Segmentation of Basal Ganglia Sub-Structures in Brain MR Images Using Denoising Based Stacked Sparse Autoencoder

Akanksha Bindal[1], Anshul Gupta[2], Anubha Gupta[3], and Chetan Arora[4]

[1]Computer Engg. Deptt., DTU, Delhi; [2]Electronics and Communication Engg. Deptt., DTU; Delhi, India
[3]SBILab, Electronics and Communication Engg. Deptt., IIIT-Delhi; [4]Computer Science & Engg. Deptt., IIIT-Delhi, India
akanksha.bindal@gmail.com; anshulkgupta93@gmail.com; anubha@iiitd.ac.in; chetan@iiitd.ac.in

Abstract: In this paper, we have applied a deep learning architecture, namely, denoising based stacked sparse autoencoder to segment the constituent sub-structures of basal ganglia in brain MR images. We implement a 2-stage segmentation wherein intensity and probability maps of 2D/3D neighbourhood of labelled voxels are used as input. Owing to symmetry between left and right halves of subcortical regions, we propose left-and-right combined region segmentation in the first pass and later segment the left and right regions based on their spherical coordinates. Since there will be tissue intensity variability and structural size variability across subjects, it is appropriate to use denoising based architecture. In addition, sparse encoder will in capturing distinct patterns with much ease. Hence, we implement denoising based stacked sparse autoencoder. Results on large public database show good segmentation performance over both normal and diseased dataset.

***Keywords*-**Basal Ganglia, MR segmentation, Sparse Autoencoder

I. INTRODUCTION

Structural Brain Imaging plays a crucial role in neurological psychological research and diagnostic science. MR image segmentation is important for medical diagnosis as it allows quantifying changes in shape, intensity, volume, etc. of anatomical structures for the purpose of disease diagnosis. Often, these structures are found to get altered in brain disorders, including Alzheimer's disease and Parkinson.

Current endeavors in MR image segmentation are dominated by multi-atlas based methods [1, 2]. To segment a query image, such methods first perform registration of query image on atlas. Similar registration transformations are applied to the atlas labels which are then propagated to segment the query image. However, these methods are computationally complex and may not give good results due to lack of pertinent reference atlas, say, in different disorders or across varying population.

Recently, MR Image segmentation has been carried out using Support Vector Machines (SVM) [3]. However, this method requires hand crafting in designing/choosing feature set. In [4], stacked convolutional independent subspace analysis network has been used over MR images captured via 7 Tesla scanner. Signal-to-noise ratio is very high in 7T compared to 1.5T or 3T machines. Hence, segmentation will be much easier over such images. Currently, 1.5T or 3T scanners are being used worldwide as standard machines. Hence, it is not appropriate to compare segmentation results of 7T images with those of 3T or 1.5T scanner images. In [5], sparse stacked autoencoder has been used to segment hippocampus region from healthy infant brains. In [6], sparse stacked autoencoder is used to segment basal ganglia sub-regions via training on healthy human brain data, while the performance is tested on diseased subject dataset. In [7], convolutional neural network is used to segment 134 brain regions of 15 adult human brains. However, the dataset is small that will lead to overfitting of classifier.

In this paper, we consider segmentation of basal ganglia sub-structures of brain, responsible for controlled movement and routine learning. Morphological changes of these regions have been associated with a number of neurological disorders, including Parkinson, Alzheimer, Schizophrenia, Huntington's disease, etc. Image analysis of basal ganglia helps during disorder diagnosis, progression monitoring and treatment. In [8, 9], authors have segmented the Caudate and Putamen regions of Basal Ganglia using artificial neural network (ANN). Problem in finding discriminative and robust features for distinguishing Basal ganglia sub-structures, particularly, the smaller regions like Accumbens and Pallidus is one of the challenges that are persuading researchers to explore deep learning architecture on MR brain segmentation.

