

UNIVERSITY OF SCIENCE AND TECHNOLOGY
COLLEGE OF GRADUATE STUDIES AND ACADEMIC
ADVANCEMENT

Faculty of Computer Science and Information Technology

A Neuro-Fuzzy Inference Model for Breast Cancer
Classification

By

Abeer Ganem Alriah Hussain

A Thesis

Submitted to the College of Graduate Studies and Academic Advancements
in Partial Fulfillment of the Requirement for the Degree of Master of
Science in Information Technology

Supervisor

Dr. Abdelhamid Salih

Sept 2016

Abstract

Breast cancer is one of the most common cancer types. It has been one of the major causes of death among women and true emergency for. The investigation of machine learning in medical diagnosis helps medical experts in diagnosis diseases. Several machine learning algorithms and techniques were used for classification breast cancer diseases. The hybrid Neuro-fuzzy system in soft computing is one of effective methods used for classification. In this research we investigated hybrid Neuro-fuzzy model with attribute selection methods to select the best attributes, the data set used in this research was collected from varies Khartoum state hospitals and by investigating ANFIS tool classified data into two classes, the first class contains all patients in early stage of Breast cancer and the other class contains patients in late stage or in danger stage Breast cancer. The classification dependent on stage (TNM) and grade of tumor, with the knowledge that all cases of patients were diagnosed as malignant tumor. The results obtained after training the data set on ANFIS model with 100 epoch, the error tolerance obtained is 0.16565 and after testing the error obtained is 0.14936. According to results obtained the model can help and improve assisted health care monitoring.

المستخلص

سرطان الثدي هو واحد من أكثر أنواع السرطان شيوعاً. يعتبر واحداً من الأسباب الرئيسية للوفاة بين النساء.

يمثل خطر حقيقي على أنظمة الرعاية الصحية في البلدان الصناعية. استخدام أدوات التعلم الآلي في المجال الطبي يساعد الخبراء الطبيين في تشخيص الأمراض. هناك العديد من التقنيات والخوارزميات التي تعتبر مفيدة في هذا المجال. يعتبر النظام الهجين بين المنطق الغامض والشبكات العصبونية في الحوسبة المرنة هي واحدة من الطرق الفعالة المستخدمة لتصنيف البيانات. في هذا البحث تم استخدام 1421 سجل من البيانات التي تم جمعها من مختلف مستشفيات ولاية الخرطوم وتم استخدام خوارزمية ANFIS لتقوم بتصنيف البيانات إلى فئتين حيث تحتوي الفئة الأولى على كافة المرضى في مرحلة مرض سرطان الثدي المبكر والفئة الأخرى يحتوي على حالات المرضى في المراحل المتأخرة أو في مرحلة الخطر. هذا التصنيف اعتمد على مرحلة الورم (tnm) ودرجة الورم، مع العلم أن جميع حالات المرضى التي تم تشخيصها مسبقاً من قبل أطباء مختصين ببيانات المرضى جميعهم مصابين بالمرض. بعد تدريب البيانات على نموذج ANFIS حوالي 100 مره تم الحصول على معدل الخطأ حوالي 0.16565 وبعد اختباره عن طريق فحص البيانات التي تم الحصول عليها كان الخطأ 0.14936

1.1 Introduction

This chapter describes the research work, states the problem, defines research objective, scope of research and describe thesis structure.

1.2 Background of the Research

Breast cancer is one of the most common cancer types. Due to the complexity of this disease, it is important to face its study with an in targeted and multilevel approach, from genes, transcripts and proteins to molecular networks, cell populations and tissues. Breast cancer has been one of the major causes of death among women and a true emergency for the healthcare systems of industrialized countries. It is the second leading cause of cancer deaths among women in the world .It is characterized by an abnormal multiplication of a cell in the human body. Not entailing serious consequences, early, cancer can be developed into a serious condition if treatment is not done on time. Due to its late diagnosis, it often causes a mutilating and expensive treatments accompanied by a high mortality rate. It has the form of lumps or tumors in the tissues of the breast[1, 2]. Tumors can either be malignant or benign. Machine learning involves adaptive mechanisms that enable computers to learn from experience, learn by example and learn by analogy. Learning capabilities can improve the performance of an intelligent system over time. Machine learning mechanisms form the basis for adaptive systems[3]. The most popular approaches to machine learning are artificial neural networks and genetic algorithms. A neural network can be defined as a model of reasoning based on the human brain. The brain consists of an interconnected set of nerve cells, or basic information-processing units, called neurons. The human brain incorporates nearly 10 billion neurons and 60 trillion connections, synapses, between them. By using multiple neurons, the brain can perform its functions much faster than the fastest computers in existence today.

An artificial neural network consists of a number of very simple and highly interconnected processors, also called neurons, which are analogous to the biological neurons in the brain. The neurons are connected by weighted links passing signals from one neuron to another. Each neuron receives a number of input signals through its connections; however, it never produces more than a single output signal[4].

An Expert system is an interactive computer-based decision tool that uses both the facts and heuristics to solve difficult decision making problems. Fuzzy logic is a new way of expressing probability. Neural Networks are eminently suited for approximating and designing of fuzzy Controllers and other types of Fuzzy Expert System. Neuro-fuzzy systems are connectionist models that allow learning as artificial neural network, but their structure can be interpreted as a set of fuzzy rules. Fuzzy logic and neural networks form the basis of the majority aided diagnostic intelligent systems. It would be interesting to combine the two approaches to exploit both advantages[5].

1.3 Problem Statement

In traditional health care in hospitals the diagnosis of the patients of breast cancer diseases, carried by experts and the data was recorded in patients file sheet, and later will be achieved at statistical department. However, the use of these file sheets shows serious limitations. The storage of such type of data is increased at a very rapid rate, due to continuous increasing the size of health care data. By using the traditional methods it becomes very difficult to analyze and classify these big data. In order to extract the meaningful information from it and discover knowledge. Thus there is a need to drive to find more effective models of providing such methods for analysis and classification of big data.

1.4 Research Questions

- i. Is the dataset usable for early investigation of breast cancer?
- ii. Can the model give accuracy results for classification of breast cancer dataset?

1.5 Objective of the Research

The purpose of classification in our research is to obtain:

- i. Build a data set of Breast cancer of Khartoum state
- ii. To investigate and analyze a neuro-fuzzy inference model by intelligent analysis using MATLAB tool to get better results to improve breast cancer classification into two classes: early and late.
- iii. To build a neuro-fuzzy inference model

iv. To evaluate the proposed model.

1.6 Scope of the Research

The scope of research is a breast cancer in Khartoum state hospitals, specifically about stage and grade of breast cancer.

1.7 Structure and Organization of the Thesis

This thesis contains six chapters ,Chapter tow gives an overall idea of fuzzy logic ,artificial neural network, hybrid neuro-fuzzy systems and comparison between Fuzzy Inference System and Hybrid neuro fuzzy Systems, Chapter three contains some related work in fuzzy systems, neural network and hybrid neuro-fuzzy systems, Chapter four describes the research methodology, Chapter five include the steps of practical in details, research results and discussion, Chapter six provide the conclusion and recommendations of the study.

