UNIVERSITI PUTRA MALAYSIA

PREDICTION OF BREAST CANCER RELAPSE TIME IN CONTINUOUS SCALE BASED ON TYPE-2 TSK FUZZY MODEL

SAYED HAMID MAHMOUDIAN

FK 2010 75

PREDICTION OF BREAST CANCER RELAPSE TIME IN CONTINUOUS SCALE BASED ON TYPE-2 TSK FUZZY MODEL

By

SAYED HAMID MAHMOUDIAN

Thesis Submitted to the School of Graduate Studies, University Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

November 2010

© COPYRIGHT UPM

Abstract of thesis presented to the senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

PREDICTION OF BREAST CANCER RELAPSE TIME IN CONTINUOUS SCALE BASED ON TYPE-2 TSK FUZZY MODEL

By

SAYED HAMID MAHMOUDIAN

November 2010

Chairman : Associate Professor Mohammad Hamiruce Marhaban, PhD

Faculty : Engineering

Recently, microarray analysis and gene expression profiles have been widely applied in diagnosis and classification of different types of cancer such as liver, colon or breast cancer. As the number of breast cancer cases increased dramatically in many countries including Malaysia in recent decades, different types of studies have been done to control the disease or reduce the cost of their treatments. Gene expression profiles, which can screen the behavior of a large number of genes simultaneously, have been used in some studies to extract the significant genes related to breast cancer. Tumor classification, Estrogen Receptor status recognition or survival analysis has been usually considered as important objectives in these studies. Due to the fact that studies in survival analysis of breast cancer can reduce the cost of treatments and side effects of the adjuvant therapy,

© COPYRIGHT UPM

different methods for predicting the outcome of the disease have been proposed by previous researcher.

The two major objectives of this research are to propose a fuzzy classifier to discriminate breast cancer tumors into two classes, which are high risk and low risk by some interpretable rules similar to linguistic words, and to predict the relapse time of breast cancer by TSK fuzzy models in continuous scale.For this reason, breast cancer dataset has been applied for training the models and two other independent samples have been used for validating the results. In addition, K-fold Cross Validation, B632 and B632+ methods have been used for error estimation.

In the first objective of the thesis, a lemma has been proven and a new hybrid algorithm based on Fuzzy Association Rule Mining has been proposed to gather some selected genes and generate fuzzy rules for classification.

In the second one, a method for generating the fuzzy rules to discriminate the samples of breast cancer into the different groups have been proposed and applied to predict the relapse time of samples in continuous scale while handling the uncertainties in linguistic terms of the rules.

© COPYRIGHT UPM

The relapse time of two available independent samples of breast cancer have been predicted by the model and the results show the superiority of the proposed model with respect to the previous study. Finally 46 significant genes and 16 fuzzy rules have been introduced which can be used in a Type-2 TSK fuzzy model as a predictor.

© COPYRIGHT UPM

ABSTRAK

Abstrak tesis dikemukakan kepada Senat Universiti Putra Malaysia sebagi memenuhi keperluan untuk ijazah Doktor Falsafah

RAMALAN MASA RELAPS DALAM SKALA SELANJAR BAGI KANSER PAYUDARA BERDASARKAN MODEL KABUR TSK JENIS-2

Oleh

SAYED HAMID MAHMOUDIAN

November 2010

Pengerusi: Profesor Madya Mohammad Hamiruce Marhaban, PhD

Fakulti: Kejuruteraan

Kebelakangan ini, analisa jujukan mikro dan profil ekspresi gen telah digunakan dengan meluas dalam diagnosis dan pengelasan beberapa jenis kanser seperti kanser hati, kolon atau payudara. Disebabkan kes kanser payudara telah meningkat bilangannya dalam dekad ini di kebanyakan negara termasuk Malaysia, pelbagai kajian telah dijalankan untuk mengawal penyakit tersebut atau mengurangkan kos rawatannya. Profil ekspresi gen berupaya untuk mencerminkan tingkah laku gen berjumlah besar dengan serentak, telah digunakan dalam beberapa kajian mengekstrak gen-gen penting yang berkaitan dengan kanser payudara. Pengelasan tumor, pengecaman status penerima estrogen atau analisa peluang hidup biasanya dianggap sebagai objektif utama untuk kajian-kajian ini. Memandangkan kajian tentang peluang hidup daripada kanser payudara boleh

© COPYRIGHT UPM

mengurangkan kos rawatan dan kesan sampingan daripada terapi adjuvan, pelbagai kaedah untuk meramal hasil penyakit tersebut telah dicadangkan oleh para penyelidik.

