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

Preeti Sharma[1] and Prof.(Dr)Sarvottam Dixit[2]
[1](Research Scholar, Mewar University, Gangarar, Rajasthan, India)
[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 usingNLPCA. 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 datainto 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, KNNclassification used to the NLPCsof the image in order to assess further improvements introduced by the dimensionality reduction offered bythe NLPCA. Results demonstrate that the proposed method is ableto improve the classification both by reducing dimensionalityand 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 factorsofImageProcessing 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 majorproblemsfor womenall over the world.Early detection is the best wayin breast cancer diagnosis. Mammography

[13] is the mosteffective screening method for early detection of breast cancer. Masses and micro calcifications are the two powerfulindicators of cancers in examining mammograms [14]. Themass detection [6] is a more challenging problem thanmicrocalcification; due to its variation in size and shape alsoit 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 nonlinearprincipal component analysis (NLPCA) [11], which is a well-knownmethod of principal component analysis(PCA) [8], using auto associative neural networks(ANN) [7] works better than the regular PCA, because it helps to find nonlineardependences in data [15].

Independent and temporal data samples can be viewed. For instance, wecan consider the dataset as a set of images, whose width and height associated are w and h pixelsrespectively.When the dimension of image islarge to analyse, a pre-processing stage is thus necessary to extract the significant informationbefore applying to the NN[9] run. For instance, a PCA [10] can be used as a dimensionality reduction algorithm, leavingthe NN to work over a fewerdimension only. In this paper, NLPCA [12] runs with linearlyreduced input patterns, thus limiting the method's potentials.

In this paper, we would like to proposenonlinearfeatures using an end-to-end network for breast cancerdetection, which is named as recurrent neuralnetwork (RNN) including LSTM and GRU concept. Although NLPCA andRNN are well-established techniques for classification problems, to the best of our knowledge, we arethe first to combine them for nonlinear classification inthe mammogram databases. Note that integrating NLPCAand RNN in an end-to-end manner has also been explored inmammogram image classification, where the NLPCA isonly used for extracting nonlinear features to build a classifier for the classification purpose. In our work, the NLPCApart transforms the input, a pair of nonlinear features,to an abstract feature representation, whereasthe RNN part is employed for modeling temporaldependency. In other words, thefeatures from the proposed RNN encapsulate informationrelated to spectral, spatial, and temporal components in mammogram images, making them useful for a holistic changecancer detection task. The proposed architecture extract a featurerepresentation of multitemporal data through learningwith a structured deep architecture and finally producing labels for the image sequence using KNN.

The remainder of this paper is organized as follows. Afterthe introductory Section 2 detailing breast cancer detection using NLPCA, Section3 is dedicated to the details of the proposed NLPCA carried out using recurrentneural 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 methodof 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 dimension 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 modellingnon-linear data. The NLPCA method uses deep recurrent neural network (DRNN) training

Proceduresto generate nonlinear features.

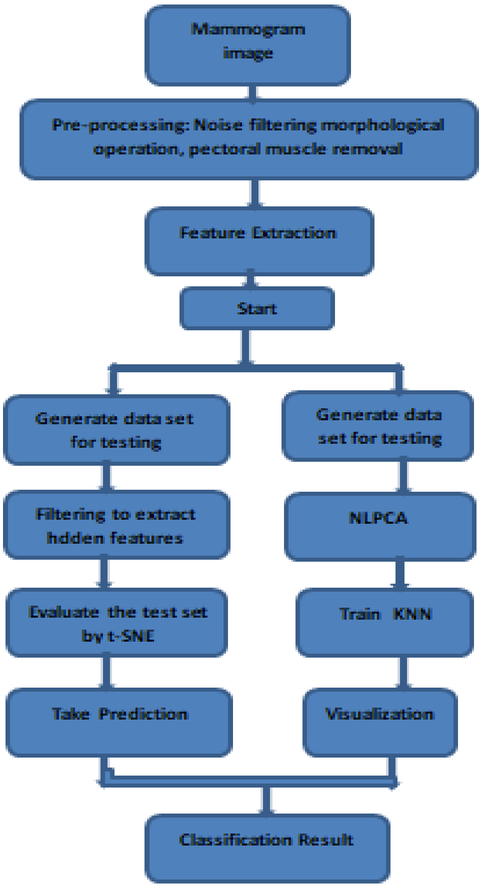


Fig. 1: Flow diagram

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 incompative 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,

