Abstract

Data mining techniques are being used extensively in healthcare sector to discover hidden pattern and relationship between patients’ record and their medical diagnosis dataset. In the concept of disease prediction, high classification accuracy can be obtained from accurately pre-processed and trained model. But existence of unimportant and irrelevant attributes in the training dataset may decrease the predictive accuracy and increase the time complexity in training phase. To increase the accuracy and efficiency, feature selection technique is frequently used in data mining. In this paper, a sequential forward selection based wrapper approach is proposed to select optimal and informative feature subset. It is known that diabetes mellitus is the most serious health problem and the complications lead to cause of death. So the aim of this research is to identify the significant attributes and classify diabetes dataset. The proposed approach is used to build the classifier models like Decision tree, K-Nearest Neighbor and Support Vector Machine produces the accuracies of 81.17%, 86.36% and 87.01% respectively. Finally, from results it is clear that the proposed model is performing better with
high accuracy comparing the similar existing models. In the research, the Pima Indian diabetes dataset is used.

References


29. Sun, Ming-an, Qing Zhang, Yejun Wang, Wei Ge, and Dianjing Guo. "Prediction of redox-sensitive cysteines using sequential distance and other sequence-based features." BMC


Index Terms

Computer Science

Biomedical

Keywords

Classification, Feature Selection, Wrapper Approach, Feature selection, SFS, Pima Indian Diabetes Dataset, C4.5, KNN, SVM