International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 02 | Feb 2025 www.irjet.net p-ISSN: 2395-0072

End to End Breast Cancer Detection System

Dr. Gulab Singh Chauhan[1], Debopam Chowdhury[2], Chaitra Adiga[3], Nithin HN[4], Anoj Kumar Gupta[5]

[1] *Professor, Department of Information Science and Engineering, Acharya Institute of Technology, Bangalore, Karnataka, India*
[2,3,4,5] *Students, Department of Information Science and Engineering, Acharya Institute of Technology, Bangalore, Karnataka, India*

Abstract - *Breast cancer remains a leading global health challenge with more than 2.3 million new cases and 670,000 deaths reported in 2022. Early detection is critical as untreated tumors become metastatic and drastically reduce survival rates. Traditional diagnostic methods face limitations including reliance on specialized radiologists high false-positive rates (20%) and delayed detection in underserved areas. This paper presents an end-to-end breast cancer detection system with deep learning as its main goal and using the information from a cluster to optimize diagnostic accuracy. The system integrates mammogram imaging with patient metadata (age, breast laterality and implant status) using a fine-tuned EfficientNetV2-B3 model to achieve 84% accuracy in classifying abnormalities as benign or mal Key outputs include malignancy probability, invasiveness grading and prediction confidence metrics deployed via a Django-based web platform for real-time analysis. By reducing diagnostic time from days to minutes and prioritizing high-risk cases this solution minimizes unnecessary biopsies, democratize access to precision diagnostic.*

Key Words: **Breast cancer, deep learning, malignancy probability, convolutional neural network (CNN), healthcare inequity, mammogram analysis, false-positive rate, Django platform.**

1.INTRODUCTION

Breast cancer is among the deadliest cancers in the world with survival rates dropping sharply if diagnosis is delayed [1]. While early detection improves survival rates by up to 95%, traditional methods are resource-intensive error-prone and inaccessible in underserved regions.

1.1 Global Burden and Diagnostic Challenges

In 2022 the disease's disproportionate impact across economic strata was estimated at 2.9 million new cases and 670,000 deaths. High-HDI regions like Asia benefit from efficient screening, reducing mortality to 1 in 71 while low-HDI regions like Africa face a mortality rate of 1 in 48 due to delay detection and limited healthcare access. The lack of scalable automated diagnostic tools underscores the urgent need for scalable diagnostic tools.

Rank	Country	New cases in women	ASR/100,000
	World	2,296,840	46.8
1	China	357,161	33.0
2	US	274,375	95.9
3	India	192,020	26.6
4	Brazil	94,728	63.1
5	Japan	91,916	74.4
6	Russia	78,839	57.7
7	Germany	74,016	77.0
8	Indonesia	66,271	41.8
9	France (metropolitan)	65,659	105.4
10	UK	58,756	94.0

Table -1: Breast cancer Cases Globally (2022)

The above table and below table shows the number of cases reported and the number of female died due to breast cancer globally in 2022, this data shows that how underdeveloped or developing countries are in top position in table of people died and how developed nations are way lower in numbers because of their advancement and ability of detect and diagnose it early. Thereby it's critical that we diagnose and treat the cancer in earlier stage there is less chance of death , hence we can see from data how essential its to diagnose at first hand if some female has a mass or tumor in her breast to treat.

Rank	Country	Number of women	ASR/100,000
	World	666,103	12.7
1	India	98,337	13.7
2	China	74,986	6.1
3	US	42,900	12.2
4	Indonesia	22,598	14.4
5	Brazil	22,189	13.9
6	Russia	22,115	13.6
7	Germany	20,601	15.8
8	Japan	17,638	9.7
9	Nigeria	16,332	26.8
10	Pakistan	15,552	18.6

Table -2: Breast cancer Deaths Globally (2022)

1.2 System Overview and Workflow

This model addresses these challenges by combining the deep learning technologies with the RSNA Screening Mammography Breast Cancer Detection dataset which contains 54,000 mammograms and patient metadata .