We implement sparse denoising-based stacked autoencoder, a variant of an unsupervised learning algorithm, which automatically learns high-level hidden representation from the input data. The contributions of the paper are below:

1. We implement sparse denoising-based autoencoder with multiple layers to capture structural patterns of subcortical brain regions. Denoising-based network helps in better capturing of a) subject data variability and b) variability across healthy and diseased MR subjects' images.
2. We use intensity and probability maps of 2D/3D neighbourhood of labelled voxels as input data. Apriori information in probability map of local neighbourhood significantly improves the segmentation performance.
3. We train our classifier on a mixed collection of healthy and diseased (Bipolar with and without Psychosis, Schizophrenia) in order to build a robust network. In general, only healthy subjects' MRI images are used for training. We use a Public dataset with 103 brain MR volumes of healthy and diseased subjects. Here, denoising based network architecture will play an important role in detecting latent relationships between diseased and healthy sets.
4. We implement 2-stage segmentation that first performs left-and-right combined segmentation of underlying structures and later segments individual structures in the second pass.

This paper is organized as follows. In section 2, we briefly review the existing segmentation methodologies and related work in deep learning. In section 3, we describe our proposed architecture and algorithm. Simulation results are explained in section 4. We conclude with a discussion in section 5.

II. PRELIMENARIES

In this section, we briefly present the theory of autoencoder, denoising based autoencoder, and sparse denoising based stacked autoencoder.

A. Autoencoder

A basic autoencoder is a minimal layer neural network with one hidden layer and one output layer. First, it takes an input vector $\mathbf{x}$ and maps it to a hidden representation $\mathbf{y}$ using encoding defined as $\mathbf{y} = g_\theta(\mathbf{x}) = S(\mathbf{W}\mathbf{x} + \mathbf{b})$, parameterised by $\theta = \{\mathbf{W}, \mathbf{b}\}$ where $\mathbf{W}$ is a matrix of hidden weight, $\mathbf{b}$ is a bias vector and S is a non-linear mapping such as a sigmoid function. The latent representation $\mathbf{y}$ is then mapped back into the reconstructed output vector $\mathbf{z}$ by using the decoding function $h_\theta(\mathbf{x}) = S(\mathbf{W}'\mathbf{y} + \mathbf{b}')$ with $\theta' = \{\mathbf{W}', \mathbf{b}'\}$. The weights of the network are constrained to minimise the average mean-squared or cross-entropy reconstruction error between input $\mathbf{x}$ and output $\mathbf{z}$.

B. Denoising Autoencoder

Denoising autoencoder performs unsupervised learning based on the idea of constructing learned features that are robust to partial corruption of the input pattern. This approach can be used to train an autoencoder and once stacked with multiple layers form deep architectures [1].

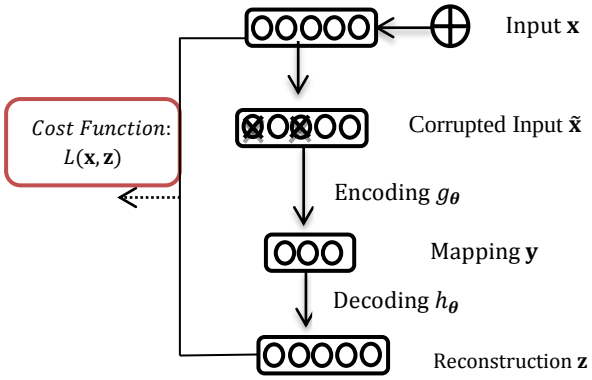

Fig1. Representation of a denoising autoencoder

C. Sparse Autoencoder and denoising based Sparse Autoencoder

For correlated input representations, a sparse autoencoder discovers interesting features by imposing sparsity constraint on the number of hidden layer nodes that will be activated for a given input pattern. This constraint allows only a small number of weights to be activated for every input vector. l^1 regularization penalizes large parameter values, by adding them to the updated cost function. This constraint optimization can be solved using LASSO (Least Absolute Shrinkage and Selection Operator) regularisation and fused within the autoencoder framework. The denoising based sparse autoencoder (SPADE) incorporates the denoising framework and employs sparsity in the activated nodes in the encoding layer or the hidden layer of the autoencoder.

