CancerGuard: Predictive Modeling for Breast Cancer Using Machine Learning

Rohini Sharma[1], Yash Sonkar[2]. Manu Kumar[3], Pranav Soni[4], Brijesh Gupta[5]
Information Technology, J.S.S. Academy of Technical Education, Noida, India

Abstract—**Mammography images can effectively be used to detect breast cancer, which remain a leading cause of mortality among women worldwide. Cancer Guard: Predictive Modelling for Breast Cancer created a framework that combines YOLO and a customized UaNet through deep learning to achieve this goal. Our Cancer Guard system improved the accuracy of detection significantly compared to conventional methods because our model utilized three generalizable datasets – In Breast, CBIS-DDSM, and MIAS. The deployed framework achieves an end-to-end pipeline, including dataset preprocessing, model training, evaluation, and web-based real-time diagnosis. CancerGuard demonstrates greater accuracy in localization and classification of lesions through a dual-model strategy. We were able to successfully place CancerGuard at the forefront of innovation for AI driven breast cancer detection by significantly improving precision and accuracy.**

I. INTRODUCTION

Breast cancer is the most common cancer and kills the most women. Survival chances increase significantly with early diagnosis because it allows for timely intervention and treatment. Commonly used diagnostic techniques such as mammography depend heavily on a radiologist's judgment skill, which is susceptible to bias and mistakes. Breast cancer detection can be automated and improved with AI and deep learning, which enable better accuracy and faster diagnoses with less need for manual interpretation.

This document describes CancerGuard: Predictive Modelling for Breast Cancer, an AI framework that improves detection of breast cancer through deep learning. Unlike traditional models that use a single dataset, CancerGuard uses three well-known mammogram datasets — InBreast, CBIS-DDSM, and MIAS. This improves the diversity and generalization of the detection system. The model follows a two-architecture strategy using YOLO (You Only Look Once) for lesion detection and a customized UaNet for feature extraction and classification.

Cancer Guard transcends detection by offering a complete pipeline comprising of dataset preprocessing, model training, evaluation, and real-time deployment all done via a web-based interface. This all-encompassing design allows it to be implemented in a clinical setting where it provides radiologists with an AI-based tool to assist them in diagnostics. In an effort to stimulate a paradigm shift in AI-enabled breast cancer detection, CancerGuard tackles the issues regarding dataset diversity, false positives, and model interpretation.

This paper studies the methodology of CancerGuard systems, assesses its performance, and contrasts it with previous models. With a lot of effort and analysis, we show how the system is able to achieve enhanced precision and recall for breast cancer diagnoses which would yield better outcomes for the patients.

II. LITERATURE REVIEW

The detection of breast cancer has been worked on with many techniques of machine learning and deep learning. Most of the older systems were built upon feature extraction and classical machine performing tasks using Support Vector Machines (SVM) and Random Forests. These methods often had generalization problems and needed a lot of feature engineering done (Spanhol et al, 2016). In deep learning, convolutional neural networks (CNN) stand out as one of the best performers and is now an industry standard for breast cancer detection (Litjens et al, 2017).

Many attempts have been made to use deep learning for mammography analysis. Shen et al. (2019)

showcased how CNNs could accurately classify malignant versus benign breast lesions and were able to achieve high results using the CBIS-DDSM dataset. Likewise, Yala et al. (2018) also described a system that predicts the risk of breast cancer using deep learning with the focus of large dataset having a positive impact on the model. These studies recognized the ability of CNNs to act as major diagnostic tools, however, many of these tools were not live deployed or tested on different datasets.

Recent studies have examined object detection models with a focus on Faster R-CNN and YOLO for lesion identification in mammograms (Ribli et al., 2018). Specifically, YOLO has become widely adopted because it can process images rapidly which is advantageous in clinical settings (Jiang et al., 2021). Nevertheless, many implementations are developed around a single dataset which reduces their usage across different imaging environments.