Dua objektif utama kajian ini ialah untuk mencadangkan satu pengelas kabur dalam membezakan tumor kanser payudara kepada dua kelas iaitu kelas berisiko tinggi dan kelas berisiko rendah, mengikut peraturan yang boleh diinterpretasi, menggunakan kata-kata linguistik serta meramalkan masa relaps bagi kanser payudara menggunakan model kabur TSK dalam skala selanjar.

Untuk itu, set data kanser payudara van't Veer telah digunakan dalam melatih model-model yang dicadangkan serta 2 lagi sampel bebas yang diterbitkan oleh van't Veer dan van de Vijver telah digunakan untuk mengesahkan keputusan yang diperolehi. Kaedah keesahan silang, B632 dan B632+ telah digunakan untuk menganggarkan ralat.

Bagi objektif pertama, satu lemma telah dibuktikan dan satu algoritma hibrid berdasarkan Perlombongan Peraturan Berkaitan Kabur telah dicadangkan untuk mengumpul beberapa gen yang terpilih dan menjana peraturan kabur untuk pengelasan.

© COPYRIGHT UPM

Dalam objektif kedua, satu kaedah untuk menjana peraturan kabur dalam membezakan sampel-sampel kanser payudara kepada kumpulan-kumpulan yang berbeza telah dicadangkan dan diaplikasikan untuk meramal masa relaps sampel-sampel tersebut dalam skala selanjar serta mengendali ketidakpastian dalam terma linguistik peraturan-peraturan.

Masa relaps bagi 2 sampel kanser payudara yang bebas telah diramalkan oleh model tersebut dan hasil menunjukkan model yang dicadangkan jauh lebih baik daripada model-model dari kajian sebelum ini. Subset gen-gen yang penting dan peraturan kabur yang dijana untuk ramalan masa relaps juga dikemukakan. Akhirnya, 46 gen penting dan 16 aturan kabur telah dicadangkan yang mana ia boleh digunakan oleh model kabur TSK Jenis-2 sebagai peramal.

© COPYRIGHT UPM

ACKNOWLEDGEMENTS

First and foremost, my praise to Allah Jalla Jalaloh, who blessed me with patience, courage, consistency and good health during this study, I would like to express my thanks to the chairman of my supervisory committee Associate Professor Dr Mohammad Hamiruce Marhaban, who helped me to be able to continue this study and gave me great guidance, suggestion and encouragement. My gratitude goes to the member of supervisory committee, Professor Dr Raha Abdul Rahim, Associate Professor Dr. Rozita Rosli and Dr. M. Iqbal Saripan for his kind guidance.

Moreover, I am thankful to the staff in the department of Electrical and Electronic Engineering that helped me to continue this study.

Finally, it is needed to express the heartfelt thanks to my wife who support me in my research and patience, care and encouragement during the study. I am very grateful for my mother, for always being there when I needed her and also my kids Nima and Raha who encourage me to work hard by their beautiful smiles. I would also like to express my gratefulness to my friends for giving me a good time here.

© COPYRIGHT UPM

I certify that a Thesis Examination Committee has met on (July 2010) to conduct the final examination of Hamid Mahmoodian on his thesis entitled "Prediction of Breast Cancer Relapse Time in Continuous Scale Based on Type-2 Fuzzy Model" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Noorhisam b. Misron, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Abdul Rahman b. Ramli, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

Syed Abd. Rahman Al-Attas, PhD
Associate Professor
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
(External Examiner)

Golshah Naghdy, PhD
Associate Professor
Faculty of Electrical Computer and Telecommunications Emgimeering
University of Wollongon
Country Australia
(External Examiner)

BUJANG KIM HUAT, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

© COPYRIGHT UPM

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Mohammad Hamiruce Marhaban, PhD
Associate Professor
Department of Electrical and Electronic Engineering
University Putra Malaysia
(Chairman)

Raha Abdul Rahim, PhD
Professor
Faculty of Biotechnology and Biomolecular Sciences
University Putra Malaysia
(Member)

Rozita Rosli, PhD
Associate Professor
Faculty of Medicine and Health Sciences
Universiti Putra Malaysia
(Member)

M. Iqbal b. Saripan, PhD
Lecturer
Department of Electrical and Electronic Engineering
Universiti Putra Malaysia
(Member)

HASANAH MOHD GHAZALI, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

© COPYRIGHT UPM

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institutions.

SAYED HAMID MAHMOODIAN

Date: 29 November 2010

© COPYRIGHT UPM

TABLE OF CONTENTS

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM