Deep Neural Networks for Segmentation of Basal Ganglia Sub-Structures in Brain MR Images

Akshay Sethi Akshat Sinha Ayush Agarwal Chetan Arora Anubha Gupta
Indraprastha Institute of Information Technology-Delhi (IIIT-D), India

ABSTRACT

Automated segmentation of brain structure in magnetic resonance imaging (MRI) scans is an important first step in diagnosis of many neurological diseases. In this paper, we focus on segmentation of the constituent sub-structures of basal ganglia (BG) region of the brain that are responsible for controlling movement and routine learning. Low contrast voxels and undefined boundaries across sub-regions of BG pose a challenge for automated segmentation. We pose the segmentation as a voxel classification problem and propose a Deep Neural Network (DNN) based classifier for BG segmentation. The DNN is able to learn distinct regional features for voxel-wise classification of BG area into four sub-regions, namely, Caudate, Putamen, Pallidum, and Accumbens. We use a public dataset with a collection of 83 T-1 weighted uniform dimension structural MRI scans of healthy and diseased (Bipolar with and without Psychosis, Schizophrenia) subjects. In order to build a robust classifier, the proposed classifier has been trained on a mixed collection of healthy and diseased MRs. We report an accuracy of above 94% (as calculated using the dice coefficient) for all the four classes of healthy and diseased dataset.

CCS Concepts

•**Computing methodologies** → **Image segmentation;** *Neural networks;* •**Applied computing** → *Imaging;*

Keywords

Basal Ganglia, MR segmentation, Deep Neural Network

1. INTRODUCTION

Structural brain imaging plays a crucial role in the field of neurological research for diagnosing various brain disorders. MRI segmentation is important to quantify the changes in shape, intensity, volume, etc in the anatomical structures of brain owing to any disorder such as Alzheimer or Dementia.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICVGIP, December 18-22, 2016, Guwahati, India

© 2016 ACM. ISBN 978-1-4503-4753-2/16/12…$15.00

DOI: http://dx.doi.org/10.1145/3009977.3010048

In this paper we focus on the problem of automated segmentation of Basal-Ganglia (BG) region of the brain in an MRI scan, into four sub-regions, namely, Caudate, Putamen, Pallidum, and Accumbens. Figure 1 shows a sample MRI scan with the 4 regions marked in different colors. The problem is important, since BG region is responsible for controlled movement and routine learning in humans. The region also plays a role in eye movement and motivation. Structural change in the anatomy of basal-ganglia region has been related to a number of brain disorders, including Parkinson, Alzheimer, Schizophrenia, Huntington's disease, etc. Analysis of basal ganglia MRI scans has been used in diagnosing, progression monitoring, and treatment of these disorders. Segmenting BG region from brain MRI scans is an important first step for these analysis.

Current endeavors in MRI segmentation are dominated by multi-atlas based methods [1, 9]. In order to segment a query image, these methods first perform registration of query image on atlas. Similar registration transformations are applied to the atlas labels that are then propagated to segment the query image. However, the methods are computationally complex and may not give good results due to lack of pertinent reference atlas for different disorders or across varying population.

Recently, MRI segmentation has been carried out using Support Vector Machines (SVM) [12]. However, the technique requires hand crafting in designing/choosing feature set restricting the generalizability and accuracy in varying datasets and conditions. In [8], 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 is much easier in such images. Considering the relative popularity of 1.5T or 3T scanners over 7T scanners, the technique have limited applicability. In [6], sparse stacked autoencoder has been used to segment hippocampus region from healthy infant brains. In [11], 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. Although the method gives good results on tested dataset, consistency in results across all regions and on diseased scans is lacking. In [10, 14], authors have segmented the Caudate and Putamen regions of Basal Ganglia using an artificial neural network. In [4], convolutional neural network has been used to segment 134 brain regions of 15 adult human brains. However, the tested dataset is small and the achieved accuracy is only 72%. In contrast we report an accuracy of more than 94%

in this paper using the proposed technique.

We pose the problem of MRI segmentation as classification of a voxel in BG as a 5 class classification problem. The 4 classes correspond Caudate, Putamen, Pallidum, and Accumbens regions in BG while the 5th corresponds to background (or none of the 4). The low contrast of voxels in the region and owing to undefined boundaries presents problem in finding discriminating and robust features for distinguishing Basal ganglia sub-structures. In particular, voxels in smaller regions such as Accumbens and Caudate are difficult to classify via automated methods and hence, motivates us to explore this problem.

