Analysis of Risk Factors for Breast Cancer Decision Support System in Egypt

Basma Emad Abd El-Fatah, Department of Biomedical Engineering and Systems, Cairo University, Giza, Egypt
Mohamed I. Owis, Department of Biomedical Engineering and Systems, Cairo University, Giza, Egypt
Manal Abdel Wahed, Department of Biomedical Engineering and Systems, Cairo University, Giza, Egypt

ABSTRACT

Breast cancer is the most common form of cancer among women. A risk factor is anything that influences the possibility of obtaining a disease as cancer. The goal of this study is creating economic method for the early detection of breast cancer using risk factors. Two approaches were followed by applying classification techniques. In the first approach, benign, malignant and normal were classified. The second approach comprised two phases. In the first phase, normal and tumor cases were detected, then in the second phase benign and malignant cases were detected. Risk factors were ranked by two different feature selection methods. The best result accuracy of the first phase of the second approach was 96.7%. This could help in the detection of normal and tumor cases without mammography giving a fast diagnosing result, training of non-expert doctors and reducing mammography harms and false diagnosis of inexperienced doctors.

KEYWORDS

Classification Techniques, Data Mining, Early Detection, Economic Method, Fast Diagnosing, MRMR, Random Forest

INTRODUCTION

A risk factor is anything influences the possibility of obtaining sickness such as cancer. Breast cancer is the leading cause of cancer death in women aged 20 to 59 years (Siegel, 2015). In Egypt female breast cancer is the most frequent cancer as estimated from the National Population-Based Registry Program of Egypt (from 2008 till 2011) where the ratio was 32.04% (Ibrahim, 2014). It is tricky where in early stages the tumor is still tiny so no symptoms appear but in the same time the early detection cases are well treatable.

Organized national mammographic screening has been adopted as the gold standard for breast cancer early detection in western countries; however, it may not be the most cost-effective approach to early detection in low and middle income countries (LMC) as it is very demanding in terms of human and financial resources (Corbex, 2012). The analysis of mammograms is affected by the experience of the crew and need a lot of crew time and exertion that cause slowing of diagnosing period.

Health care resources are limited in a lot of nations. The expenses connected with mammographic screening approach are as high as purchasing new devices, radiologists and radiographers training, running costs, etc. for example the cost of the WHOP fully equipped mobile unit is about 6.75 million Egyptian Pounds (EGP) and its annual running costs is about 1.6 million EGP (Rylands-Monk, 2012). The unwanted impacts of mammography include the secondary effect of the test itself as radiation and delay of diagnosis. Disadvantages from false positive examination include the extra testing, invasive diagnostic steps and anxiety (Corbex, 2012). From a systematic survey of the related researches, around

DOI: 10.4018/IJBCE.2017010102

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
half of the ladies suffer from pain at mammography examinations where the breasts are compressed between two plates although that the duration of the test takes a moment (Gotzche, 2012).

Despite being less costly, free of radiation, widely available and better tolerated by patients than most imaging modalities, Ultrasound Scan is operator dependent and demands time and skill (Drukteinis, 2013). MRI is expensive, taking more time and not available in every hospital in Egypt. And the objective of this work is screening of applicant women for the further expensive examinations.

The most accurate way for breast cancer diagnosis is biopsy. However, biopsy may cause patient pain, bleeding and contamination in addition to unnecessary expenses spent for examinations. Thus, reducing unnecessary biopsy is an important goal and it is necessary to acquire successful identification mode to recognize breast cancer.

Breast cancer prevention is still a remote objective; accordingly, auxiliary prevention through early detection is the main possible methodology now. The goals of this study are creating an economic method for the early detection of breast cancer with no pain to patient, giving a fast diagnosing result in addition to training of non-expert doctors and trying to reduce mammography harms and false diagnosis of inexperienced doctors. High quality mammography images have high resolution and big size, consequently handling these images need high computational capacities. Thus, this study focuses on risk factors instead of mammography images. The objective is to build a computer system that supplies beneficial analysis tools that can assist radiologists in diagnosing for better performance.

Yoo and colleagues (Yoo,2012) collected some recent studies of classification applications; it was observed that the decision tree algorithm is the most public classification algorithm as a result of its ability to visually illustrate how classification decisions can be made.

Keles and Colleagues (Keles, 2011) used mammography mass data set to predict the severity of a mammographic mass lesion and decided if it is benign or malignant, fuzzy rules acquired by neuro-fuzzy technique were used in their expert system. The performance evaluation of their system for breast cancer diagnosing was 97% specificity, 76% sensitivity.

Healthcare data are utilized for classification techniques comparison. Decision tree method (CART), Discriminant Analysis, logistic regression model and artificial neural network were used in that comparison. CART executed the optimum overall accuracies and considered the second speedy technique within them (Harper, 2005).

In a recent study (Papageorgiou, 2015) a Fuzzy Cognitive Map based familial breast cancer risk management system was developed. The resulted accuracy for 40 patient records was 95%.

The goal of this research study is creating economic method for the early detection of breast cancer using risk factors. We followed two approaches; in the first approach, we classified the three cases, benign, malignant and normal by applying classification techniques of decision tree and random forest. On the other hand the second approach comprised two phases. In the first phase, we detected two categories of normal and tumor cases, applying decision tree and random forest. In the second phase, we detected benign and malignant cases using decision Tree and random forest. We ranked risk factors by two different feature selection methods: Fisher linear discriminant (FLD) and minimal redundancy-maximal relevance criterion (MRRM). We used in classification techniques 10 folds cross-validation method to decrease the bias of results. The approaches and their results are presented in the next sections.

METHODS

Data Collection and Preparation

The data set was provided from Women Health Outreach Program (WHOP). WHOP is the first Egyptian national examination program for early detection of Egyptian women breast cancer (Salem, 2008). Data were collected from 2007 till 2013, containing some important risk factors of breast cancer for each examined woman. Data preparation is consisting of data selection, data cleansing and data construction.
Neural Network Based Automated System for Diagnosis of Cervical Cancer

Networkcentric Healthcare and the Entry Point into the Network
[www.igi-global.com/chapter/networkcentric-healthcare-entry-point-into/26208?camid=4v1a](www.igi-global.com/chapter/networkcentric-healthcare-entry-point-into/26208?camid=4v1a)