

Brain Disease Classification Using Deep Learning Technique

¹K.Venu, ²N.Sasipriya, ³K.Narendran, ⁴S.Rajkumar, ⁵R.Revanth

¹Assistant Professor, ²Assistant Professor, Kongu Engineering College
^{3,4,5}UG Scholar, Kongu Engineering College

Abstract

Alzheimer's disease (AD) is the most common fatal and progressive neurological disorder that was assumed to result in dementia, learning disabilities and gait inconsistencies. Although the currently existing treatments cannot stop the disorder from becoming more serious, they can temporarily slow the worsening of symptoms and improve the quality of life of the victims and their caregivers, if identified earlier. For the diagnosis of Alzheimer's, brain Magnetic Resonance Image is normally taken. Deep neural networks provide the efficient result in processing the medical image data compared to the traditional machine learning algorithms. In this paper, the combination of Convolution neural network and deep belief network is introduced for the diagnosis of AD on MRI images. The performance of this automated system is validated by taking samples from Alzheimer's disease Neuroimaging Initiative. The parameters such as precision, sensitivity, specificity and accuracy are calculated to qualitatively validate the system.

Keyword: Alzheimer's disease, deep learning, Classification, MRI images

Introduction

The neurological disorders result in the changes of behavioural abilities, memory impairment, and gait inconsistencies which could be fatal to the patients if not detected, diagnosed and treated early (Mattson et al. 2001). Alzheimer's disease is a gradually progressing irreversible neurological disorder which is found to be the most common among dementia. It results in various levels of cognitive impairments of an individual (Prince et al. 2015; Ferreira et al. 2014). AD is initiated by the accumulation of amyloid-beta and tau proteins in the brain. The abnormal presence of amyloid-beta proteins causes amyloid plaques in the MTL and cortex regions causing damage to the neurons and disrupting the communications among the brain regions (Brunnstrom&Englund 2010). The abnormal deposition of tau proteins forms Neurofibrillary Tangles (NFTs) which leads to functional breakdown of neurons resulting in the cellular death in the brain tissues. The progressive death of neuronal cells leads to the morphological variation of brain regions and finally results in brain shrinkage.

The present research intends to find out the most suitable classification technique among the available ones and a new one based on deep learning approaches, which is a multi-resolution technique and is not yet used for AD detection. Therefore the combination of CNN and DBN is applied to develop an automatic AD classification with good accuracy. The whole system will be used to diagnose the Alzheimer's disease at the early stage. This can be used as a biomarker to detect memory related diseases. Since it is automatic, it is user friendly and accurate. The performance is analyzed by calculating the Precision, Recall and accuracy. The texture features and the number of features for classification is reduced using convolution neural network. The Deep belief network is used to classify the disease. The classifier is validated by calculating precision, sensitivity, specificity, and accuracy.

Related works

Functional MRI (fMRI) is a non-invasive imaging methodology which tends to assess the functional and cognitive abilities of the human brain. The spatial resolution of fMRI is poor compared to structural imaging technique (Varghese et al. [1]). It has also been reported that techniques like magnetic resonance spectroscopy, diffusion tensor imaging and molecular imaging have also aided in the diagnosis of AD. The imaging biomarkers are widely proved to act as a potential indicator to detect and diagnose atrophy. They also help in deciding the appropriate therapies to slow down the disease progression. Pathological reports confirmed the presence of senile plaque and NFTs on the hippocampus 15 years before the initialization of clinical incidence of AD. This could aid in early diagnosis and provide treatments to slow down the progression of AD.

Gaussian filtering is used for noise smoothing in images during segmentation using level set methods. The gradient information obtained during diffusion process is used as the edge stopping criterion in curve evolution. Gaussian diffusion results in smoothening of edge information in the images thus affecting the segmentation process (Suganthi&Ramakrishnan [2]).

