

Research on some supervised machine learning methods and their application in supporting breast cancer diagnosis

Noysinakhone Douangboudy

Abstract: This study investigates several supervised machine learning methods and their application in supporting the diagnosis of breast cancer. Given the critical importance of early and accurate detection, various classification algorithms—including Support Vector Machines (SVM), Random Forests, k-Nearest Neighbors (k-NN), and Logistic Regression—were implemented and evaluated on a standard clinical dataset. The research methodology involved comprehensive data preprocessing, feature selection, and hyperparameter tuning to optimize model performance. Comparative analysis was conducted using key metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The experimental results demonstrate that ensemble methods, particularly Random Forest, achieved superior predictive performance, with an accuracy exceeding 97% and high sensitivity in identifying malignant cases. Furthermore, the study explores the interpretability of these models to enhance clinical trust. The findings confirm that supervised machine learning can serve as a reliable, non-invasive decision-support tool for radiologists and oncologists, potentially reducing diagnostic errors and improving patient outcomes in breast cancer screening.

Keywords: Machine learning, supervised learning, breast cancer diagnosis, medical diagnosis