

Title : Enhancing Early Breast Cancer Detection Through Advanced Data Analysis**Publisher: IEEE****Author Names:** Md. Atiqur Rahman; Mohamed Hamada; Shayla Sharmin; Tanzina Afroz Rimi; Atia Sanjida Talukder; Nafees Imran

Abstract: In recent years, breast cancer, originating from breast tissue, has become one of the significant global health challenges for women worldwide, with early detection crucial for improved survival rates. Researchers have proposed numerous detection techniques, and recently, machine learning-based methods have gained considerable attention due to their reusability and speed. Despite various models proposed by researchers for breast cancer detection, there is an ongoing need for more accurate models. This study proposes an enhanced machine-learning approach for breast cancer detection using the Wisconsin Breast Cancer (Diagnostic) (WDBC) dataset. We applied several data preprocessing techniques, including hypothesis testing, feature engineering, scaling, and feature selection. We trained 14 classifiers by selecting the 13 most significant features using a gradient boosting regressor with Bonferroni correction. Our proposed eXtreme Gradient Boosting model demonstrated superior performance, achieving 99.12% accuracy, 0.9767 precision, 1.0 recall, 0.9861 specificity, and 0.9882 F1-score. These results surpass those of previous studies, underscoring the model's potential for early and accurate breast cancer diagnosis. Furthermore, evaluations based on training time and Kappa score indicate that our eXtreme Gradient Boosting model is faster and more reliable.

Key words: Breast cancer detection; Machine learning; Wisconsin Breast Cancer Dataset (WDBC); eXtreme Gradient Boosting (XGBoost);

DOI: [10.1109/ACCESS.2024.3483095](https://doi.org/10.1109/ACCESS.2024.3483095)

<https://ieeexplore.ieee.org/document/10720756>