In this paper we propose a Deep Neural Network model for the classification of BG voxels in one of the five constituent classes. The specific novelties and contributions of this paper are:

1. We propose a novel deep neural network model for the classification model. Given the limited of training data available, the network is pretrained using stacked sparse autonecoder in an unsupervised fashion and then fine tuned using available ground truth labels.

2. We use intensity and probability maps of 2D neighborhood as input to our deep network model. We show that a-priori information in probability map of local neighborhood significantly improves the segmentation performance.

3. We propose to perform a combined left and right segmentation of anatomical basal ganglia structures first. This increases number of training samples as well as reduces number of classes. Classification into left and right region is trivial owing to clear spatial separation between the regions.

4. We use a large public dataset comprising of 103 brain MR volumes of healthy and diseased subjects. In order to build a robust classifier, we train our classifier on an assorted collection of healthy and diseased (Bipolar with and without Psychosis, Schizophrenia). This is in contrast to [11], where only healthy subjects' MRI images are used for training.

This paper is organized as follows. In Section 2, we briefly present the related theory. In Section 4, we describe the data set and its preprocessing. In Section 3, we present the proposed deep neural network model. We present the details of the experiments and results in Section 5 and give concluding remarks in Section 6.

2. PRELIMINARIES

We propose a deep neural network based classification model in this paper. As has been commonly prescribed by various researchers, we use autoencoder by pre-training for initializing the weights of the proposed network. In this section, we briefly review the autoencoder and its variant, sparse stacked autoencoder, used for the pretraining.

2.1 Autoencoder

Autoencoders are a technique used for the unsupervised learning of representations of the underlying data [3, 2]. Autoencoders are a special kind of artificial neural network in which the input itself is presented as the target output as well. The objective of training is to learn an approximation to the identity function in the hidden layers of autoencoder which can reconstruct the input reliably. The basic autoencoder model contains one hidden and one output layer. Conceptually an autoencoder learns an encoder and decoder function described as follows:

Encoder: The mapping g_θ that transforms an input vector x into hidden representation y is called the encoder. It takes an input x and maps it to a hidden representation y via Equation (1), parameterised by $\theta = \{W, b\}$ where W is a matrix of hidden weights, b is a bias vector and s is a non-linearity such as the *sigmoid*, *tanh* or $ReLU$.

$$y = g_\theta(x) = s(Wx + b) \tag{1}$$

Decoder: Decoder maps resulting hidden representation y from encoder to a reconstructed vector z in input space such that, $z = h_\theta(y)$. The mapping h_θ is as given by Equation (2) with $\theta' = \{W', b'\}$. The weight matrix W can be optionally constrained by $W' = W^T$, in which case the autoencoder is said to have 'tied weights'.

$$z = h_\theta(y) = s(W'y + b') \tag{2}$$

The parameters of the autoencoder model are constrained to minimize the average reconstruction error:

$$(\theta^*, \theta'^*) = \arg\min_{\theta,\theta'} \frac{1}{n} \sum_{j=1}^{n} L(x^{(i)}, z^{(i)}),$$

where L is a user specified loss function. Common choices for the loss function include:

- The traditional mean squared error $L(x, z) = ||x - z||^2$ for real valued x,
- The cross entropy error $L(x, z) = -\sum_{k=1} x_k \log z_k + (1 - x_k) log(1 - z_k)$ for binary input.

The optimization is typically carried out by a variant of the gradient descent algorithm known as scaled conjugate gradient descent. Unsupervised *pre-training* is performed with a goal of bringing the network's hidden weights and biases to a region of the parameter space that constitutes a better starting point than random initialisation of weights. The pre-training is followed by a supervised training stage, also known as *fine-tuning*.

2.2 Sparse Autoencoders

Standard autoencoders are effective at learning a representation that is smaller in dimensions than the input but when there are more hidden units than input units (an over complete representation), a standard autoencoder can copy the inputs in hidden layers instead of learning meaningful features. To avoid this problem sparse autoencoders have been developed which put an extra cost on the hidden layer to create sparse representation or restrict the number of allowed non-zero activations [13].

