IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 11, November 2015*

# Approach of Jordan Elman Neural Network to Diagnose Breast Cancer on Three Different Data Sets

**S.Swathi[1], P. Santhosh Kumar[2], P.V.G.K. Sarma[3]**

Research Scholar, Department of Computer Science, S.V. University, Tirupati, India[1]

Research Scholar, Department of Biotechnology, SVIMS University, Tirupati, India[2]

Assoc. Professor, Department of Biotechnology, SVIMS University, Tirupati, India[3]

**Abstract:** Breast cancer is the second leading cause of cancer deaths worldwide, occurs in one out of eight women .still there is no known way of preventing this pathology. Early detection of this disease can greatly enhance the chances of long-term survival of breast cancer victims. Artificial Neural Network is a branch of Artificial intelligence, has been accepted as a new technology in computer science. Neural Networks are currently a 'hot' research area in medicine, particularly in the fields of radiology, urology, cardiology, oncology and etc. It has a huge application in many areas such as education, business; medical, engineering and manufacturing. Neural Networks has been widely used for cancer prediction and prognosis. This paper highlights on Jordan Elman neural networks approaches to solve breast cancer diagnosis, using three different database of breast cancer viz. Wisconsin, WDBC and WPBC. We also introduce recurrent neural network technology as Jordan Elman neural network. To diagnose problems Jordan Elman neural network is successful on three different breast cancer data set is major feature of this paper.

**Keywords:** recurrent network, benign, malignant, WDBC, WPBC, mammography, FNA, mean square Error, Correlation, Sensitivity, Specificity, ROC.

## INTRODUCTION

In recent years machine learning methods have been widely used in prediction, especially in medical diagnosis. Medical diagnosis is one of major problem in medical application. Several research groups are working world wide on the development of neural networks in medical diagnosis. Neural networks are used to increase the accuracy and objectivity of medical diagnosis. ‘Neural networks’ research and application have been studied for a half of hundred years A detailed study on Artifical Neural Network (ANN) can be seen in "Neural and Adaptive Systems: Fundamentals Through Simulations "by Principe, Euliano, and Lefebvre(2000). Paulo J. Lisboa and Azzam F.G. Taktak (2006) had done a systematic review on artificial neural networks in decision support in cancer. The number of research works conducted in the area of breast cancer detection, classification. Many university centers, research centers and commercial institutions are focused on this issue because of the fact that breast cancer is becoming the most common form of cancer disease of today’s female population. Thus, the construction of a fully automatic cancer detection system supporting a human expert has become a challenging and difficult task. Breast cancer is a malignant tumor that starts from cells of the breast. Sometimes cancer cell is benign one but physician fatigue treats them as malignant one. Patient suffers unnecessary biopsy for noninvasive cancer [4]. It is observe that thousands of breast cancer diagnose are in the entire world [1][2][3]. There are three ways to diagnose cancer viz. mammography, fine needle aspirant and surgery. This paper uses fine needle aspirant data sets on artificial neural network (ANN). ANN is collections of mathematical models that emulate some of the observed properties are biological nervous systems and draw on the analogies of adaptive biological learning. The key elements of ANN paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to synapses [6].ANN is used to predict and learn from a given sets of data.

Survival estimations are currently performed by clinicians using the statistical techniques of survival analysis. In this way, artificial neural networks are shown to be a powerful tool for analyzing datasets where there are complicated non-linear interaction between the input data and the information to be predicted for breast cancer relapse [5]. The results show that the proposed system is a useful tool to be used by clinicians to search through large datasets seeking subtle attributes in decision making factors and that may further assist the selection of appropriate adjuvant treatments for the individual patient.

## BACK GROUND

Zheng Yang et.al.[4], Raymond Fang and M. Riedmiller, et.al[7], Sawarkar et.al [8] used Probabilistic Neural Network(PNN), Back propagation and Self organizing Map(SOM) respectively to diagnose breast cancer. Neural network aided breast cancer diagnosis gives promising

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

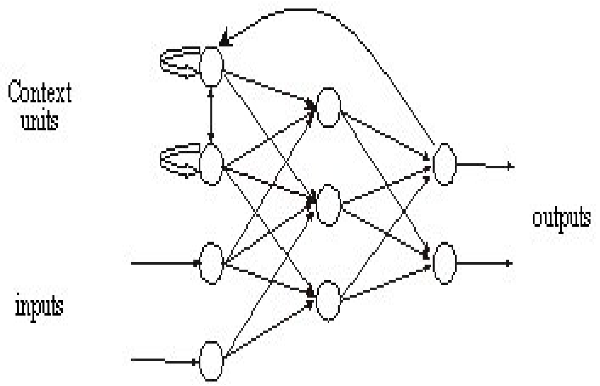
results. It can be a good supplement to the conventional clinical diagnosis system.

