

Title: Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI

Author Names: Taminul Islam, Md. Alif Sheakh, Mst. Sazia Tahosin, Most. Hasna Hena, Shopnil Akash, Yousef A. Bin Jardan, Gezahign FentahunWondmie, Hiba-Allah Nafidi & Mohammed Bourhia

Abstract: Breast cancer has rapidly increased in prevalence in recent years, making it one of the leading causes of mortality worldwide. Among all cancers, it is by far the most common. Diagnosing this illness manually requires significant time and expertise. Since detecting breast cancer is a time-consuming process, preventing its further spread can be aided by creating machine-based forecasts. Machine learning and Explainable AI are crucial in classification as they not only provide accurate predictions but also offer insights into how the model arrives at its decisions, aiding in the understanding and trustworthiness of the classification results. In this study, we evaluate and compare the classification accuracy, precision, recall, and F1 scores of five different machine learning methods using a primary dataset (500 patients from Dhaka Medical College Hospital). Five different supervised machine learning techniques, including decision tree, random forest, logistic regression, naive bayes, and XGBoost, have been used to achieve optimal results on our dataset. Additionally, this study applied SHAP analysis to the XGBoost model to interpret the model's predictions and understand the impact of each feature on the model's output. We compared the accuracy with which several algorithms classified the data, as well as contrasted with other literature in this field. After final evaluation, this study found that XGBoost achieved the best model accuracy, which is 97%.

Key words: Breast cancer detection Machine learning Explainable AI (XAI) XGBoost

<https://www.nature.com/articles/s41598-024-57740-5>