Feature Extraction Using NLPCA Neural Networks for Detecting Breast Cancer in Mammogram Image

Preeti Sharma\textsuperscript{1} and Prof.(Dr) Sarvottam Dixit\textsuperscript{2}
\textsuperscript{1}(Research Scholar, Mewar University, Gangarar, Rajasthan, India)
\textsuperscript{2}(Research Guide, Mewar University, Gangarar, Rajasthan, India)

Abstract

This paper explains the result of Non-Linear Principal Component Analysis (NLPCA) for classification of mammogram images. An algorithm is proposed for the feature extraction using NLPCA. NLPCA is a technique for nonlinear feature extraction that is commonly implemented by a Feed forward Neural Network. This has been analysed to have enhanced performance, compared with linear principal component analysis (PCA), in difficult problems where the associations between the variables are not linear. In this paper, the NLPCA techniques are used to classify data into one of two classes: Normal and Abnormal. For classification purposes, we use only two nonlinear features of the data. The distribution of these features is shown using recurrent neural network (RNN). Finally, KNN classification used to the NLPCs of the image in order to assess further improvements introduced by the dimensionality reduction offered by the NLPCA. Results demonstrate that the proposed method is able to improve the classification both by reducing dimensionality and features. Classification obtained using NLPCA-derived data is more accurate.

Keywords: NLPCA, classification, mammogram, feature extraction, pre-processing

1. Introduction

Over a past decade, the image processing was used in various fields such as medical fields, remote sensing, pattern recognition, environmental monitoring, etc. The computer-aided diagnosis plays an adequate part of explicating the extracting of features in an image classification. Image classification is one of the main factors of image processing and machine learning [3]. There are large networks [2], such as perceptron [4], convolution neural networks [1], and deep neural network [5]. Breast cancer is one of the major problems for women all over the world. Early detection is the best way in breast cancer diagnosis. Mammography
is the most effective screening method for early detection of breast cancer. Masses and micro calcifications are the two powerful indicators of cancers in examining mammograms. The mass detection is a more challenging problem than microcalcification; due to its variation in size and shape also it exhibits poor image contrast.

The participation of noise filtering is tempted as a precise feature in the characterization of the selected image classification. The necessary explanation behind the commotion dropping methodology is to expel speckle noise by holding the basic features of the image. Noise processing serves as an important tool to attain and process the unspecified noises present in the characterization of an image. Subsequently, the image processing is performed with the accomplishment of various trends of classification streams to manipulate the collected dataset. Henceforth, it is inadequate to extract the classification of particularized features from the image and personalize the undetermined data in the image dataset. Accordingly, it is relevant to induce an efficient methodology to determine the hidden specification of the corresponding dataset.

According to M. Kramer, the nonlinear principal component analysis (NLPCA), which is a well-known method of principal component analysis (PCA), using auto associative neural networks (ANN) works better than the regular PCA, because it helps to find nonlinear dependences in data. Independent and temporal data samples can be viewed. For instance, we consider the dataset as a set of images, whose width and height associated are w and h pixels respectively. When the dimension of image is large to analyse, a pre-processing stage is thus necessary to extract the significant information before applying to the NN run. For instance, a PCA can be used as a dimensionality reduction algorithm, leaving the NN to work over a few dimension only. In this paper, NLPCA runs with linearly reduced input patterns, thus limiting the method’s potentials.

In this paper, we would like to propose nonlinear features using an end-to-end network for breast cancer detection, which is named as recurrent neural network (RNN) including LSTM and GRU concept. Although NLPCA and RNN are well-established techniques for classification problems, to the best of our knowledge, we are the first to combine them for nonlinear classification in the mammogram databases. Note that integrating NLPCA and RNN in an end-to-
end manner has also been explored in mammogram image classification, where the NLPCA is only used for extracting nonlinear features to build a classifier for the classification purpose. In our work, the NLPCA part transforms the input, a pair of nonlinear features, to an abstract feature representation, whereas the RNN part is employed for modeling temporal dependency. In other words, the features from the proposed RNN encapsulate information related to spectral, spatial, and temporal components in mammogram images, making them useful for a holistic change cancer detection task. The proposed architecture extracts a feature representation of multitemporal data through learning with a structured deep architecture and finally producing labels for the image sequence using KNN.

