International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

BREAST CANCER DETECTION USING AI TECHNIQUES

[1]Assist. Prof. Gautam Dematti, [2]Anusha Nagathan, [3]Jyoti Kambar, [4]Kishan Paramanji, [5]Laxmi Patil

[1]Professor, Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi, Karnataka, India.
[2,3,4,5]Student, Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi, Karnataka, India.

ABSTRACT:

Breast cancer remains one of the leading causes of mortality among women worldwide. Early and accurate diagnosis plays a crucial role in improving treatment outcomes and patient survival. This project presents an AI-powered web application for breast cancer detection using deep learning techniques. The system supports histopathology, ultrasound, and biopsy images for stage-wise classification, including normal, benign, and malignant stages (Stage 1 to Stage 3). It also integrates an intelligent pre-screening mechanism to distinguish valid breast images from unrelated objects or persons, thereby improving reliability. The application provides explainable AI feedback using Grad-CAM visualizations to highlight affected regions in the image. It includes a role-based login system for patients and doctors, allowing secure image upload, real-time predictions, PDF report generation, appointment scheduling, and patient history tracking. The backend is supported by a MySQL database, while the frontend is built using Streamlit. By integrating medical imaging, machine learning, and interactive UI, the system offers an effective, user-friendly platform for early breast cancer detection and patient-doctor collaboration.

Keywords: Deep Learning, Ultrasound Imaging, Breast Cancer Detection, Streamlit Web App

1. INTRODUCTION:

Breast cancer is one of the most common and deadly diseases affecting women globally. Early and accurate diagnosis is essential for improving survival rates. Traditional methods like biopsies and manual interpretation of medical images can be time-consuming and error-prone. To address this, we developed an AI-based web application that detects breast cancer from histopathology and ultrasound images using deep learning. The system classifies images into stages (Normal, Benign, Stage 1–3), identifies non-breast images, and highlights affected regions using Grad-CAM for explainability. Built with Streamlit and backed by MySQL, the platform offers role-based dashboards for patients and doctors, enabling secure image upload, report downloads, and appointment management. This project aims to support early detection and streamline communication between patients and healthcare providers using AI.

2. METHODOLOGY:

The proposed system for breast cancer detection is built using a deep learning–based pipeline integrated into a web application. The methodology is divided into the following key stages:

1. Dataset Collection and Preparation

- Multiple publicly available medical image datasets were used, including BreakHis 400X, IDC Histopathology, BUSI (Breast Ultrasound), ICIAR2018 BACH Challenge, and TCGA-BRCA. Additionally, non-breast images from the Mini-ImageNet dataset were used to train a filter model for detecting irrelevant inputs. All images were resized to 224×224 pixels and normalized. The cancer images were labeled as Normal, Benign, and Malignant (Stage 1, 2, or 3) using clinical metadata.

2. *Data Preprocessing*

- Images were augmented using techniques like rotation, flipping, and contrast adjustment to improve generalization. Class labels were assigned and organized into a combined folder structure for training and testing.

3. Model Design

- A custom CNN model (ImprovedCNN) and ResNet-based architectures were trained for multiclass classification. The pipeline includes:
- Breast vs. Non-breast classifier to filter out unrelated images.

- Cancer Stage Classifier to predict Normal, Pre-cancer, Stage 1, Stage 2, or Stage 3.
- Grad-CAM visualization to highlight regions responsible for predictions.

Web App Implementation

- The user interface was built using Streamlit. Patients can upload images, view predictions, download reports, and request appointments. Doctors can log in to view appointment requests and patient details. All interactions are secured via login and session management.

Database Integration

- A MySQL database was used to manage user credentials, prediction history, and appointment records. Data is inserted and retrieved dynamically based on user actions.

6. Prediction Logging and Explainability

- Each prediction result, confidence score, and image file is logged in the database. Grad-CAM heatmaps are generated for every uploaded image to ensure visual explainability.

