Chapter 12

Severity of Breast Mass Prediction in Mammograms Based on an Optimized Naive Bayes Diagnostic System

Abeer S. Desuky

https://orcid.org/0000-0003-1661-9134

Al-Azhar University, Egypt

ABSTRACT

Mammography is the most effective tool for breast mass screening. It is a special CT scan technique used only to detect breast tumors early and accurately. Detecting tumors in its early stage has improved the survival rate for breast cancer patients. Computer-aided diagnostic systems help physicians to detect breast cells abnormalities earlier than other traditional procedures. The main aim of this chapter is to increase physicians' ability to determine the severity of a mammographic mass lesion from the BI-RADS features and the patient's age using the bio-inspired chicken swarm optimization (CSO) algorithm for Naive Bayes (NBC) classifier. The mammographic mass dataset is used to analyze the proposed method (CSO-NBC). The dataset is preprocessed and divided to train the CSO-NBC system and test it by the 5-fold cross-validation technique. The performance of the proposed classification system is compared with the results from other research to show the efficiency of the system in predicting the severity of breast tumors with the highest accuracy.

INTRODUCTION

Cancer is one of the top leading causes of death. It is caused by uncontrolled growth of cells which invade and spread around the body, often resulting in death. Cancer is the second leading cause of death globally and was responsible for 8.8 million deaths in 2015[1]. Globally, nearly 1 in 6 deaths is due to cancer according to the WHO (World Health Organization). Per the US National Cancer Institute, 60 percent of the world's new cancer cases happen in Asia, Africa, Central, and South America, and 70 percent of

global cancer deaths occur in those same regions as well. Moreover, breast cancer death rate was 571 000 equivalently 6.5 percent of all deaths in 2015[2]. Because of this fact, early detection and true diagnosis is an important issue and plays a key role to reduce mortality of this disease.

Mammography is an efficient imaging appliance for early breast cells abnormality detection. Mammography exists in two types: first type, *screening mammography* which is breast X ray used to test the changes in breast area for early detection of breast tumors in women with no symptoms of cancer. It can also detect tiny calcium deposits (micro calcification) that is one of cancer manifestations. The second type, *Diagnostic mammography* which is a breast X ray to check the signs of breast cancer after detecting a mass or other symptom. symptoms include breast size/shape change, skin thickening, pain or nipple discharge (Sickles et al, 2002; Ramani & Vanitha, 2014).

Significant improvements can be made in the lives of breast cancer patients by detecting cancer early and avoiding delays in care, computer-aided diagnosis (CAD) or data-driven clinical decision support systems (CDSS) can help the physicians to do this task faster and more accurate than traditional procedures.

During the last decade with development of machine learning approaches in diagnostic systems, breast cancer detection has improved. The aim of using machine learning approaches is to minimize mistakes that may occur by specialists in diagnosis (Güzel et al., 2013).

Recently, intelligent techniques and systems have been highlighted in many papers by researchers. Different computer-aided diagnosis and decision support have included studies for tumors diagnosing, breast tumors have special concern in these studies. In some studies, the focus was for proposing a single technique, while a combination of techniques was proposed by others to obtain the best reliable results based on different data. Ramani and Vanitha (2014) used weighted histogram algorithms to select features from mammogram data then classified the selected features using random forest, naive bayes and ANN algorithms. Another method proposed in (Sahar & Alaa, 2013) based on missing values imputation and three models support vector machine SVM with polynomial kernel, ANN with pruning parameters and DT with Chi-squared interaction detection were derived for prediction.

Alickovic and Subasi (2017) proposed a system of two stages. In the first stage, the genetic algorithm was used to extract the most useful and essential features from the used data set which containing 569 and 32 cases and features respectively. Second, the authors applied an individual and hybrid data mining techniques to find the best accurate result. The best accuracy obtained by the Random Forest method obtained with 99.48% rate.

In (Wang et al., 2018) a hybridization of Support Vector Machine (SVM) algorithms and the Weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE) were performed to propose a novel method for increasing the performance accuracy and reducing different breast cancer diagnoses. The proposed method evaluated using three data sets, two of them are published and one is real. 97.68% was the highest accuracy rate which was achieved for the (WDBC) data set after applying 10 cross validation folds. Joshi and Mehta (2018) have tested the performance accuracy of their proposed K Nearest Neighbor (KNN) diagnosing technique in the R environment using a data set with 569 and 32 cases and features respectively. They used the most significant features selection methods: Principal Component Analysis (PCA) and (LDA) to employ their technique. Linear Discriminant Analysis was the method that achieved the better accuracy with 97.06% rate.

Another new intelligent CDSS model has been proposed by Liu et al. (2019). It is based on the hybridization of the information gain as a features ranking algorithm and the wrapper of simulated annealing genetic algorithm as an optimizer for the features selection. The optimal subset of features that has been selected inferred the cost-sensitive support vector machine classifier. The hybrid proposed model was

Related Content

Understanding the Legacy of Herbert Simon to Decision Support Systems
Jean-Charles Pomeroland Frédéric Adam (2008). *Encyclopedia of Decision Making and Decision Support Technologies (pp. 930-938).*
www.irma-international.org/chapter/understanding-legacy-herbert-simon-decision/11338

Markov Processes in Finance With Application to Stock Markets
(2018). *Alternative Decision-Making Models for Financial Portfolio Management: Emerging Research and Opportunities (pp. 141-253).*
www.irma-international.org/chapter/markov-processes-in-finance-with-application-to-stock-markets/188286

Determination of the Number of Clusters in a Data Set: A Stopping Rule × Clustering Algorithm Comparison
Derrick S. Boone (2011). *International Journal of Strategic Decision Sciences (pp. 1-13).*
www.irma-international.org/article/determination-number-clusters-data-set/60528

Morphology, Processing, and Integrating of Information from Large Source Code Warehouses for Decision Support
Jörg Rech (2006). *Processing and Managing Complex Data for Decision Support (pp. 122-152).*
www.irma-international.org/chapter/morphology-processing-integrating-information-large/28150

Predicting Credit Rating Migration Employing Neural Network Models
Michael D'Rosarioand Calvin Hsieh (2018). *International Journal of Strategic Decision Sciences (pp. 70-85).*
www.irma-international.org/article/predicting-credit-rating-migration-employing-neural-network-models/215354