



ORIGINAL ARTICLE

Predicting 5-Year Survival Status of Patients with Breast Cancer based on Supervised Wavelet Method

Maryam Farhadian^a, Hossein Mahjub^{b,*}, Jalal Poorolajal^c,
Abbas Moghimbeigi^d, Muharram Mansoorizadeh^e

^aDepartment of Epidemiology and Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

^bResearch Center for Health Sciences and Department of Epidemiology and Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

^cModeling of Noncommunicable Diseases Research Center, Department of Epidemiology and Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

^dModeling of Noncommunicable Disease Research Center, Department of Biostatistics and Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

^eDepartment of Computer Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamadan, Iran.

Received: August 4,
2014
Revised: September
15, 2014
Accepted: September
22, 2014

KEYWORDS:

breast cancer,
microarray data,
supervised wavelet,
support vector machine

Abstract

Objectives: Classification of breast cancer patients into different risk classes is very important in clinical applications. It is estimated that the advent of high-dimensional gene expression data could improve patient classification. In this study, a new method for transforming the high-dimensional gene expression data in a low-dimensional space based on wavelet transform (WT) is presented.

Methods: The proposed method was applied to three publicly available microarray data sets. After dimensionality reduction using supervised wavelet, a predictive support vector machine (SVM) model was built upon the reduced dimensional space. In addition, the proposed method was compared with the supervised principal component analysis (PCA).

Results: The performance of supervised wavelet and supervised PCA based on selected genes were better than the signature genes identified in the other studies. Furthermore, the supervised wavelet method generally performed better than the supervised PCA for predicting the 5-year survival status of patients with breast cancer based on microarray data. In addition, the proposed method had a relatively acceptable performance compared with the other studies.

Conclusion: The results suggest the possibility of developing a new tool using wavelets for the dimension reduction of microarray data sets in the classification framework.

*Corresponding author.

E-mail: mahjub@umsha.ac.ir

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Metastatic breast cancer is a stage of breast cancer where the disease has spread to distant organs or tissues. Treatments against metastasis exist, but usually further treatments after surgery can have serious side effects and involve high medical costs [1]. An important task to optimize the adjuvant chemotherapy of metastasis related to breast cancer is to diagnose the risk of metastasis accurately [2–4].

Classification of cancer patients into different risk classes is very important in clinical applications. Traditional methods for patient classification were mainly based on a series of clinical and histological features [3]. It is estimated that the advent of high-dimensional gene expression data could improve patient classification [5]. Gene expression profiles of breast tumor samples could be used to predict relapse and metastatic patterns in breast cancer patients that could be potential candidate targets for new treatments [4]. It is reasonable to assume that any difference between the two tumors should be represented by some difference in gene expression. However, in microarray studies, the number of samples is relatively small compared to the number of genes per sample. Furthermore, from the biological aspect, only a small portion of genes have predicted the power for phenotypes. If all or most of the genes are considered in the predictive model, they can induce substantial noise and thereby lead to poor predictive performance [6]. Thus, in order to obtain good classification accuracy, a crucial step towards the application of microarray data is the dimensional reduction from the gene expression profiles. In recent years, both feature selection and feature extraction methods have been widely used for classifying gene expression data [7]. Bair and Tibshirani [8] and Bair et al. [9] explored the use of supervised principal component analysis (PCA), which is similar to conventional PCA except that it uses a subset of the predictors selected based on their association with the outcome. Wavelet-based methods have also been used to solve the dimension reduction problem. The primary intuition for applying wavelets in the case of gene expression is that genes are often coexpressed in groups. Therefore, it would be useful to treat the group as a single variable, akin to the motivation behind methods such as PCA [10]. One-dimensional discrete wavelet transform (DWT) is frequently used for feature extraction in the analysis of high-dimensional biomedical data [11]. Studies showed that this method has an acceptable performance in the field of feature extraction in the classification framework [11–15].

The current study aimed to introduce a dimension reduction strategy for transforming the high-dimensional gene expression data in a low-dimensional space based on wavelet transform (WT) in order to predict metastasis

of breast cancer. Accordingly, a predictive support vector machine (SVM) model was built upon the reduced dimensional space. Then, the proposed novel supervised wavelet method of feature extraction was compared with the supervised PCA.

2. Materials and methods

The proposed method was applied to three publicly available microarray data sets related to breast cancer.

2.1. Data

2.1.1. Breast cancer data from van't Veer (NKI_97)

The first data set is reported by van't Veer et al [2] and referred to as NKI_97. The original van't Veer data consists of gene expression profiles and clinical information for 97 samples of primary breast cancer tumors, and each case is described by the expression levels of 24,481 genes. Fifty-one patients remained free from metastasis for at least 5 years and were metastasis-negative, and 46 cancer patients developed metastasis within 5 years and were metastasis-positive. All patients were <55 years old and were lymph node-negative. They had no tumor cells in local lymph nodes [2]. The data used in this study is a filtered version of the van't Veer data including gene expression values of 4948 genes in 97 tumor samples [2]. The data are publicly available at the “cancer data” R package (<http://www.bioconductor.org/packages/release/data/experiment/html/cancerdata.html>).

2.1.2. Breast cancer data from van de Vijver (NKI_295)

The second data set is reported by van de Vijver et al [4] and referred to as NKI_295. The data set provides the gene information for 295 primary breast cancer patients, of which 234 patients were new and the remaining 61 patients were involved in the first data set. Of the total 295 patients, 194 patients were metastasis-negative and 101 patients were metastasis-positive. Of the 234 new patients, 164 patients were metastasis-negative and 70 patients were metastasis-positive. Of the 61 patients involved in the first data set, 30 were metastasis-negative and 31 patients were metastasis-positive. The data is a filtered version of the van de Vijver data including gene expression values of 4948 genes in 295 tumor samples [4]. The data are publicly available at the “cancer data” R package.

2.1.3. Breast cancer data from the Wang study (VDX_286)

The last data set, reported by Wang et al [16] and referred to as VDX_286, contains 286 lymph node-negative breast cancer patients who had not received any adjuvant systemic treatment [16]. Among them, 106

Download English Version:

<https://daneshyari.com/en/article/4202013>

Download Persian Version:

<https://daneshyari.com/article/4202013>

[Daneshyari.com](https://daneshyari.com)