2.3 Regularization

Typical deep network models consisting of large number of parameters often give good prediction accuracy for the training set well but fail to generalize to the new data. A popular solution to regularize the network weights is to use

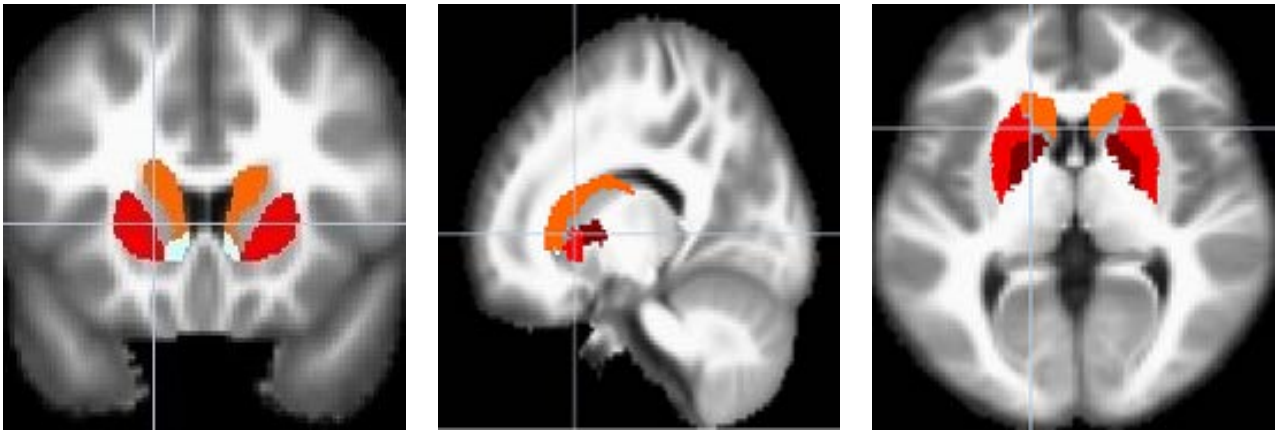

Figure 1: A posterior lateralized view of the basal ganglia (best seen in color). Left and right regions are coloured as: Caudate in bright red, Accumbens in white, Putamen in Orange and Pallidus in dark red. From L-R: Coronal view, Sagittal view, Axial view

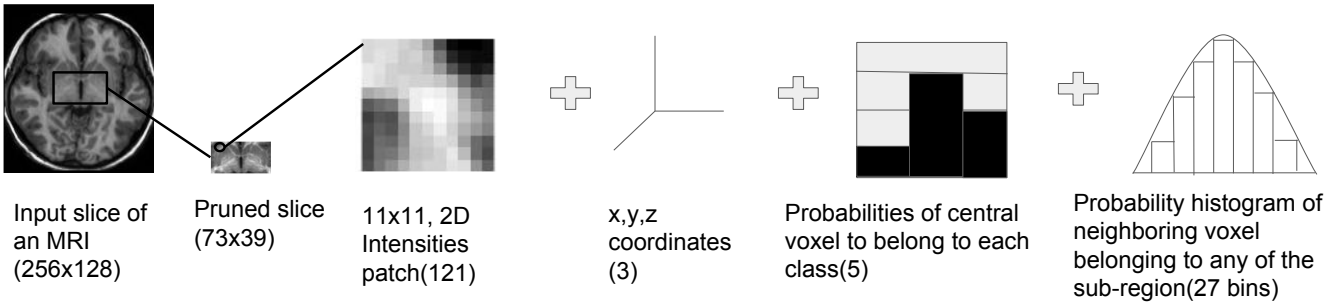

Figure 2: We construct a 156 dimensional feature vectors for each voxel to be fed to our deep neural network model. The feature uses cues from raw pixel intensities as well as probability patterns of 11×11 neighborhood patch around the voxel of interest.

L_1 and L_2 regularization to the cost function thus penalizing large weight values. Following the notation of [13], let us denote by $\theta^{(l)}$, the matrix of weights from layer l to layer $l+1$. If the network has s_l units in layer l and s_{l+1} units in layer $l+1$, then $\theta^{(l)}$ is of dimension $s_{l+1} \times (s_l + 1)$ (also considering the extra bias unit). L_1 regularization updates the cost function as:

$$J_{reg}(\theta) = J(\theta) + \lambda \sum_{l=1}^{L-1} \sum_{i=1}^{sl+1} \sum_{j=1}^{sl} |\theta_{ij}^{(l)}|.$$

L_2 regularization, also known as weight decay regression, is more popular and adds the squared weights to the cost function:

$$J_reg(\theta) = J(\theta) + \lambda \sum_{l=1}^{L-1} \sum_{i=1}^{sl+1} \sum_{j=1}^{sl} |\theta_{ij}^{(l)}|^2.$$

The magnitude of the regularization can be manipulated by the regularization parameter λ, whose value is dependent on the model selection. Both L_1 and L_2 regularization have been extensively compared in [13]. L_1 regularization tends to produce a sparse model, which often sets many parameters to zero, effectively declaring the corresponding attributes to be irrelevant. L_2 regularization gives no preference to zero weights.

