Diabetes Detection using Deep Learning Approach

Ms. T P. Kamble
Student
Vishwakarma Institute of Technology, Pune

Dr. S. T. Patil
Professor
Vishwakarma Institute of Technology, Pune

Abstract

Deep learning is branch of Artificial Intelligence. It has same architecture as neural network but has extra hidden layers. Deep learning had shown more data processing power or capability than the shallow architecture. It had shown more accuracy in results. The Deep network has Restricted Boltzmann machine as basic building block but restricted Boltzmann machine can be used as classifier, feature extractor [1].Today due to modern life style the people have more stress of work, less physical activities, changed eating habits due to this reason people facing many health related problem. The diabetes is one reason behind the death of people. Diabetes may lead to kidney, eye problems heart problem also [2]. Hence it is better to detect Diabetes in early stage to avoid other health risks. In proposed System Deep learning based Restricted Boltzmann machine approach is used to detect whether patient is diabetic or not as Restricted Boltzmann machine is popular for classification and recognition purpose. To detect either patient is having type 1 or type 2 diabetes decision tree technique used.

Keywords: Deep learning, Restricted Boltzmann machine, Decision Tree Algorithm

I. INTRODUCTION

Deep learning model is of human brain type model. It has ability to solve the complicated problems like human brain. Deep network can handle large set of data, learn complex functions as well as it can handle unlabelled data with small human inputs [3].Basic theme of deep learning is that it generates the high level representation of raw data. Deep learning is a learning method with the deep architecture and algorithms which can learn the features which are having no labels .i.e. unsupervised data. This motivates the idea that we must do a lot of unsupervised learning .As the deep learning follows the principle of human brain. This ability, together with the efficient learning algorithms that can ensure this ability, it is fast in processing and it can handle high dimensional data as input just one step ahead of artificial intelligence. Deep learning approach is more popular because of its deep architecture [4].According to world health organization the percentage of having diabetes is very high. Today diabetes is major reason behind the death of peoples. Hence we can say that Diabetes is a chronic disease. Diabetes occurs either when the body does not produce enough insulin or when the body cannot effectively use the insulin which is produced by pancreas. Usually doctor takes blood sample of patient then he checks the sugar concentration in blood sample. This process takes more time and there are many other factors which needs to be considered when we are trying to detect whether patient is diabetic or not .The other factors such age body mass index, age, family history. If the ancestors showing the diabetic presence then there are chances that next generation will have diabetes [2].There is no application or tool which can detect whether patient is having Type 1 or Type 2 diabetes. So there is need of tool which can detect patient is diabetic or not and also which will detect whether patient is having Type 1 or Type 2 diabetes. There are various symptoms behind the type 1 and type 2 diabetes usually doctor check these symptoms then conclude whether patient is having type 1 or type 2 diabetes but this whole process takes more time. Hence we need some automated work which will give the decision about presence of diabetes and the type of the diabetes.

II. RELATED WORK

In past Large number of work has been done to find out efficient methods of medical diagnosis for various diseases. Similarly lots of research has been done to check whether patient is diabetic or not. Our work is an attempt to predict efficiently diagnosis of Diabetes with reduced number of attributes which are based on symptoms that occurs early stages of Diabetes Mellitus. In 2012 Rahmat Zolfaghari proposed a three-layer hierarchy multi-classifier. And min max normalization is applied on the data to avoid the numerical difficulties in calculations. First Layer detects the features then next layer separately classify using back propagation and svm then in next layer combine output from both layers the ensemble means combination or fusion of 2 or more algorithms. [5].

In 2014 K. SRIDAR and Dr. D. SHANTHI used the back propagation and Apriori algorithm for the diabetes detection purpose. First the input is taken from the user. The input from user is glucose level which was detected in blood sample then back propagation algorithm is applied on the inputs then the output of back propagation is applied as input to association rule mining. Final output shows percentage of diabetes in patient [6].

