CHAPTER 6
CONCLUSION AND FUTURE WORK

This chapter begins by summarizing the research. A conclusion is drawn here with a summary of findings. The entire work is evaluated and areas that require further research are also recommended. It then provides suggestions for further investigation.

6.1 CONCLUSION

Extracting user-centric information from the largest repository efficiently and effectively is becoming increasingly imperative. The objective of this research was to evaluate the information retrieval techniques in order to identify the one that will offer the best performance. The work began by reducing from the high dimensional data set. As a first process in gene classification, the high dimensionality of the microarray gene data is reduced using LPP. The LPP has been utilized for the dimensionality reduction and the SVM has been trained for effectual gene classification. For the purpose of retrieving information from a microarray gene expression, an effective gene classification technique is implemented. The technique has been tested by classifying the microarray gene expression data of human acute leukemias and colon cancer data.

The technique has classified the AML and ALL gene expressions well. The comparative results show that the new technique LPP possesses better accuracy and lesser error rate than the ANN-based and PCA-based gene classification techniques. This research has demonstrated that the SVM ensemble was utilized for the classification process in order to classify more than a single class. Here three SVMs for classifying ALL, AML and Lymphoma genes were utilized for this process, initially these three diseases equivalent gene data was trained with the SVMs separately. The dimensionality of the microarray data has been reduced
with the aid of PCA mechanism. In the testing process any of these genes given
and they have identified in which the class they belongs to. The research work
obtains better result and they are compared with another dimensionality reduction
algorithm i.e. PLS.

The final phase of the research was the evaluation of the techniques LPP
and PCA. The performance metrics used to evaluate these techniques were:
specificity, sensitivity, accuracy and error rate. The comparative results have
shown that the new technique LPP possesses better accuracy and lesser error rate
than PCA technique. The LPP has the ability of preserving locality of
neighborhood relationship. When technique like LPP is chosen for the
dimensionality reduction of microarray gene data there is effective retrieval. The
blending of LPP and SVM leads to effectual and powerful classification. Hence,
this means of gene classification have paved the way for effective information
retrieval in the microarray gene expression data.

This phase of the research was further extended to an effective microarray
gene data clustering technique. As LPP has good positive features in their task of
dimensionality reduction, the dimensionality of the microarray data has been
reduced with the aid of LPP mechanism. The technique has been tested by
clustering the microarray gene expression data of human acute leukemia and colon
cancer data. This research has demonstrated by experiment and the comparative
results have shown that the technique FCM has provided more accuracy,
correlation and less distance and error rate rather than the k-means gene clustering
techniques. We have achieved improvement in the quality of the results by using
FCM. The objectives outlined at the beginning of the work have been achieved.
Although the nature of the dataset used in this work is one of the most difficult
data mining datasets to deal with, it was still managed and good results have been
obtained.
6.2 FUTURE WORK

This research has identified the best technique that will offer the most efficient performance when applied to genetic information retrieval in the field of bioinformatics. In theory, this has proven to yield better results. It is also recommended that this is applied in the field and monitored to validate these results. It would be interesting to apply this technique in any data mining tool in medical domain and compare the results obtained from this work. This work will be more useful to implement any kind of dataset to deal with. A number of avenues for future work exist that may greatly improve the effectiveness of this approach.

DNA Microarray Technology: Although all of the cells in the human body contain identical genetic material, the same genes are not active in every cell. Studying which genes are active and which are inactive in different cell types helps scientists to understand both how these cells function normally and how they are affected when various genes do not perform properly.

Design Treatment Strategies: This research can be further widened to understand human genetic variation, vowing to create a new type of gene map that may propel medical research forward by explaining such common ailments as heart attacks, diabetes and obesity. It will be possible for the researchers to further classify the types of cancer on the basis of the patterns of gene activity in the tumor cells.

Pharmacogenomics: This will tremendously help the pharmaceutical community to develop more effective drugs as the treatment strategies will be targeted directly to the specific type of cancer.