2.3.1 KL divergence

Another method of enforcing sparsity is using Kullback-Leibler(KL) divergence. KL divergence is a distance to measure how different two distributions are. The KL divergence between two random Bernoulli distributios with mean p and q,is given by:

$$KL(p||q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}.$$

KL divergence is 0 when $p = q$ and positive otherwise.

Let $a_j^{(2)}(x)$ denote the activation of the hidden unit j in the network when it is given a specific input x. The average activation for a unit j for the whole training set is then given by $\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} a_j^{(2)}(x^{(i)})$. Enforcing a sparsity constraint, ρ, is equivalent to enforcing the average activation of a unit, $\hat{\rho}_j$, to ρ. The same can be done by adding a KL divergence based penalty in the cost:

$$J_{reg}(W,b) = J(W,b) + \beta \sum_{j=1}^{s_2} KL(\rho||\hat{\rho}_j).$$

Here s_2 is the number of neurons in the hidden layer. The regularization term adds a positive value to the term for $\hat{\rho}_j \neq \rho$ and converges to the global minimum for suitable β. Thus, it supports a sparse model. This process is statistically better than L_1 regularization but it requires complex

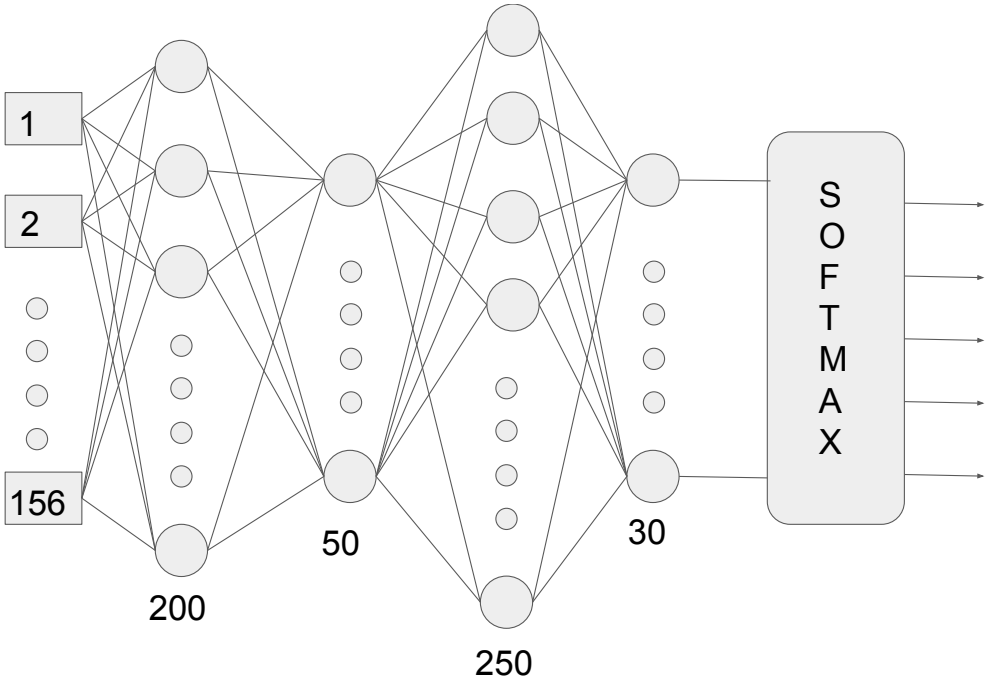

Figure 3: We propose a deep neural network comprising of 4 fully connected hidden layers for the classification task. The size of the each layer is 200, 50, 250 and 30 respectively. The sawtooth structure of layers ensures that the successive layers first learn the sparse features and then drop the non-essential features.

changes to backpropagation as shown in [13]. In our model, we achieve sufficient results by exploiting the parameters of regularization.

