CHAPTER 5

CONCLUSION AND SCOPE OF FUTURE WORK

In the field of Applied Sciences the integration of medicine disciplines and Information Technology which is known as medical informatics provides more significant improvement in both quality of dataset and accuracy in the prediction of disease. The rapid growth of dataset in the medical field due to many clinical reports, the difficult arises in retrieving the useful information and gaining knowledge from it. The advancement in the field of knowledge discovery leads to the way of using data mining for pattern finding and digging hidden knowledge from the huge volume of data. The adaption of data mining in the medical field produces the medical data mining which is greatly helpful to provide effective diagnostic capabilities in the earlier stages of the diseases. This may assist the doctors to fasten their process and give guidance relating the symptoms of the diseases.

This research work consists of two different views of data preprocessing to enrich the medical datasets. The real world dataset is always incomplete, redundant and presented with the irrelevant information, so using them as such for pattern recognition and knowledge discovery will lead to improper result of classification and detection of disease. The occurrence of error in medical field is very dangerous and cost effective.

The first phase of this work handles the problem of missing values in the disease dataset by imputing the enhanced fuzzy k-NN algorithm for producing the complete dataset. The experimental result shows that the proposed method outperforms the existing mean imputation, knn and weighted knn in the process of missing value handling.

The voluminous dataset in disease diagnosis is very common and retrieving useful and relevant information is essential and it is an important factor to be considered in the data preprocessing process. This research work eliminates the irrelevant information and redundant data using the evidential approach of Dempster Shafer theory which improves the performance of the classifier by producing the set of features from the whole feature set which contributes more in recognition of the pattern among instances of disease dataset.
The simulation result clearly shows that the proposed Dempster Shafer based feature subset selection in the case of uncertainty greatly handles the problem of predicting the disease very accurately and effectively.

In near future, this research work can be extended in the following dimensions:

- The evolving algorithms based clustering techniques can be adapted for discovering the hidden information in medical datasets.
- The expert systems based on intuitionistic fuzzy logic and the generalization of fuzzy logic concepts can be extended for classification of diseases in the presence of uncertainty among the data instances.
- Dimensionality reduction with optimal search algorithms can be utilized for finding the potential attributes.
- Parallel processing based algorithms can be used to reduce the time and memory usages.
- Biological behavior based determination of pattern interestingness can be discovered in the disease diagnose datasets.