## METHODOLOGY

The data: this work grew out of the desire by Dr.Wolberg diagnose breast masses based solely on a fine needle aspiration (FNA). There are total 699 instances, having 16 Missing attributes. Total benign cases are 458 and 241cases malignant. This is winconins database be used by Zheng Yang, Raymond Fang, M. Riedmiller, Sawarkar to diagnose breast cancer. Additionally JEN tested on additionally two more breast cancer data set like WDBC and WDPC which is having 32 attributes. WDBC have 569 instances with benign (B) and malignant (M) is output. WPBC have total 198 instances, in which 151 non recurrent (NR) and 47 recurrent (R) with in 24 monthperiod[12]. The neural network aided breast cancer diagnosis gives promising results. It can be a good supplement to the conventional clinical diagnosis system. In this paper we solve problem with the help of Jordan Elman neural network.

## RECURRENT NETWORK

Recurrent Neural Networks (RNN) has a closed loop in the network topology. Theyare developed to deal with the time varying or time-lagged patterns and are usable forthe problems where the dynamics of the considered process is complex and themeasured data is noisy. Specific groups of the units get the feedback signals from theprevious time steps and these units are called *context* unit [10]. The RNN can be either fullyor partially connected. In a fully connected RNN all the hidden units are connectedrecurrently, whereas in a partially connected RNN the recurrent connections are omittedpartially.Examples of recurrent neural networks are Hopfield networks, Regressivenetworks, Jordan-Elman networks, and Brain-State-In-A-Box (BSB) networks.




**Recurrent Neural Network Architecture**

All types of recurrent neural networks are normally trained with the back-propagationlearning rule by minimizing the error by the gradient descent method. Mostly they usesome computational units which are called associative memories or context units, thatcan learn associations among dissimilar binary objects, where a set of binary inputs isfed to a matrix of resistors, producing a set of binary outputs [11]. The outputs are '1' if thesum of the inputs is above a given threshold, otherwise it is zero. The weights (whichare binary) are updated by using very simple rules based on *Hebbian*learning. These arevery simple devices with one layer of linear units that maps N inputs (a point in Ndimensional space) onto M outputs (a point in M dimensional space). However, theyremember the past events.The simple recurrent network is a single hidden layer feed forward neural network [13]. However, it has feedback connections from the outputs of the hidden layer neurons to the input of the network. This network is similar to an architecture proposed by Jordan but difference is that it has feedback connections from the output layer neuron to input of the network. The Jordan – Elman (JEN) was part of recurrent network where context units save result while doing processing of operation on different entites.

**Jordan-Elman Networks**

Jordan and Elman networks combine the past values of the context unit with the presentinput ($x$) to obtain the present net output. The Jordan context unit acts as a so called low *pass* filter, which creates an output that is the weighted (average) value of some ofits most recent past outputs. The output ($y$) of the network is obtained by summing thepast values multiplied by the scalar parameter $\tau n$. The input to the context unit is copiedfrom the network layer, but the outputs of the context unit are incorporated in the netthrough their adaptive weights.

$$Y(n)=\sum_{I=0}^{n} X(N)JIn - Ieq\text{-----}(1)$$

In these networks, the weighting over time is inflexible since we can only control thetime constant (i.e. the exponential decay). Moreover, a small change in time is reflectedas a large change in the weighting (due to the exponential relationship between the timeconstant and the amplitude). In general, we do not know how large the memory depthshould be, so this makes the choice of $\tau$problematic, without having a mechanism toadopt it.In linear systems, the use of past input signals creates the moving average (MA)models. They can represent signals that have a spectrum with sharp valleys and broadpeaks [16]. The use of the past outputs creates what is known as the *autoregressive* (AR)models. These models can represent signals that have broad valleys and sharp spectralpeaks. The Jordan net is a restricted case of a non-linear AR model, while the configuration with context units fed by the input layer is a restricted case of non-linear MA model. Elman's net does not have a counterpart in linear system theory. These two topologies have different processing power.