3. PROPOSED WORK

In this section we describe the proposed deep neural model for the classification task and the autoencoder based pre-training technique used by us to train the proposed model.

3.1 Combined Left and Right Segmentation

We propose a left-and-right combined segmentation technique for segmentation of basal ganglia structures (refer to Figure 1). The proposal is potentially useful because:

1. There is symmetry in the intensity structures of left and right regions,
2. Since these structures are very small, combining the left and right increases the training data, and
3. The left and right structures can be easily separated, post classification, with the help of spatial coordinates.

3.2 Training Set Generation

We have used public dataset available at [7] for our experiments. We divide 83 MRI scans available in the dataset into a training set of 50 MRIs that include 20 healthy control and 10 MRIs of each of the disease class. From this training set of 50 MRIs, 500 voxels of each class (4 Basal ganglia sub-regions and one background), for all 50 MRIs, are randomly chosen. In order to build a robust classifier that works good on both healthy and disease class, we have included the disease MRIs as well in the training data.

3.3 Input Feature Vector

Since Basal Ganglia occupies a small region in the brain, we first select the voxels having non-zero probability of being contained in any of the sub-regions, across all the 50 training MRI scans. The selection, essentially, picks the voxels which were seen belonging to one of the sub-regions in atleast one of the training MRI scans. This pruning of data is carried out to balance the number of samples for various classes. Without pruning the number of samples corresponding to background class explode. Hence, simple approach of selecting an ROI and pruning samples outside has been implemented. Note that the training set may still include voxels lying in the vicinity of BG, since such voxels were found to be contained in a sub-region in one or more of the 50 training MRI scans and were outside the BG region in rest of the MRIs.

After the pruning step, we construct a feature vector for each remaining voxel. The feature vector contains:

- Intensities of a 2D patch of size 11×11 (centered on the voxel of interest).
- Probability vector of length 5 of the central voxel as belonging to each of the 5 classes (4 sub-regions and one background) computed from the training data,
- 3-D co-ordinates of the central voxel,
- Probability histogram of the patch: We compute the probability, for each of the 120 neighboring voxels, of it being lying in one of the sub-regions. This is computed by simply summing the probabilities corresponding to each of the sub-regions. We then create a 27 bin histogram of the 120 obtained probability values.

Hidden Layer No.	l^2 parameter	Sparsity regularizer	Sparsity proportion	Max epochs used
Layer 1	0.015	4	0.36	1000
Layer 2	0.010	4	0.36	1000
Layer 3	0.005	4	0.36	1000
Layer 4	0.002	4	0.36	1000

Table 1: Parameters used to pretrain the proposed model using stacked sparse autoencoder.

Data Set	Caudate	Putamen	Pallidus	Accumbens	Background
Healthy Control	0.9583	0.9632	0.9716	0.9572	0.9763
Bipolar with Psychosis	0.9437	0.9345	0.9569	0.9554	0.9491
Bipolar without Psychosis	0.9415	0.9353	0.9696	0.9562	0.9609
Schizophrenia	0.9497	0.9390	0.9753	0.9632	0.9698
Overall Average	0.9483	0.9430	0.9684	0.9580	0.9640

Table 2: Accuracy obtained, in Dice Coefficient, on various controlled groups of the test dataset. We achieve significantly higher accuracy than any of the results reported so far. Further we obtained uniformly good results on healthy as well various disease groups showing the robustness and generalizability of the proposed model.

The obtained feature vector is of 156 dimensions: 121 scalars from intensities, 5 from voxel probability vector, 3 from coordinates and 27 from the histogram. Figure 2 shows the feature vector graphically.

For each MRI scan chosen for training we randomly sample 2500 voxels for the training, resulting in train set of 125K voxel samples. We create the test datasets for each disease and healthy control group from remaining 33 MRIs.

3.4 Pretraining

We propose a deep neural network comprising of 4 fully connected hidden layers for the classification task. The size of the each layer is 200, 50, 250, and 30, respectively (refer to Figure 3). We pretrain the network model using stacked sparse autoencoder. The sawtooth structure of layers ensures that the successive layers first learn the sparse features and then drop the non-essential features for the structure. Given the sawtooth structure of the model, sparsity is also an essential constraint during the training to avoid learning trivial identity transformation at the intermediate layers.

We do a layerwise training of the autoencoder as suggested in [11]. Sparsity proportion and l^2 regularisation parameter for training the model are also set to as suggested by [11]. Since our model is deeper than [11], we have increased the sparsity regularization to 4 in all the layers of our architecture. Further we also decreased the l^2 regularization parameter in each layer.