He et al. [3] have proposed inverse P-M model by modifying traditional P-M model for noise suppression and enhancement of edges. The second order Partial Differential Equation (PDE) has achieved good trade-off between noise suppression and preservation of edge information but results in the blocky effects which could affect the segmentation process. ShaikBasheera and M Satya Sai Ram [4] developed the AD diagnosis system using convolution neural network (CNN). The Gaussian filter and skull stripping algorithm is applied for voxels enhancement. The performance of this study is evaluated with the metrics accuracy, recall and precision and achieved 90.47%, 86.66%, and 92.59% respectively.

The regional and edge level set methods are prone to improper segmentation due to poor identification of weak edges and considering the high intensity noise pixels as edge pixel. Therefore, hybrid methods were proposed to overcome the setbacks of the regional and edge-based level set method by incorporating both the edge and regional information into the curve evolution terms for efficient segmentation. The hybrid level set method has the advantages of lower computation time and effective segmentation performance compared to traditional regional and edge level set methods (Jiang et al. [6]).

Ruoxuan Cui et al [7] introduced the RNN based AD diagnosis system using the CNN as a feature extractor. Spatial features of MR images are learned by building the CNN network. Then these are inputted to RNN network for classification process. The study achieved 91.33% accuracy.

Proposed Methodology

This section describes the AD diagnosis using DBN classification approach. The convolution neural network learns the spatial features from the input images. In the present study both CNN and DBN learns the spatial and longitudinal features for the disease classification. The overall workflow of the study is illustrated in figure 1. This methodology is evaluated on ADNI dataset.