III. DATA DESCRIPTION AND SET-UP

We utilized public dataset available at [10] and were acquired by [11] at the McLean Hospital Brain Imaging Center on a 1.5 Tesla General Electric (GE) scanner. The manually segmented regions are stored in corresponding label images [11]. It comprises of 103 T1-weighted MRI volumes from four diagnostic groups: healthy control, Bipolar Disorder without Psychosis, Bipolar Disorder with Psychosis and Schizophrenia. The subjects are children and adolescents with ages ranging from 6 to 17 years, both male and female. Data registered has been carried out on Talairach coordinate space [11] along with bias field correction. Region segmentation was performed with a semi-automated intensity contour algorithm [11]. More information on acquisition details can be found in [11].

MRI volumes need to be registered to make them as similar as possible to a common template prior to any further processing. The data of [10] is provided after motion realignment, slice time correction, and normalization. In our pipeline, we first converted MRI images of NIFTI (Neuroimaging Informatics Technology Initiative) data format into a 3D Matrix. All MR volumes with a resolution different than 256x256x128 voxels were discarded, leaving only 81 valid MRIs. These MR volumes were then divided into a training set of 20 healthy control + 10 of each brain disorder (total 30 volumes). From this training set of 50 MRI volumes, 2.5 percent of samples are extracted from each class (No ROI and 4 ROIs) of around 46,000 voxels. A validation data of around 30,000 voxels was also extracted from these 50 MRI volumes. Rest of the 31 brain volumes were used for testing purpose.

In [6], the author trains his network on healthy sets treating the diagnostic MRIs as outliers. We include the diagnostic MRI feature sets in our network architecture to build a robust network capable of learning representations of both healthy and diseased MRIs.

We have utilized Theano framework in Python for implementing denoising-based sparse stacked autoencoder [12]. We extracted features in MATLAB. BSMVIEW toolbox [13] in SPM12 [14] is used to visualize ROIs in consideration. All programs were executed on a machine equipped with GeForce GTX 980 GPU accelerator and 16 GB RAM memory.

IV. PROPOSED WORK

Image analysis of basal ganglia is essential to understand many neurological disorders, including Parkinson, Alzheimer, Schizophrenia, Huntington's disease, etc., monitor their progression, and evaluate possible treatments. In this paper, we focus on eight regions of interest (ROIs) of Basal Ganglia – the left and right Caudate, left and right Globus Pallidus, left and right Putamen forming the Striatum also known as the sensory-motor functional block, and the left and right Nucleus Accumbens belonging to the ventral system. The Nucleus Accumbens is positioned at the intersection of the head of the Caudate and the anterior portion of the Putamen, with no visible anatomical separation from these structures in the MR image, though having different functionality. Thus, it is a challenging structure for segmentation purpose and is included in the region of interests of our study. Please refer to Fig. 2 that show these regions.

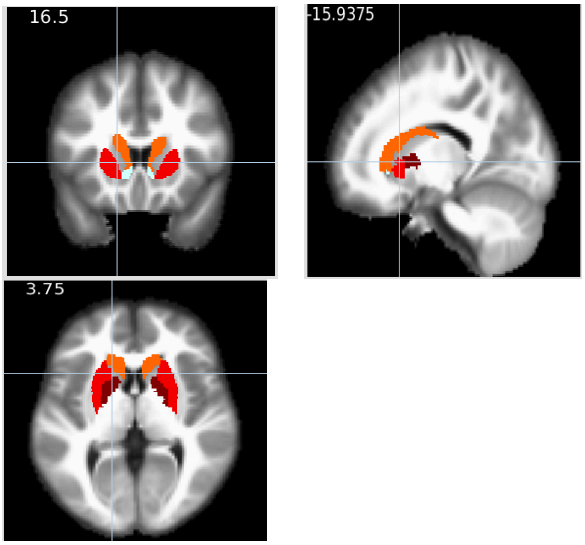

Fig2. A posterior lateralized view of the basal ganglia. Coordinates (-15.9375, 16.5, 3.75) correspond to a voxel in the left Caudate Nucleus. Left and right regions are coloured as: Caudate in bright red, Accumbens in white, Putamen in Orange and Pallidus in dark red.