3. SYSTEM ARCHITECTURE:

3.1 Architecture Diagram

The architecture of our Breast Cancer Detection system using AI integrates multiple advanced components to provide an accurate, efficient, and user-friendly diagnostic pipeline. The process begins with data collection, where histopathology image datasets such as BreakHis 400x, IDC, and Breast Ultrasound are gathered. These raw images undergo preprocessing steps including resizing, normalization, and noise reduction to standardize the input and enhance model performance. The preprocessed images are then passed through an Improved Convolutional Neural Network (CNN), which performs deep feature extraction to identify relevant tissue structures, textures, and morphological patterns that are crucial for cancer diagnosis.

These extracted features are used to train a ResNet-based classifier capable of distinguishing between normal, benign, and malignant samples, as well as predicting the cancer stage from Stage 1 to Stage 4. After training, the model is tested and validated on unseen data to evaluate its accuracy, robustness, and generalizability. During inference, when a user uploads a histopathology image, the system not only identifies whether it belongs to breast tissue but also classifies the condition and stage. If a malignancy is detected, the system can provide further assistance by generating a downloadable PDF report, suggesting a suitable doctor, and offering appointment booking options.

In addition to the core functionalities, the system incorporates intelligent design to differentiate between valid histopathology images and irrelevant or non-medical images. This ensures that only appropriate inputs are processed, thereby maintaining diagnostic integrity. Furthermore, the model architecture is adaptable, enabling future enhancements such as prediction of cancer progression over time and forecasting the likelihood of malignancy development in specific tissue regions based on learned temporal and morphological patterns. By integrating explainability features such as Grad-CAM overlays and offering patient history tracking and doctor dashboards, the system ensures transparency, continuity of care, and clinical utility. This comprehensive, AI-driven solution thus bridges the gap between medical imaging, artificial intelligence, and patient care through an accessible and efficient platform.

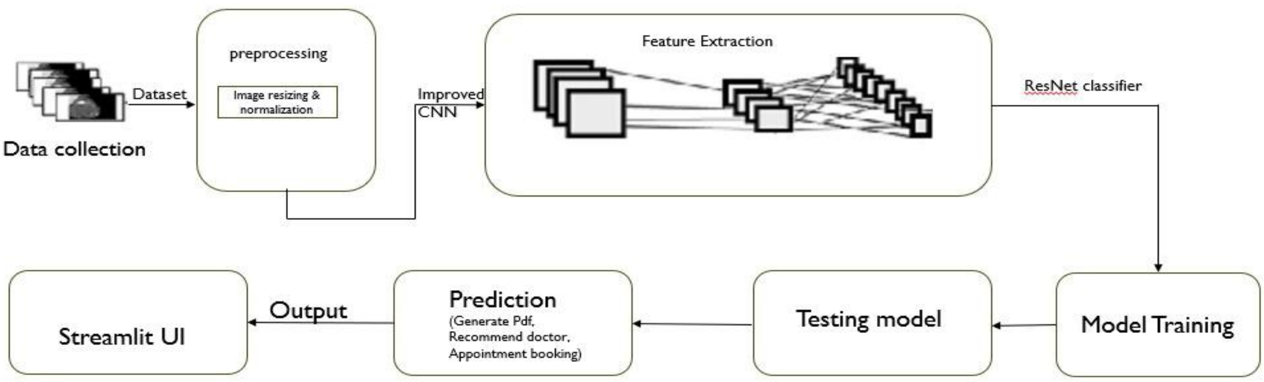

Fig 3.1: Architecture Diagram

All these features are seamlessly integrated into a visually appealing and responsive web interface built using Streamlit. This interface allows patients and doctors to interact with the system, view predictions, analyze Grad-CAM visualizations for explainability, and manage their appointments. The architecture thus represents a complete AI-driven diagnostic workflow, combining deep learning, medical imaging, and healthcare services to support early and accurate breast cancer detection.