$$X' = A(T_n) + C \quad \ldots.(1)$$

Where, X is the original sampling data, $T_n = [t_1, t_2, \ldots. t_k]$ 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 hasbeen predominantly achieved by feed forward neural networkarchitectures like CNN. In such networks, we implicitly assumethat all inputs are independent of each other. RNNs are a kindof neural networks that extend the conventional feed forwardneural networks with loops in connections. Unlike a feed forwardnetwork, an RNN is capable of dealing with dependent,sequential inputs by having a recurrent hidden state whoseactivation at each time step depends on that of the previoustime. By doing so, the network can exhibit dynamic temporalbehavior, which is in line with our purpose; i.e., modelingtemporal dependency between the sample data (T_1 and T_2). To thisend, three types of RNN architectures, namely, fully connectedRNN, LSTM, and gated recurrent unit (GRU), are used toconstruct 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,

$$h_t = \begin{cases} 0 & if\ t = 0 \\ \varphi(h_{t-1}, f^{T_t}) & else \end{cases} \quad ....(2)$$

Where, φ is a nonlinear activation function, such as a hyperbolictangent function or logistic sigmoid function. The recurrentlayer will output a sequence $h = h_1, h_2$. For our task, weonly need the last one as input to the fully connected layersfor classifier.In the fully connected RNN model, the update of therecurrent hidden state in Eq. (2) is implemented as,

$$h_t = \varphi(Uh_{t-1} + Wf^{T_t} \quad \ldots (3)$$

Where U and W are the coefficient matrices for the activationof recurrent hidden units at the previous time step and for theinput at the present time, respectively.Fully connected RNN is the concisest RNN model, and itcan reflect the essence of RNNs; i.e., an RNN is capable ofmodeling a probability distribution over the next element ofthe sequence data, given its present state h_t, by capturinga distribution over sequence data. Let $p(f^{T_1}, f^{T_2})$ be thesequence probability, which can be decomposed into,

$$p(f^{T_1}, f^{T_2}) = p(f^{T_1})p(f^{T_2} \mid f^{T_1}) (4)$$

Then, the conditional probability distribution can be modeledwith an RNN:

$$p(f^{T_2} \mid f^{T_1}) = \varphi(h_2) \quad (5)$$

where h_2 is obtained from Eq. (2). Our motivation in this workis apparent here: mammogram images act as true sequentialdata instead of a simple difference image or stacked imageand, therefore, an RNN can be used to model the temporaldependency.

LSTM: As shown in Eq. (2),recurrent hidden units in a fully connected RNN simplycompute a weight sum of inputs and then apply a nonlinearfunction. In contrast, an LSTM-based recurrent layer maintainsa series of memory cells c_t at time step t. The activation of LSTM units can be calculated by,

$$h_t = o_t \tanh(c_t) \quad (6)$$

Where, tan function is the hyperbolic tangent function and o_t is the output gates that control the amount of memory contentexposure. The output gates are updated by,

$$o_t = \sigma(W_{oi} f^{T_t} + W_{oh}h_{t-1} + W_{oc}c_t) \ldots (7)$$

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

The memory cells c_t are updated by partially discardingthe present memory contents and adding new contents of thememory cells $\tilde{c}_t$:

$$c_t = i_t \cdot \tilde{c}_t + f_t \cdot c_{t-1} \qquad (8)$$

Where, $\cdot$ is an element wise multiplication? The new memory contents are,

$$\tilde{c}_t = \tanh(W_{ci} f^{T_t} + W_{ch} h_{t-1}) \qquad (9)$$

Where W_{ci} is input-memory weight matrix and W_{ch} representshidden-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 memoryinformation is added to the memory cell, whereas the lattercontrols the degree to which contents of the existing memorycells are forgotten. Specifically, gates are computed as follows:

$$i_t = \sigma(W_{ij} f^{T_t} + W_{ih} h_{t-1} + W_{ic} c_{t-1}) \qquad (10)$$

$$f_t = \sigma(W_{fi} f^{T_t} + W_{fh} h_{t-1} + W_{fc} c_{t-1)} \qquad (11)$$

GRU:Similarly to LSTM, a GRU makes use of a linearsum 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 informationwill 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 $\tilde{h}_t$:

$$h_t = (1 - u_t) h_{t-1} + u_t \tilde{h}_t \qquad (12)$$

Where the update gates u_t determine how much GRUs updatetheir activations or contents. Update gates can be computedby,