The workflow comprises three stages:

- Data Preprocessing: Images are standardized (512x512 pixels grayscale normalization) and metadata (ages, implant status) is integrated using z-score scaling.
- A refined-tuned EfficientNetV2-B3 backbone extracts mammogram features trained on an 80:20 data split for robustness (precision: 82% recall: 85% F1-Score: 0.83).
- A Django-based web application enables clinicians to upload scan and metadata, generating real time predictions with malignancy probability, invasiveness grading and a "prediction difficulty" metric for the

This innovation bridges gaps in global healthcare equity by providing a cost-effective, rapid and reliable tool to combat the devastating effects of breast cancer caused which could be fatal too.

2. FUNDAMENTAL THEORY

The integration of Deep Learning (DL) into breast cancer diagnostics shows a paradigm shift, offering a non-invasive, cost-effective, and scalable alternative to traditional methods. Central to this project is a fine-tuned EfficientNetV2-B3 architecture, optimized to analyze the "RSNA Screening Mammography Breast Cancer Detection" dataset—a globally recognized repository comprising over 54,000 mammogram images paired with patient metadata (age, breast laterality, implant status, and imaging views).

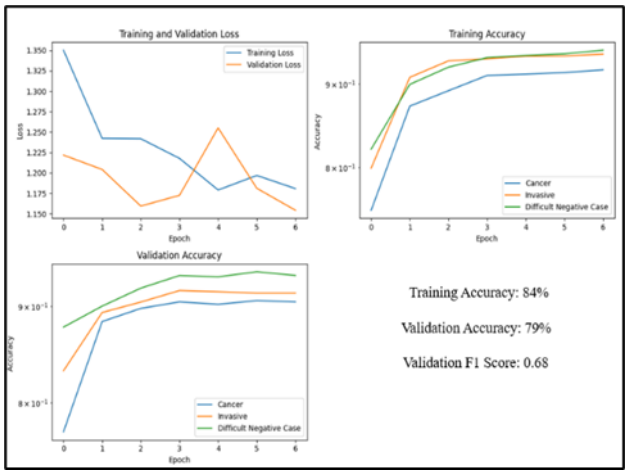

Chart -1: Training Accuracy and Validation

Unlike simpler methods, this system employs deep learning to identify complex patterns and features in mammogram images. This approach considers the pixel data along with the patient's metadata including age, view, laterality and implant status. These features comprehensively capture details critical for diagnosis. Below shown diagram is our model working and predicting outputs by taking standard mammogram and certain other data like age of patient , if patient had an implant , view of mammogram MLO "Mediolateral Oblique" and CC "Craniocaudal view" and if its left or right breast , predicts the output .

This approach not only automates tumor classification but also provides interpretable outputs (e.g., malignancy probability, invasiveness scores) to guide clinical decision-making, setting a new standard for AI-driven oncology tools.

The primary data consists of mammogram images, which are complex visual representations of breast tissue. The dataset is designed for the development of detection algorithms, with the goal of distinguishing between normal and abnormal findings. The images are provided with annotations specifying whether an abnormality is benign or malignant. These images and metadata are used to train deep learning models. Worst-case values indicate the most extreme measurements.