**Architecture**

The Elman network commonly is a two-layer network with feedback from the first-layer output to the first-layer input. This recurrent connection allows the Elman network to both detect and generate time-varying patterns.

The Elman network has tansig neurons in its hidden (recurrent) layer, and purelin neurons in its output layer. This combination is special in that two-layer networks

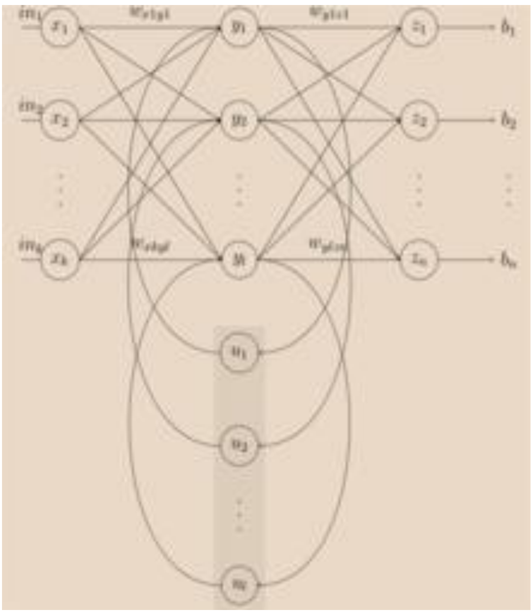
IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 11, November 2015*

with these transfer functions can approximate any function (with a finite number of discontinuities) with arbitrary accuracy [17]. The only requirement is that the hidden layer must have enough neurons. More hidden neurons are needed as the function being fitted increases in complexity [18].

Note that the Elman network differs from conventional two-layer networks in that the first layer has a recurrent connection. The delay in this connection stores values from the previous time step, which can be used in the current time step.

Thus, even if two Elman networks, with the same weights and biases, are given identical inputs at a given time step, their outputs can be different because of different feedback states.Because the network can store information for future reference, it is able to learn temporal patterns as well as spatial patterns [12][13]. The Elman network can be trained to respond to, and to generate, both kinds of patterns.

A two-layer Elman network is shown below.The diagram shows Jordan Elman network and its context units. It receives n inputs and single layer of n hidden nodes to perform operation. Here input start with 1 to K while hidden layer having 1 to L and output unit is
1 to n.



Jordan – Elman Network

**Algorithm:**

1. Begin training
2. Set initial value of weight and bias
3. Set initial value for state units
4. Set state units weight connection
5. Do while stop condition

Apply an input pattern

Do while all pattern at stage 1
Stage2 Feed Forward
Stage3 Backward
Stage4 Weight updating
Apply next pattern
Copy output of output units to state units

6. Loop
7. Check condition
8. Loop
9. End

**Data set:**

The original data is present in the form of analog values ranging from 0-10. The given data sets every neuron operates independently, processing the input receives, adjusting weights and propagating its computed output thus a neuron is a natural level of parallelization for neural networks. Every neuron is treated as a parallel process.

The Wisconsin breast cancer diagnosis, prognostic and original data base, the result of the effort made at the university of Wisconsin hospital for accurately diagnosing breast masses. Nine individually assessed characteristics of the databases and assigned on the integer value between 1 and 10. The measured variables are as follows [1]. The diagnostic, prognostic and WBCD data base were furnished by specialist in the oncologist field.

**The Attribute Information:**
**Wisconsin's Diagnostic Data Set:**
**Attribute Information:**

Ten real-valued features are computed for each cell nucleus:

a) radius (mean of distances from center to points on the perimeter)
b) texture (standard deviation of gray-scale values)
c) perimeter
d) area
e) smoothness (local variation in radius lengths)
f) compactness (perimeter^2 / area - 1.0)
g) concavity (severity of concave portions of the contour)
h) concave points (number of concave portions of the contour)
i) symmetry
j) fractal dimension ("coastline approximation" - 1)