We implement a machine learning architecture, a denoising-based stacked sparse autoencoder, whereby a given training set, comprising of labelled ROIs in MR images, is used to train the network for classifying each voxel into its corresponding anatomical structure. Knowledge of the ROIs is extracted from manually labelled 3D Brain MRIs.

First data is processed according to methodology described in subsection 4(A) below. Proposed features used to train the network are discussed in subsection 4(B). The details on denoising based stacked sparse autoencoder as used in this paper are presented in subsection 4(C).

A. Data Pre-processing

We propose a left-and-right combined segmentation technique for segmentation of basal ganglia structures. The idea behind performing left and right combined segmentation of anatomical basal ganglia structures is to recognise the clear spatial separation between the left and right symmetrical regions of the basal ganglia structure as depicted in Figure 3 and to learn abstractions from the combined set of voxels belonging to the same anatomical structure. First, we mask all voxels outside ROIs to extract training data (*Pruning Step*) and train our network using features proposed in the next section.

Later, we implement a code that post-processes the predicted voxels. We calculate the centroid of each of the combined classes by computing the mean of all points in the subspace belonging to that particular class and assign the concerned voxel to left or right region based on its spatial position, thus, leading to 8 ROIs (left and right Caudate, left and right Putamen, left and right Pallidus, left and right Accumbens) segmentation.

B. Feature Extraction

In this paper we have utilized three types of features at the input layer of the autoencoder- Probabilty map, neighbourhood patch intensity, and position coordinates.

(i) Probability Map:

We compute the 3x3x3 probability map of neighbourhood of each of the voxel. This requires computation of probability value at each voxel. This is determined for voxel belonging to each of the four ROIs and is calculated by computing the frequency, i.e., the number of times each voxel (position) belongs to a particular ROI out of total MRI volumes used in the training set. For example, if a voxel (position) belongs to ROI-1 in 30 MRI training volumes out of 50 MRI volumes used for training, then the probability for this voxel being in ROI-1 is 3/5.

Two probability map regions are considered: 1. probability of every voxel belonging to each of the four ROIs is given by *map_roi* which is a 4-diemnsional vector; 2. Probability of 3x3x3 neighbourhood of voxel of interest (i.e., 26 3D neighbours), labelled as *map*. This *map* stores binary values corresponding to each neighbour as belonging to any ROI or belonging to no ROI. For example, let us say only 2 of the 26 3D neighbours of voxel v belong to ROIs and rest 24 do not belong to any of the 4 ROIs, then this map will have two 1's and 24 0's.

The pruning step mentioned in pre-processing section above makes use of this probability map and filters out only those voxels that have non-zero probability of belonging to any ROI.

(ii) Intensity Patch Size

Intensity patches of 2D and 3D are considered centred on the voxel. We show a comparison between 4 ROI and 8 ROI for the first part, where patch size of 5x5 and 11x11 is considered in each of the three orthogonal planes to give stacked 2D feature set. Feature set comprises of 2D intensity patch size. We find that 4 ROI shows comparable result with the mean dice coefficient in the 8 ROI case for Caudate, Putamen and Pallidus regions, with slight improvement in the case of the smaller Nucleus Accumbens.