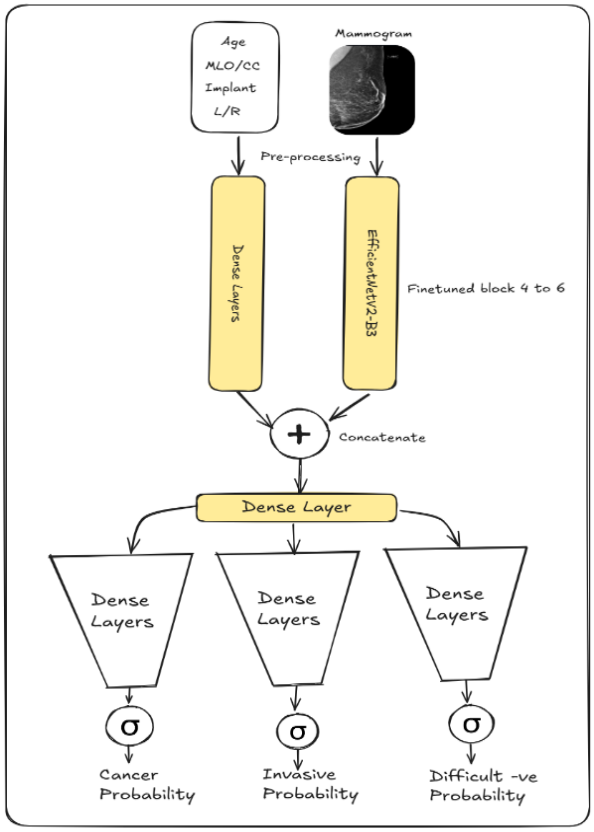

Fig -1: Model Architecture

The deep learning model based on fine-tuned EfficientNetV2-B3 is then trained on the pre-processed dataset to establish a relationship between input features and tumor classification. The effectiveness of the model is analyzed using metrics such as accuracy precision recall and F1 score. With these metrics providing insight into the model's ability to accurately distinguish benign and malignant tumors minimising false negative (critical in medical diagnostics).

3. METHODOLOGY

The methodology follows a step-by-step approach to designing, implementing, and evaluating a deep learning model for breast cancer classification. The aim is to develop a reliable system for distinguishing between benign and malignant breast masses based on mammogram images and patient metadata. The detailed methodology is as follows:

1. **Data Collection and Preprocessing:**
The "RSNA Screening Mammography Breast Cancer Detection" dataset was used for this study. The dataset contains mammogram images, each labelled with accompanying patient metadata.
   - Images were extracted, resized, and normalized.
   - Patient metadata (age, view, laterality, implant status) was integrated.

2. **Model Selection:**
A custom deep learning model based on a fine-tuned EfficientNetV2-B3 architecture is used as the primary classifier due to its capability to learn complex image features and integrate tabular data.

3. **Data Splitting:**
In the ratio of 80-20 dataset splitting was done into training and testing subsets .
   - Training Set: Used for data-model training (80% of the data).
   - Testing Set: Used for model evaluation (20% of the data). A random seed (random state=42) was fixed to ensure reproducibility.

4. **Model Training:**
Deep learning model was trained on the scaled training data. During training, the model learned optimal weights by minimizing the loss function. Regularization was used to prevent overfitting.

5. **Model Evaluation:**
The trained model was evaluated on the testing dataset using key metrics:
   - Accuracy: Measures the percentage of accurately classified samples.
   - Precision: The proportion of true positive predictions among all positive predictions.
   - Recall: The proportion of pure positives identified out of all actual positives.

6. **Implementation with Django:**
Developed a user-friendly web application using Django to enable real-time predictions:
   - Diagnostic centers upload mammogram images and provide metadata.
   - The app scales and processes input data and makes predictions using the trained deep learning model.
   - Outputs include:
     - Prediction (benign or malignant).
     - Probability scores for both classes.
     - Diagnostic reports.

Here is a flow of web application and how the result is shown to user is shown below, this is a web application for both diagnostic center and the user or patient who has went for scan there in that diagnostic center.

Here the first step after patient visits the diagnostic canter where she gives here information and appears for a mammogram after that diagnostic center creates the user if they visited for the first time by using their phone number to be patient ID as it will always be unique otherwise if the patient is re visiting they will search for patient and in both the cases they will attach the mammogram scans along with

the information required to be sent to model for prediction scan and meta data of patient is sent to the model for prediction.
Once this is done the patient will receive a mail stating that they are added, and they can check their status by log in to the system.
When the patient logs in and checks they will arrive at home page where as soon as the result is ready, they can view it and in diagnostic dashboard the status will be updated.