In 2014 Gaganjot Kaur, Amit Chhabra proposed Classification which used Decision tree algorithm to predict class whether patient is diabetic or not. The labelled data is feed as input. The leaf of the tree j48 acts as the class labels. The information gain is calculated for each attribute. Then the gain in information is calculated that would result from a test on the attribute [7].

III. PROBLEM

A. Perceptron learning:

Perceptron is first generation of the neural network.it was invented in 1960 by frank Rosenblatt. It was having one input layer of features and output layer and weight vectors in network. Limitations: The limitation of Perceptron learning was that it was only able to handle linearly separable functions such as and, or. It was not able to handle functions like xor which are linearly not separable.

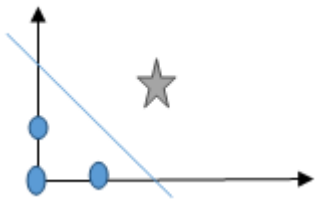

Fig. 1: Boolean and linearly separable

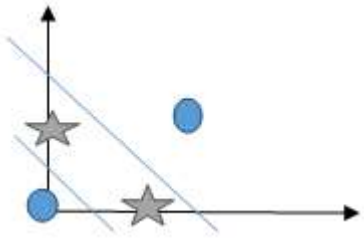

Fig. 2: Boolean Linearly Non separable

It was able to handle only binary values as input and output. There was need to add one more layer to handle linearly non separable functions such as x-or [8].

B. Back Propagation:

Back propagation is an artificial neural network which uses the supervised learning approach. It was having three layers input layer, output layer in between input and the output the layer is present called hidden layer. The threshold of output is set to some value. The weights between input to hidden and hidden to output layer are updated so that the error is negligible. When there is major difference in threshold set and the output then error is sent back to inner layers hence the name was given back propagation algorithm

C. Limitations of Back Propagation:

The many limitations of back propagation are local minima, slow convergence and network paralysis.

1) Local minima-

The network always changes its weights hence the error is reproduced it does not get reduced and algorithm gets stuck and error will not reduce further.

2) Network paralysis-

The weights are kept to very large values the large values force most of the units of network to operate in extreme values in such region derivation of activation functions is very small.

The number of times back propagation requires the input pattern presentation for which weights need to update so that network will settle down to its optimal solutions [9].

D. Support vector machine:

Vapnik developed algorithm for support vector machines the support vector machine task is to find the plane which separates the data in to 2 classes such as plane is called hyper plane. The hyper plane margin should be more to classify data.

1) Limitations of SVM:

1) The svm performance is mainly depend upon the kernel function we have chosen.2. The speed of svm is slow while training and testing phases [10].

IV. METHODOLOGY

In proposed system as shown in Figure 2 the steps are performed. First the feature is selected then that are applied to restricted Boltzmann machine for classification. Classifier detects whether diabetic or not. If classifiers output is diabetic then it is further applied to Decision Tree algorithm to check whether is Type 1 or Type 2 diabetes.

