CHAPTER 7

CONCLUSION AND FUTURE WORK

This chapter summarizes the content of the thesis and also highlights the future research work. In this proposed work the high-dimensional data by a large number of features problems are solved by formulating a hybrid feature subset selection with eadaboost algorithm using various decision tree classifiers.

7.1 SUMMARY OF THE RESEARCH

The feature subset selection technique helps the researchers to generate an accurate predictive model. The approach supports the researchers by selecting features that will provide good or better accuracy. The hybrid Feature subset selection solves the problem of high-dimensional data categorized by a large number of features and selects less amount of features and extremely increases the performance of learning algorithms. Ensemble method of effective AdaBoost algorithm implemented to reduce the error rate when compared with the existing methods and generates better classification accuracy by reweighting each feature for further process. The decision tree based classifiers is used because the decision trees are leading learning approaches which are able to organize the information extracted for a training dataset in a hierarchical structure.
The hybrid feature subset selection using ensemble methods of eAdaBoost is formulated with different decision trees based classification algorithm such as C4.5, Decision Stumps, Naive Bayes Tree, and Random Forest with ten-fold cross-validation. The major determination of the hybrid feature selection with ensemble methods of eAdaBoost is to improve the classification accuracy, prediction and to reduce the execution time for various datasets.

The objective of the research work is to build the hybrid feature subset selection for multiple datasets using various boosting algorithms for decision tree based classifiers. The specific objective is also listed as to design hybrid model by combining the subsets to form an aggregate subset. To develop an effective Adaboost algorithm and to reduce the error rate by weighting the features. To develop a hybrid feature selection model and ensemble model to analyze the classifier performance. Therefore the hybrid feature subset selection in ten-fold cross validation helps in improving the prediction performance of the analysts, shows faster and more cost-effective analysts and contributes to a better thought of the primary procedure that generated the data.

7.2 SCOPE AND LIMITATIONS OF STUDY

The scope and limitations of the proposed study are as follows:

- The proposed study only considers the correlation based feature selection, information gain and reliefF, whereas there are many other methodologies that could also be considered.
The empirical results of the analytical modelling of the proposed method have considered performing with intersection and exclusive OR, but not with another predictable model.

The statistical test for the hypothesis of the proposed model is considered with performing all pairwise comparison, whereas there are many other arithmetical methods that could also be considered.

The ensemble methods of the boosting algorithms for the proposed study have compared with AdaBoost, Real AdaBoost and MultiBoostAB, but not with other boosting algorithms. Other boosting algorithm could also be considered.

7.3 FUTURE SCOPE

This research work visualizes many concepts in which the work can be performed to improve the feature selection from large databases and classification for multiple datasets using big data. Because the world is adopting a data-driven approach, and as far as future the big data is concerned it is growing exponentially. Big data is for sure, getting significant in fundamental research. The potential outcomes of scientific review are currently growing, and researchers are gaining access to massive new data sets and analysis tools. Therefore the big dataset with huge number of features and instances along with large memory size of datasets can be considered. The proposed work can be extended by using fuzzy set theory concept as a hybrid model and might be incorporated in future.