© May 2025 | IJIRT | Volume 11 Issue 12 | ISSN: 2349-6002

OncoSage: AI-Powered Breast Cancer Prediction Based Web Application

Prajakta Deshpande, Prajakta Narhare, Shweta Kenjale, Nikita Singh
Department of Electronics and Communications Engineering, MIT ADT University, Pune

***Abstract*- Breast cancer remains a leading cause of mortality among women worldwide. Early detection through non-invasive imaging can significantly improve prognosis and treatment outcomes. In this paper, we present OncoSage, a robust, deep learning-based web application designed to assist radiologists and clinicians in classifying breast cancer types using ultrasound images. By leveraging transfer learning with the ResNet18 convolutional neural network (CNN) and advanced data augmentation techniques, our system achieves high accuracy in distinguishing between benign, malignant, and normal breast tissues. The model is integrated into a user-friendly Flask-based interface, which provides real-time predictions and confidence scores. We demonstrate the effectiveness of our approach through extensive training on the public BUSI dataset, yielding a classification accuracy of 98%. OncoSage aims to support clinical decision-making and enhance diagnostic efficiency in medical imaging.**

***Index terms*- Breast Cancer, BUSI Dataset, CNN, Machine Learning, Resnet18, Web Application**

I.INTRODUCTION

Breast cancer is one of the most commonly diagnosed cancers and a major contributor to cancer-related deaths globally. Traditional diagnostic methods, including mammography and biopsies, although effective, have limitations in terms of accessibility, invasiveness, and operator dependence. Ultrasound imaging offers a non-invasive and cost-effective alternative, especially in low-resource settings. However, interpretation of ultrasound images is subject to human error and variability.

Deep learning particularly convolutional neural networks(CNNs), has emerged as a powerful tool for image classification tasks in medical diagnostics. This research introduces Oncosage, system designed to automate breast cancer detection from ultrasound images. The proposed platform combines the predictive power of transfer learning with a cloud-enhanced web interface, making it accessible, scalable, and practical for clinical environments.

II.RELATED WORK

Recent advances in computer-aided diagnosis (CAD) have significantly contributed to the early detection and classification of breast cancer using medical imaging. Numerous studies have employed convolutional neural networks (CNNs) to classify ultrasound, mammography, and MRI scans.

The work of Al-Dhabyani et al. introduced the BUSI dataset, providing a foundation for supervised learning in breast ultrasound classification. He et al.'s ResNet architecture has been widely adopted due to its ability to mitigate vanishing gradients using residual connections, making it suitable for deep medical networks.

Researchers have also experimented with other CNN architectures such as VGGNet, DenseNet, and Inception, achieving varying levels of success in tasks like tumor segmentation and classification. Transfer learning has proven effective in scenarios with limited medical data, enabling pretrained models to adapt to specialized domains.

Despite these successes, many existing solutions lack user accessibility and integration into real-time clinical workflows. Few systems provide both high accuracy and a user-friendly interface. OncoSage addresses this gap by not only achieving strong predictive performance but also integrating seamlessly into a web-based interface that supports usability for both clinicians and patients. Furthermore, the inclusion of a conversational assistant distinguishes OncoSage from prior work, offering an interactive and educational dimension to AI-assisted diagnosis.

III.METHODOLOGY

A.Dataset Acquisition

For the development and evaluation of OncoSage, we used the publicly available BUSI (Breast Ultrasound

Images) dataset provided by the University of Cairo. This dataset is widely used for research on breast cancer diagnosis using ultrasound images.

The dataset contains a total of 780 grayscale ultrasound images along with corresponding ground truth annotations. These are categorized into three classes:

1) Benign: 437 images
2) Malignant: 210 images
3) Normal: 133 images

Following are a few samples of ultrasound images from this dataset.