$$u_t = \sigma(W_{ui} f^{T_t} + W_{uh} h_{t-1}) \qquad (13)$$

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

$$\tilde{h}_t = \tanh(U(r_t \cdot h_{t-1}) + W f^{T_t}) \qquad (14)$$

Where, r_t is the set of reset gates. When reset gates aretotally off (i.e., r_t is 0), GRUs will completely forget theactivation of the recurrent layer at previous time and onlyreceive existing input. When open, reset gates will partiallykeep the information of the previously computed state. Resetgates are calculated similarly to update gates:

$$r_t = \sigma(W_{ri} f^{T_t} + W_{rh} h_{t-1}) \tag{15}$$

Where, W_{ri} is the input-reset weight matrix and W_{rh} representsthe 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 forclassifying objects based on closest training examples in thefeature space. In K-NN, anobject is classified by a majority vote of its neighbors, withthe object being assigned to the class most common amongstits k nearest neighbors (k is a positive integer, typicallysmall). If k = 1, then the object is simply assigned to the classof its nearest neighbor. The neighbors are taken from a set ofobjects for which the correct classification is known. This canbe thought of as the training set for the algorithm, though noexplicit 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 theparameters to evaluate the performance of classificationalgorithm. Confusion matrix helps to get information aboutboth actual and predicted class classification.The TPR and FPR are used plot the ROC curve. The TPRis used to calculate correctly classified malignant ROIs fromall available malignant ROIs. The FPR parameter cancalculate incorrectly classified benign ROIs amongst the totalnumber of benign ROIs. At the end Accuracy, Precision,Sensitivity and

Specificity parameters are calculated to assessthe system performance.

3.3 Results and discussion

True Positives (TP), False Positives (FP), Truenegatives (TN) and False Negatives (FN) are four differentpossible outcomes of a single prediction for a two class case.Accuracy, Sensitivity, specificity and ROC curve with AUCscore are statistical parameters that help to evaluate theperformance. Sensitivity measures the proportion of realpositives which are properly recognized when themammogram contains malignancies tissues in it. Specificityquantifies the proportion of negatives which are properlyrecognized when cancer is not present in the mammogram.

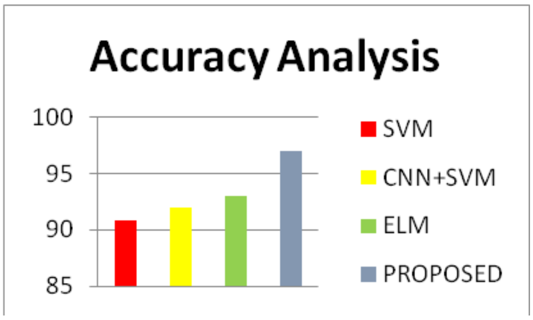


Figure 8: Accuracy analysis

Figure 6 shows the analysis and comparison of accuracy. The existing classifiers such as SVM, CNN+SVM and ELM are considered for comparison. The proposed NLPCA based classification achieves better accuracy.

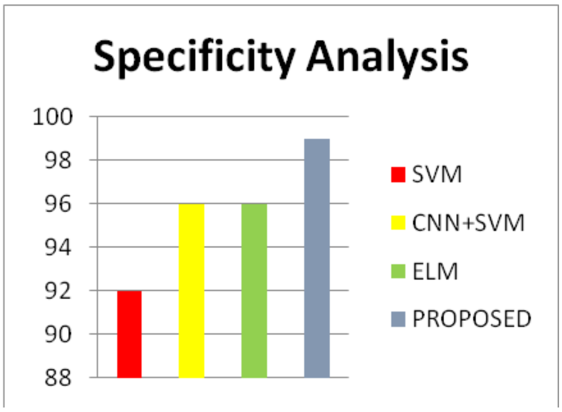
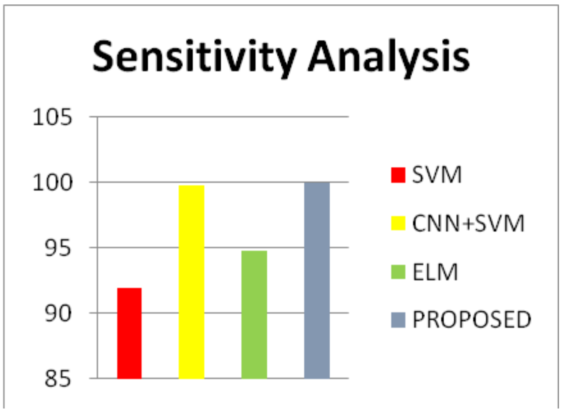


Figure 9:
S



pecificity analysis

Figure 10: Sensitivity analysis

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

[1]. Spanhol, Fabio Alexandre, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. "Breast cancer histopathological image classification using convolutional neural networks." In Neural Networks (IJCNN), 2016 International Joint Conference on, pp. 2560-2567. IEEE, 2016.

