



A Polynomial Neural Network Model for Prognostic Breast Cancer Prediction

Shweta Saxena¹, Vishnu Pratap Sing Kirar², Kavita Burse³

¹M. Tech Scholar, Oriental College of Technology, Bhopal (M. P.), India, shweta_sxn@yahoo.com

²Department of Electronics & Communication, Truba Group of Engineering & Information Technology, Bhopal (M. P.), India, vishnupskirar@live.com

³ Director, Oriental College of Technology, Bhopal (M. P.), India, kavitaburse14@gmail.com

Abstract: The second leading cause of death among women is breast cancer. Once a patient is diagnosed with breast cancer, the prognosis gives the anticipated long-term behavior of the ailment. In this research breast cancer prognosis is done using Polynomial Neural Network (PNN) on Wisconsin Prognostic Breast Cancer (WPBC) dataset. The data pre-processing technique named Principal Component Analysis (PCA) is used as a features reduction transformation method to improve the accuracy of PNN.

Key words: breast cancer prognosis, principal component analysis, polynomial neural network, Wisconsin prognostic breast cancer dataset.

INTRODUCTION

The second leading cause of death among women is breast cancer, as it comes directly after lung cancer [1]. Cancer is a disease in which cells become abnormal and form more cells in an uncontrolled way. With breast cancer, the cancer begins in the tissues that make up the breasts. The cancer cells may form a mass called a tumor. They may also invade nearby tissues and spread to lymph nodes and other parts of the body [2]. The prognosis is the principal factor in determining the treatment that will immediately follow the diagnosis of the disease [3]. There is much research on breast cancer prognosis. In [4] the authors addressed the breast cancer prognosis problem by employing two neural network architectures with WPBC dataset. The accuracy of the neural classifiers reaches 92%. The prognostic recurrence predictions were further evaluated using survival analysis through the Kaplan-Meier approximation method and compared with other techniques from the literature. In [5] a data mining framework that incorporates the task of learning patterns and rules that will facilitate the formulation of decisions in new cases are designed for WPBC dataset. The paper also highlights the performance of feature reduction and classification algorithms on the training dataset. In [6] the authors focused on the prediction of the prognosis in lymph node negative breast cancer (without apparent tumor cells in local lymph nodes at diagnosis). The outcome is defined as a variable that can have two values: poor prognosis or good prognosis. Poor prognosis is corresponding to recurrence within 5 years after diagnosis and good prognosis is corresponding to a disease free interval of at least 5 years. If these two groups can be distinguished, patients will be

treated more optimally thus eliminating over or under treatment. In [7] a comparison among the different classifiers decision tree (J48), Multi-Layer Perception (MLP), Naive Bayes, Sequential Minimal Optimization (SMO) and instance based for K-Nearest neighbor (IBK) is presented on WPBC dataset by using classification accuracy and confusion matrix based on 10-fold cross validation method. The experimental results show that the classification using fusion of three different classifiers MLP, J48, SMO and IBK is superior to the other classifiers.

In this research breast cancer prognosis is done with help of PNN on WPBC data. Data is preprocessed before training and testing using PCA. This paper is organized as follows: Section 2 presents the dataset description. Section 3 presents breast cancer prognosis model. Section 4 discusses simulation and results and section 5 concludes the paper.

DATASET DESCRIPTION

The description of WPBC data is given in Table I [5]. These are consecutive patients seen by Dr. Wolberg [8]-[10] since 1984.

Table 1: Attributes of Wisconsin prognostic dataset

Attribute name	Significance	Attribute ID
ID	Unique ID of patient	1
Outcome	Nature of the case (R- recurrent/ N- Non- recurrent)	2
Time	TTR (Time to recur)/ DFS (Disease free survival)	3
Radius 1,2,3	Mean of distances from centre to points on the perimeter	4, 14, 24
Texture 1, 2,3	Standard deviation of gray scale values	5, 15, 25
Perimeter 1,2,3	Perimeter of the cell nucleolus	6, 16, 26
Area 1,2,3	Area of the cell nucleolus	7, 17, 27
Smoothness 1,2,3	Local variation in radius lengths	8, 18, 28
Compactness 1,2,3	Perimeter ² / area - 1.0	9, 19, 29
Concavity 1,2,3	Severity of concave portions of the contour	10, 20, 30
Concave points 1,2,3	Number of concave portions of the contour	11, 21, 31
Symmetry 1,2,3	Symmetry of the cell nuclei	12, 22, 32
Fractal dimension 1,2,3	Coastline approximation - 1	13, 23, 33
Tumour size	Size of the tumour	34
Lymph node status	Status of the lymph node	35