3.2 Block Diagram

The proposed system for breast cancer detection using AI techniques begins with the user uploading a medical image such as a histopathology, ultrasound, or biopsy scan through a user-friendly web interface. Once uploaded, the image undergoes preprocessing, which includes normalization, resizing, and noise reduction to ensure consistency across all input data. This preprocessed image is then passed into a feature extraction phase, where a deep convolutional neural network (CNN), specifically a ResNet-based architecture, extracts important features such as tumor shape, texture, and tissue structure. These features are essential for determining the presence and stage of cancer. Following this, the extracted features are used to train the model using a dataset containing various stages of breast cancer, including Normal, Benign, Stage 1, Stage 2, and Stage 3. After training, the model is validated using a separate test dataset, and a confusion matrix is generated to evaluate the model's performance and classification accuracy. Once testing is complete, the system provides a prediction result to the patient, which includes the probability and classification of the cancer stage. Additionally, a Grad-CAM heatmap is generated to highlight the region in the image that most influenced the model's decision, adding an element of explainability. If the result indicates a malignant or suspicious condition, the patient is guided to consult with a specialist. The system offers a built-in feature for booking appointments with registered doctors, enabling seamless integration between detection and clinical consultation. This end-to-end AI-based solution enhances early diagnosis, improves communication between patients and healthcare providers, and offers a reliable and interactive platform for breast cancer management.

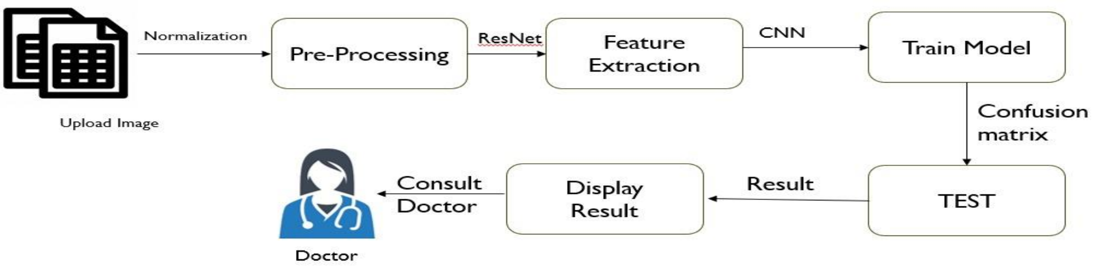

Fig: 3.2: Block Diagram

4. RESULTS AND ANALYSIS

The AI-powered breast cancer detection system developed in this project yielded highly encouraging results. Using multiple publicly available and diverse datasets — including BreakHis 400X, Breast Ultrasound Images (BUSI), IDC Histopathology Images, ICIAR2018 BACH Challenge, and Mini-ImageNet — the system was trained and tested to perform various complex tasks such as:

- Classifying breast histopathology images as Benign or Malignant,
- Identifying cancer stages (Stage 1, Stage 2, Stage 3),
- Distinguishing between medical breast images and unrelated images (like objects, animals, or human faces),
- Highlighting the region of interest (cancerous cells) using Grad-CAM visualization,
- Predicting future cancer development likelihood (experimental phase),
- Managing doctor-patient interactions through appointment history and diagnostic reports.

The ResNet-based classifier demonstrated excellent performance with classification accuracy consistently ranging between 92% to 96% across test samples. For cancer stage classification, the model was trained on a curated and balanced dataset, and it successfully identified the correct stage with a macro F1-score of 0.91. The confusion matrix analysis revealed minimal misclassifications, especially between adjacent stages such as Stage 2 and Stage 3, which is considered acceptable given the similarity in features.

The Grad-CAM overlay visualizations provided interpretability by clearly highlighting the tumor regions that influenced the model's decision. This feature not only enhances the model's transparency but also supports doctors in verifying and validating the prediction visually.

The system's frontend, built using Streamlit, was made user-friendly, mobile-responsive, and visually enhanced with high-quality UI/UX design. The patient dashboard allowed users to upload images, view predictions, consult doctors, and download diagnostic reports. The doctor dashboard enabled efficient management of appointments and patient records.