**Wisconsin original data set: (WBCD)**
**Attribute Information:**

1. Clump Thickness: 1 - 10
2. Uniformity of Cell Size: 1 - 10
3. Uniformity of Cell Shape: 1 - 10
4. Marginal Adhesion: 1 - 10
5. Single Epithelial Cell Size: 1 - 10
6. Bare Nuclei: 1 - 10
7. Bland Chromatin: 1 - 10
8. Normal Nucleoli: 1 - 10
9. Mitoses: 1 - 10
10. Class: (2 for benign, 4 for malignant)

**Wisconsin Prognostic Data Set:**
**Attribute Information:**

1. Outcome (R = recur, N = non recur)
2. Time (recurrence time if field 2 = R, disease-free time if field 2 = N)

Ten real-valued features are computed for each cell nucleus:

a) radius (mean of distances from center to points on the

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 11, November 2015*

perimeter)
b) texture (standard deviation of gray-scale values)
c) perimeter
d) area
e) smoothness (local variation in radius lengths)
f) compactness (perimeter^2 / area - 1.0)
g) concavity (severity of concave portions of the contour)
h) concave points (number of concave portions of the contour)
i) symmetry
j) fractal dimension ("coastline approximation" - 1)

The Jordan Elman neural network can be trained with Momentum, quick prop delta bar delta, Leven beg Marqua and conjugate gradient method. Momentum learning rule is used because the momentum provides the gradient descent with some inertia, so that it tends to move along a direction that is the average estimate for down. The amount of inertia i.e; how much of the past to average over, is dictated by the momentum parameter.

| Performance Criteria | DATABASE of Breast Cancer | | |
|---|---|---|---|
| | Wisconsin's | WDBC | WDPC |
| Sensitivity (%) | 100 | 98.92 | 94.78 |
| Specificity (%) | 98.03 | 100 | 87.86 |
| Accuracy (%) | 98.75 | 97.25 | 89.725 |
| Mean square Error | 0.046 | 0.08 | 0.732 |
| Correlation vector | 0.95 | 0.93 | 0.41 |

The higher the momentum, the more it smoothes the gradient estimate and the less effect a single change in the gradient has on the weight change.

The major benefit is the added ability to break out of local minima that a step component might otherwise get caught in. the momentum parameter is the same for all weights of the attached component. An access point has been provided for the step size and momentum allowing access for adaptive and schedule learning rate procedures.

## RESULTS AND SIMULATION

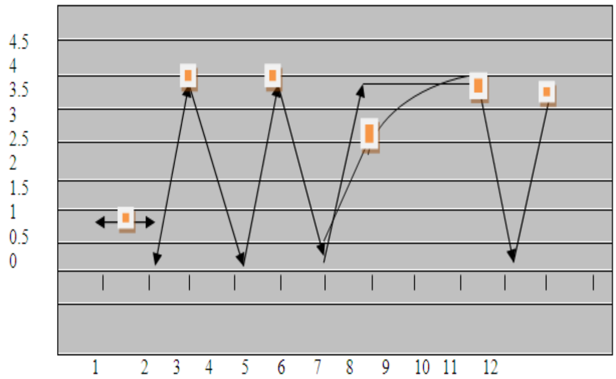
The different attribute in the data set I, II and III to be tested, Aim is to find the role of Jordan Elman algorithm in deciding the efficiency for breast cancer detection and diagnosis. Neuro solution is used to solve problem. Context unit used is 0.8. Transfer function for context used is 0.8. Transfer function for context used is integrator axon. Momentum is assign to be 0.7, step size is 0.1. Hidden layers used are 1, number of neurons is hidden layer 4, number of epochs are 10000.

The training time and testing time result of respective 3 different sets has been shown in table:

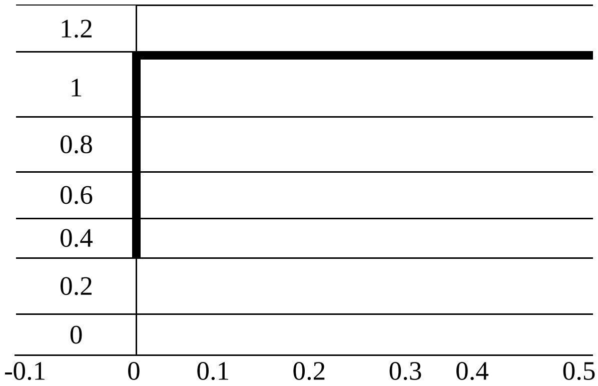
| | DATABASE of Breast Cancer | | | | | |
|---|---|---|---|---|---|---|
| | Wisconsin's | | WDBC | | WDPC | |
| | B | M | B | M | NR | R |
| Samples | 458 | 241 | 357 | 212 | 151 | 57 |
| Diagnose | 456 | 237 | 350 | 208 | 107 | 41 |

The neural network is trained and tested for individual attribute as input to the neural network. After experimentation it is observed that minimum MSE obtained is 0.046 for Wisconsin's database, sensitivity is 100%, specificity is also 100% and accuracy is also 98.75% for momentum learning.