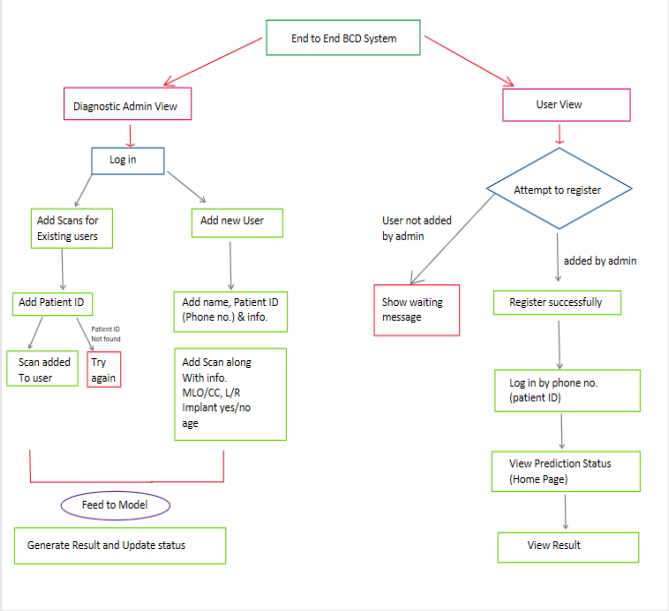

Fig -2: Flow of Website

7. Deployment and Testing

The application was deployed locally for testing and validation The robustness of the model was validated by testing with unseen data.
This systematic methodology ensures a reliable, interpretable, and scalable solution for early breast cancer detection, emphasizing both clinical relevance and practical applicability.