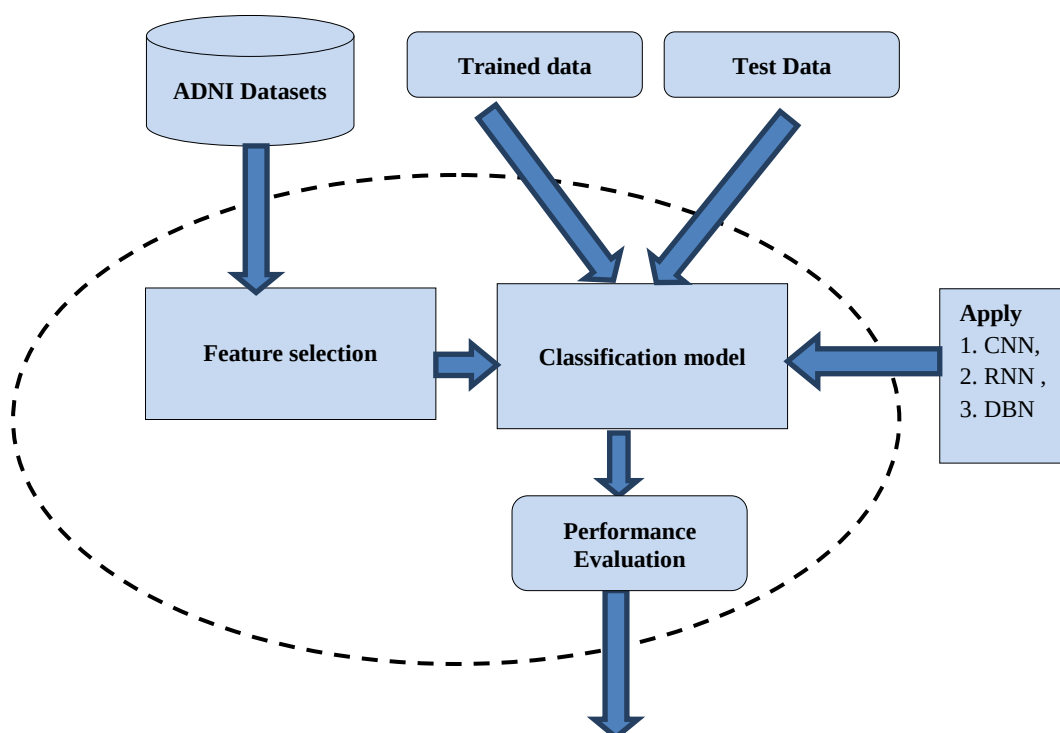


Figure 1. Proposed Methodology

Convolutional Neural Networks

CNNs utilized for feature selection process in AD diagnosis. Unlike neural networks, convolutional neural networks are executed through a combination of convolution, pooling and fully connected layers. The Cnn model is build based on these layers by different numbers of layers. Different architectures employ different combinations of these layers plus activation units and other mechanisms such as normalization and regularization. It takes $w \times h \times d$ (w-width, h-height, d-depth of an image) array of pixels as input over which a $k \times k$ size window, known as filter or kernel is slid across its width and height such that it covers all the area of the input image. While sliding through the image, each pixel value in the input image is multiplied with the values in the filter element by element and summed up to give one pixel value of the output image. Each layer outputs a set of activation maps or feature maps, one for each filter and this is fed as input to the next convolution layer.

Deep Belief Network

The most common deep learning architecture is considered a Deep Belief Network or DBN. DBN is a generative graphical model which contains several hidden units. When training sets along with output class labels are used to train DBN, then probabilistic reconstruction of input is learned by it. After this, the layers can be used as the feature detectors. In the next step, supervision training is provided to DBN by updating the weights to perform classification. In weight updating process the learning rate is used to diminish the function of time, it is relevant to the error rate in each epoch, and also determines how much epoch is required for network training that is based on the values of the current weights. Several Restricted Boltzmann Machines (RBM) could be trained and stacked in a greedy manner to produce DBN architecture. RBM is an unsupervised learning method which contains two layers: input and hidden layers. Each and every layer was constructed with nodes. The input nodes process the user dataset. The hidden nodes extract the multilevel features of the dataset. The weighted parameter W is used to denote the connection between hidden and input layers. Training data can be extracted in a hierarchical representation that is achieved by the graphical models of DBN. Figure 2 represented the Structure of DBN.

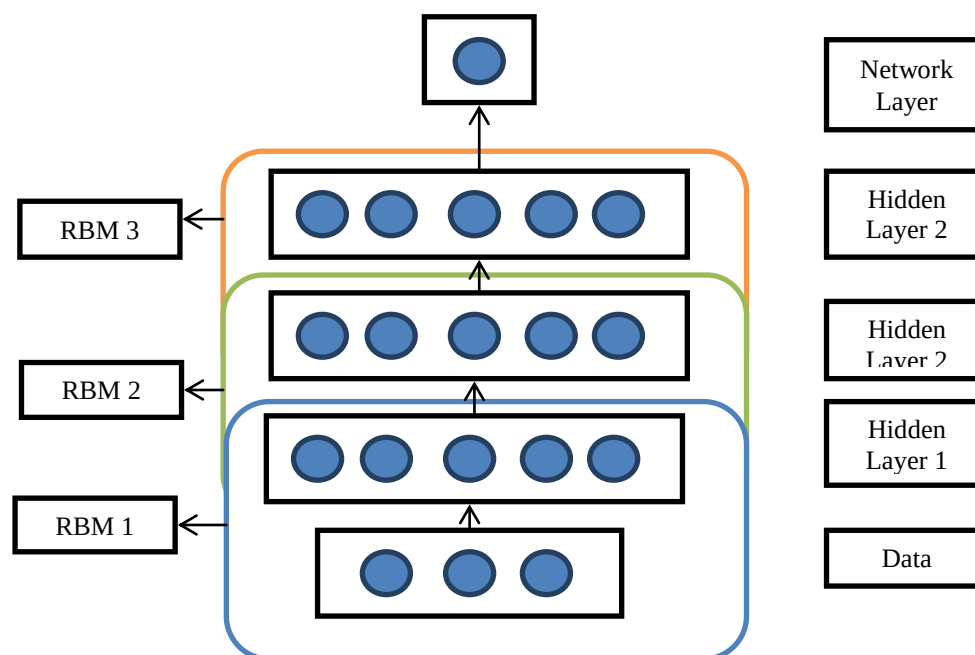


Figure 2. Structure of DBN

The contrastive divergence learning procedure is used to train the RBMs. The hyper parameters values are set through this procedure; these include visible and hidden unit numbers, initial weight values, as well as weight's learning rate, momentum, and cost.

