Abstract

Breast cancer is one of the most severe types of cancers and is the most common cause of death among the female cancer patients. In order to ease the process of decision making and financial arrangements, it is essential to be aware of survivability of patients. In recent years, effective data-mining techniques have been employed to predict the 5-year survivability of cancer patients, showing reasonable accuracy. The efficiency of these models can be improved by making them accessible on smartphones. In order to achieve this, it is essential to reduce the maximum required memory occupied by the prediction models, since a smartphone has a limited available memory. This issue, which is still an open area of research, is the concern of the present study. A hybrid method is enhanced by combining synthetic minority over-sampling technique (SMOTE), information gain attribute evaluation (InfoGainAttributeEval), AdaBoost.M1 algorithm and a decision tree. The more effective attributes are selected using InfoGainAttributeEval and the less effective nodes are removed by decision tree pre-pruning during the tree building. The hybrid method is further simplified by employing the post-pruning technique on the decision tree after its creation. The proposed method was subjected to a
An Efficient Method for Predicting the 5-year Survivability of Breast Cancer

5-year cancer survivability dataset, showing considerable reduction in the maximum required memory while maintaining the accuracy of prediction.

References

16. C. Edeki and S. Pandya, "Comparative Study of Data Mining and Statistical Learning..."
An Efficient Method for Predicting the 5-year Survivability of Breast Cancer


Index Terms

Computer Science  Information Sciences

Keywords

Breast cancer, Decision tree, Synthetic minority over-sampling technique, Information gain attribute evaluation, maximum required memory, smartphones, hybrid method.