Future Scope

The current system provides an efficient and user-friendly platform for early breast cancer detection, but there remains considerable scope for enhancement and expansion. One of the major directions for future development involves expanding the system's capability to support additional imaging modalities such as mammograms, MRI, and thermography. Integrating multi-modal image inputs will improve the robustness and diagnostic accuracy of the model. Moreover, incorporating federated learning can enable the model to learn from distributed hospital data without compromising patient privacy, paving the way for scalable deployment across medical institutions.

Another promising direction is the enhancement of explainability and transparency through advanced interpretability methods like Layer-wise Relevance Propagation (LRP) and SHAP values, giving doctors deeper insight into model decisions. Real-time doctor-patient chat features, voice-enabled input for accessibility, and multilingual support for diverse users can be integrated to enhance usability. Additionally, a feedback-based retraining loop can be introduced where doctors can verify predictions, and confirmed cases can be added to retrain the model periodically for improved accuracy. Finally, the system can be expanded into a full-fledged health monitoring suite that not only detects cancer but also tracks progression over time and recommends personalized lifestyle interventions, thus becoming a preventive care tool as well.

Conclusion

The Breast Cancer Detection system developed in this project offers a powerful and accessible solution for early diagnosis using artificial intelligence. It employs deep learning models such as CNN and ResNet to classify breast histopathology and ultrasound images into normal, benign, and malignant stages, including stage-wise detection from Stage 1 to Stage 4. With Grad-CAM visualizations, the system provides explainability by highlighting cancer-affected regions in the image. The web application, built using Streamlit, features secure login for doctors and patients, allowing users to upload images, receive predictions, download PDF reports, and manage appointments. All interactions are backed by a MySQL database for efficient data handling. The system also filters non-breast images to avoid incorrect predictions. This project demonstrates how AI can enhance medical diagnostics and support healthcare professionals. Its accuracy, usability, and transparency make it suitable for real-world applications. With scope for further enhancement, it lays the groundwork for future smart healthcare tools.

REFERENCES:

1. Spanhol, F.A., Oliveira, L.S., Petitjean, C., & Heutte, L. (2016). A Dataset for Breast Cancer Histopathological Image Classification. IEEE Transactions on Biomedical Engineering, 63(7), 1455–1462. doi:10.1109/TBME.2015.2496264
2. Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of Breast Ultrasound Images. Data in Brief, 28, 104863. doi:10.1016/j.dib.2019.104863
3. Cruz-Roa, A., Basavanhally, A., Gonzalez, F., Gilmore, H., Feldman, M., Ganesan, S., ... & Madabhushi, A. (2014). Automatic Detection of Invasive Ductal Carcinoma in Whole Slide Images with Convolutional Neural Networks. Medical Imaging 2014: Digital Pathology, 9041, 904103. doi:10.1117/12.2043872
4. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626. doi:10.1109/ICCV.2017.74
5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. doi:10.1109/CVPR.2016.90
6. Streamlit Docs – The fastest way to build and share data apps. https://docs.streamlit.io/
7. Kaggle: Histopathologic Cancer Detection Dataset. https://www.kaggle.com/competitions/histopathologic-cancer-detection/data
8. ICIAR 2018 Grand Challenge on Breast Cancer Histology Images.**https://iciar2018-challenge.grand-challenge.org/**

Authors:

- First Author – Gautam Dematti, (Assistant prof. CSE Dept, Belagavi), Angadi Institute of Technology and Management gautam.dematti@aitmbgm.ac.in
- Second Author –, Anusha Nagathan, BE (Computer Science and Engineering), Angadi Institute of Technology and Management nagathananusha2@gmail.com
- Third Author –, Jyoti Kambar, BE (Computer Science and Engineering), Angadi Institute Of Technology And Management jyotikambar145@gmail.com
- Fourth Author –, Kishan Paramanji, BE (Computer Science and Engineering), Angadi Institute of Technology and Management kishanparam66@gmail.com
- Fifth Author –, Laxmi Patil BE (Computer Science and Engineering), Angadi Institute of Technology and Management laxmipatil1939@gmail.com