World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:05, 2016

Breast Cancer Survivability Prediction via Classifier Ensemble

Authors : Mohamed Al-Badrashiny, Abdelghani Bellaachia

Abstract : This paper presents a classifier ensemble approach for predicting the survivability of the breast cancer patients using the latest database version of the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. The system consists of two main components; features selection and classifier ensemble components. The features selection component divides the features in SEER database into four groups. After that it tries to find the most important features among the four groups that maximizes the weighted average F-score of a certain classification algorithm. The ensemble component uses three different classifiers, each of which models different set of features from SEER through the features selection module. On top of them, another classifier is used to give the final decision based on the output decisions and confidence scores from each of the underlying classifiers. Different classification algorithms have been examined; the best setup found is by using the decision tree, Bayesian network, and Na¨ıve Bayes algorithms for the underlying classifiers and Na¨ıve Bayes for the classifier ensemble step. The system outperforms all published systems to date when evaluated against the exact same data of SEER (period of 1973-2002). It gives 87.39% weighted average F-score compared to 85.82% and 81.34% of the other published systems. By increasing the data size to cover the whole database (period of 1973-2014), the overall weighted average F-score jumps to 92.4% on the held out unseen test set.

Keywords : classifier ensemble, breast cancer survivability, data mining, SEER
Conference Title : ICAIDM 2016 : International Conference on Artificial Intelligence and Data Mining
Conference Location : London, United Kingdom
Conference Dates : May 23-24, 2016