The remainder of this paper is organized as follows. After the introductory Section 2 detailing breast cancer detection using NLPCA, Section 3 is dedicated to the details of the proposed NLPCA carried out using recurrent neural network. Section 4 then provides data set information, network setup, experimental results, and discussion. Finally, Section 5 concludes the paper.

2. Proposed methodology

A novel method for breast cancer classification using a combination of NLPCA, RNN is proposed. The method consists of four phases

- Acquisition of Raw mammogram image (RMI)
- Pre-processing of RMI
- Feature extraction
- Classification of RMI

In this paper, we propose a method of classification using a combination of NLPCA and KNN. The method consists of four steps as given in Figure 1. Nonlinear principal component analysis is a technique for reducing high dimensional data, similar to the well-known method of principal component analysis. In this paper, NLPCA is carried out using deep recurrent neural network (DRNN) including the concept of LSTM and GRU which are capable of modelling nonlinear data. The NLPCA method uses deep recurrent neural network (DRNN) training procedures to generate nonlinear features.
2.1 Mammogram Image Pre-processing:

Mammogram pre-processing is the operations done in the image at the lowest level of abstraction. It helps for image enhancement by suppress the unwanted distortions from the image. The mammograms are disconcert by noise, non-mass regions and background of the mammogram. In the pre-processing stage, the incompatible images are improved as a highly enhanced image for further process. The artifacts and noise are removed from the image, to enhance the quality of the image.

In the pre-processing step first the noise is removed. The noises such as straight lines, alphanumeric values are removed from the image. The noise is removed by using the 2D median filter. This nonlinear digital filter is often used to remove the noise from the image. The median filter works through each pixel by pixel and replace the pixel value by the median value of nearest neighbouring pixels. Here the median value is calculated by sorting all the pixel values and replace the pixel by the middle pixel value.

After the removal of noise the mammogram may contain artifacts such as labels, markers etc. This may affect the detection of mass region from the image. These artifacts are removed before detection of mass from the breast region. The morphological operations such as erosion and dilation are performed to remove the artifacts from the image. The morphological operations processed on the image based on their shape. In the dilation the output pixel is maximum of all the input pixel’s neighbourhood and in the erosion the output pixel is minimum of all input pixel’s neighbourhood.

Finally, the pectoral muscles are removed from the mammograms. Because the
pectoral muscles appear with the same density as like the dense tissues in the mammogram images. The pectoral muscles are removed by using Single Seeded Region Growing Algorithm.

2.2 Feature Extraction:

After pre-processing the features are extracted for image classification.

**NLPCA:**

The multilayer neural network is the class which is mainly focussed by the nonlinear PCA and is used to reduce dimension of the data and to extract the nonlinear features of the data. The NLPCA detects both the linear and nonlinear interdependencies between the data. The NLPCA mainly focused to minimize the mean square error between the original features of the image and the features that obtained after the dimensionality reduced.

The main objective of NLPCA is to extract both the linear and nonlinear features between the data and is expressed as,

\[ X' = A(T_n) + C \quad ....(1) \]

Where, X is the original sampling data, \( T_n = [t_1, t_2, ..., t_k] \) is the matrix of nonlinear principal component points, C is the matrix of residuals, \( X' \) is the estimated value of X and \( A(*) \) is the nonlinear function expressed by the input training recurrent neural network.

2.3 Recurrent neural network

Sequential data is processed by using RNN. NLPCA is based on a multi-layer perceptron which is commonly referred as Auto associative neural network. In this work, RNN is considered for model the features using feed forward connections and sigmoidal activation functions in each node. The impressive success of recent deep learning systems has been predominantly achieved by feed forward neural network architectures like CNN. In such networks, we implicitly assume that all inputs are independent of each other. RNNs are a kind of neural networks that extend the conventional feed forward neural networks with loops in connections. Unlike a feed forward network, an RNN is capable of dealing with dependent, sequential inputs by having a recurrent hidden state whose activation at each time step depends on that of the previous time. By doing so, the network can exhibit dynamic temporal behavior, which is in line with our purpose; i.e., modeling temporal dependency between the sample data (\( T_1 \) and \( T_2 \)). To this end, three types of RNN architectures, namely, fully connected RNN, LSTM, and gated
recurrent unit (GRU), are used to construct the recurrent network.

