Accounting for Cancer Patients with Severe Outcomes: An Anomaly Detection Perspective

Yang Yan[1] Christopher Lominska[2] Gregory N Gan[2] Hao Gao[2] Zhong Chen[1]

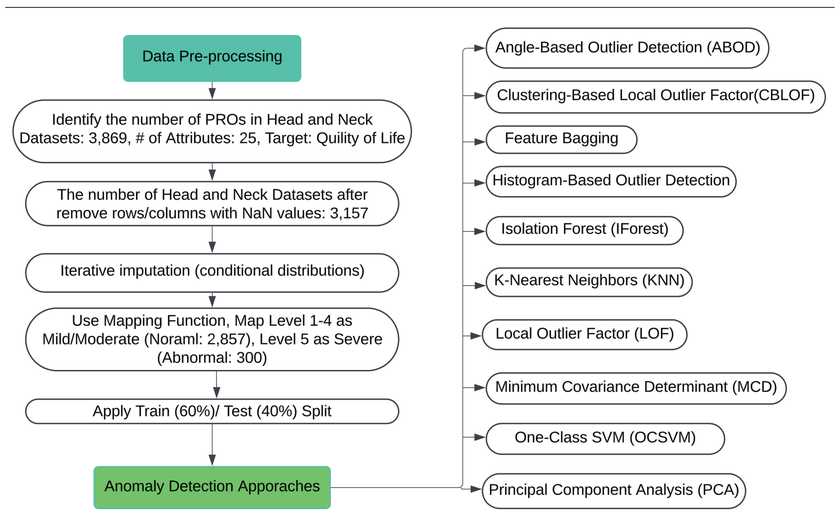
[1]School of Computing, Southern Illinois University

[2]Department of Radition Oncology, University of Kansas Medical Center

Abstract

Health outcomes and radiation-induced toxicities for cancer patients undergoing radiation therapy are influenced by several factors, including the disease process, treatment plan, and various symptoms produced by both the disease and treatment. Accurate prediction and assessment of a patient's health status are pivotal for precision and personalized healthcare, especially for patients suffering severe outcomes such as pain, depression, and sleep disorders. Motivated by the issue of extreme class imbalance, this study investigates a set of unsupervised anomaly detection approaches to classify patients with mild/intermediate and severe outcomes using patient-reported outcomes (PROs) datasets. We found that the HBOS method demonstrated superior performance for the minority class, representing cancer patients with severe health statuses. Moreover, IForest and KNN showcased good potential by effectively considering both majority and minority patients. This study may provide insightful guidelines for clinical practice.

