression, was used to develop an accurate and efficient model. The results show decision tree classifier predictions and the actual classification of patients, with ones representing malignant (cancerous) cases and zeros representing benign (non-cancerous) cases. This model can predict a greater number of correct values than incorrect values. By detecting breast cancer at an early stage, patients may have better treatment options and avoid some painful surgeries. The overall computational time for preprocessing was approximately 3.5 seconds, while the processing stage took around 5 seconds for the dataset considered. These times may vary depending on the size of the selected dataset.

REFERENCES

[1] Loomans-Kropp H.A., Umar A. Increasing Incidence of Colorectal Cancer in Young Adults. J. Cancer Epidemiol. 2019
[2] Eroglu Y., Yildirim M., Çinar A. Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. Comput. Biol. Med. 2021
[3] Karthik R., Menaka R., Kathiresan G., Anirudh M., Nagharjun M. Gaussian Dropout Based Stacked Ensemble CNN for Classification of Breast Tumor in Ultrasound Images. IRBM. 2021.
[4] Lakhani S.R., Audretsch W., Cleton-Jensen A.M., Cutuli B., Ellis I., Eusebi V., Greco M., Houslton R.S., Kuhl C.K., Kurtz J., et al. The management of lobular carcinoma in situ (LCIS).
[5] Is LCIS the same as ductal carcinoma in situ (DCIS)? Eur. J. Cancer. 2006
[6] Visser L.L., Groen E.J., Van Leeuwen F.E., Lips E.H., Schmidt M.K., Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses Prognostic Factors for DCIS: What Do We Really Know? Cancer Epidemiol. Biomark. Prev. 2019
[7] Din N.M.U., Dar R.A., Rasool M., Assad A. Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. Comput. Biol. Med. 2022
[8] Khalifa F., Soliman A., El-Baz A., El-Ghar M.A., El-Diasty T., Gimel'Farb G., Ouseph R., Dwyer A.C. Models and methods for analyzing DCE-MRI: A review. Med. Phys. 2014
[9] Alahmari F., Naim A., Alqahtani H. IoT-Enabled Convolutional Neural Networks: Techniques and Applications. River Publishers; Aalborg, Denmark: 2023. E-Learning Modeling Technique and Convolution Neural Networks in Online Education.
[10] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks.
[11] UK Statistics Authority. Cancer statistics registrations: registrations of cancer diagnosed in 2004, England. London: UK Statistics Authority; 2007.