CHAPTER 7
CONCLUSION

The primary objective of the work is to predict the cancer cells using only very few genes with help of well trained data. The proposed integrated gene ranking, feature selection and classification models are evaluated on microarray gene expression datasets. The observed results are tabulated and proved that the proposed hybrid models gives appreciable results than traditional algorithms. Moreover the following findings are observed from this study.

7.1 FINDINGS OF THE RESEARCH

The findings of this research are listed below

- Microarray data classification is a learning task that can predicts the diagnostic category of a sample from its expression array phenotype.
- Gene selection method frequently chooses the top-ranked genes according to their individual discriminative control without conduct the elevated degree of redundancy among the genes.
- Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product.
- Feature selection technique is used to select the most possibly cancer related genes from huge microarray gene expression data. It aims to achieve improved classification performance by comparing huge number of genes to the number of tissue samples.
- ANOVA technique is efficient method for gene ranking which will ranks the genes according to the relevancy between genes. This will be simplified the feature selection process in the classification.
- The newly developed integrated (gene ranking, feature selection and classification) Extreme Learning Machine (ELM) is standard algorithm for the cancer diagnosis area mainly it is used for direct multi-category classification problems.
• Proposed hybrid Genetic Algorithm and Extreme Learning Machine classifier is the best alternative for the machine learning based classification.
• Proposed hybrid Grey Wolf optimizer (GWO) and Extreme Learning Machine classifier is a best method for evolutionary based hybrid classifier for cancer classification.

Discussion of this Research

Based on this research it is conclude that cancer classification is a significant field of research for cancer based on microarray gene expression profile. Also, most cancer research in microarray expression profiling have proved a good performance. Microarray is considered an efficient technique for cancer classification and also for diagnosis, prognosis and treatment purposes. Over the decade DNA microarray technique has gained more attraction in both scientific and in industrial fields and to determine the informative genes which cause the cancer to improve early diagnosis and to provide chemotherapy treatment. Classifying cancer datasets based on microarray gene expression data is a challenging task because microarray is high dimensional low sample dataset with lots of noisy or irrelevant genes and missing data. The inherent presence of a large number of irrelevant genes increases the difficulty of the classification task influencing the discrimination power of relevant features. So the proposed hybrid classifiers are provided the best solution.

7.2 CONCLUSION

Cancer classification is important for treatment and drug discovery. Cancer is one of the significant characteristic in the biomedicine field and classification is important for treatment and drug discovery. Diagnosis of multi-category cancer problem has become increasing active research area. Exact calculation of numerous cancer types has higher value for improved treatment and toxicity it decrease on the patients. In traditional, cancer classification is usually depends on morphological and clinical analysis. This research work
reported the recent developments in algorithms and applications of classifiers. Microarray or gene expression profiling is applied to compare and determine the gene expression level pattern for different cell types or tissue samples in a single experiment. In this research work evaluation of gene expression has been carried out by using various algorithms. The overall performance can be assessed using the performance Area Under Curve (AUC). The obtained AUC values for LDA with Successive Feature Selection (SFS), Extreme Learning Machine (ELM), Proposed Feature Selection with SFS and ELM, Hybrid Genetic Algorithm (GA) – ELM and Hybrid Grey Wolf Optimization - ELM are 0.65, 0.72, 0.85 and 0.94. It is observed that the proposed algorithm finds a small gene subset which provides high classification accuracy on several DNA microarray gene expression datasets.

**Microarray Gene Expression on ELM**

The Chapter 4 concentrates on ANOVA based gene ranking, Extreme Learning Machine (ELM) classifier is combined with Successive Feature Selection (SFS) method for better diagnosis. The need for cancer classification has become essential because the number of cancer sufferers is increasing. To overcome problems of multi-category cancer classification microarray gene expression data classification using fast Extreme Learning Machine is used. It rectify problems which are faced by iterative learning methods such as local minima, improper learning rate and over fitting and the training completes with high speed. The experimental results shows that proposed ranking, feature selection with SFM and ELM gives high accuracy than other traditional algorithms. To improve the effectiveness of nodes in the hidden layer in ELM a hybrid Genetic algorithm with ELM technique is implemented.

**Hybrid Genetic Algorithm-ELM**

The Chapter 5 concentrates on novel Hybrid Genetic algorithm is combined with ELM to classify the gene expression data. This work deals with finding a small subset of informative genes from gene expression microarray
data which can maximize the classification accuracy by ignoring the noisy features, the noisy features are retained due to poor fitness functions which are generated by bad chromosome blocks are eliminated in this phase. The experimental results shows that hybrid GA-ELM which improves the fitness functions and managing the imbalanced data of classification. To rectify this issue grey wolf optimization based technique is introduced.

**Grey Wolf Optimization Based on ELM**

The Chapter 6 concentrates on integrating a Grey Wolf Optimization (GWO) and Extreme learning machine. GWO method is used for selecting the particular features and ELM classifier used for extracting the functional relationship between the selected features and class labels. The Experimental results shows that GWO-ELM is best for selecting genes and it makes classification easy. The proposed technique is used to choose the top genes. Then the classifier is trained with highly correlated samples to minimizing errors and it proved that GWO-ELM has higher classification accuracy, precision, recall and area under curve.

### 7.3 FUTURE WORK

By constantly improving, gene expression technology it provides the ability to measure the expression levels of thousands of genes in single chip. The most vital process for gene is identification and classification of cancer gene expression data is expected to significantly aid in the development of efficient cancer diagnosis and classification platforms. Many of the feature selection algorithms produced fault for their ranked gene performance. To prevent this method produces the better accuracy by producing a feature selection algorithm in gene expression data analysis of sample classifications.

Future works will also continue to compare the use of linear and non-linear functions in ranking, as well as to investigate the performance of different parameters of the algorithm. The future works by applying the proposed methodology to more practical problems and bioinformatics problems, which is
beyond the scope of this review and plan to implement this method in a parallel way with the aid of high performance tools. A new domain statistical method would be considered. The statistical method has a capability to indicate a classification accuracy on test data and also it can be extended to cancer images such as MRI, CT-Scan features etc.