Let there be a binary vector training set which can be designed using RBM that is a two-layer network. A joint configuration (V, H) represents the visible and hidden units' energy as:

$$E(V, H) = - \sum_{i \in \text{visible}} x_i v_i - \sum_{j \in \text{hidden}} y_j h_j - \sum_{i,j} v_j h_i W_{i,j}$$

Where i and j represents the visible and hidden units, $v_j h_i$ represents the i's and j's binary states, their biases are represented by $x_i y_i$ and $W_{i,j}$ represents the weight among them.

Every visible and hidden vector pair is allotted with probability by the network by using energy function as

$$P(V, H) = \frac{1}{S} e^{-E(V, H)}$$

Where S represents the partition function that can be calculated by the sum of every each visible and hidden vector pair as:

$$P(V) = \frac{1}{S} \sum_H e^{-E(V, H)}$$

By lowering the energy of the particular dataset through adjusting the biases and weights so that the training dataset probability can be raised which is then assigned by the network. Training vector's log probability derivative w.r.t. weight is given by:

$$\frac{\partial \log p(v)}{\partial w_i} = \langle V_i h_j \rangle_{\text{data}} - \langle V_i h_j \rangle_{\text{model}}$$

The above equation helps in simple learning rule used to perform the training data's log probability for the stochastic gradient.

$$\Delta_{w_{i,j}} = \epsilon (\langle V_i h_j \rangle_{\text{data}} - \langle V_i h_j \rangle_{\text{model}})$$

Where, ϵ is a learning rate. In an RBM the hidden units are not connected directly, therefore an unbiased sample of $\langle V_i h_j \rangle_{\text{data}}$ can be prepared easily. Assume a training set which is selected randomly, where V represents the binary states, h_j represents the hidden unit j that is having a probability value of 1.

$$P(h_j = 1 | V) = \sigma(y_j + v_i W_{i,j})$$

Where $\sigma(x)$ represents the logistic sigmoid function and therefore, an unbiased sample is represented by

$$\frac{1}{(1 + \exp(-x))}, v_j h_i$$

Similarly, the RBM's visible units are also not connected directly, therefore, for a visible unit state the unbiased sampling can be achieved using a hidden vector as:

$$P(h_j = 1 | V) = \sigma(y_j + v_i W_{i,j})$$

The change in weight is then given by

$$\Delta_{w_{i,j}} = \epsilon (\langle V_i h_j \rangle_{\text{data}} - \langle V_i h_j \rangle_{\text{recon}})$$

By setting the probability of every v_i as 1, a reconstruction to the above equation can be produced. The flow chart of the DBN is given in figure 3.

DBN is trained by a collection of RBM. In the RBM, the weight updating process is carried out by adding the weights, learning rate and by multiplying the positive and negative phase. To fix the number of the parameter while training the network the learning rate is used.

The learning rate

The error may increase if the learning rate is high and due to this the weights may get blasted. The reconstruction error will get reduced automatically if, the learning rate is decreased during the normal learning process.

Updating the hidden states

By utilizing the logistic function, the hidden unit's (j) probability can be calculated as:

$$\sigma(x) = \frac{1}{(1 + \exp(-x))}, \text{ to its total input } P(h_j = 1) = (y_j + \sum_i v_i W_{i,j})$$

Furthermore, whenever the hidden unit probability is more than any random amount which is allocated uniformly among 0-1, then these units are turned on.

Weigh Decay

For correcting the large weights, weight decay is used with the normal gradient. The $L2$ being the simplest penalty function is represented by half of the sum of the square weights called the weight cost. The weight cost coefficient typically ranges between 0.01- 0.00001 for $L2$ weight decay.

