Abstract

Data mining has been widely adopted in recent years in many fields, especially in the medical field. This paper highlights the prediction of unknown primary tumors in the dataset. The multiclass classifier with Random forest is used for classification of multiclass dataset as it gives much higher accuracy than binary classifiers. SMOTE method for this imbalanced dataset with Randomize technique is applied during preprocessing for reducing the biasness among classes. These all evaluations and results are carried out using WEKA 3.6.10 as a data mining tool.
Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Health Informatics.

Index Terms

Computer Science
Data Mining

Keywords
SMOTE WEKA Primary tumor Multiclass classifier Random forest