Testing Time Graph for Wisconsin Database




ROC graphs have long been used in signal detection theory to depict trades between hit rate and false error rate.




## CONCLUSION

The last decade has witnessed major advancements in the methods of the diagnosis of breast cancer. It was found that the use of ANN increases the accuracy of most of the methods and reduces the need of the human expert. Neural network aided breast cancer diagnosis gives promising results. It can be a good supplement to the conventional clinical diagnosis system provide the medical experts with a second opinion thus removing the need for biopsy, excision and reduce the unnecessary expenditure.. This study shows the decision taking ability of artificial neural network model.

For diagnosis the efficiency of Jordan Elman neural network, it shows that it can support the doctors or physicians to consider it as a second opinion of the learning machine to prevent biopsy. In addition, these neural network based clinical decision support systems avoid unnecessary excision and expenses. Malignant cancer cell can be efficiently diagnosed with caring such attributes.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 11, November 2015*

## REFERENCES

[1]. December, 2003 Breast Cancer Facts American Cancer Society national cancer institute Komen Foundation.
[2]. 2003: Image (NBCC) National Breast Cancer Center website: www.nbcc.org
[3].Understanding breast cancer website: http:// www. Breastcancer.org/ symptoms
[4]. K. ZhengRong Yang, Weiping Lu, Dejin Yu, Robert Harrison, Detecting false benign in breast cancer diagnosis, IEEE-INNS-ENNS international joint Conference on Neural Networks ( IJCNN'00)- Volume 3
[5]. Scarff RW, Torloni H: Histological typing of breast tumors. International histological classification of tumors. World Health Organization 1968, 2(2):13-20.
[6] Elston CW, Ellis IO: Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991, 19:403-410.
[7]. M.Riedmiller, and H. Braun, 'A Comprehensive Foundation, MacMillan College Publishing Co., NewYork , 1994.
[8]. Sudhir D. Sawarkar, Ashok A.Ghatol, " predictoion of breast cancer malignancy using self-organizing maps" NCSIT.
[9]. Jordan M.I & Jacobs R.A(1994) Hierarchical mixtures of experts and the EM algorithm b'euml computation.
[10]. C. Demir and B.Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," Rensselaer Polytechnic Institute, Tech. Rep., 2005.
[11] .A. Tutac, D. Racoceanu, T. Putti, W. Xong, W.-K. Leow, and V. Cretu, "Knowledge-guided semantic indexing of breast cancer histopathology images," in *Proc. Int. Conf. on Biomedical Engineering and Informatics*, 2008.
[12]. S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal gradedifferentiating parameters for breast cancer," *BMC Medical Imaging*, vol. 6, no. 14, 2006.
[13]. S. Doyle, M. Hwang, M. Feldman, and J. Tomaszeweski, "Automated grading of prostate cancer using architectural and textural image features," in *Proc. of 4th IEEE Int. Symp. on Biomedical Imaging*, 2007, pp. 1284 – 1287.
[14] .H. Soltanian-Zadeh and K. Jafari-Khouzani, "Multiwavelet grading of prostate pathological images," *IEEE Trans. on Biomedical Engineering*, vol. 50, pp. 697–704, 2003.
[15]. R. Owen Rigers. A framework for parallel data miningusing neural networks. Technical report , Queen'sUniversity, Canada, 1997.
[16].SimonHaykin. Neural Networks –A Comprehensive Foundation. Pearson Education, 2001.
[17]. K. Anil Jain, Jianchang Mao and K.M. Mohiuddin.Artificial Neural Networks: A Tutorial. IEEE Computers,1996, pp.31-44.
[18]. George Cybenko. Neural Networks in ComputationalScience and Engineering. IEEE Computational Science andEngineering, 1996, pp.36-42.