We used, l^2 and KL divergence regularizers used for sparsity. l^2 weight was set at 0.015 and decreased by a constant factor of two-thirds at each layer. Similarly the Sparsity proportion parameter was set to 0.36 on the first layer and decreased by the same constant factor.The sparsity proportion was set at 4 for all the layers. Table 1 gives all the parameters used during pre-training.

3.5 Classifier Design and Training

Once the autoencoder is trained, it is chopped at the encoder output. This encoder output is fed to a fully connected softmax layer. The labeled data from the training set is fed to this classifier and fine tuning of the hidden layers (encoder layers of pre-trained autoencoder) is carried out using back propagation. This leads to an autoencoder based classifier design.

4. DATA DESCRIPTION AND PRE-PROCESSING

We have used public dataset available at [7] for our experiments. The dataset was acquired by [5] 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 [5]. 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 registration has been carried out on Talairach coordinate space [5] along with bias field correction. Region segmentation was performed with a semi-automated intensity contour algorithm [5].

For training proposed DNN model, we have utilized the Neural Network toolkit available in MATLAB 2015B. The features extraction and visualization of ROIs has been done in MATLAB using the ANALYZE-NIFTI toolkit. All programs were executed on a system with 20 cores of Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50 GHz with 66 GB RAM equipped with Tesla K40 GPU accelerator and 64-bit operating system.

Usually, MRIs are registered to make them as similar as possible to a common template prior to any further processing. However the data described in [7] is provided after motion realignment, slice time correction, and normalization. Original MRI loaded in NifTI format is transformed into a 3D Matrix and loaded in MATLAB for further processing. All scans with a resolution different than $256 \times 128 \times 256$ voxels were discarded, leaving only 83 valid MRIs.

5. RESULTS

A test set of 33 MRIs was used for testing our trained model. The test dataset was divided into 18 MRIs of disease 1 group, 8 MRIs of disease 2 group, 4 MRIs of healthy control group and 3 MRIs of disease3 group. We used dice coefficient to measure the accuracy of our results as used in [1] and [11]. Dice Coefficient is calculated as:

$$Dice = \frac{2 * TP}{2 * TP + FP + FN}, \tag{3}$$

where TP is the number of voxels which are correctly pre-

Method	Training Dataset	Technique	Sample Size	Caudate	Putamen	Pallidus	Accumbens	Background
Proposed	Healthy+Disease	Stacked Sparse Autoencoder	83 T1-Weighted	0.9483	0.9430	0.9684	0.9680	0.9640
[11]	Only Healthy	Shallow Sparse Autoencoder	81 T1-Weighted	0.8916	0.9042	0.8376	0.7395	0.9169
[1]	Disease Only	Classifier Fusion and Labeling	270 T1 weighted	0.8809	0.8984	0.8187	0.7582	NA
[1]	Disease Only	Bayesian appearance model	270 T1 weighted	0.8560	0.8640	0.7947	0.6873	NA
[1]	Disease Only	Profile active appearance model	270 T1 weighted	0.8343	0.8633	0.7933	0.6769	NA

Table 3: Comparison of Dice coefficients obtained on BG segmentation with methods in literature. 'NA' implies 'Not Available'

Data Set	Caudate	Putamen	Pallidus	Accumbens	Background
Proposed model	0.9483	0.9430	0.9684	0.9580	0.9640
Intensities(121) only	0.8792	0.8777	0.9313	0.9132	0.8882
Intensities(121)+Coordinates(3)	0.9331	0.9282	0.9604	0.9512	0.9304
Intensities(121)+Probabilities(5)	0.8838	0.8786	0.9344	0.9144	0.8821
Intensities(121)+Probability Histogram(27)	0.9384	0.8944	0.9447	0.9287	0.9604

Table 4: Accuracy obtained, in Dice Coefficient, with intensities only (baseline) and combination of intensity feature with each of the other individual feature.

dicted, FP is the number of voxels wrongly classified as part of the subregion and FN is the number of voxels wrongly classified.

The accuracy obtained for various groups of test dataset in as given in Table 2. We achieve significantly higher accuracy than any of the results reported so far. Further, we obtained uniformly good results on healthy as well various disease groups showing the robustness as evident from Table 3. This indicates the generalizability of the proposed model. We speculate that the reason behind better performance on various disease sets is due to inclusion of both healthy control and diagnostic group MRIs during the training process.