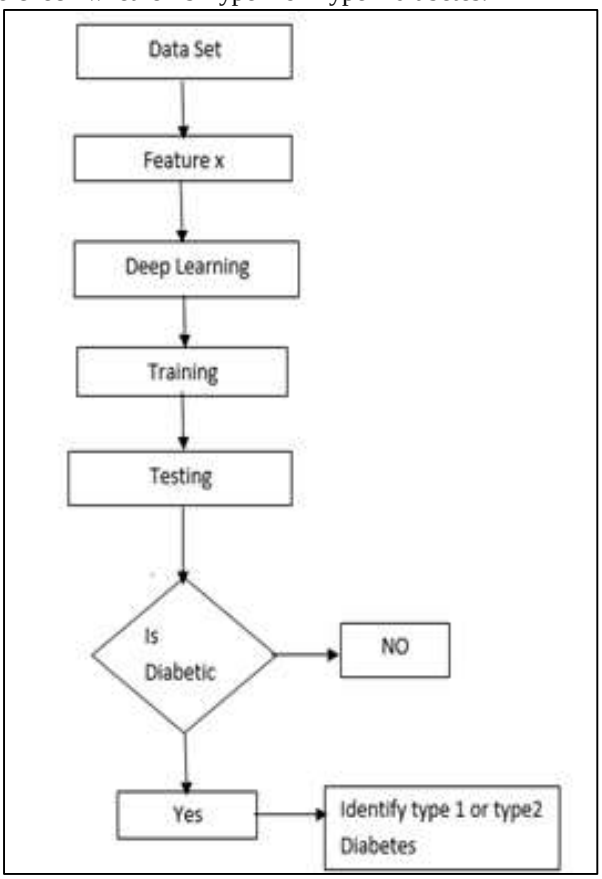

Fig. 3: Proposed System architecture.

A. Data Preprocessing:

In data pre-processing step first only the three features are selected from the file. Then that features data set is normalized to min max normalization. To get input vector range between 0 and 1 and avoid computation complexity. And next the dataset is divided into training dataset and test dataset such as 80% of training dataset and 20% of test dataset. Upper bound (UB) =1 and lower bound (LB) =0.equation number (1) is used to find min max normalization of data [11].

$$X_{normalized} = (X-X_{Min})/ (X_{Max}-X_{Min}) (UB –LB) \quad (1)$$

B. Restricted Boltzmann Machine for Classification:

Restricted Boltzmann machine is model which is having structure like bipartite graph and it is energy based model. Its graphical structure have undirected graph. It has visible layer and the hidden layer [12].The input is directly clamped with the visible layer. It has a layer of visible units connected to a layer of hidden units but no connections within a layer Typically, RBMs use binary units for both visible and hidden variables. But the real valued data can be applied to RBM which is having structure as Gaussian Bernoulli type architecture.[12]. To model real-valued data, a modified RBM with binary logistic hidden units and real-valued Gaussian visible units can be used [13].

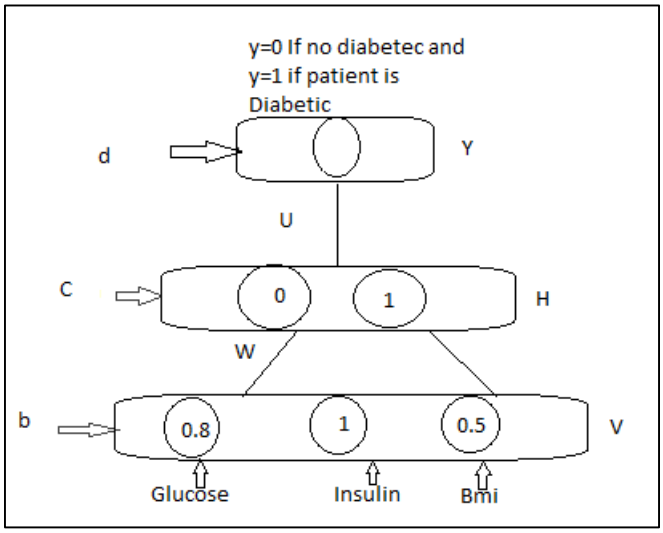

Fig. 4: Architecture for restricted Boltzmann machine for classification.

Arestricted Boltzmann machine is aparticular type of a Markov random field which has two layers. One layer of an RBM consists of visible input units, v is ranging from 0 to 1, which are connected to the other layer of hidden stochastic unitsis h either 0 or 1. Figure 4 shows the basic structure of an RBM.Thedistributionofstate{v,h}ofanRBMisspecified by the following energyfunction[13]

$$E(v,h) = \sum_j \frac{(vi - bi)^2}{\sigma_i^2} - \sum_{i,j} \frac{vi}{\sigma\, i} * w_{i,j} * h_j - \sum_j c_j h_j \quad (2)$$

where W represents visible-to-hidden weight matrix consisting of weights wij of connections between neurons vi and hj, b represents a visible bias vector, and a represents a hidden bias vector. A set of all parameters can be denoted by θ = {W,b,a}. The conditional distribution of the hidden vector h given the visible vector v i.e. The joint probaility distribution is computed using formula(2).The Distribution computation energy is kept minimum to obtain maximum probability as negative energy is increase in probability and vice versa.

