Pattern Recognition 42 (2009) 1845 – 1852

Contents lists available at ScienceDirect

Pattern Recognition

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

# A novel soft cluster neural network for the classification of suspicious areas in digital mammograms

Brijesh Verma[a,*], Peter McLeod[a], Alan Klevansky[b]

[a]School of Computing Sciences, Central Queensland University, Rockhampton, QLD 4701, Australia
[b]Radiology Department, Gold Coast Hospital, Gold Coast, QLD 4215, Australia

ARTICLE INFO

*Article history:*
Received 30 April 2008
Received in revised form 21 November 2008
Accepted 20 February 2009

*Keywords:*
Pattern classification
Neural networks
Clustering algorithms

ABSTRACT

This paper presents a novel soft cluster neural network technique for the classification of suspicious areas in digital mammograms. The technique introduces the concept of soft clusters within a neural network layer and combines them with least squares for optimising neural network weights. The idea of soft clusters is proposed in order to increase the generalisation ability of the neural network by providing a mechanism to more aptly depict the relationship between the input features and the subsequent classification as either a benign or malignant class. Soft clusters with least squares make the training process faster and avoid iterative processes which have many problems. The proposed neural network technique has been tested on the DDSM benchmark database. The results are analysed and discussed in this paper.



## 1. Introduction

The analysis of medical images represents an important, yet challenging part of pattern recognition and artificial intelligent systems [1]. Computer aided diagnostic (CAD) systems can help a physician in diagnosing a patient. CAD systems are used in the classification task where certain features (clinical findings) are used to assign a case to a particular pattern which represents a diagnosis. In particular neural networks have demonstrated their efficacy in the clinical domain with diseases such as cancer where there is a weak relationship between the class patterns forming a benign or malignant diagnosis [2,3]. Breast cancer diagnosis is particularly challenging due to the underlying cause of the disease not being known and the similarities that exist between benign and malignant masses. The seriousness of the disease means that breast cancer is the leading cause of cancerous deaths in women in the 20–59 age group [4]. This situation is then further compounded by mammograms being a fairly low contrast representation of the breast. On top of this anatomical anomalies, distortion of the breast during screening and background noise add further complications to the process. The early stages of breast cancer may only have subtle indications which can be varied in appearance, making physical examination ineffective and making diagnosis difficult even for experienced radiologists [5,6].

Various studies [7,8] have shown that the size and appearance of a lesion can affect the chance that a Radiologist may miss or incorrectly diagnose a cancer. Subtle masses and small lesions are more often missed [9] than larger ones. With an estimated 11–25% of breast cancers being missed [10] during routine screening mammography there is still room for improvement. Although different studies have found that certain lifestyle factors can increase the chance of contracting breast cancer [11], the underlying cause of the disease has not been discovered. The overall survival rate in the United States for early stage breast cancer is 98%. If cancer has spread to the regional lymph nodes the survival rate decreases to 84%, and if metastases have spread to distant organs the chance of survival declines to 28% [11]. This means that in the short term the likelihood of developing a cure is low indicating that early and accurate diagnosis is the surest mechanism of reducing the mortality rate.

The use of CAD can improve the performance of the radiologists in terms of reducing the number of missed diagnosis [9,10,12,13] and reducing the time taken to reach a diagnosis. Not only can CAD lead to improvements in accuracy but CAD systems can lead to faster turnaround of results as well as improved storage and retrieval of radiographic images [14]. This can also facilitate the transmission of these images to other medical specialists. CAD systems can provide a valuable tool for the training of radiologists.

The remainder of this paper is organised into seven sections. Section 2 reviews the existing techniques. Section 3 details the theoretical underpinnings, architecture and training of the soft cluster neural network (SCNN). Section 4 presents the research methodology. Section 5 details the results obtained with Section 6 providing a discussion of the results together with a comparative analysis of

* Corresponding author.
*E-mail addresses:* b.verma@cqu.edu.au (B. Verma), peterm@practical.com.au (P. McLeod).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.02.009

the findings. In Section 7, conclusions are drawn and future research directions are addressed.

## 2. Related work

Medical diagnosis is a pattern classification dilemma where a set of input features are used to determine if a patient has a particular disorder. Breast cancer diagnosis is a classification problem in terms of distinguishing between a malignant and benign mass from a suspicious lesion or region. Such problems have leant themselves to the application of CAD systems over the last few decades. A myriad of approaches such as artificial neural networks, genetic algorithms, wavelet filters and statistical transforms have been employed in order to solve this problem.

