An Evaluation of Machine Learning Classifiers and Ensembles for Early Stage Prediction of Lung Cancer

Muhammad Imran Faisal, Saba Bashir, Zain Sikandar Khan, Farhan Hassan Khan

Abstract— Researchers have widely used statistical and machine learning techniques to construct prediction models in several domains such as prediction of software faults, spam detection, disease diagnosis, and financial fraud identification. The prediction of patients prone to lung cancer can help doctors in their decision making regarding their treatments. In this regard, this research paper attempts to evaluate the discriminative power of several predictors in the study to increase the efficiency of lung cancer detection through their symptoms. A number of classifiers including Support Vector Machine (SVM), C4.5 Decision tree, Multi-Layer Perceptron, Neural Network, and Naïve Bayes (NB) are evaluated on a benchmark dataset obtained from UCI repository. The performance is also compared with well-known ensembles such as Random Forest and Majority Voting. Based on performance evaluations, it is observed that Gradient-boosted Tree outperformed all other individual as well as ensemble classifiers and achieved 90% accuracy.

For the published version of record document, go to:
http://dx.doi.org/10.1109/ICEEST.2018.8643311