Published in final edited form as:

Mach Learn Med Imaging. 2014 ; 8679: 68–76. doi:10.1007/978-3-319-10581-9_9.

Deep Learning for Cerebellar Ataxia Classification and Functional Score Regression

Zhen Yang¹, Shenghua Zhong¹, Aaron Carass¹, Sarah H. Ying², and Jerry L. Prince¹

¹Johns Hopkins University, Baltimore, USA

²Johns Hopkins School of Medicine, Baltimore, USA

Abstract

Cerebellar ataxia is a progressive neuro-degenerative disease that has multiple genetic versions, each with a characteristic pattern of anatomical degeneration that yields distinctive motor and cognitive problems. Studying this pattern of degeneration can help with the diagnosis of disease subtypes, evaluation of disease stage, and treatment planning. In this work, we propose a learning framework using MR image data for discriminating a set of cerebellar ataxia types and predicting a disease related functional score. We address the difficulty in analyzing high-dimensional image data with limited training subjects by: 1) training weak classifiers/regressors on a set of image subdomains separately, and combining the weak classifier/regressor outputs to make the decision; 2) perturbing the image subdomain to increase the training samples; 3) using a deep learning technique called the stacked auto-encoder to develop highly representative feature vectors of the input data. Experiments show that our approach can reliably classify between one of four categories (healthy control and three types of ataxia), and predict the functional staging score for ataxia.

1 Introduction

Cerebellar ataxia is a progressive neuro-degenerative disease that preferentially affects the cerebellum. This relatively rare spectrum of diseases has multiple genetic versions, each with a characteristic pattern of anatomical degenerations that yields distinctive motor and cognitive problems. Despite the significant impact on the lives of patients, the current standard of diagnosis, prognosis, and treatment of ataxia is inadequate. The clinical evaluations are mostly indirect, by use of clinical motor and cognitive testing. There are no accurate methods to predict the character and timing of likely functional losses. MR image analyses provide potentials to improve the evaluation of cerebellar neuro-degeneration by revealing the structural changes of the cerebellum. Fig. 1 shows example coronal sections of the cerebellum from healthy control (HC), spinocerebellar ataxia type 2 (SCA2), spinocerebellar ataxia type 6 (SCA6), and ataxia-telangiectasia (AT). We can see that all of the three ataxia types show cerebellar atrophy compared to the HC. However, SCA2 shows significant atrophy of the corpus medullare (central white matter of the cerebellum and the deep cerebellar nuclei) while SCA6 shows more atrophy in the posterior-inferior regions of the cerebellum. Besides discriminating degeneration patterns, it may be possible to quantitatively study the correlation between the amount of structural change and degree of functional loss.

Various approaches have been proposed for studying the correlation between the structural changes of the brain and the clinical measurements. According to the type of features used to characterize the structural changes, they fall into two categories: 1) using low-dimensional carefully designed features, e.g., volumetric measurement of manually delineated region of interests (ROIs) [1, 2]; 2) high dimensional features with the same order as the input images, e.g., brain morphology changes represented by deformation field from a template [3, 4]. The latter group of approaches has gained popularity for: 1) less involvement of manual design and delineation and 2) being able to capture the complex patterns of structural changes. However, the high-dimensional input (up to millions) of a typical medical image and the small sample size (often several hundred) that can be acquired makes the problem challenging. Various proposed to encode the high-dimensional input into a relatively small number of features that are both representative of the data and discriminative for classification purposes [5, 6].

In this work, we present a learning framework for MR image based classification and regression of cerebellar ataxia degeneration patterns. We address the problem of analyzing high-dimensional data with limited training samples with a series of strategies. Instead of classification/regression directly on the whole image volume, we train weak classifiers/regressors on a set of image subdomains separately, and then learn a classifier/regressor to combine the weak decisions. Based on the local smoothness properties of medical images, we perform a local perturbation to generate more training samples. Stacked auto-encoder (SAE) [7], a deep learning techniques, is used to develop highly representative feature vectors of the input data. Experiments show that our approach can reliably classify four categories, (HC and three types of ataxia), and predict the functional staging score for ataxia (FSFA). This is the first machine learning approach with MR image input for studying the correlation between cerebellar structural change and degeneration patterns of cerebellar diseases.