**Fully connected RNN:** Given feature vectors $f^T_1$ and $f^T_2$ learned from the NLPCA, a fully connected RNN updates its recurrent hidden state $h_t$ by,

$$
 h_t = \begin{cases} 
 0 & \text{if } t = 0 \\
 \varphi(h_{t-1}, f^T_t) & \text{else} 
\end{cases}
$$

... (2)

Where, $\varphi$ is a nonlinear activation function, such as a hyperbolic tangent function or logistic sigmoid function. The recurrent layer will output a sequence $h = h_1, h_2$. For our task, we only need the last one as input to the fully connected layers for classifier. In the fully connected RNN model, the update of the recurrent hidden state in Eq. (2) is implemented as,

$$
 h_t = \varphi(Uh_{t-1} + Wf^T_t) 
$$

... (3)

Where $U$ and $W$ are the coefficient matrices for the activation of recurrent hidden units at the previous time step and for the input at the present time, respectively. Fully connected RNN is the concisest RNN model, and it can reflect the essence of RNNs; i.e., an RNN is capable of modeling a probability distribution over the next element of the sequence data, given its present state $h_t$, by capturing a distribution over sequence data. Let $p(f^T_k, f^T_{k+T})$ be the sequence probability, which can be decomposed into,

$$
p(f^T_k, f^T_{k+T}) = p(f^T_k) p(f^T_{k+T} | f^T_k)
$$

... (4)

Then, the conditional probability distribution can be modeled with an RNN:

$$
p(f^T_{k+T} | f^T_k) = \varphi(h_2)
$$

... (5)

where $h_2$ is obtained from Eq. (2). Our motivation in this work is apparent here: mammogram images act as true sequential data instead of a simple difference image or stacked image, and therefore, an RNN can be used to model the temporal dependency.

**LSTM:** As shown in Eq. (2), recurrent hidden units in a fully connected RNN simply compute a weight sum of inputs and then apply a nonlinear function. In contrast, an LSTM-based recurrent layer maintains a series of memory cells $c_t$ at time step $t$.

The activation of LSTM units can be calculated by,

$$
 h_t = o_t \tanh(c_t)
$$

... (6)

Where, $\tanh$ function is the hyperbolic tangent function and $o_t$ is the output gates that control the amount of memory content exposure. The output gates are updated by,

$$
 o_t = \sigma(W_{ot} f^T_t + W_{oh} h_{t-1} + W_{oc} c_t) 
$$

... (7)
where the W terms represent coefficient matrices; e.g., \( W_{oi} \) and \( W_{oc} \) are the input-output weight matrix and memory output weight matrix, respectively.

The memory cells \( c_t \) are updated by partially discarding the present memory contents and adding new contents of thememory cells \( \tilde{c}_t \):

\[
c_t = i_t \cdot \tilde{c}_t + f_t \cdot c_{t-1}
\]  

(8)

Where, \( \cdot \) is an element wise multiplication?

The new memory contents are,

\[
\tilde{c}_t = \tanh(W_{ci} f^T_t + W_{ch} h_{t-1})
\]  

(9)

Where \( W_{ci} \) is input-memory weight matrix and \( W_{ch} \) represents hidden-memory coefficient matrix. The \( i_t \) and \( f_t \) are input gates and forget gates, respectively. The former modulates the extent to which the new memory information is added to the memory cell, whereas the latter controls the degree to which contents of the existing memory cells are forgotten.

Specifically, gates are computed as follows:

\[
i_t = \sigma(W_{di} f^T_t + W_{dh} h_{t-1} + W_{dc} c_{t-1})
\]  

(10)

\[
f_t = \sigma(W_{fi} f^T_t + W_{fh} h_{t-1} + W_{fc} c_{t-1})
\]  

(11)

**GRU:** Similarly to LSTM, a GRU makes use of a linear sum between the existing state and the newly computed state. It, however, directly exposes whole state values at each timestep, instead of controlling what part of the state information will be exposed.

