



دانشگاه علوم پزشکی کرمان

دانشکده بهداشت



پایان نامه مقطع کارشناسی ارشد رشته آمار زیستی

عنوان:

پیش بینی مرگ و میر بیماران سرطان پستان با استفاده از مدل مخفی مارکف

توسط: فرزانه مددی زاده

استاد راهنما: دکتر عباس بهرام پور

استاد مشاور: خانم میترا منتظری

سال تحصیلی: ۱۳۹۴ - ۱۳۹۳



Prediction of Breast Cancer mortality by Hidden Markov Model

A Thesis

Presented to

The Graduate Studies

By

Farzan Madadizadeh

In Partial Fulfillment

Of the requirements for the Degree

Master of Science in:

Biostatistics

Kerman University of Medical Sciences

January 2015

مقدمه و هدف :

سرطان پستان یکی از شایع ترین علل مرگ و میر در میان زنان جهان است. هر چند پیشرفت های چشم گیری در حوزه علوم پزشکی اتفاق افتاده و عوامل خطر سرطان پستان شناخته شده اند اما تشخیص مرگ و میر سرطان پستان برای متخصصان کم تجربه امری زمان بر است. بنابراین نیاز ضروری به مدل های پیشگو جدید است تا هم با دقت عمل پیش بینی را انجام دهند و هم صرفه جویی در زمان صورت گیرد. روش های پیش بینی و طبقه بندی ابزارهای مفید و صرفه جو در زمان جهت یاری به متخصصان در تشخیص سریع تر بیماری هستند. در سالهای اخیر به مدل های آماری جهت طبقه بندی داده های پزشکی بر طبق بیماری های مختلف و برآمد های آنها توجه شایانی شده است. مدل مخفی مارکف به صورت موفقیت آمیزی جهت تشخیص الگو و پیش بینی در مطالعات بالینی به کار رفته است. هدف اصلی این پژوهش ارزیابی توانمندی مدل مخفی مارکف در پیش بینی مرگ و میر سرطان پستان است.

مواد و روش ها : داده های مربوط به سرطان پستان که در این پژوهش استفاده شده اند از مرکز ثبت سرطان استان کرمان کسب شده اند. داده ها شامل اطلاعات ۹۰۰ بیمار سرطان پستان ۱۵ الی ۹۰ ساله و عوامل خطر مربوط به آنها از سال ۱۳۷۸ تا سال ۱۳۸۶ بوده است. مدل مخفی مارکف بر اساس ۷۵ درصد داده ها (اطلاعات ۶۷۵ بیمار) آموزش یافته و بر اساس ۲۵ درصد داده ها (اطلاعات ۲۲۵ بیمار) ارزیابی شده است. در ساخت مدل مخفی مارکف تعداد ۵ حالت با ۳۰۰ تکرار در الگوریتم EM به کار رفته است. علاوه بر این مدل نهایی برای دقت بیشتر ۱۰۰ بار تکرار شده است. ناحیه زیر منحنی راک، حساسیت، ویژگی و دقت به عنوان معیارهای ارزیابی کارایی مدل در پیش بینی وضعیت مرگ و میر بیماران استفاده شده است.

یافته ها : در این مطالعه، حساسیت، ویژگی، دقت و سطح زیر منحنی مشخصه عملکرد (ROC) برای مدل مخفی مارکف به ترتیب برابر ۰/۹۸۹، ۰/۹۹۲، ۰/۹۳۹ و ۰/۹۶۴ بود.

نتیجه گیری : بر طبق چهار معیار ارزیابی مدل مخفی مارکف عملکرد خوبی ارائه می دهد و یک ابزار مناسب برای پیش بینی مرگ و میر سرطان پستان می باشد.

واژگان کلیدی : مدل مخفی مارکف، الگوریتم EM، سرطان پستان، بقا

Abstract

Background & Objective: Breast cancer is one of the most widespread causes of death among women worldwide. Although significant progress has been made in the field of medical science, and prognostic risk factors of breast cancer have been identified, diagnosis of breast cancer mortality is time consuming for novice specialist. So there is an urgent need for new predictive models to do careful forecasting and save time. Statistical prediction and classification methods are a helpful and time saving tool for aiding technicians to diagnose the diseases more quickly. In the recent years, considerable attention has been to the statistical models for classification of medical data according to various diseases and its outcomes. Hidden Markov Models have been successfully used for pattern recognition and prediction due to in the some clinical studies. The main objective of this research was to evaluate empirically predictive ability of hidden markov model in prediction of breast cancer mortality.

Methods: The breast cancer data set used in this investigation was acquired from Cancer Registry Organization of Kerman University of medical science in Kerman province, Iran. This data includes the information of 900 breast cancer patients aged 15 to 90 years, and its associated risk factors among since 1999 to 2007. Hidden markov model was made based on a set of 675 patients information (75% dataset) and they were validate in a test set of 225 patients informations (25% dataset). In the hidden markov model five state with 300 iteration in EM algorithm was employed furthermore the final model for more accuracy was repeated 100 times. . The Area Under the ROC Curve (AUC), sensitivity, specificity and accuracy used as measures of validate of the efficiency of model in prediction of patients mortality status.

Results: In this study, the sensitivity, specificity, accuracy and the area under the ROC curve of the hidden markov model was 0.989, 0.99, 0.939 and 0.964, respectively.

Conclusions: According to the four evaluation criteria, hidden markov model give good performance and that is an appropriate tool for prediction breast cancer mortality.

Keywords: Hidden Markov Model (HMM), EM algorithm, Breast cancer, survival