We also performed detailed experiments with individual features, results of which are presented in Table 4. From the table, it is evident that only intensities as features provided us the accuracy of 89% compared to 95% obtained with the proposed model. In addition, we also note that probability feature vector (length 5) is least significant, while both coordinates (length 3) and probability histogram (length 27) provided a significant boost over the baseline accuracy obtained with intensities only. This shows that probability feature (length 5) is perhaps redundant because coordinates have already been accounted for in the model.

We also experimented by designing a support vector machine (SVM) classifier with the same 156 length feature vector and observed 88.9% accuracy. This establishes the efficacy of the proposed work.

6. CONCLUSIONS

In this paper we propose a deep neural network architecture to segment Basal Ganglia region in human brain from an MRI scan. We use stacked sparse autoencoder to train the propose model using a public dataset of 103 T1-weighted images. The paper proposed a novel idea of jointly segmenting left and right sun-regions of the basal ganglia. This allows the proposed network to focus on inter-region differences without getting distracted by small intra-region differences. Including the probability histogram of neighboring patch allows the network to learn local context. All the novel contributions jointly enable the deep neural network to learn distinguishing features for various sub-regions leading to state of the art accuracy of over 94% measured in terms of dice coefficient. In future, we would like to extend the idea of deep neural networks further to segment other important regions of the brain anatomy.

7. REFERENCES

[1] K. O. Babalola, B. Patenaude, P. Aljabar, J. Schnabel, D. Kennedy, W. Crum, S. Smith, T. Cootes, M. Jenkinson, and D. Rueckert. An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage*, 47(4):1435–1447, 2009.

[2] P. Baldi. Autoencoders, unsupervised learning, and deep architectures. *ICML unsupervised and transfer learning*, 27(37-50):1, 2012.

[3] Y. Bengio, Y. LeCun, et al. Scaling learning algorithms towards ai. *Large-scale kernel machines*, 34(5), 2007.

[4] A. de Brebisson and G. Montana. Deep neural networks for anatomical brain segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 20–28, 2015.

[5] J. A. Frazier, S. M. Hodge, J. L. Breeze, A. J. Giuliano, J. E. Terry, C. M. Moore, D. N. Kennedy, M. P. Lopez-Larson, V. S. Caviness, L. J. Seidman, et al. Diagnostic and sex effects on limbic volumes in early-onset bipolar disorder and schizophrenia. *Schizophrenia bulletin*, 34(1):37–46, 2008.

[6] Y. Guo, G. Wu, L. A. Commander, S. Szary, V. Jewells, W. Lin, and D. Shen. Segmenting hippocampus from infant brains by sparse patch matching with deep-learned features. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 308–315. Springer, 2014.

[7] D. N. Kennedy, C. Haselgrove, S. M. Hodge, P. S. Rane, N. Makris, and J. A. Frazier. Candishare: a resource for pediatric neuroimaging data. *Neuroinformatics*, 10(3):319–322, 2012.

[8] M. Kim, G. Wu, and D. Shen. Unsupervised deep learning for hippocampus segmentation in 7.0 tesla MR images. In *International Workshop on Machine Learning in Medical Imaging*, pages 1–8. Springer, 2013.

[9] A. Klein, B. Mensh, S. Ghosh, J. Tourville, and J. Hirsch. Mindboggle: automated brain labeling with multiple atlases. *BMC medical imaging*, 5(1):1, 2005.

[10] V. A. Magnotta, D. Heckel, N. C. Andreasen, T. Cizadlo, P. W. Corson, J. C. Ehrhardt, and W. T. Yuh. Measurement of brain structures with artificial neural networks: Two-and three-dimensional applications 1. *Radiology*, 211(3):781–790, 1999.

[11] A. Martínez González. Segmentation of brain MRI structures with deep machine learning. 2012.

[12] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson. Comparison of adaboost and support vector machines for detecting alzheimerâĂŹs disease through automated hippocampal segmentation. *IEEE transactions on medical imaging*, 29(1):30, 2010.

[13] A. Ng. Sparse autoencoder. *CS294A Lecture notes*, 72:1–19, 2011.

[14] S. Powell, V. A. Magnotta, H. Johnson, V. K. Jammalamadaka, R. Pierson, and N. C. Andreasen. Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. *Neuroimage*, 39(1):238–247, 2008.