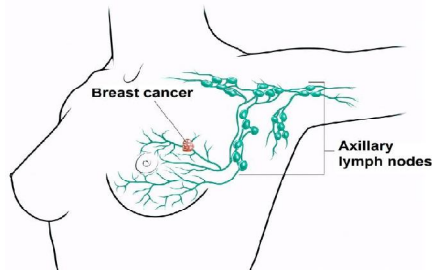


Fig 1: Axillary lymph nodes near a breast with cancer

The outcome in the above dataset is the target attribute (class label) and all other attributes (except ID) are predictor attributes whose values determine the result. Attribute no 3 to 33 are ten real-valued features computed for each cell nucleus from a digitized image of a fine needle aspirate (FNA) of a breast mass. The mean, standard error, and/or largest (worst case-mean of the three largest values) of these features were computed for each image, resulting in 30 features. The thirty fourth attribute is tumor size and the thirty fifth is the lymph node status. Tumour size is the diameter of the excised tumour in centimetres. Lymph node status is the number of positive auxiliary lymph nodes observed at time of surgery. The lymph nodes in the armpit are the first place breast cancer is likely to spread as shown in Fig. 1[7]. Lymph node-negative means the lymph nodes do not contain cancer and lymph node-positive means the lymph nodes contains cancer.

BREAST CANCER PROGNOSIS MODEL USING PNN AND PCA

The model consists of two phases: data pre-processing by PCA and classification by PNN. The Wisconsin Prognostic Breast Cancer dataset is downloaded from the UCI Machine Learning Repository website [8] and saved as a text file. This file is then imported into Excel spreadsheet and the values are saved with the corresponding attributes as column headers. We use PCA to deal with missing values. The ID of the patient cases does not contribute to the classifier performance. Hence it is removed and the outcome attribute defines the target or dependant variable thus reducing the feature set size to 33 attributes. PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables [7]. PCA involves feature selection and feature extraction. Feature selection is the process of finding a subset of the original variables, with the aim to reduce and eliminate the noise dimension. Feature extraction is a technique to transform high-dimensional data into lower dimensions [11]. After pre processing the WPBC data is applied to PNN which classifies the data into two sets recurrent and non-recurrent. In PNN an error back propagation based learning using a norm-squared error function is used which can be described as follows [12]. The aggregation function is considered as a product of linear functions in different dimensions of space. A bipolar sigmoidal activation function is used at each node. This kind

of neuron itself looks complex in the first instance but when used to solve a complicated problem needs less number of parameters as compared to the existing conventional models. PNN is a type of feed forward NN. Fig. 2[13] shows a schematic diagram of a generalized single multiplicative or polynomial neuron. The operator Ω is a multiplicative operation as in (1) and (2) with the weights w_i and biases b_i being the parameter θ of the operator. Fig. 3[13] shows architecture of PNN.

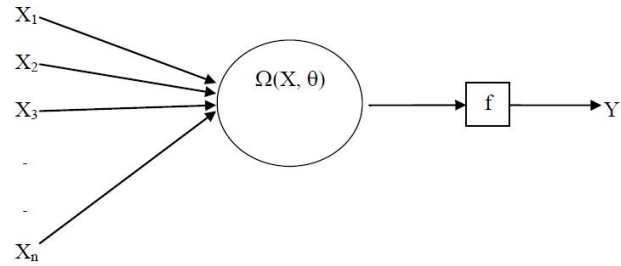


Fig 2: Node Structure of PNN

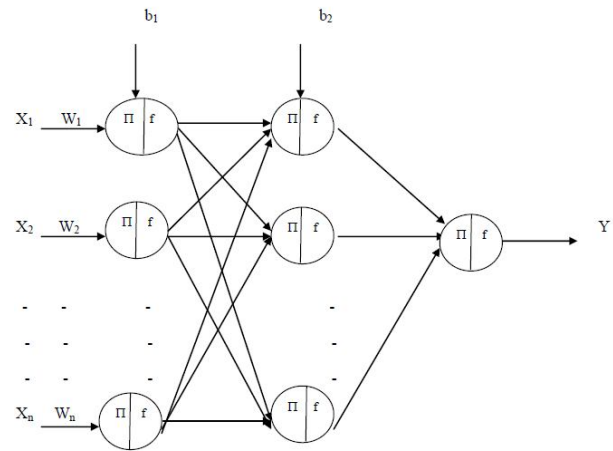


Fig 3: Polynomial Neural Network

Aggregation u before applying activation function is given by-

$$u = \Omega(X, \theta) \quad (1)$$

$$u = \prod_{i=1}^n w_i x_i + b_i \quad (2)$$

The output at the node y is given by-

$$y = f(u) = \frac{1 - e^{-u}}{1 + e^{-u}} \quad (3)$$

The mean square error is given by-

$$E = \frac{1}{2N} \sum_{p=1}^N (y^p - y^d)^2 \quad (4)$$

Where, p is the number of input patterns. The weight update equation for the split complex back propagation algorithm is given by-

$$\begin{aligned} \Delta w_i &= -\eta \frac{dE}{dw_i} \\ &= -\frac{1}{2} \eta (y - d)(1 + y)(1 - y) \frac{u}{w_i x_i + b_i} x_i \end{aligned} \quad (5)$$

Where, η is the learning rate and d is the desired signal.

The bias is updated as-

$$\Delta b_i = -\eta \frac{dE}{db_i} = -\frac{1}{2} \eta (y - d)(1 + y)(1 - y) \frac{u}{w_i x_i + b_i} \quad (6)$$

$$w_i^{new} = w_i^{old} + \Delta w_i \quad (7)$$

$$b_i^{new} = b_i^{old} + \Delta b_i \quad (8)$$

The weights are updated after the entire training sequence has been presented to the network once. This is called learning by epoch. The algorithm is extended to train PNN.

SIMULATION AND RESULTS

PCA data preprocessing is used in this dataset to deal with missing data. The MSE for normalized and PCA preprocessed WPBC data are compared in Table 2 and 3. The corresponding graphs are given in Fig. 4 and 5. We observed that MSE is substantially reduced for PCA processed data.

Table 2: Training performance

Number of Training Patterns	MSE for Normalization	MSE for PCA
100	0.0050	3.5910e-08
198	0.0025	2.1737e-08

Table 3: Testing performance

Number of Testing Patterns	MSE for Normalization	MSE for PCA
100	0.0022	6.9075e-08
198	0.0019	4.6077e-11

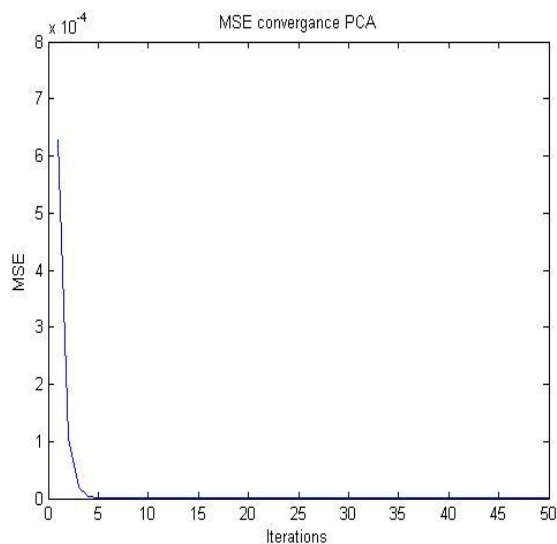


Fig 4: MSE Convergence for PCA

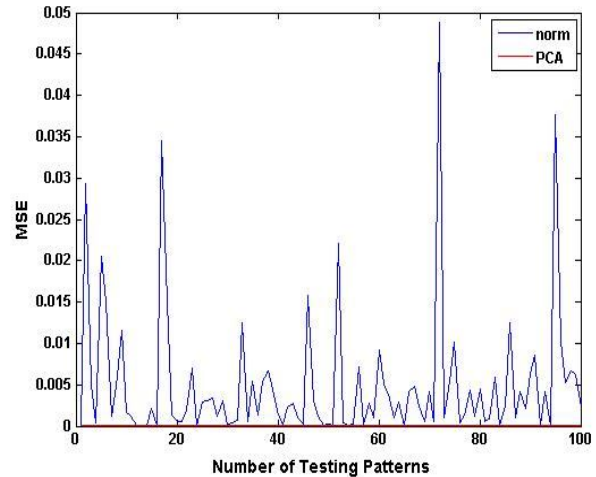


Fig 5(a): Testing Performance Comparison for 100 Data

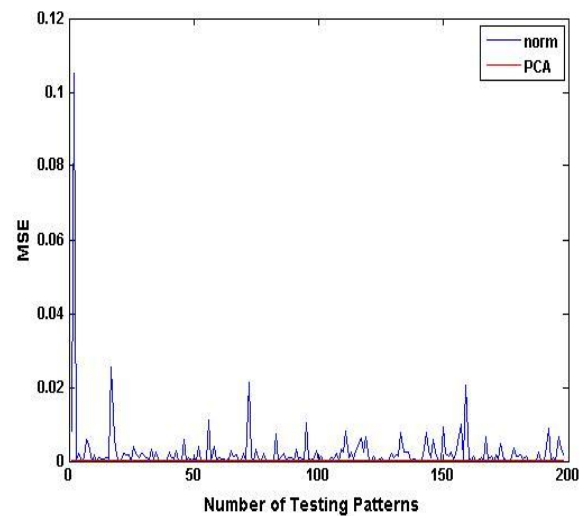


Fig 5(b): Testing Performance Comparison for 198 Data

CONCLUSION

In this research WPBC data is classified with help of PNN. The results are compared for various training and testing patterns. Data pre-processing technique named PCA is used as data preprocessing. Experimental results show that the prognosis by PNN using PCA gives more accurate results than normalization.

REFERENCES

- [1] U.S. cancer statistics working group. United States cancer statistics: 1999–2008 incidence and mortality web-based report. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, 2012[Online]. Available: <http://www.cdc.gov/uscs>.
- [2] P. Pantel, Breast cancer diagnosis and prognosis [Online]. Available: www.citeseerx.ist.psu.edu.
- [3] Breast cancer [Online]. Available: <http://www.womenshealth.gov/breast-cancer/what-is-breast-cancer>.

- [4] I. Anagnostopoulos, C. Anagnostopoulos, A. Rouskas, G. Kormentzas, and D. Vergados "The Wisconsin Breast Cancer Problem: Diagnosis and DFS time prognosis using probabilistic and generalised regression neural classifiers", Draft version of paper to appear at the Oncology reports, special issue Computational Analysis and Decision Support Systems in Oncology, last quarter 2005.
- [5] S. G. Jacob and R. G. Ramani, "Efficient Classifier for Classification of Prognostic Breast Cancer Data through Data Mining Techniques", in *Proc. World Congress on Eng. and Comput. Sci., 2012*, San Francisco, USA.
- [6] O. Gevaert¹, F. D. Smet, D. Timmerman, Y. Moreau, and B. D. Moor, "Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks", *Bioinformatics*, vol. 22 , no. 14, pp. 184-190, 2006,
- [7] G. I. Salama, M. B. Abdelhalim, and M. A. Zeid, "Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers", *Int. J. of Comput. and Inform. Technology*, vol. 1, no. 1, pp. 2277 – 0764, Sept. 2012.
- [8] A. Frank and A. Asuncion (2010). UCI machine learning repository[Online]. Available: <http://archive.ics.uci.edu/ml>. Irvine, CA: University of California, School of Information and Computer Science.
- [9] W.H. Wolberg, W.N. Street, and O.L. Mangasarian, "Image analysis and machine learning applied to Breast cancer diagnosis and prognosis", *Analytical and Quantitative Cytology and Histology*, vol. 17, no. 2, pp. 77-87, April 1995.
- [10] W.H. Wolberg, W.N. Street, D.M. Heisey, and O.L. Mangasarian, "Computer-derived nuclear grade and breast cancer prognosis", *Analytical and Quantitative Cytology and Histology*, vol. 17, pp. 257-264, 1995.
- [11] R. W. Sembiring and J. M. Zain "The design of pre-processing multidimensional data based on component analysis", *Comput. and Inform. Sci* , vol. 4, no. 3, pp. 106-115, May 2011.
- [12] K. Burse, R. N. Yadav, S. C. Srivastava, and V. P. S. Kirar, "A compact Pi network for reducing bit error rate in dispersive FIR channel noise model," *Int. J. of Elect. and Comput. Eng.*, vol. 3, no. 3, 2009, psp. 150-153.
- [13] R.N. Yadav, P.K. Kalra, and J. John, "Time series prediction with single multiplicative neuron model", *Applied Soft Computing*, vol. 7, pp. 1157-1163, 2007.