Optical Fourier transforms [15] and wavelet transforms are effective on micro calcification where the high spatial frequency in the Fourier spectrum allows for their identification, but may not be as accurate on masses where a lower spatial frequency exists similar to the surrounding tissues. Fuzzy logic [16], Bayesian networks [17], case based systems [18] and artificial neural networks [2,3,19,20] have been utilised to varying degrees of success. Artificial neural networks have demonstrated their capabilities in this area and have been widely adopted due to their generalisation capabilities [21]. Georgiou [21] utilised morphological features with a support vector machine (SVM) to obtain 91.54% classification accuracy. Halkiotis et al. [22] used the MIAS database and a multi layer perceptron (MLP) type neural network to obtain a classification rate of 94.7% with an average of 0.27 false positives per image. Brem et al. [13] used the second look CAD system (version 3.4) to determine the performance of CAD systems on different sized lesions and micro calcifications to achieve an overall sensitivity of 89%. Their investigation was to try and determine if lesion size would adversely impact upon the performance of a CAD system. Abdalla et al. [23] used textual features with a SVM classifier to achieve a classification accuracy of 82.5% on mammograms from the Digital Database of Screening Mammography (DDSM) [24]. Rangayyan et al. [25] noted that several methods have good sensitivity ( $>85\%$) for the identification of masses but also have a high false positive rate.

As can be seen from the above, neural network based techniques have been used for the classification of suspicious areas into benign and malignant classes, however there are several disadvantages with existing neural network techniques. Neural networks must undergo a training phase which can take a while to complete. The training phase needs to be repeated if new knowledge is to be added to the system making it hard to add new knowledge. The optimisation of a neural network relies on the setting of various parameters which need to be adjusted during the training phase which can greatly affect the performance of the system if not correctly tuned. The tuning of the network parameters often requires multiple iterations to determine the optimal configuration. In other situations certain neural networks may suffer from network paralysis, or excessive use of computational resources [26] making them infeasible for use in a clinical situation. Neural networks are often tasked with handling a large number of features, many of which are only weakly correlated or are of limited diagnostic value with the class pattern that they are attempting to classify. This creates a problem for training the network and produces low accuracy on test data. In this paper, we introduce a new concept of soft clusters with least squares and propose a novel neural network which can avoid some of the above mentioned problems with improved classification accuracy.

## 3. Soft cluster neural network

SCNN is based on an idea that in a classification problem, each class can have more than one cluster called soft clusters and that incorporating soft clusters output values into the learning of neural network weights will improve the learning process and the overall classification accuracy. The use of soft clustered neural network for the classification of benign and malignant patterns can lead to an improvement in classification accuracy over other approaches. To validate the proposed claim, we define the null and alternative hypotheses as follows.

**$H_0$:μ** The use of soft clustered neural networks makes no difference to the classification of benign and malignant classes when compared to traditional networks such as k-means or SVM.

**$H_1$:μ** The use of soft clustered neural networks results in an improvement in the classification of benign and malignant classes when compared to traditional approaches such as k-means or SVM.

The theoretical underpinnings, network architecture and training methods for the SCNN are presented in the following sections.

### 3.1. Theoretical underpinnings

Let $P=[p_1{}^{(j)},p_2,\ldots,p_n{}^{(j)}]$ be the set of $j$ input patterns with $n$ attributes and $O=[o_1{}^{(j)},o_2{}^{(j)}]$ be the output classes.

Let $J=\sum_{j=1}^{k}\sum_{i=1}^{n}||p_i^{(j)}-c_j||$be the objective function for clustering where $k$ is the number of clusters, $||p_i^{(j)}-c_j||^2$is a chosen distance measure between a data point $p_i{}^{(j)}$ and the cluster centre $c_j$, which is an indicator of the distance of the $n$ data points from their respective cluster centres (centroids).

The clustering based on objective function mentioned above can cluster $P=[p_1{}^{(j)},p_2{}^{(j)},\ldots,p_n{}^{(j)}]$ inputs into two clusters/classes $O=[o_1{}^{(j)},o_2{}^{(j)}]$. However hard clustering with one single hard boundary creates a lot of problems. Therefore, many soft clusters are created and the output of only those clusters which strongly associate with one of the classes, are used for weight optimisation. The weak and zero clusters are removed from influencing the network parameters in the wrong directions which will reduce the number of false classifications. It will also reduce the input data for weight optimisation and provide faster training of the network.

Let $J=[j_1 j_2,\ldots j_k]$ be the output of $k$ clusters, where $j_1 j_3,\ldots j_{n-1}$ may belong to $o_1$ and $j_2 j_8,\ldots j_n$ may belong to $o_2$. The output of the neural network can be optimised using the following output functions.

output $=f(J'*W)$, where $J'\in(j_1 j_{3,}\ldots j_n)$ and $(j_1 j_{3,}\ldots j_n)\in$ output of all clusters.

output $=f(J*i)$, where $J\in(j_2,\ j_{3,}\ j_4)$ and $(j_2,\ j_{3,}\ j_4)\in$ output of non-zero and strong clusters where the first output function utilises all the weights from the clustering process for the network training process while the second output prunes the underperforming weights from inclusion in the training process of the neural network. The actual details of these functions are expanded in Sections 3.2 and 3.3.

### 3.2. SCNN architecture

An overview of the proposed SCNN architecture is shown in Fig. 1. It has $n$ inputs, $m$ outputs and $k$ soft clusters.

The error minimisation function for SCNN is defined as follows. SCNN is trained to minimise the following error:

$$E=\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{m}(y_j^i-t_j^i)^2$$

where $p$ is the number of training samples, $m$ is the number of outputs, $y$ is the network output and $t$ is the target (desired) output.