DOI: https://doi.org/10.36094/sc.v92.2026.A_Comparative_Study_of_ML_and_DL_Based.Ghosh.631

ARTICLE

Sci. and Cult. 92 (9–10) : 631-637 (2026)

A COMPARATIVE STUDY OF ML AND DL BASED FRAMEWORK FOR BREAST CANCER DIAGNOSIS WITH STATISTICAL SIGNIFICANCE ANALYSIS

POULOMI GHOSH[1]*, SUMAN DAS[1],
SUDIP GHOSH[1] AND ANIRBAN KUNDU[1]

The early and correct diagnosis of breast cancer has been shown to greatly enhance the survival rates of patients. This paper provides a comprehensive comparative study of six machine learning classifiers, namely Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Artificial Neural Network (ANN), for the diagnosis of breast cancer using the Wisconsin Diagnostic Breast Cancer database. The proposed framework provides a statistically validated benchmark for breast cancer prediction systems. Here The dataset contains 569 patient records with 30 medical features each. These features describe characteristics of cell nuclei in breast cancer samples. The dataset is used for classification into two categories: benign and malignant. Out of 569 patients, 357 are benign and 212 are malignant. The performance of the classifiers was compared using accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and precision-recall curve. The classifiers were compared using stratified 80-20 train-test split and 5-fold cross-validation. XGBoost had the highest test accuracy of 97.37%, followed closely by Random Forest and SVM. However, the McNemar statistical test and Friedman test showed that there are no statistically significant differences in the performance of the classifiers ($p > 0.05$). The analysis of feature importance enabled the selection of the most important predictors for the diagnosis of breast cancer. The results of this study indicate that although ensemble-based classifiers slightly outperform traditional classifiers, all classifiers have similar diagnostic capabilities.