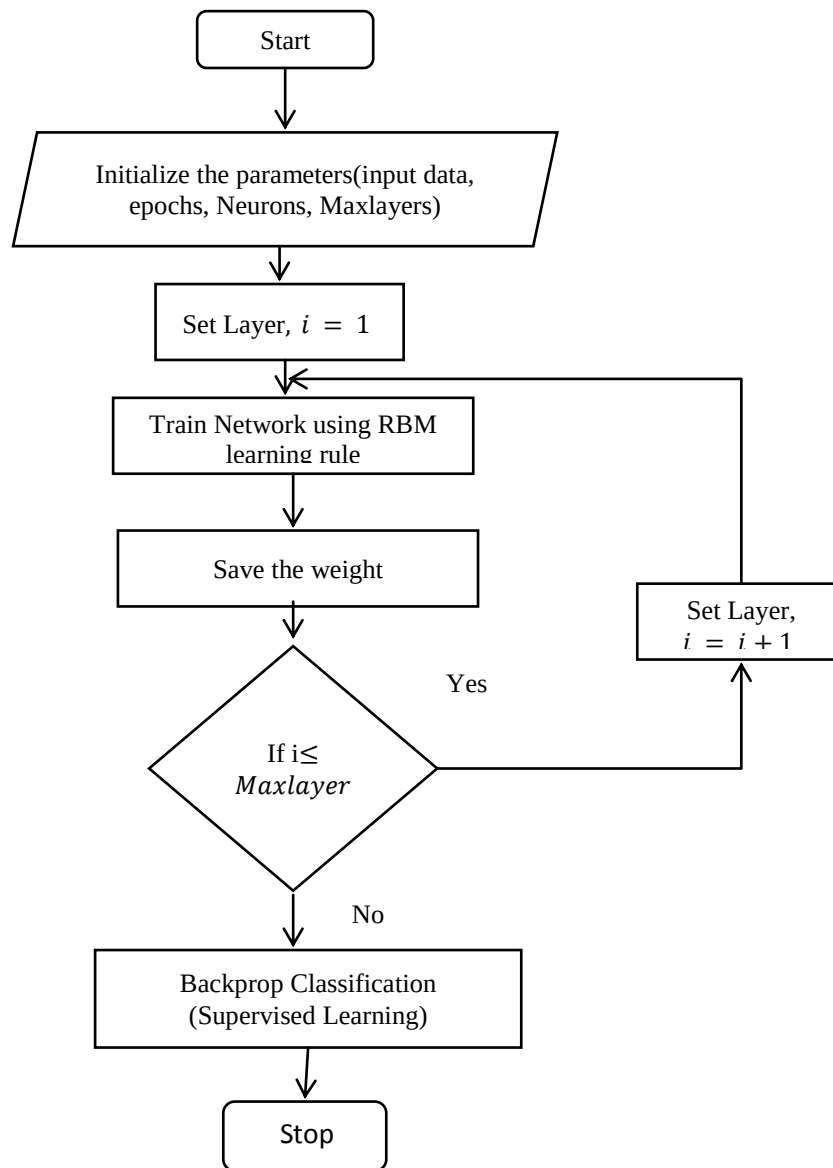


Figure 3. Flowchart of the DBN training process

Experimental Setup

To test and validate the proposed method, totally 279 data are selected and downloaded. Out of 279 data, 133 belong to AD and 146 belong to Normal. To train and test the classifier, out of 133 AD data, 88 data belonging to AD features are randomly assigned for training and the remaining 45 data are assigned for testing. Out of 146 normal data, 97 data belonging to normal are assigned for training and the remaining 49 are assigned for testing. Test and validations are done with five analyses. All experiments are validated with the help of traditional performance metrics given in table 1.

Table1. Events that assign TP, FN, TN and FP

	Actually AD	Actually Normal
Classified as AD	TP	FP
Classified as Normal	FN	TN

The CNN [5], RNN [6] and DBN, classifiers are used. The classification accuracy, average precision and average recall as shown in Tables 2 to 3 and Figures 2 to 5. The performance evaluation values of the validation parameters are given below

Table 2. Classification Accuracy for DBN.

