Expert Systems With Applications 46 (2016) 139–144

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Breast cancer classification using deep belief networks

Ahmed M. Abdel-Zaher*, Ayman M. Eldeib

Department of Systems and Biomedical Engineering, Cairo University, Giza, Egypt

ARTICLE INFO

Keywords:
Breast cancer diagnosis
CAD
Classification
Deep learning based classifier
Pattern recognition

ABSTRACT

Over the last decade, the ever increasing world-wide demand for early detection of breast cancer at many screening sites and hospitals has resulted in the need of new research avenues. According to the World Health Organization (WHO), an early detection of cancer greatly increases the chances of taking the right decision on a successful treatment plan. The Computer-Aided Diagnosis (CAD) systems are applied widely in the detection and differential diagnosis of many different kinds of abnormalities. Therefore, improving the accuracy of a CAD system has become one of the major research areas. In this paper, a CAD scheme for detection of breast cancer has been developed using deep belief network unsupervised path followed by back propagation supervised path. The construction is back-propagation neural network with Liebenberg Marquardt learning function while weights are initialized from the deep belief network path (DBN-NN). Our technique was tested on the Wisconsin Breast Cancer Dataset (WBCD). The classifier complex gives an accuracy of 99.68% indicating promising results over previously-published studies. The proposed system provides an effective classification model for breast cancer. In addition, we examined the architecture at several train-test partitions.

1. Introduction

Breast cancer is the most common cancers among women with nearly 1.7 million new cases diagnosed in 2012 (Centers for disease control and prevention, cancer prevention control, 2014) (World cancer research fund, 2014). Breast cancer represents 18.3% of the total cancer cases in Egypt. A percentage of 37.3% of breast cancer could be fully healed especially in case of early detection (Salama, Abdelhalim, & Zeid, 2012). In Egypt and Arab countries, the breast cancer targets women in the age of 30 and represents 42 cases per 100 thousand of the population (Salama et al., 2012).

An accurate classifier is the most important component of any CAD scheme that is developed to assist medical professionals in early detecting mammographic lesions. CAD systems are designed to support radiologists in the process of visually screening mammograms to avoid miss-diagnosis because of fatigue, eyestrain, or lack of experience. The use of an accurate CAD system for early detection could definitely save precious lives. In this study, back propagation neural network initialized by weights from a trained deep belief network with similar architecture (DBN-NN) was used to diagnose the breast cancer. Our data source is the Wisconsin Breast Cancer Dataset (WBCD) taken from the University of California at Irvine (UCI) machine learning repository (Wisconsin breast cancer dataset (WBCD) (original), 2014).

2. Background

A variety of classification techniques were developed for breast cancer CAD systems. The accuracy of many of them was evaluated using the dataset taken from the UCI machine-learning repository. For example, Goodman, Boggess, and Watkins, tried different methods that produced the following accuracies: optimized learning vector quantization (optimized-LVQ) method's performance was 96.7%, big-LVQ method reached 96.8%, and the last method, they proposed AIRS, which depending on the artificial immune system, obtained 97.2% of classification accuracy (Goodman, Boggess, & Watkins, 2002).

Quinlan reached 94.74% classification accuracy using 10-fold cross validation with C4.5 decision tree method (Quinlan, 1996). Abonyi and Szeifert used Supervised Fuzzy Clustering (SFC) technique and obtained 95.57% accuracy (Abonyi & Szeifert, 2003). Salama et al. (2012) performed an experiment on WBC dataset and results showed that the fusion between MLP and J48 classifiers with feature selection (PCA) is superior to the other classifiers.

Hamilton, Shan, and Cercone (1996) with RIAC method obtained 96% accuracy. Polat and Günes (2007) examined the robustness of the least square Support Vector Machine (SVM) by using classification accuracy, analysis of sensitivity and specificity, k-fold cross-validation method, and confusion matrix. They obtained classification accuracy of 98.53%.

Nauck and Kruse (1999) obtained 95.06% with neuro-fuzzy techniques. Pauline and Santhakumaran used Feed Forward Artificial Neural Networks and back propagation algorithm to train the network (Pauline, 2011).The performance of the network is evaluated

* Corresponding author. Tel.: +20 1114488419.
E-mail addresses: ahmedallah.m.s@gmail.com (A.M. Abdel-Zaher), eldeib@ieee.org (A.M. Eldeib).

http://dx.doi.org/10.1016/j.eswa.2015.10.015
0957-4174/

using Wisconsin breast cancer dataset for various training algorithms. The highest accuracy of 99.28% is achieved when using Levenberg Marquardt algorithm.

The accuracy obtained by Pena-Reyes & Sipper (1999) was 97.36% using fuzzy-GA method. Akay (2009) combined SVM with feature selection obtaining highest classification accuracy (99.51%) for SVM model that contains five features. Moreover, Setiono (2000) was reached 98.1% using the Neuro-rule method. Übeyli (2007) used SVM and obtain 99.54% accuracy at 37% train and 63% test partition.

Mert, Kılıç, Bilgili, and Akan (2015)., explored features reduction properties of independent component analysis (ICA) on breast cancer decision support system. They proofed that a one-dimensional features vector obtained from (ICA) causes Radial Bases Function Neural Network (RBFNN) classifier to be more distinguishing with the increased accuracy from 87.17% to 90.49%.