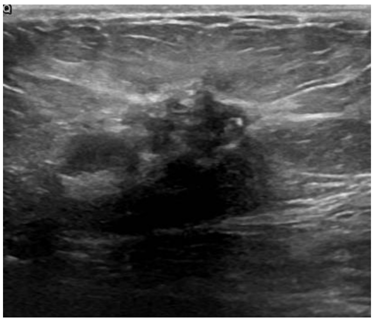

Fig 3.1-Malignant

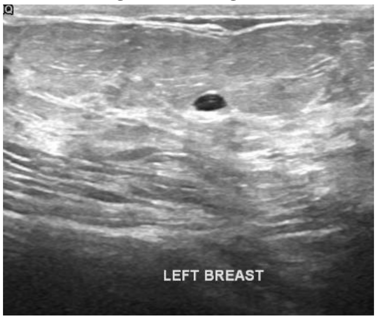

Fig 3.2-Benign

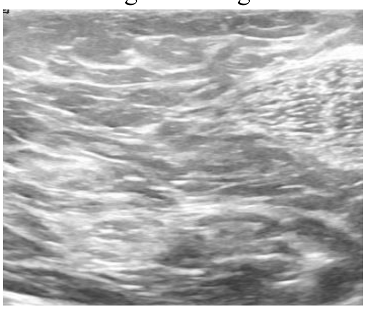

Fig 3.3Normal

Each image varies in size and resolution and includes visible characteristics of different types of breast tissues. The dataset also includes mask images for segmentation, although this project focuses solely on classification. Prior to training, all images were resized to 224×224 pixels and normalized using Albumentations to ensure consistency and improve training performance. Data augmentation was applied to artificially increase dataset variability and reduce overfitting.

B. Project Environment

The OncoSage project was developed in a Python-based deep learning environment using PyTorch and related libraries. All experiments were conducted on a Windows machine with the following specifications:

1) Operating System: Windows 11 (64-bit)
2) Processor: Intel Core i5 / i7/ i9
3) RAM: 16 GB
4) GPU: NVIDIA GTX/RTX series (if CUDA available)
5) Python Version: 3.10+

Key Libraries and Tools Used:

1) PyTorch: For model definition, training, and evaluation (ResNet18 from torchvision.models)
2) Torchvision: For model architecture and image transformation utilities
3) Albumentations: For robust data augmentation (flipping, rotation, brightness contrast)
4) scikit-learn: For computing class weights, evaluation metrics, and KFold cross-validation
5) Matplotlib & Seaborn: For visualization of loss curves and confusion matrix
6) OpenCV (cv2): For image processing during prediction
7) Flask: For building the web-based user interface
8) HTML/CSS/JS: For frontend development
9) Font Awesome: For icon support in the web UI

The training was performed over 50 epochs using a batch size of 32 and an Adam optimizer with a learning rate of 0.001. Model checkpoints and performance graphs were logged to facilitate reproducibility and iterative refinement.

C. Flow of the project

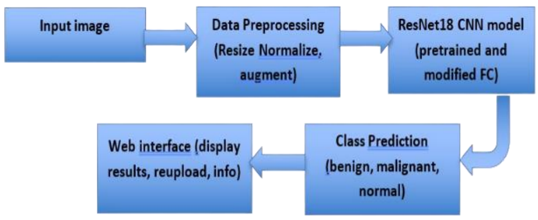

Fig 3.4- Block diagram

1) Breast Ultrasound Image Input
2) Preprocessing & Augmentation
    a) Resize to 224x224
    b) Normalize
    c) Flip, Rotate, Brightness/Contrast
3) Feature Extraction via ResNet18

a) Convolutional Layers (Pretrained on ImageNet)
b) Transfer Learning (Fine-tuned FC Layer for 3 classes)

4) Classification Layer
   a) Output: Benign / Malignant / Normal
   b) Confidence Score Calculation (Softmax)
5) Evaluation
   a) 5-Fold Cross Validation
   b) Metrics: Accuracy, Precision, Recall, F1-score
6) Flask Web Application
   a) Upload Interface
   b) Result Display (Prediction + Confidence)

D. Data Preprocessing

To prepare these images for training and ensure robustness of the model, we apply the following preprocessing and augmentation techniques:

1) Resize :Images are resized to 224x224 pixels to fit ResNet18 input dimensions.
2) Normalization: Pixel intensities are scaled to match the pretrained ResNet18 expectations.
3) Augmentation (Albumentations): Includes horizontal/vertical flips, brightness-contrast tweaks, and random rotations to mimic real-world variability.

These transformations increase the diversity of the dataset and improve generalization.

E. Feature Extraction via ResNet

After preprocessing, images are passed to a ResNet18 CNN model, which has been pretrained on ImageNet. This model captures complex patterns in medical images via multiple convolutional layers.

We replace the final fully connected layer of ResNet18 to adapt it to our classification task with 3 output classes. The convolutional layers extract spatial and textural features, such as:

1) Mass shapes
2) Edges and margins
3) Echo patterns
4) Shadowing typical of malignant tumors.

By leveraging transfer learning, we reuse pretrained knowledge and fine-tune only the final layers for ultrasound image classification.

F. Classification and Confidence prediction

The final layer of the model outputs logits, which are passed through a softmax function to derive:

1) Predicted class (Benign, Malignant, Normal)
2) Prediction confidence score (as a percentage)

This helps provide users with not only a classification but also a measure of how confident the model is in its decision.