The activation \( h_t \) of GRUs at time step \( t \) is a linear interpolation between the previous activation \( h_{t-1} \) and the candidate activation \( \tilde{h}_t \):

\[
h_t = (1 - u_t) h_{t-1} + u_t \tilde{h}_t
\]  

(12)

Where the update gates \( u_t \) determine how much GRUs update their activations or contents. Update gates can be computed by,

\[
u_t = \sigma(W_{ui} f^T_t + W_{uh} h_{t-1})
\]  

(13)

Where, \( W_{ui} \) and \( W_{uh} \) are the input-update coefficient matrix and hidden-update weight matrix, respectively. The candidate activation function \( \tilde{h}_t \) is computed similarly to that of the fully connected RNN (cf. Eq. (2)) and as follows:

\[
\tilde{h}_t = \tanh(U(r_t \cdot h_{t-1}) + W^T f_t)
\]  

(14)

Where, \( r_t \) is the set of reset gates. When reset gates are totally off (i.e., \( r_t = 0 \)), GRUs will completely forget the activation of the recurrent layer at previous time and only receive existing input. When open, reset gates will partially keep the information of the previously computed state. Reset gates are calculated similarly to update gates:
\[ r_i = \sigma(W_{ri}f^T + W_{rh}h_{i-1}) \]  \hspace{1cm} (15)

Where, \( W_{ri} \) is the input-reset weight matrix and \( W_{rh} \) represents the hidden-reset coefficient matrix.

### 2.4 KNN classification

The matrix representation of RNN is input to KNN for classification. The k-nearest neighbor algorithm (K-NN) is a method for classifying objects based on closest training examples in the feature space. In K-NN, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor. The neighbors are taken from a set of objects for which the correct classification is known. This can be thought of as the training set for the algorithm, though no explicit training step is required.

### 3. Experiments

For research purpose only the MIAS database is obtainable and it comprises 322 mammography images which is attained in mediolateral view. In the beginning, scanned from the film with the determination of 50 microns and the entire images were diminished to 200 microns and clipped / padded thus that they were suitable for a 1,024 x 1,024 bounding box. In the portable network graphics (PNG), the image files are obtainable and the format and interpreted with the below information: a database reference number indicating left and right breast, character of background tissue, pathology, class of lesion present and coordinates and the size of these lesions. In this paper, as we focus more on classification of breast cancer into benign or malignant. Each of benign and malignant cases has a metadata that shown the location and size of the breast cancer. The complete proposed approach shown in the system architecture diagram is implemented in MATLAB 2017a software.

### 3.1 Performance Evaluation

Confusion matrix, ROC curve with AUC score is the parameters to evaluate the performance of classification algorithm. Confusion matrix helps to get information about both actual and predicted class classification. The TPR and FPR are used to plot the ROC curve. The TPRs is used to calculate correctly classified malignant ROIs from all available malignant ROIs. The FPR parameter can calculate incorrectly classified benign ROIs amongst the total number of benign ROIs. At the end Accuracy, Precision, Sensitivity and
Specificity parameters are calculated to assess the system performance.

### 3.3 Results and discussion

True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) are four different possible outcomes of a single prediction for a two class case. Accuracy, Sensitivity, specificity and ROC curve with AUC score are statistical parameters that help to evaluate the performance. Sensitivity measures the proportion of real positives which are properly recognized when the mammogram contains malignancies tissues in it. Specificity quantifies the proportion of negatives which are properly recognized when cancer is not present in the mammogram.

### 4. Conclusion

In this paper we introduced the use of NLPCA as a pre-processing tool to improve classification of mammogram images. The use of NLPCA is effective in reducing the dimensionality of mammogram images without losing relevant information. However, classification results obtained using only the NLPCs, seem to improve the classification accuracies of some classes, although a definition loss is noticeable.
This problem could be to take into account also spatial information.

References