$$\mathrm{P(v,h)} = \frac{\exp(-\mathrm{E(v,h)})}{\mathrm{z}} \quad (3)$$

Where Z is partition function given summing all the pairs of visible and hidden unit given by below equation[16].

$$Z = \sum_{v,h} e^{-E(v,h)} \quad (4)$$

The conditional probability distribution of each unit is given by the logistic sigmoid activation function of the input it receives using below formual:

$$P(h_j|v) = sigm(\sum_i w_{i,j} v_i + c_j) \quad (5)$$

$$P(v_i|h) = \mathbb{N}(\sum_j w_{i,j} h_j + b_i, \sigma_i) \quad (6)$$

$$\mathrm{sigm(x)} = \frac{1}{1+\exp(-\mathrm{x})} \quad (7)$$

Computing p(y, x) is intractable, but it is possibleto compute p(y|x), sample from it, or choose the most probable class under this model,for reasonable numbers of classes C this conditional distribution can be computed exactly and efficiently[13], below formula is used to compute energy distribution when there is classification related problem:
Energy function function defines Distribution as

$$\mathrm{P(v,h,y)} = \frac{\exp(-\mathrm{E(v,h,y)})}{\mathrm{z}} \quad (8)$$

Using the distribution given in equation(8) the classes are predicted such as to check patient is having diabetes or not having diabetes.

Steps in training restricted boltzman machine are as follows[14]:

1) Take trainnig data set clamp directly data set to visible unit.
2) To update the hidden unit states use the sigmod activation function equation number 7.
3) For ith hidden unit compute activation energgy using equation number 5.
4) Set visible unit value to 1 using formula 3 and to 0 usnig equation number 6.
5) Compute positive statistics for edge (eij) = vi*hj

6) Reconstruct the visible units using similar technique. For each visible unit, compute the activation energy equation number 5and update the state.
7) Now update hidden units again, and compute (eij) = vi*hj which is negative statistics for each edge.

In restricted boltzman machien the visible unit are used to find distribution of the hidden units and hidden units are used to compute the distribution of visible unit until the stable state is obtained.Markov chain is important in RBM as it helps to get samples from probality distribution such as gibbs ditribution.

Markov Chain is prcocess where the next state of system is depend on the current state of the system not on the next states of systems[16].

Figure 5 shows the one step in gibbs sampling

Steps in Gibbs sampling[15]:

1) Start updating at random state of visible unit.
2) Then update all hidden units in parallel.
3) Update or reconstruct visible unit in parallel.
4) Repeat all process for all training example.

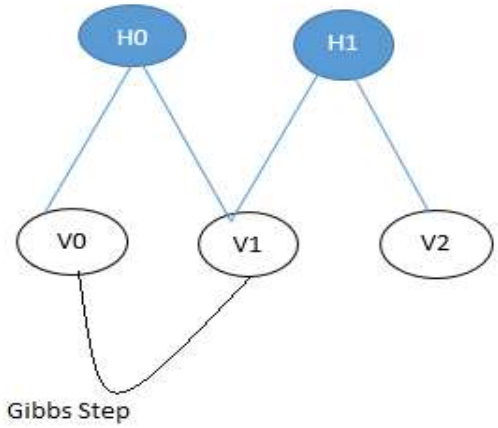

Fig. 5: Gibbs sampling step

Once the Sample probabilites of hidden layer computed using restrcited boltzman machine then the logistic regression is applied to classify the data as diabetic or not diabetic.In the Logistic regression binary decesion tassk function equation (9) is used to take decesion whether data belongs to class 1 or class 0.

$$0<=h(x)<=1 \qquad (9)$$

Sigmod activation function is used to interprete result of logistic regression.