The designed framework for CancerGuard overcomes these issues by combining several datasets under a dual-model framework in which lesions are detected through the YOLO model while classification is done using UaNet, enhancing accuracy and generalization. It builds on existing work while incorporating a complete pipeline for practical use of breast cancer diagnosis.

III. METHODOLOGY

The CancerGuard: Predictive Modelling for Breast Cancer framework describes a windowed approach to breast cancer detection from mammogram images using deep learning. The methodology encompasses sophisticated stages starting from dataset collection and cleaning, model training and testing, to system integration and application. The next part describes the steps taken when developing and acting upon the system.

A. Dataset Acquisition and Preprocessing

To create a robust and universally applicable model for breast cancer detection, CancerGuard combines three widely accepted mammography datasets:

InBreast: A digital full-field mammogram that contains high-resolution images of well-characterized lesions.

CBIS-DDSM (Curated Breast Imaging Subset of DDSM): A popular collection of digitized film mammograms that have confirmed biopsied lesions.

MIAS (Mammographic Image Analysis Society): A low-quality mammogram dataset marked with different types of abnormalities.

Cancellation by integration aids in eliminating dataset bias and enhances model performance for varied imaging conditions due to the integration of multiple datasets.

Data Preprocessing

In order to ensure smoothness and optimization of model performance, the images are supplemented with the following steps:

Rescaling and Normalization: All datasets are resized to the same resolution because mammograms have different resolution and intensity distribution, pixel values are also normalized with a certain range. This step is known as Rescaling and Normalization.

Contrast Enhancement: Subtle mammography lesions are enhanced using histogram obfuscation; the same is done using contrast-limited obfuscation of adaptive histograms. This step is called Contrast Enhancement.

Data Augmentation: Methods like random rotations, flipping, cropping, and brightness modifications are employed to broaden the dataset artificially, which

IV. LITERATURE REVIEW

Multiple prior works have explored machine learning and deep learning for breast cancer detection. Our proposed system builds on these but introduces several innovations in model design, dataset diversity, deployment, and usability. The following table presents a comparative summary:

Numerous studies have explored the use of machine learning and deep learning in breast cancer detection, such as the work of Spanhol (2016) which relied on traditional classifiers like SVM, and Litjens (2017) who confirmed the value of CNNs in medical imaging. Others like Yala (2018) and Ribli(2018) leveraged deep learning models such as YOLO or risk predictors on single datasets. However, most of these studies either focused on a single modality (image or tabular) or lacked live deployment capability. In contrast, our approach presents a dual-model system—using VGG19 for image classification and XGBoost for clinical parameter analysis—trained on a unified and diverse dataset composed of InBreast, CBIS-DDSM, and MIAS. We further enhanced practical usability by deploying the system as a Flask-based web application with Firebase-backed session and authentication via Google OAuth, making it both clinically relevant and user-ready. This full-stack integration with multi-source training data marks a significant advancement over prior methods.

Area	Existing Work	Our Innovation
Model Variety	Typically uses either CNN or object detection	Uses dual models: VGG19 (images), YOLO (detection)
Dataset	Trained on one dataset	Combined 3 datasets: InBreast, CBIS-DDSM, MIAS
Accuracy	Up to ~85-90% on individual models	80% (image model) + 98% (parameter model)
Deployment	Rarely deployed in production	Fully deployed Flask app with real-time inference
Usability	Primarily academic	Frontend + Backend with user-friendly UI
Feature Integration	Image or tabular data only	Uses both image and patient parameters
Authentication	Not considered	Secure access via Firebase + Google OAuth

decreases overfitting risk and enhances the model's generalization capabilities.

Dataset Conversion: The images and their corresponding annotations are converted to YOLO and COCO formats for ease of training on various architectures. This step creates the bounding box coordinates and lesion labels required for localization and classification.

B. Model Architecture and Training

The CancerGuard framework follows a dual model approach with one network based on object detection for lesion localization and segmentation-classification network for precise diagnosis.

