

Chapter 1

Introduction

Chapter 1

Introduction

This thesis mainly introduces several novel techniques to enhance cancer classifiers accuracy. It first introduces a novel feature selection method called MLFS that is based on integrating both deep and active learning. Then, it introduces new semi-supervised cancer classification methods that integrate both labeled and unlabeled samples along with miRNA and gene expression profiles.

The following sections discuss the motivation, the objectives to be achieved and the organization of the thesis.

1.1 Motivation

Our main motivation is to enhance cancer classification using different methods. First, cancer classification enhancement was explored from the feature selection perspective, how to select the smallest number of genes/miRNAs that will lead to enhancing cancer classifiers. In order to enhance feature selection approaches, our proposed method focuses on selecting genes based on gene grouping behavior, mainly how they perform together. It focuses also on integrating both deep and active learning while maintaining the ability to identify the selected genes/miRNAs. Moreover, cancer classification enhancement was explored from the classification perspective by integrating both labeled and unlabeled sets and integrating both miRNA and gene expression profiles using semi-supervised machine learning (self-learning and co-training). Also, biology relation between genes and miRNAs were explored and used in both approaches by using miRNA-target gene biology relation available in databases

like miRanda [4].

1.2 Research Objectives

The objectives of the thesis can be summarized as given below:

1. Propose new feature selection and classification techniques that improve sample cancer classification accuracy.
2. Integrate both gene and miRNA expression profiles to enhance cancer classifiers.
3. Integrate both labeled and unlabeled cancer expression sets to enhance cancer classifiers.
4. Select genes in feature selection method based on gene grouping behavior, mainly how they perform together.
5. Maintain the ability to identify genes/miRNAs that are output of the feature selection method to give it to biologists to analyze.

1.3 Thesis Organization

This thesis is divided into 5 chapters. The remaining chapters are organized as follows:

Chapter 2: presents the background knowledge and related work discussion.

Chapter 3: describes our proposed feature selection method (MLFS) and its performance evaluation results.

Chapter 4: describes our proposed miRNA and gene expression based cancer classification using self-learning and co-training approaches. At the end performance evaluation results of the approaches are presented.

Chapter 5: presents thesis conclusion and suggestions for future work.