Cancer classification is one of the most challenging studies for scientists today. With the advent of DNA microarrays technology, constructing gene expression profiles for different cancer types has already become a promising means for cancer classification. However, it offers a challenge for current machine learning research. Microarray datasets are characterized by high dimension and small sample size. Over-fitting is a major problem due to the high dimension, while the small data size makes it very difficult to obtain good performance from machine learning algorithms.

Support vector machine (SVM) is a statistical classification algorithm that classifies data by separating two classes with the help of a functional hyper plane. SVM is known for good performance on noisy and high dimensional data such as microarray. One main disadvantage of using SVMs is that the performance of classifier depends on setting of parameters. In this thesis, we do classify gene expression data sets, namely, colon, leukemia, lung, ovarian, prostate and breast. The experimental results show that the classification accuracy rates of the proposed method are competitive to that of other existing methods. It can be used as an efficient computational tool for microarray data analysis.

We employ the combination of PSO and simulated annealing (SA) to determine proper setting of SVM parameters which can improve the quality of SVM model. Our approach is a combination of methods. The motivation is to bring out an effective classification method for cancer by utilizing the strength of various techniques and compensating for their weaknesses. The proposed approach is tested on six benchmark cancer datasets and the results show that the proposed method can improve the classification accuracy rates of SVMs when compared to existing methods.
classification accuracy (%) .................................................................................... 24

Table 3-6. The comparison of our method with others ................................................ 25

REFERENCES