(iii) Position Coordinates

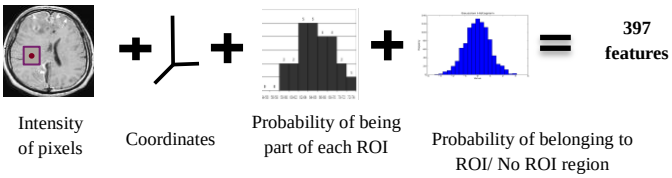

Fig3. Visual explanation of feature sets for 11x11 stacked 2D patch for 4 ROI classifications. 11x11 patches are considered in each of the orthogonal planes and appended to the input feature vector.

V. SIMULATION RESULTS

We implement a denoising based stacked sparse autoencoder with 5 layers comprising of 1000 nodes in each layer except for the middle layer that included 1500 nodes. The network fine-tuning learning rate was empirically found to be 10^{-5}. The pre-training learning parameter is fixed at 10^{-8}. Additive Gaussian noise is added with corruption rates, progressively varying from 0.1 to 0.4 in the layers. Sparsity constraint is weighed down on the network with l^1 parameter = 10^{-5}. Concave penalisation of l^1 term is used to enforce better sparsity in the network.

We use the dice coefficient metric for assessment of the accuracy of our results as used in [22] and [Alberto]. Dice Coefficient is calculated as:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

where TP is the number of voxels which are correctly predicted, FP is the number of voxels wrongly classified as part of the ROI and FN is the number of voxels wrongly classified outside the ROI. It can be shown that the Dice Coefficient corresponds exactly to the Statistical measure commonly known in Machine Learning as F_1 Score, defined in terms of Precision and Recall,

$$F_1 = \frac{2Precision.Recall}{Precision + Recall} \quad (5)$$

where precision is defined as $Precision = {TP}/{TP + FP}$ and recall is defined as $Recall = {TP}/{TP + FN}$. All these measurements range from 0 to 1 with 1 being the highest.

We first perform an experiment segmenting the MRI into 8 classes (left and right Caudate, left and right Putamen, left and right Pallidus, left and right Accumbens) using 5x5 and 11x11 stacked 2D patches in each orthogonal plane for both 4 ROI (using combined left-and-right 4 classes input feature set) and simple 8 ROI input feature set. We find that combined left-and-right approach shows comparable results with an improvement in the Nucleus Accumbens region which is smallest region amongst all ROIs selected.

ROI	Caudate		Putamen		Pallidus		Accumbens	
	Left	Right	Left	Right	Left	Right	Left	Right
4 ROI (5x5)	0.87	0.86	0.78	0.77	0.656	0.66	0.625	0.62
8 ROI (5x5)	0.838	0.84	0.79	0.77	0.66	0.646	0.58	0.565
4ROI (11x11)	0.9	0.88	0.86	0.845	0.79	0.80	0.74	0.722
8ROI (11x11)	0.91	0.87	0.87	0.85	0.81	0.79	0.725	0.71

Table1: Dice Coefficient based comparison on 8 ROIs vs 4 ROIs (left and right regions combined) using triplanar patches in our proposed network

Having established the improvement achieved by our approach, we then demonstrate an experiment to determine the best intensity patch size for our problem statement. We carry out experiments for patch sizes 5x5x5, 11x11x3, 11x11x5, 11x11x11 and 17x17x3. In Table 2, we display the results obtained for 4 ROI classifications in Dice Coefficient metric. The best results are obtained for stacked 11x11 stacked 2D Patches and 3D patch 11x11x3. We show the results in Table 3 for Schizophrenic diseased patients in the test set comprising of Schizophrenic, Bipolar with Psychosis and Bipolar without Psychosis disease.