G. Evaluation framework

To assess model robustness, we applied:

1) Cross-Entropy Loss with class weighting (to address imbalance)
2) 5-Fold Cross Validation, where the dataset is split into 5 subsets and training is repeated 5 times
3) Performance metrics: Accuracy, Precision, Recall, F1-score, Confusion Matrix

The model achieved strong results across all folds, as reflected in the confusion matrix and classification report, as provided further in this report.

H. Web Deployment via Flask

Once trained, the model is saved and integrated into a Flask web application, forming the interactive OncoSage platform:

1) Users can upload ultrasound images
2) The model returns predictions with confidence scores

Results are displayed with warnings to avoid misuse. The interface is styled to be intuitive and mobile-friendly for the ease of use.

IV.RESULTS

A. Classification Report

To assess the effectiveness of the OncoSage classification model, we evaluated it using standard classification metrics: accuracy, precision, recall, and F1-score. These metrics were computed on the test set using the final trained ResNet18 model.

Classification Report:

	precision	recall	f1-score	support
benign	0.98	0.99	0.99	175
malignant	0.98	0.98	0.98	88
normal	1.00	0.98	0.99	53
accuracy			0.98	316
macro avg	0.99	0.98	0.98	316
weighted avg	0.98	0.98	0.98	316

Fig.4.1- Classification report

These results indicate the model performs consistently across all categories, with particularly high accuracy in detecting benign and normal cases. The F1-scores being close to 1.0 reflect a strong balance between precision and recall, meaning the model not only identifies tumors correctly but also minimizes false positives and false negatives.

1) The precision of 1.00 for the *normal* class implies zero false positives, a desirable trait in screening tools.
2) The recall of 0.99 for *benign* and 0.98 for *malignant* shows the model's ability to capture most relevant cases.
3) The overall accuracy of 98% confirms the model's generalization capability over diverse samples.

The macro average (simple mean of metrics for all classes) and weighted average (takes class imbalance into account) both reinforce the model's balanced performance.

B. Confusion Matrix Analysis

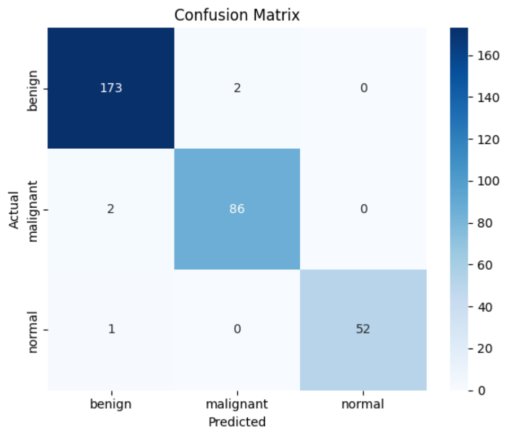

Fig.4.2- Confusion Matrix

To further evaluate model performance beyond aggregate metrics, we analyzed the confusion matrix (Fig.4.2) of OncoSage on the test dataset. The matrix outlines the number of correct and incorrect predictions made by the model across all classes — benign, malignant, and normal.

Actual\Predicted	Benign	Malignant	normal
Benign	173	2	0
Malignant	2	86	0
Normal	1	0	52

Table.4.1

Key Observations:

1) The model correctly predicted 173 out of 175 benign cases, with only 2 misclassified as malignant.

2) It identified 86 out of 88 malignant samples correctly, with 2 misclassified as benign.

3) Out of 53 normal images, 52 were accurately detected, with only a single instance misclassified as benign.

This matrix demonstrates high classification accuracy across all categories. The low number of false positives and false negatives is critical in medical applications, as misdiagnosing malignant cases or producing unnecessary alarms can have serious implications. The model's strong performance in correctly identifying malignant cases emphasizes its potential utility in assisting early detection and reducing diagnostic burden on healthcare professionals.

C. GUI of the web application

The GUI of the web application looks as follows.