Techniques	Classification accuracy
CNN	90.47
RNN	91.33
DBN	93.7

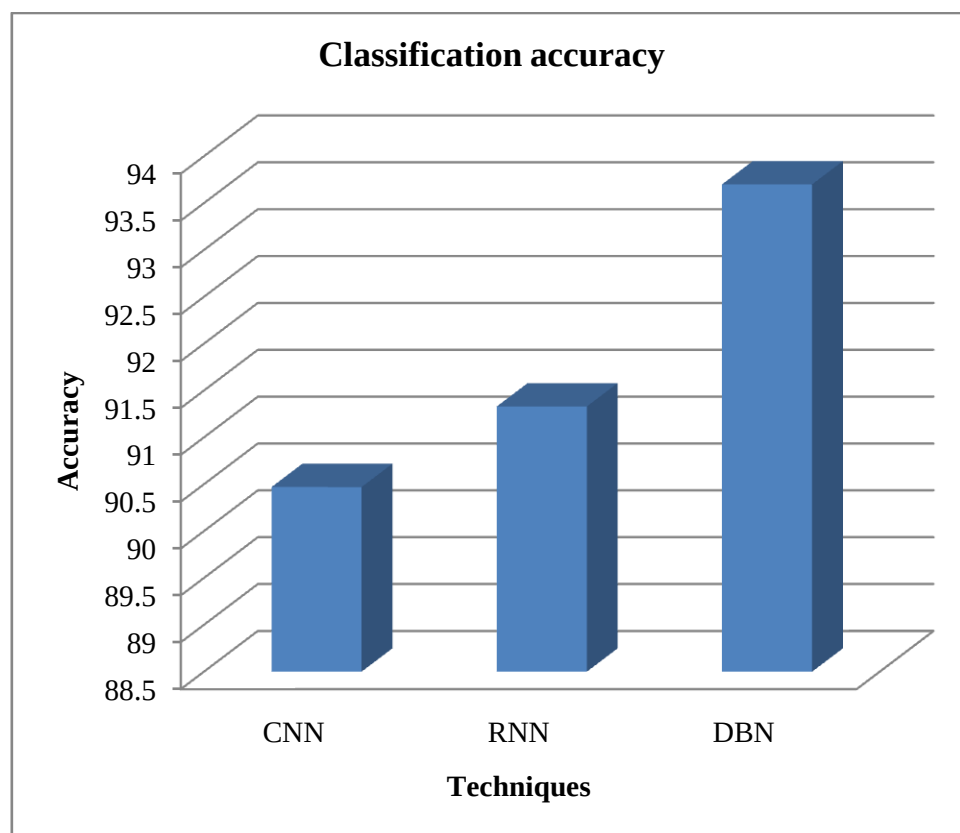


Figure 4. Average accuracy of the classifier

From the Figure 4, it can be observed that the DBN has higher average accuracy of 90.47%, CNN has achieved the next higher accuracy of 91.33% and RNN got 93.7%.

Table 3. Classification Recall value for DBN.

Techniques	Classification Recall
CNN	86.66
RNN	87.34
DBN	90.2

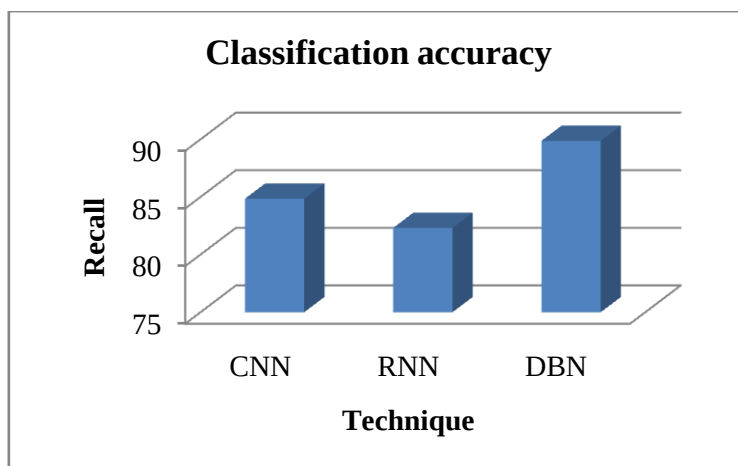


Figure 5. Average Recall of the classifier

From the Figure 5, it can be observed that the DBN has higher average Precision of 84.85%, CNN has achieved the next higher recall of 82.32% and CNN got 89.87%.

Table 4. Classification precision for DBN.