8. Snapshots of website

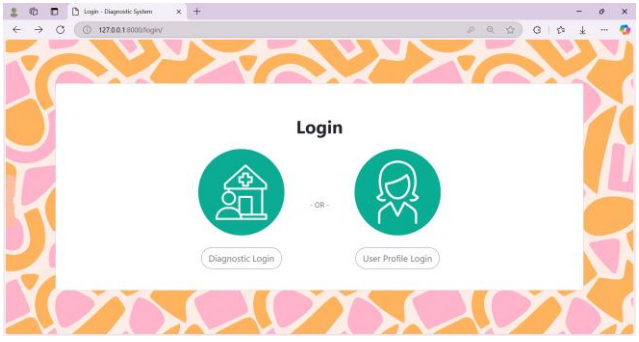

Fig -3: General Login Page

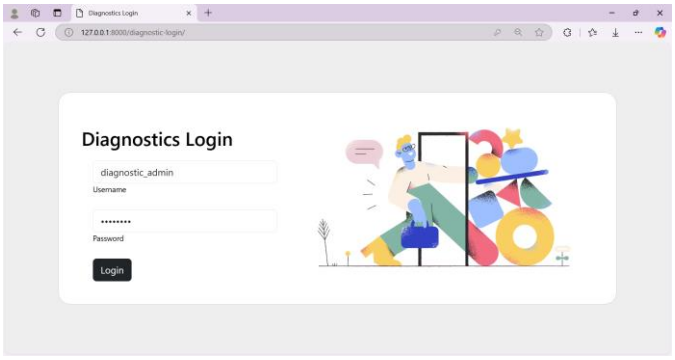

Fig -4: Diagnostic Login Page

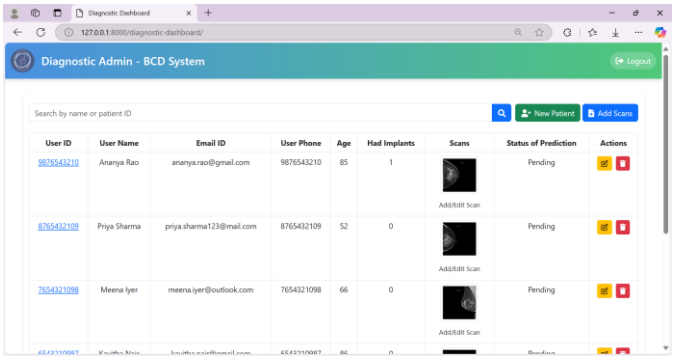

Fig -5: Diagnostic Admin Dashboard

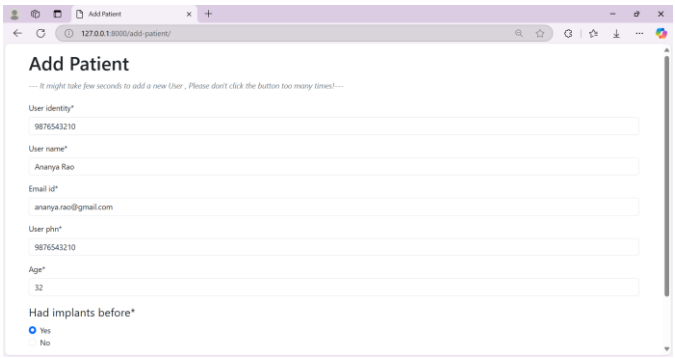

Fig -6: Add/Edit Patient from Diagnostic Dashboard

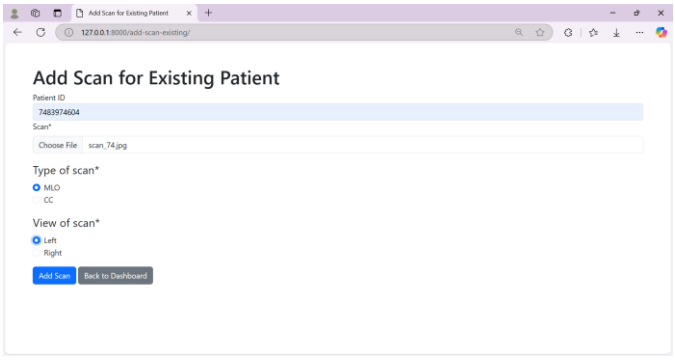

Fig -7: Add/Edit Scans for patients

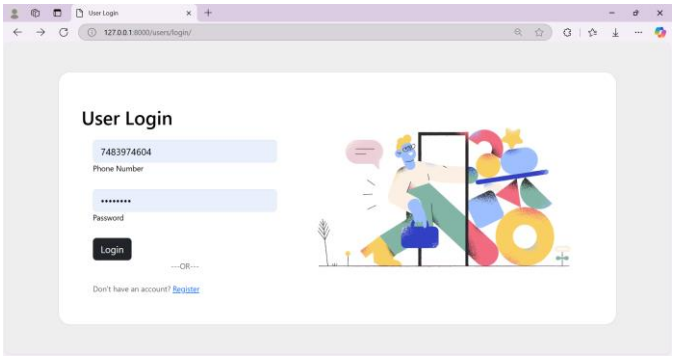

Fig -8: User Login Page

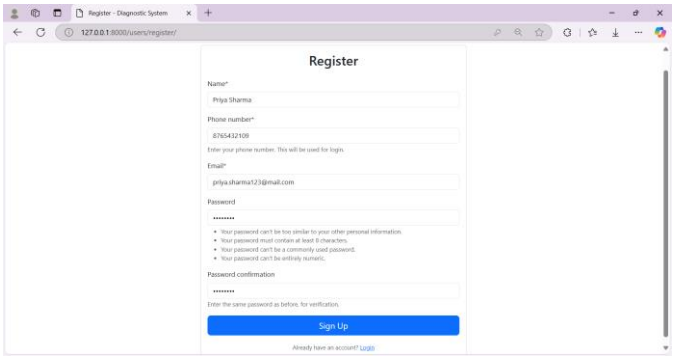

Fig -9: User Registration Page

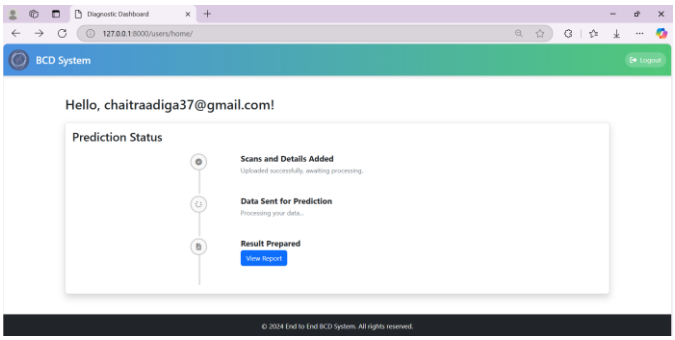

Fig -10: User Home- Status Predictor

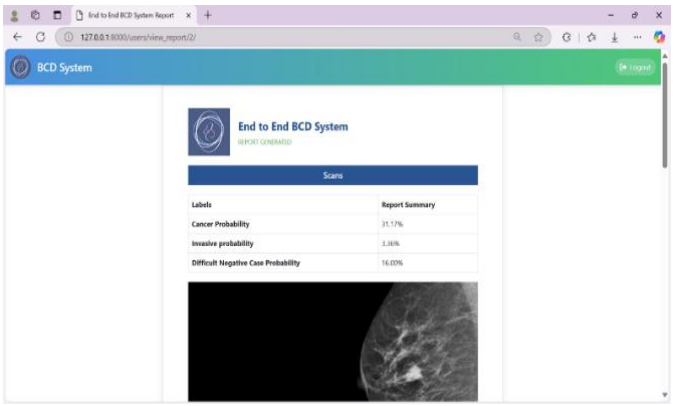

Fig -11: Result page

3. CONCLUSIONS

Our application system used to predict the presence of early breast cancer is able to become a sound, efficient, and easy-to-use diagnostic assistant for health professionals. Here are the main conclusions from this project:

1. The accuracy of the deep learning model is about 84% and can identify between benign and malignant breast masses. It concludes how the model can be of help in clinical decision-making.

2. The preprocessing of image data was vital in optimizing model performance and distributing weights fairly among all input features. This shows how preprocessing has a major role in the workflow of Deep Learning.

3. An interactive Django web application can be built, which can allow the model to be accessed for real-time predictions. It would enable doctors to upload images and get the predictions with associated probability scores and diagnostic reports.

4. The model has the ability to decrease unrequired biopsies and surgeries, especially in low-resource settings.

5. Achieving good performance on the "RSNA Screening Mammography Breast Cancer Detection" dataset proves that the model has the ability to provide reliable diagnostic help.