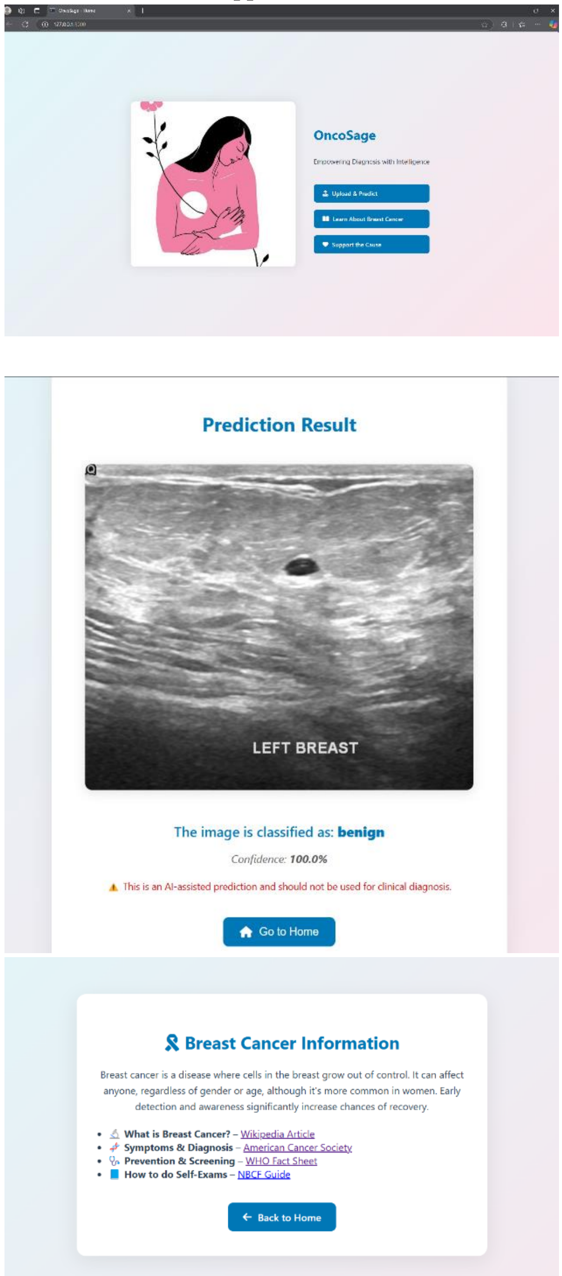

Fig 4.3-GUI of Web application

The design is user friendly and easy to navigate. Along with the prediction, it also provides information about breast cancer and donation links.

V. CONCLUSION

This study introduces OncoSage, a deep learning-driven breast cancer classification system that integrates the predictive power of ResNet18 with a cloud-accessible, interactive web interface. By leveraging transfer learning and advanced data augmentation techniques, our model achieved a high classification accuracy of 98% on the BUSI dataset, demonstrating robust performance across all three diagnostic categories: benign, malignant, and normal. Unlike traditional diagnostic models, OncoSage prioritizes accessibility and user engagement through a web-based interface, enabling both clinicians and patients to interact with AI-driven insights in real time. This contributes to early detection efforts and provides supplementary diagnostic support in resource-limited environments.

The platform's success lies in its holistic pipeline—from preprocessing and data balancing to training with cross-validation and real-world deployment—proving that AI can be practically embedded into clinical workflows.

In the future, OncoSage can be extended by incorporating additional imaging modalities (e.g., mammograms, thermography), multilingual support in the chatbot, and clinical integration with hospital management systems (HMS). Furthermore, conducting longitudinal studies and user-centric evaluations will be key in transitioning OncoSage from a research prototype to a reliable clinical decision support tool.

VI. ACKNOWLEDGEMENT

We express our profound thanks to Guide Prof. Dr. Sanjaykumar Nipanikar from ECE Department, MIT School of Engineering, MIT-ADT University, Pune for his/her expert guidance, encouragement and inspiration during this project work.
We are very thankful Prof. Dr. Sanjaykumar Nipanikar, UG Project Co-ordinator, Electronics and Communication Engineering for his support and all kinds of help to complete our work.
We also express thanks to Prof. Dr. Sachin Takale, Program Head-ECO and Prof. Dr. Gopal Gawande, HoD, Electronics and Communication Engineering Department, being constant source of inspiration throughout the execution.
We are grateful to Prof. Dr. Virendra V. Shete, Director, School of Engineering and Sciences, MIT-ADT University, Pune, for providing us all the facilities to carry out this work.
We also thank all the faculty members in the department for their support and advice. We very much grateful to our Parents for their immense love, affection, help, cooperation and encouragement to complete this course.
Finally, a special thanks to all our classmates and friends who helped and encouraged us to complete successfully our project work.

REFERENCES

[1] Sumeet Mathur (2024). Supervised Machine Learning-Based Classification and Prediction of Breast Cancer. International Journal of Intelligent Systems and Applications in Engineering.

[2] Richa Gupta, Mohsina Bilal (2023). Breast Cancer Screening Using Artificial Intelligence Techniques: Enhancing Biochemical Insights and Diagnostic Accuracy. Chemistry Proceedings

[3] Schiffman, J. D., et al. (2021). Ethical considerations in Artificial Intelligence and machine learning. Journal of Healthcare Management. 66(2), 117-124.

[4] Al-Dhabyani et al., "Dataset of Breast Ultrasound Images," Data in Brief, 2020.

[5] He et al., "Deep Residual Learning for Image Recognition," CVPR, 2016.

[6] Buslaev et al., "Albumentations: Fast and Flexible Image Augmentations," Information, 2020.

[7] Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," Medical Image Analysis, 2017.