Lesion Localization Using YOLO

Lesions in mammographic images are detected using YOLO (You Only Look Once) model. With YOLO, an image is processed as a single whole instead of in fragments or parts, as is the case with traditional region-based detectors. Because of this, real-time inference becomes possible with only one pass over the image. The steps below are taken during the YOLO training process:

1. Labeling and Annotations: The datasets require annotated bounding boxes to show the locations of the lesions.
2. Model Selection: Detectors of YOLO version 8, and improved version, are used because accuracy is greatly enhanced.
3. Training Execution: The model is trained using: 'yolo train data = dataset.yaml model =yolov8n'
4. Loss Function and Optimization: It employs cross- entropy loss for classification and mean-squared-error for bounding box regression. The Adam optimization algorithm is used for faster convergence.

Feature Extraction and Classification Using UaNet

CancerGuard incorporates a refined classification technique that integrates UaNet as a mask classifier, which is developed from the U-Net architecture. This technique is specifically trained to classify and segment lesions as benign or malignant. Its architecture includes:

Encoder-Decoder Structure: The encoder extracts a hierarchy of features, and the decoder reconstructs the segmented lesion regions.

Skip Connections: These connections assist in preserving spatial context and improving the accuracy of segmentation.

Final Classification Layer: The classification layer categorizes the segmented lesion region based on the provided attributes.

The diagnostic accuracy of the system is improved by training the UaNet model independently to serve as a complement to the YOLO detection output

C. Model Evaluation

To evaluate the effectiveness of the concept system for Breast cancer diagnosis, CancerGuard, its functionality is assessed through rigorous evaluation using multiple performance metrics.

Mean Average Precision (mAP): It determines the ability of the YOLO model to detect lesions within the images.

Sensitivity and Specificity: Determines how well the classification of cases by UaNet as malignant or benign is performed.
F1 score: Enables evaluation of the system that achieves good precision and recall at the same time.
COCO Evaluation Metrics: This is for the accuracy the system locates lesions in a series of mammograms processed.
Models are trained and validated on separate test data to check for robustness and generalization across various patient cohorts.

D. Deployment and Web Application

To support practical use, CancerGuard is implemented as a web applicatioin enabling radiologists and health practitioners to ergonomically work with the system. The deployment process involves:
Backend Development: In this stage, the system is designed as a Flask app that supports inference of the learned model on the uploaded mammograms.
Frontend Development: A web interface facilitating easy image upload and result display is developed.
Real-time Computing: The uploaded mammograms are processed instantly by the pre-trained models of YOLO and UaNet, who then return detection and classification results without delay.
To activate the web application, the user runs the code:‘Python app.py’
From this point, the application can be accessed through the link http://127.0.0.1:5000. This serves as a functional and easily extendable solution for the detection of breast cancer.

V. RESULT AND DISCUSSION

The Breast CancerGuard: Predictive Modelling Systems for Breast Cancer’s lesions delineating and identification classification competency were tested on three benchmark datasets InBreast, CBIS-DDSM, and MIAS. The evaluated problem is detection and classification of lesions. The result from the dual-model method where YOLO was used for lesion detection and UaNet for classification showed positive improvement in both accuracy and robustness over current approaches.

Quantitative Performance Analysis
The CancerGuard system was evaluated based on several performance metrics and score indices, such as mean Average Precision (mAP), sensitivity, specificity, and F1 score. The findings reveal that the model strikes a balance between lesion detection precision and classification reliability.

Model	Dataset Used	mAP (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
Faster R-CNN	CBIS-DDSM	85.2	82.1	84.5	83.3
YOLOv4	InBreast	87.5	85.3	86.8	86.0
UaNet (Proposed)	Multiple Datasets	**92.3**	**90.4**	**91.5**	**90.9**

Nonetheless, these new findings show that the model CancerGuard achieves better results than other models
with its multi-dataset approach and hybrid architecture.
Comparative Analysis

Feature	Existing Models	*CancerGuard* (Proposed)
Dataset Diversity	Single dataset	Multi-dataset (InBreast, CBIS-DDSM, MIAS)
Lesion Localization	Faster R-CNN, YOLO	YOLOv8 (Improved Accuracy)
Classification Model	Basic CNNs	UaNet (Advanced Segmentation)
Real-time Deployment	Limited	Flask-based Web Application
Generalizability	Moderate	High (Cross-Dataset Performance)

This confirms the contributions of CancerGuard and compares its advantages with the existing ones based on several aspects including the model’s dataset diversity, performance, and its practicality in real world scenarios.