If h(x)>=0.7 then we can say it is class 1 data point Else if h(x)<0.7 then the data point belog to class 0

C. *Decision Tree algorithm to identify type 1 or type 2 diabetes:*

To find out whether patient is having Type1 or Type 2 diabetes decision tree is used. The Decision Tree finds frequency or count of the rows which is having the matching condition of the symptoms of the type 1 and type 2 diabetes. The features selected to detect whether Type 1 or Type 2 diabetes are BMI, age, insulin level, glucose level and the diabetic pedigree function[17].

The Type 1 diabetes occurs in children and teenagers also. The person having normal BMI may have Type 1 diabetes. The glucose level is less than 140mg/dL in type 1 diabetes patient. The family history doesn't matter for the Type1 diabetes. It can occur if no one ancestor has diabetes in family. A person is said to have a normal response when the 2-hour glucose level is less than or equal to 110 mg/dL [2]

The type 2 diabetes occurs in age of 40 also the insulin levels of patient are very high. When the patient is having glucose level above 140mg/dl then it is treated s type 2 diabetes The type 2diabetes occur when the patient have BMI above 30 means when patient is overweight, the diabetes pedigree function value between equals 0.500 when the relative. Is a parent or full sibling, equals 0.250 when the relative. is a half sibling, grandparent, aunt or uncle, and equals 0.125 when the relative, is a half aunt, half uncle or first cousin below the 0.125 is considered as no family history is there of diabetic person[2].

Following steps are performed in decision tree algorithm [17]:

1) Step 1: Divide Dataset into training set and test set.
2) Step 2: Choose a feature which differentiate the instances in training set.
3) Step 3: Create tree node whose value is chosen attribute
4) Step 3a: Create child links from this node where each link represents unique value for chosen attribute.
5) Step 3b: Use child link values to subdivide the instances into sub classes.
6) Step 4: For each sub class created in step 3

7) Step 4a: If the instances in the subclass satisfy the predefined criteria or if the set of remaining attribute choices for this path is null specify the classification for new instances following this decision path.
8) Step 4b: If the subclass does not satisfy the criteria and there is at least one attribute to further subdivide the path of the tree, let T be the current set of subclass instance and return to step 2

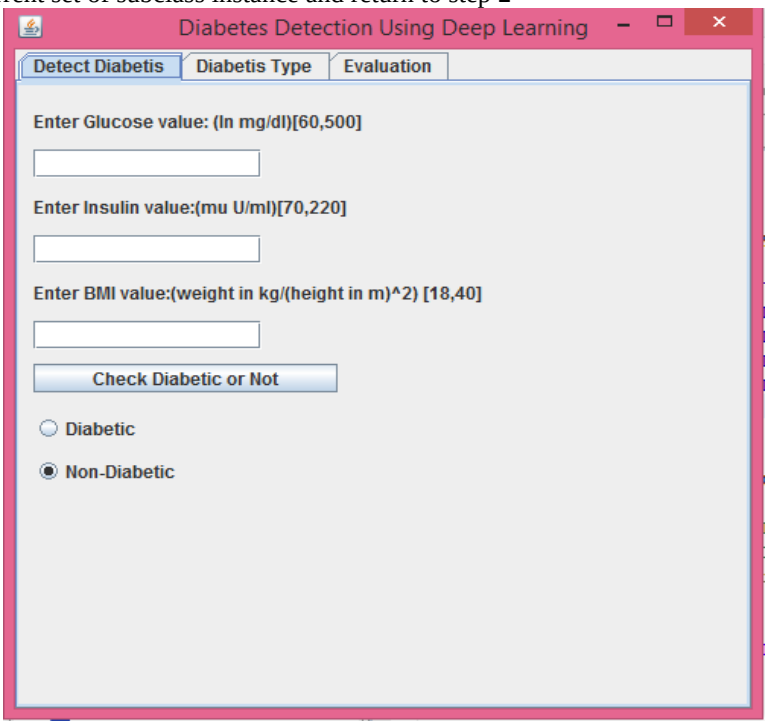

Fig. 7: GUI for Diabetes detection

To detect Diabetic or not the GUI is used as shown in figure 7 if patient is diabetic then diabetic radio button will be on else non diabetic button will be on and on button click the Feature extraction using RBM and logistic regression is executed to find whether patient is diabetic or not.

