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
Developing Diabetes Disease Classification Model using Sequential Forward Selection Algorithm

high accuracy comparing the similar existing models. In the research, the Pima Indian diabetes
dataset is used.

References

1. Karthikeyani, V., I. Parvin Begum, K. Tajudin, and I. Shahina Begam. "Comparative of
data mining classification algorithm (CDMCA) in diabetes disease prediction." International
2. Raghavendra, S., and M. Indiramma. "Classification and Prediction Model using Hybrid
and diagnosis of diabetes mellitus." International journal of computer applications 95, no. 17
4. Almarabeh, Hilal, and Ehab F. Amer. "A Study of Data Mining Techniques Accuracy for
Patients to Heart Diseases using Support Vector Machines." International Journal of Computer
6. Vispute, Nilesh Jagdish, Dinesh Kumar Sahu, and Anil Rajput. "An Empirical Comparison
by Data Mining Classification Techniques for Diabetes Data Set." International Journal of
patients using naive bayes method." International Journal of Computer Applications 24, no. 3
8. Tambade, Shital, Madan Somvanshi, Pranjali Chavan, and Swati Shinde. "SVM based
Diabetic Classification and Hospital Recommendation." International Journal of Computer
Classification of Diabetes Mellitus." International Journal of Computers and Applications 5, no. 5
11. Sumathy, Mythili, Mythili Thirugnanam, Praveen Kumar, T. M. Jishnujit, and K. Ranjith
Kumar. "Diagnosis of Diabetes Mellitus based on Risk Factors." International Journal of
12. Karthikeyan, T., and K. Vembandadsamy. "An Analytical Study on Early Diagnosis and
Classification of Diabetes Mellitus." International Journal of Computers and Applications 5, no. 5
9-17.
wrapper feature selection on hybrid prediction model for analysis of high dimensional data." In
Industrial and Information Systems (ICIIS), 2014 9th International Conference on, pp. 1-6. IEEE,
2014.


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