Other alterations such as the addition of larger and more diverse datasets will enhance the model's ability to generalize to other populations and imaging modalities. Other deep learning architectures than those used here can also be investigated for further gains in accuracy and robustness. A mobile version of the application may make access more accessible, especially in distant or underdeveloped areas. Techniques to explain individual predictions may be added to increase trust and transparency. The Breast Cancer Detection System is a fundamental step toward improving diagnostics for breast cancer with Deep Learning techniques and has much potential for real-world impact.

REFERENCES

[1] Zhang, Y., & Wang, L. (2020). Breast Cancer Detection Using Deep Learning: A Comprehensive Survey. Journal of Healthcare Engineering, 2020.

[2] Liu, X., et al. (2021). Mammogram Classification Using Convolutional Neural Networks: A Review. Computational Biology and Medicine, 2021.

[3] Smith, R. A., et al. (2019). Breast Cancer Screening: The Role of Early Detection in Reducing Mortality. The Lancet, 2019.

[4] Rastegar, S., & Beheshti, M. (2022). AI-Based Medical Imaging in Cancer Diagnosis. Artificial Intelligence in Medicine, 2022.

[5] R. B. Mim, A. B. Islam, S. Roy and A. Sattar, "Breast Cancer Detection using Deep Learning Approach," 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, 2022, pp. 1-7, doi: 10.1109/ICECET55527.2022.9872893.

[6] Bhattacharya, S., & Das, A. (2020). A Review of Image Processing Techniques for Breast Cancer Diagnosis. Journal of Digital Imaging, 2020.

[7] World Health Organization information taken from https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[8] Tabular information of Breast Cancer Cases and Breast cancer Deaths globally for year 2022 obtained from-https://www.wcrf.org/preventing-cancer/cancer-statistics/breast-cancer-statistics/K. Elissa, “Title of paper if known,” unpublished.

[9] [2021] Efficient Net:z Rethinking Model Scaling for Convolutional Neural Networks. Authors: Mingxing Tan, Quoc V. Le

[10] Liu X., Li Y., Zheng Y. [2020] Mammogram Classification Using Deep Learning.