Discussion
The CancerGuard framework achieves a notable leap improvement against integrated models’ accuracy metrics, based on multi dataset incorporation and dual model architecture. But at the same time, deep learning models require significantly heightened expenses for it to be functional, which makes it difficult to scale up. Moreover, while real time false positive rate is a concern, YOLO allows for accurate lesion localization on the fly. Features such as enhanced generalizability, automated diagnosis, and web interfaces for the user make the system advantageous. However, such dependent sources of high-quality labeled data may induce model bias. In the future, research will be conducted on how to improve model performance while also adding diverse clinical reasoning capabilities to increase overall trust in the system.

VI. CONCLUSION

Demanding as it may sound, Breast cancer detection is one of the most crippling obstacles faced in the domain of medical imaging, as the margin of error is very slim and prompt as well as accurate diagnosis can greatly enhance the patient's prognosis. This study proposed CancerGuard: Predictive Modelling for Breast Cancer, an embedded system aimed at detecting, and classifying lesions using state of the art deep learning technology with a dual model approach. When coupled with UaNet classification, the YOLO lesion localization model outperforms the conventional systems single handedly. The inclusion of multiple datasets, InBreast, CBIS-DDSM, and MIAS offers improved generalizability and robustness across various conditions of imaging.

The results show that CancerGuard have better model accuracy, sensitivity, and specificity than currently available breast cancer detection models. Its web-based deployment further adds to its clinical usability by making it an automated and scalable solution. However, other practical issues like resource constraints, false positive rates, and numerous others need to be solved to improve practical utilization of the system.

Later studies will focus on enhancing the efficiency of the model while minimizing the computational burden, and applying techniques for making the model more explainable in order to enhance trustworthiness in clinical settings. The possibility of integrating additional real world mammographic images accompanied by multi-modal imaging techniques promises high diagnostic accuracy. As a whole, CancerGuard is an exciting first step towards implementing AI in reliable breast cancer detection in clinical practice.

REFERENCES

[1] Maistry, B., & Ezugwu, A. E. (2023). Breast Cancer Detection and Diagnosis: A Comparative Study of State-of-the-Art Deep Learning Architectures.

[2] Ezugwu, A. E., & Maistry, B. (2023). Deep Learning-Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Directions.

[3] Liu, Y., Chen, P. H. C., Krause, J., & Peng, L. (2019). How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature.

[4] [4 McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Suleyman, M. (2020). International Evaluation of an AI System for Breast Cancer Screening.

[5] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A Guide to Deep Learning in Healthcare.

[6] Jiang, Y., Yang, M., Wang, S., Li, X., & Sun, Y. (2021). Emerging Role of Deep Learning-Based Artificial Intelligence in Tumor Pathology.

[7] G. Litjens et al.t: A Survey on Deep Learning in Medical mage Analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017.

[8] Y. Jiang et al., "YOLO-Based Breast Cancer Detection Using Mammography Images," Computers in Biology and. Medicine, vol. 139, 104987, 2021

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation, in Proc. Medical Image Computing and Computer-Assisted Assisted Intervention (MICCAI), 2015, pp. 234–241.

[10] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, and D. L. Rubin,"Curated Breast Imaging Subset of DDSM A Standardized Dataset for Breast Imaging Research," Scientific Data, vol. 4, no. 170177, 2017.

[11] E. Valle, et al., "Data, Deep Learning and Diagnostics: A Review on Artificial Intelligence in Medical Imaging," Computers in Biology and Medicine, vol. 132, 104339, 2020.