4 ROI	No ROI	Caudate	Putamen	Pallidus	Accumbens
[1]	NA	0.88±0.02	0.89±0.02	0.81±0.04	0.75±0.07
[6]	0.91±0.008	0.9±0.015	0.9±0.015	0.83±0.02	0.74±0.04
5x5x5	0.807±0.02	0.84±0.01	0.82±0.03	0.69±0.05	0.635±0.04
11x11x3	0.89±0.01	0.88±0.02	0.87±0.01	0.82±0.045	0.76±0.035
11x11x5	0.88±0.01	0.87±0.01	0.89±0.01	0.79±0.05	0.72±0.02
11x11x11	0.9±0.01	0.86±0.02	0.825±0.02	0.70±0.03	0.67±0.04
17x17x3	0.87±0.02	0.85±0.03	0.84±0.02	0.72±0.06	0.74±0.01

Table 2: Dice Coefficient on healthy test data using 3D patches on 4 ROI based proposed network

	No ROI	Caudate	Putamen	Pallidus	Accumbens
Tri-planar 11x11	0.87±0.01	0.89±0.01	0.85±0.03	0.79±0.03	0.69±0.035
3D 11x11x3	0.88±0.01	0.875±0.015	0.88±0.01	0.8±0.03	0.71 ± 0.05

Table 3: Dice Coefficient on Schizophrenic test data using best patch size on 4 ROI based proposed network

From the above table we find that patch sizes of stacked 2D 11x11 features and 3D 11x11x3 features show comparable accuracy for larger regions Caudate and Putamen. However for smaller regions, 11x11x3 seemingly represents the volume of a subcortical region better than do 2D patch sets.

We considered patch sizes of 11x11x3, 11x11x5 and 11x11x11 to test our hypothesis of the third dimension, along the time axis being thinner in comparison to the other dimensions. We find that 11x11x3 gives the best results, which confirms the hypothesis that the region of interest volume is thinner along the time axis. Finally we consider patch size of 17x17x3 and find similar results to patch size of 11x11x3 with some decline in the results for Nucleus Accumbens and Globus Pallidus.

Discussion: In [6], the author proposes a stacked sparse autoencoder network, implemented on the MATLAB Deep Learning Toolbox with 2 layers for the purpose of segmenting 8 regions of the Basal Ganglia. We build upon the author's method by showing that implementing an autoencoder with deep nature (more number of layers) helps in learning better abstractions of the input data. One key motivation for this is the use of an unsupervised training criterion to perform a layer-by-layer initialization: each layer is at first trained to produce a higher level (hidden) representation of the observed patterns, based on the representation it receives as input from the layer below, by optimizing a local unsupervised criterion as described in [15].

We find that implementing an autoencoder with deep nature (more number of layers) helps in learning better abstractions of the input data. One key motivation for this is the use of an unsupervised training criterion to perform a layer-by-layer initialization: each layer is at first trained to produce a higher level (hidden) representation of the observed patterns, based on the representation it receives as input from the layer below, by optimizing a local unsupervised criterion as described in [15].

VI. CONCLUSION

In this paper, we have implemented a deep learning technique to segment the basal ganglia region from MR images that possess significant inter-subject variation in shape and size of anatomical structures. We see that a denoising based stacked sparse autoencoder network effectively manages to represent abstractions present in the input feature set. This may be because the neural activity in the brain is seemingly of sparse nature. An interesting observation is that initial combination of left and right regions of the said anatomical structure manages to increase accuracy for smaller region, Accumbens, in the final

segmentation. Due to explicit spatial separation between the left and right regions this seems like a probable conclusion. Also, some anatomical structures may possess vital information in their symmetric parts which helps in their segmentation. We find that Pallidus boundary is increasingly being misclassified in bigger patch sizes and thus accuracy falls in Column 4 of Table 2. In future we would like to improve the existing method of computing the mean over all predicted voxels by appending another hidden layer to the existing network to carry out this computation in a single pass.

