

# **Enhanced Breast Cancer Diagnosis Using Machine Learning on Patient Data and Deep Learning**

## **IEEE BASE PAPER TITLE:**

### **Toward Improving Breast Cancer Classification Using an Adaptive Voting Ensemble Learning Algorithm**

#### **IEEE BASE PAPER ABSTRACT:**

Over the past decade, breast cancer has been the most common type of cancer in women. Different methods were proposed for breast cancer detection. These methods mainly classify and categorize malignant and Benign tumors. Machine learning is a practical approach for breast cancer classification. Data mining and classification are effective methods to predict and categorize breast cancer. The optimum classification for detecting Breast Cancer (BC) is ensemble-based. The ensemble approach involves using multiple ways to find the best possible solution. This study used the Wisconsin Breast Cancer Diagnostic (WBCD) dataset. We created a voting ensemble classifier that combines four different machine learning models: Extra Trees Classifier (ETC), Light Gradient Boosting Machine (LightGBM), Ridge Classifier (RC), and Linear Discriminant Analysis (LDA). The proposed ELRL-E approach achieved an accuracy of 97.6%, a precision of 96.4%, a recall of 100%, and an F1 score of 98.1%. Various output evaluations are used to evaluate the performance and efficiency of the proposed model and other classifiers. Overall, the recommended strategy performed better. Results are directly compared with the individual classifier and different recognized state-of-

the-art classifiers. The primary objective of this study is to identify the most influential ensemble machine learning classifier for breast cancer detection and diagnosis in terms of accuracy and AUC score.

### **OUR PROPOSED ABSTRACT:**

The "Enhanced Breast Cancer Diagnosis Using Machine Learning on Patient Data and Deep Learning" project introduces a robust, dual-modality diagnostic system that leverages both patient data and histopathology image analysis for improved breast cancer detection and classification. Developed in Python, with a user-friendly front end built using HTML, CSS, and JavaScript, the system is deployed via the Flask web framework, providing a comprehensive interface for clinical diagnostics. This hybrid approach combines traditional Machine Learning with Deep Learning techniques, enabling accurate predictions from both structured patient data and biopsy images.

For the Machine Learning component, the system utilizes the Wisconsin Breast Cancer Dataset (WBCD) Diagnostic, comprising 569 records with 32 diagnostic features, including tumor characteristics like radius, texture, perimeter, area, and compactness. Two models were trained: a Stacking Classifier, which achieved a 99% accuracy on training and 94% on test data, and a CatBoost Classifier, which yielded a perfect 100% training accuracy and a test accuracy of 98%. These models provide reliable predictions from the structured dataset, assisting in diagnostic decision-making based on key tumor metrics.

In parallel, the Deep Learning module applies the DenseNet201 architecture to analyze high-resolution biopsy images from the Breast Cancer Histopathological 400X (BreakHis 400X) dataset. The dataset contains 1,693 microscopic images, classified as benign and malignant, with 371 benign and 777 malignant images

used for training, and 176 benign and 369 malignant images for testing. DenseNet201 achieved a training accuracy of 98% and a validation accuracy of 88%, demonstrating strong capability in distinguishing benign from malignant cases in histopathology images.

The combination of these models provides a comprehensive diagnostic tool that leverages both structured data and image data, enhancing diagnostic accuracy and supporting healthcare professionals in early breast cancer detection. This dual approach addresses the diagnostic challenges of breast cancer by offering insights from multiple data perspectives, which improves confidence in the diagnosis, and thus potentially contributes to better patient outcomes through timely and accurate cancer detection.

## **SYSTEM REQUIREMENTS:**

### **HARDWARE REQUIREMENTS:**

- System : Pentium i3 Processor.
- Hard Disk : 500 GB.
- Monitor : 15” LED.
- Input Devices : Keyboard, Mouse.
- Ram : 8 GB.

### **SOFTWARE REQUIREMENTS:**

- Operating System : Windows 10 / 11.
- Coding Language : Python 3.12.0.
- Web Framework : Flask.
- Frontend : HTML, CSS, JavaScript.

## **REFERENCE:**

AMREEN BATOOL AND YUNG-CHEOL BYUN, “Toward Improving Breast Cancer Classification Using an Adaptive Voting Ensemble Learning Algorithm”, in IEEE Access, vol. 12, pp. 12869-12882, 2024.