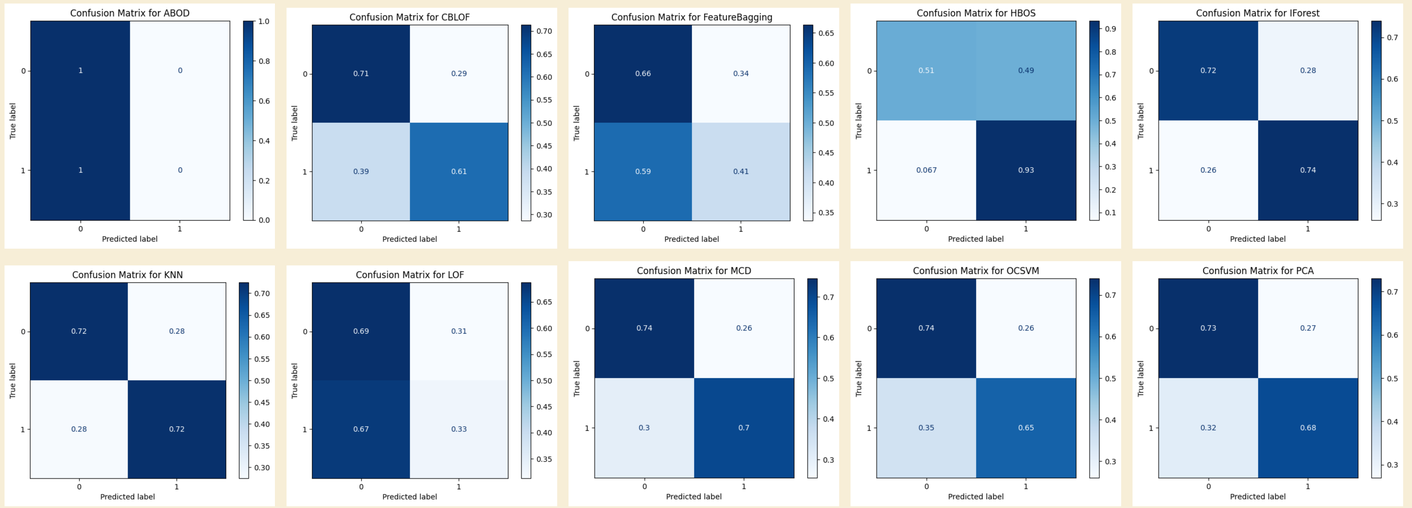
Methodology

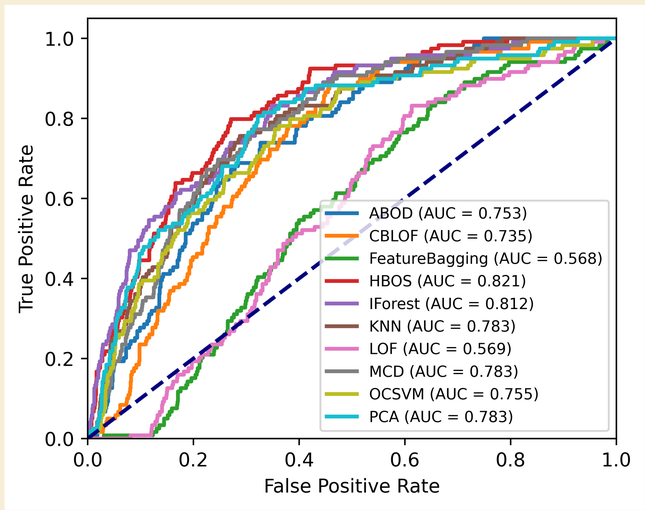


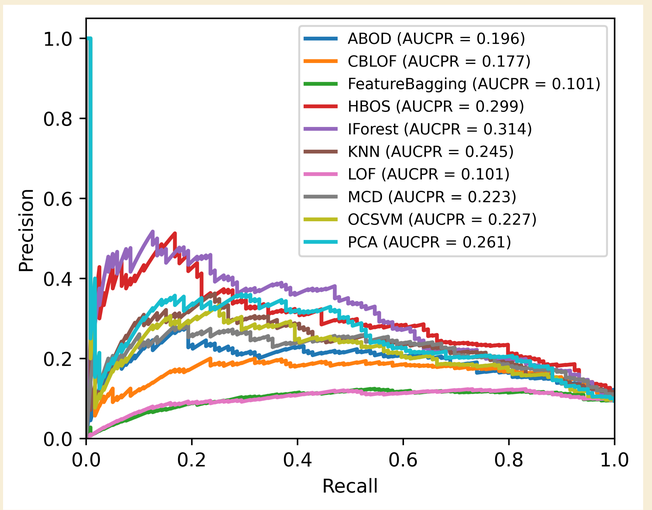
Conclusion

Based on the PRO datasets, in this study, we investigate a series of unsupervised anomaly detection approaches for classifying patients' health status with mild/intermediate and severe outcomes. According to the confusion matrix and ROC/AUCPR curve evaluation, IForest, KNN, and PCA have showcased good potential by considering the patients with mild/intermediate and severe health status.

Experimental Results







A. Confusion Matrix Comparison: The confusion matrices (Fig. 1) show that ABOD performs poorly by misclassifying all cases as severe. HBOS achieves high accuracy (0.93) for detecting severe cases but performs poorly for mild cases (0.51). IForest and KNN offer a good balance, accurately classifying both majority and minority classes with fewer false positives and false negatives.

B. ROC/AUCPR Curves and Comparison: The ROC and AUCPR curves (Fig. 2) indicate that HBOS has the highest AUC (0.821), while IForest achieves the best AUCPR. IForest, KNN, and PCA consistently perform well, effectively classifying both mild and severe cases. HBOS, despite a high AUC, struggles with majority class detection.

yang.yan@siu.edu