GUI For the system is as follows to detect Type 1 or Type 2 diabetes as shown in figure 8.if patient is type 1 diabetic then the type 1 radio button will be on if type 2 diabetic then type 2 diabetic button will be on and on button click the classification based decision tree algorithm is executed find type of diabetes.

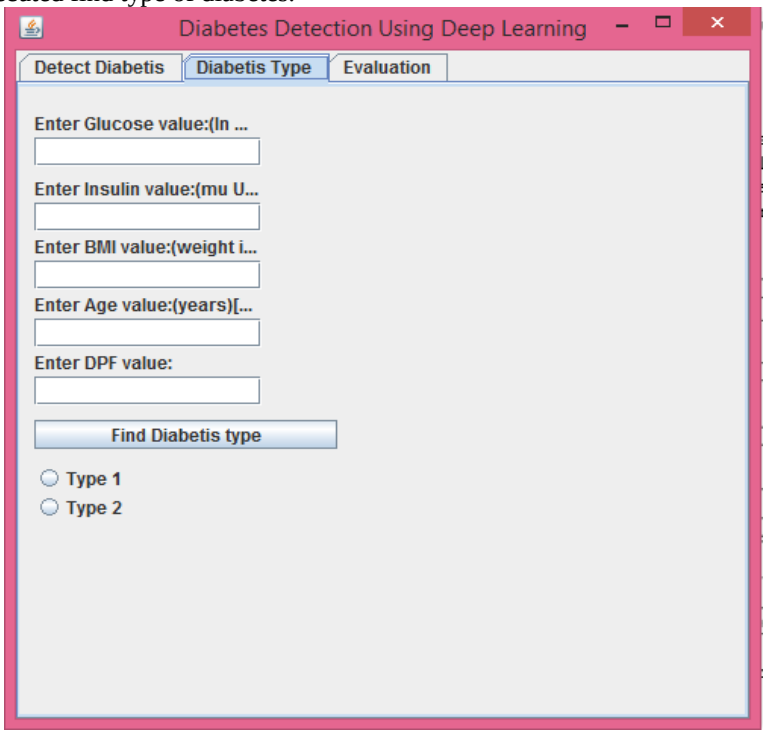

Fig. 8: GUI for type 1 or type 2 diabetes detection

V. RESULT

Initial Result obtained after applying logistic regression on data generated from restricted Boltzmann machine
Correctly Classified Instances 64 80%
Incorrectly Classified Instances 16 20%
Kappa statistic 0.5046
Mean absolute error 0.3571
Root mean squared error 0.434
Relative absolute error 81.1103 %
Root relative squared error 92.6516 %
Coverage of cases (0.95 level) 96.25 %
Mean rel. region size (0.95 level) 86.25 %
Total Number of Instances 80

Table – 1
Showing the correctly classified instances and incorrectly classified instances using RBM.

Predicted Classes			
0	*1*		*Actual Classes*
50	*4*	*0*	
12	*14*	*1*	

Precision for Non Diabetic i.e.0:=0.8064
Precision for Diabetic i.e.1:=0.7777
Recall for Non Diabetic i.e.0:=0.9259
Recall for Diabetic i.e.1:=0.53846
Error Rate: =0.2
In Diabetes Type Detection Result obtained after applying Decision tree (J48) are as follows.
Correctly Classified Instances 101 94.3925 %
Incorrectly Classified 6 5.6075 %
Kappa statistic 0.8711
Mean absolute error 0.0891
Root mean squared error 0.2005
Relative absolute error 19.6464 %
Root relative squared error 42.1447 %
Coverage of cases (0.95 level) 100 %
Mean rel. region size (0.95 level) 85.514 %
Total Number of Instances 107

Table – 2
Showing the correctly classified instances and incorrectly classified instances using J48..