Techniques	Classification accuracy
CNN	84.85
RNN	82.32
DBN	89.87

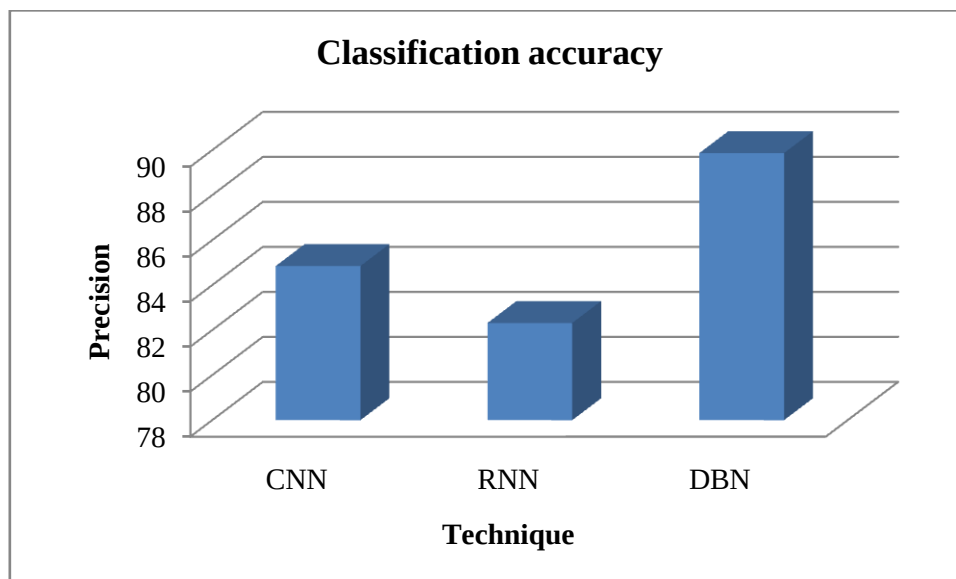


Figure 6 Average Recall of the classifier

Conclusion

The prime objective of AD detection is to diagnose the illness at the early stage very accurately, quickly and economically. To achieve this objective, a proper selection of pre-processing and right segmentation method and correct texture features are necessary. This work is focused on the development of the automatic detection of AD from MRI using the best image processing. DBN is used as a classifier to detect the AD. To improve the efficiency of detection, it is carried out using different methods and the best one is selected and monitoring the parameters such as Precision, Sensitivity, Specificity and Accuracy. The

results are also presented by displaying the confusion matrix and ROC curve. The experimental results show that the texture features extracted using CNN give better classification rate.

The disease detection accuracy can be further increased by adding other features such as volume and shape. To measure the volume and shape, 3- dimensional hippocampus segmentation is needed. The segmentation scheme proposed in this research can be extended for 3-dimensional segmentation without additional constraints.

References

1. Varghese, T, Sheelakumari, R, James, JS & Mathuranath, PS 2013, 'A review of neuroimaging biomarkers of Alzheimer's disease', *Neurology Asia*, vol.18, no. 3, pp. 239-248.
2. Suganthi, SS & Ramakrishnan, S 2014, 'Anisotropic diffusion filter based edge enhancement for segmentation of breast thermogram using level sets', *Biomedical Signal Processing and Control*, vol. 10, pp. 128-136.
3. He, Z, Wang, Y, Yin, F, & Liu, J, 2016, 'Surface defect detection for high-speed rails using an inverse PM diffusion model', *Sensor Review*, vol. 36, no. 1, pp. 86-97.
4. Basheera, Shaik, and M. Satya Sai Ram. "Convolution neural network-based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation." *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 5 (2019): 974-986.
5. Balla-Arabé, S, Gao, X, & Wang, B, 2013, 'GPU accelerated edgeregion based level set evolution constrained by 2D gray-scale histogram', *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2688-2698.
6. Jiang, X, Zhou, Z, Ding, X, Deng, X, Zou, L, & Li, B, 2017, 'Level set based hippocampus segmentation in MR images with improved initialization using region growing', *Computational and mathematical methods in medicine*, vol. 2017, pp. 1-11.