[2]. Wallender, W. W., and M. Islam. "Image Classification using Artificial Neural Networks." In AGU Fall Meeting Abstracts. 2001.

[3]. Tang, Jiexiong, Chenwei Deng, and Guang-Bin Huang. "Extreme learning machine for multilayer perceptron." IEEE transactions on neural networks and learning systems 27, no. 4 (2016): 809-821.

[4]. Zhang, Yudong, Yi Sun, Preetha Phillips, Ge Liu, Xingxing Zhou, and Shuihua Wang. "A multilayer perceptron based smart pathological brain detection system by fractional Fourier entropy." Journal of medical systems 40, no. 7 (2016): 173.

[5]. Singh, Akansha, and Krishna Kant Singh. "Satellite image classification using genetic algorithm trained radial basis function neural network, application to the detection of flooded areas." Journal of Visual Communication and Image Representation 42 (2017): 173-182.

[6]. Kashyap, Kanchan Lata, Manish Kumar Bajpai, and Pritee Khanna. "An efficient algorithm for mass detection and shape analysis of different masses present in digital mammograms." Multimedia Tools and Applications 77, no. 8 (2018): 9249-9269.

[7]. González, Juan Pablo Nieto. "Vehicle fault detection and diagnosis combining an AANN and multiclass SVM." International Journal on Interactive Design and Manufacturing (IJIDeM) 12, no. 1 (2018): 273-279.

[8]. Lin, Chun-Han, and Mark A. LaBarge. "The Influence of Tissue Architecture on Drug Response: Anticancer Drug Development in High-Dimensional Combinatorial Microenvironment Platforms." In Biomarkers of the Tumor Microenvironment, pp. 433-447. Springer, Cham, 2017.

[9]. Jena, Manaswini, and Sashikala Mishra. "Review of Neural Network Techniques in the Verge of Image Processing." In International Proceedings on Advances in Soft Computing, Intelligent

Systems and Applications, pp. 345-361. Springer, Singapore, 2018.

[10]. Artoni, Fiorenzo, Arnaud Delorme, and Scott Makeig. "Applying dimension reduction to EEG data by Principal Component Analysis reduces the quality of its subsequent Independent Component decomposition." NeuroImage 175 (2018): 176-187.

[11]. Piscini, A., and G. Licciardi. "Application of Non-Linear Principal Component Analysis to Hyperspectral data for noise filtering using artificial neural networks." In Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2014 6th Workshop on, pp. 1-4. IEEE, 2014.

[12]. Su, Chong, Yue Gao, Yuxiao Xie, Yong Xue, Lijun Ge, and Hongguang Li. "A hybrid classifier based on nonlinear-PCA and deep belief networks with applications in dysphagia diagnosis." Computer Assisted Surgery 22, no. sup1 (2017): 135-147.

[13].Caumo, Francesca, Giovanna Romanucci, Kylie Hunter, Manuel Zorzi, Silvia Brunelli, Petra Macaskill, and Nehmat Houssami. "Comparison of breast cancers detected in the Verona screening program following transition to digital breast tomosynthesis screening with cancers detected at digital mammography screening." Breast cancer research and treatment 170, no. 2 (2018): 391-397.

[14].Hazarika, Manasi, and Lipi B. Mahanta. "A Novel Region Growing Based Method to Remove Pectoral Muscle from MLO Mammogram Images." In Advances in Electronics, Communication and Computing, pp. 307-316. Springer, Singapore, 2018.

[15]. Lahdhiri, Hajer, IlyesElaissi, OkbaTaouali, Mohamed FaouziHarakat, and HassaniMessaoud. "Nonlinear process monitoring based on new reduced Rank-KPCA method." Stochastic Environmental Research and Risk Assessment 32, no. 6 (2018): 1833-1848.

[16].Preeti Sharma, DR.R..Saxena "Nonlinear Feature Learning Via NLPCA Neural tworksfor Breast Cancer Detection in Mammogram Database" Journal of Computational Information Systems 14: 5 (2018) 45-55.