Predicted Classes			
T2DM	*T1DM*		*Actual Classes*
70	*0*	*T2DM*	
6	*31*	*T1DM*	

Precision for T2DM:=0.9210526315789473
Precision for T1DM:=1.0
Recall for T2DM:=1.0
Recall for T1DM:=0.8378378378378378
Error Rate: =0.056074766355140186

VI. CONCLUSION

In proposed system used the restricted Boltzmann machine one of the algorithm of deep learning approach for the diabetes classification purpose such as to check whether patient is diabetic or not. Then System used decision tree to check whether it is type 1 or type 2 diabetes. Deep architectures help deep learning by trading a more complicated space for better performance, in some cases, even for less computation time. Currently result are generated on 300 data points in future dataset size is kept to be large.

REFERENCES

[1] Dandan Mo, "A survey on deep learning: one small step toward AI" December 4, 2012.

[2] Jennifer Lamb,"Diabetes", MATH 3220 Final Project
[3] Yoshua Bengio, "Learning Deep Architectures for AI", volume 2. 2009.
[4] Li Deng and Dong Yu, "Deep Learning Methods and Applications", Foundations and Trends® in Signal Processing 7:3-4.
[5] Rahmat Zolfaghari, "Diagnosis of Diabetes in Female Population of Pima Indian Heritage with Ensemble of BP Neural Network and SVM",IJCEM International Journal of Computational Engineering & Management, Vol. 15 Issue 4, July 2012
[6] K. Sridar,Dr. D. Shanthi, "Medical diagnosis system for the diabetes Mellitus by using back propagation-Apriori Algorithms",Journal of Theoretical and Applied Information Technology 10th October 2014. Vol. 68 No.1
[7] Gaganjot Kaur, "Improved J48 Classification Algorithm for the Prediction of Diabetes",International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014
[8] "Rosenblatt's Perceptron" chapter 1
[9] Saduf, Mohd Arif Wani ,"Comparative Study of Back Propagation Learning Algorithms for Neural Networks",International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 3, Issue 12, December 2013 ISSN: 2277 128X
[10] M.A. Wiering, M. Schutten, A. Millea, A. Meijster, and L.R.B. Schomaker "Deep Support Vector Machines for Regression Problems",Institute of Artificial Intelligence and Cognitive Engineering
[11] yogendra kumar jain , santosh kumar bhandare,"Min Max Normalization Based Data Perturbation Method for Privacy Protection"International Journal of Computer & communication Technology Volume-2 Issue-VIII, 2011
[12] Jakub M. Tomczak, "Application of Classification Restricted Boltzmann Machine with discriminative and sparse learning to medical domains", Institute of Computer Science Wroclaw University of Technology.
[13] Gijs van Tulder and Marleen de Bruijne "Learning Features for Tissue Classification with the Classification Restricted Boltzmann Machine", springer international publishing Switzerland 2014
[14] Indraneel Pole,Student developer at Daenet GmbH "Restricted Boltzmann Machine - A comprehensive study with a focus on Deep Belief Networks".
[15] Asja Fischer and Christian Igel,"An Introduction to Restricted Boltzmann Machines"CIARP 2012, LNCS 7441, pp. 14–36, 2012. Springer-Verlag Berlin Heidelberg 2012
[16] Asja Fischer Université de Montréal,"Training Restricted boltzman machine"16 PUBLICATIONS 80 CITATION (Thesis).
[17] Amit Bharti December, 2004 "A Decision Tree Approach to Extract Knowledge for Improving Satellite Image Classification"