REFERENCES

1. K. O. Babalola et al., "An evaluation of four automatic methods of segmenting the subcortical structures in the brain," *NeuroImage*, vol. 47, pp. 1435–1447, 2009.
2. A. Klein, B. Mensh, S. Ghosh, J. Tourville, and J. Hirsch, "Mindboggle: automated brain labeling with multiple atlases,"*BMC medical imaging*, vol. 5, no 1, pp. 1-14, 2005.
3. J. H. Morra et al., "Comparison of Ada Boost and Support Vector Machines for Detecting Alzheimer's Disease through Automated Hippocampal Segmentation," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 30 - 43, 2010.
4. M. Kim, G. Wu, and D. Shen, "Unsupervised Deep Learning for Hippocampus Segmentation in 7.0 Tesla MR Images," *MICCAI* (LNCS 8184), pp. 1 - 8, 2013.
5. Y. Guo1 et al., "Segmenting Hippocampus from Infant Brains by Sparse Patch Matching with Deep-Learned Features." *MICCAI*, vol. 17 no. 2, pp. 308–315, 2014.
6. A. M. González, L. I. Muñoz. "Segmentation of Brain MRI Structures with Deep Machine Learning." *Master in Artificial Intelligence (UPC-URV-UB)*, pp. 1-56, 2012.
7. A. d. Brebisson, G. Montana. "Deep Neural Networks for Anatomical Brain Segmentation." *CVPR*, pp 1-9, 2015.
8. V. A. Magnotta et al, "Measurement of brain structures with artificial neural networks: Two- and three-dimensional applications1.*Radiology*, vol. 211, no. 3, pp. 781-790, 1999.
9. S. Powell. "Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures." *Neuroimage*, vol. 39 no. 1, pp. 238-247, 2008.
10. D. Kennedy and C. Haselgrove. "CANDIShare: A Resource for Pediatric Neuroimaging Data", *Neuroinformatics*, vol. 10 no.3, pp. 319–322, 2012.
11. J. A. Frazier et al, "Diagnostic and sex effects on limbic volumes in early-onset bipolar disorder and schizophrenia." *Schizophrenia Bulletin*, vol. 34, no. 1, pp. 37–46, 2008.
12. J. Bergstra et al, "Theano: a CPU and GPU Math Expression Compiler". In *Proceedings of the Python for Scientific Computing Conference(SciPy)*, Texas, USA, June 2010.
13. Bpsmview: https://github.com/spunt/bspmview.
14. J. Ashburner et. al, SPM12 manual, May 2015.
15. Pascal Vincent et al, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion", *Journal of Machine Learning Research*, vol. 11, pp. 3371\-3408, 2010.

Differences in Approach

	Our Approach	**2012 approach**
1.	Denoising based sparsed stack autoencoder	Applied 4 networks: 1. Softmax 2. Neural Network with 1 & 2 Hidden layers 3. Sparse Stacked Autoencoder with 2 hidden layers 4. SVM
2.	Hidden layers – 5 (500,500,1000,500,500)	Hidden layers – 1 & 2 (100,150)
3.	Considered 5 classes(No ROI, Caudate, Putamen, Pallidus, Accumbens) in 4 ROI classification. Left and right parts of same class are considered as one.	9 class classification(No ROI, left and right Caudate, left and right Putamen, left and right Pallidus, left and right Accumbens)
4.	Intensity patch: stacked 5x5 and stacked 11x11 Tri-planar 11x11x3, 11x11x5, 17x17x3 3D 11x11x11 Coordinates: X,Y,Z Probability Map region: Probability of belonging to each Roi (No of ROI) + extended map of 3x3x3 demonstrating if voxel belongs to ROI or No ROI.	2D & 3D features 2D patch size – 2x2, 3x3, 5x5,7x7,11x11 3D patch size – 3x3x3,5x5x5, 7x7x7 + 8 (8 Roi information from probability map) Uses same concept of probability map but probability map values taken only for centre voxel in consideration
5.	Training sample space: 2.5 % data set sampled randomly from 50 MRIs. Approx 36,000 voxels	5 % data set. Approx 35,000 voxels.
6.	Test sample space: 32272 voxels from Schizophrenic Disorder MRI	Same