Nahato, Nehemiah, and Kannan (2015), used a rough set indiscernibility relation method with back propagation neural network (RS-BPNN). This work has two stages. The first stage handles missing values to obtain a smooth data set and to select appropriate attributes from the clinical dataset by indiscernibility relation method. The second stage is classification using back propagation neural network. The accuracy obtained from the proposed method was 98.6% on breast cancer dataset.

Dheeba, Singh, and Selvi (2014), investigated a new classification approach for detection of breast abnormalities in digital mammograms using Particle Swarm Optimized Wavelet Neural Network (PSOWNN). The proposed abnormality detection algorithm is based on extracting Laws texture energy measures from mammograms and classifying the suspicious regions by applying a pattern classifier. They achieved 93.671%, 92.105% and 94.167% for accuracy, specificity, and sensitivity, respectively.

In our study, we applied deep belief network (DBN) in an unsupervised phase to learn input features statistics of the original WBCD dataset. Then, we transferred the obtained network weight matrix of DBN to back propagation neural network with similar architecture to start the supervised phase. In supervised phase, we tested both conjugate gradient and Levenberg-Marquardt algorithm for learning back propagation neural network.

3. From back propagation (BP) to deep belief network (DBN)

In 1985, the second-generation neural networks with back propagation algorithm have emerged. However, the learning algorithm struggle to adjust network weights so that output neurons state y represent the learning example t. A common method for measuring the discrepancy between the expected output t and the actual output y is using the squared error measure:

$$E = (t - y)^2 \tag{1}$$

The change in weight, which is added to the old weight, is equal to the product of the learning rate and the gradient of the error function, multiplied by -1:

$$\Delta w_{ij} = -\frac{\partial E}{\partial w_{ij}} \tag{2}$$

where almost all data is unlabeled. However, back propagation neural network requires a labeled training data. Therefore, the biggest issue with back propagation NN appears as its possibility to get stuck in poor local optima and the learning time is huge with multiple hidden layers.

In 1963, Vapnik et al. invented the original support vector machine (SVM) algorithm. Boser, Guyon, and Vapnik (1992) suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes. In classification task, the weight of each feature

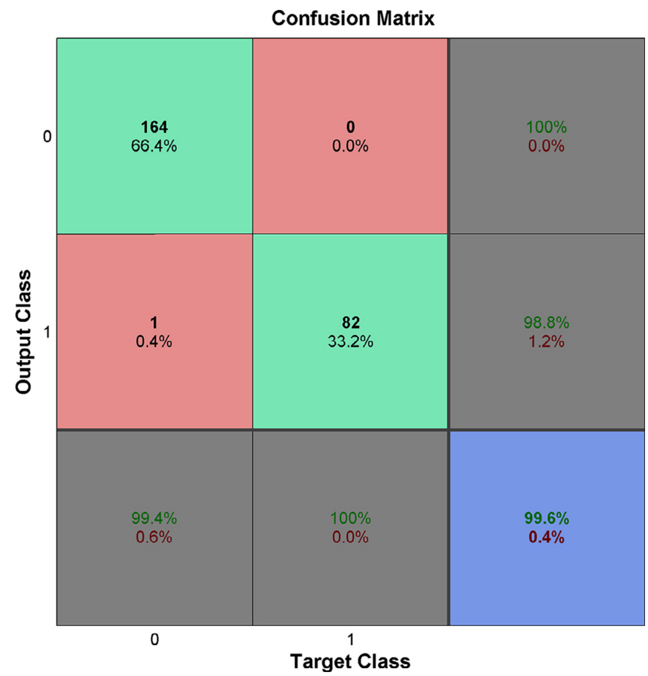

Fig. 1. Confusion matrix of DBN-NN.

is computed by optimization technique. In non-linear classification, SVMs can efficiently perform the task using what is called the kernel trick by mapping their inputs. The non-linear classification task converted to linear classification problem in high-dimensional feature spaces. The biggest limitation of SVM approach lies in choice of the kernel. In practice, the most serious problem with SVMs is the high algorithmic complexity and extensive memory requirements of the required quadratic programming in large-scale tasks (Suykens, Horvath, Basu, Micchelli, & Vandewalle, 2003).

In recent years, the attention has shifted to deep learning. Deep learning is a set of algorithms in machine learning that attempts to model high-level abstractions in data by using model architectures composed of multiple non-linear transformations (Bengio, Courville, & Vincent, 2013; Schmidhuber, 2014). Restricted Boltzmann Machine (RBM) is a generative stochastic artificial neural network that can learn a probability distribution over its set of inputs. On the other hand, Deep Belief Network (DBN) is a generative graphical model, or alternatively a type of deep neural network, composed of multiple layers of latent variables (“hidden units”), with connections between the layers but not between units within each layer (Hinton, 2009b).

From Hinton’s perspective, the DBN can be viewed as a composition of simple learning modules each of which is a restricted type of RBM that contains a layer of visible units. This layer represents the data. Another layer of hidden units represents features that capture higher-order correlations in the data. The two layers are connected by a matrix of symmetrically weighted connections (W) and there are no connections within a layer (Hinton, 2009b).

The key idea behind DBN is its weight (w), learned by a RBM define both $p(v|h, w)$and the prior distribution over hidden vectors $p(h|w)$ (Hinton, 2009b). The probability of generating a visible vector, can be written as

$$p(v) = \sum_h (p(h|w)\ p(v|h, w)) \tag{3}$$

As the learning of DBN is a computational intensive task, Hinton showed that RBMs could be stacked and trained in a greedy manner to form the DBN (Hinton, Osindero, & Teh, 2006). He introduced a fast algorithm for learning DBN. The weight update between visible v and