2 Method

2.1 Pre-processing

Our data consists of T1-weighted MPRAGE images of 168 subjects, 61 HCs and 107 patients with various types of ataxia. 120 of the subjects completed a series of neurological tests that emphasizes mobility and were assigned a functional staging score for ataxia (FSFA), which provides a score from 0 to 6. The FSFA rating scale is a subset of the Unified Ataxia Disorders Rating Scale (UADRS) [8]. A higher FSFA value indicates more functional losses. The MR images were preprocessed with Freesurfer software [9] to generate a masked and intensity-normalized image that contains only the cerebellum. The masked images are registered to a template, allowing only rigid and scale transformations. The images are then cropped to the same bounding box to tightly contain the cerebellum, see Fig. 1. We use I to denote the MR image after the above processing.

2.2 Method outline

As shown in Fig. 2(b), we train K weak classifiers/regressors on K image subdomains separately, and then learn a classifier/regressor C to combine the decisions from weak

classifiers. As shown in Fig. 2(a), each of the K image subdomains is a square plane of the same size, but with different locations and orientations. The image subdomains are chosen so that they are distributed evenly across the whole volume. Here we use six coronal planes (three on each hemisphere), and three sagittal planes, so $K = 9$. Let $\pi_k(u, v)$, $u \in [0, 1]$, $v \in [0, 1]$ be the parametric equation of the square plane at the k^{th} image subdomain, $k = 1, 2, \dots, K$.

Training weak classifier/regressor R_k —Let I^i be the preprocessed image for the i^{th} training subject. A 2D image patch $S_k^i(u, v) = I^i(\pi_k(u, v))$ can be generated by evaluating I^i on $\pi_k(u, v)$. $S_k^i(u, v)$ is discretized to a 32×32 image and vectorized. We will use S_k^i to represent the vector. In order to increase the training samples, we perturb π_k by changing its center and orientation by small amounts, and produces a set of perturbed planes, see Fig. 3. Now each subject can generate multiple image patches by evaluating I^i on the set of

perturbed planes, resulting in a set of image patches $\Omega_k^i = \{S_{k,m}^i\}$, $m = 1, 2, \dots, M$. A stacked autoencoder (SAE) is trained on the image patches thus generated from all subjects,

i.e. $\{S_{k,m}^i\}$, $m = 1, 2, \dots, M$, $i = 1, 2, \dots, N$, and outputs a low dimensional feature vector

S_{k,m_j}^i for each input image patch S_{k,m_j}^i . Finally, the set of feature vectors $\{S_{k,m}^i\}$, $i = 1, 2, \dots, M$, is used to train the weak classifier R_k . Let $\mathbf{f}_k(\cdot)$ be the outputs from R_k , which is a vector specifying the class probabilities in the case of classification and the predicted functional score in the case of regression.

Training combined classifier/regressor C —To generate a training sample for C , for each image subdomain k , an image patch S_{k,m_j}^i is randomly selected from Ω_k^i . The selection is carried out independent among different image subdomains. S_{k,m_j}^i is in turn input into SAE to generate a feature vector S_{k,m_j}^i and weak classifier/regressor R_k to generate an output $\mathbf{f}_k(S_{k,m_j}^i)$. The feature vector for a training sample is formed by concatenating the outputs from all R_k , i.e. $[\mathbf{f}_1(S_{1,m_j}^i), \mathbf{f}_2(S_{2,m_j}^i), \dots, \mathbf{f}_K(S_{K,m_j}^i)]$. We generate multiple training samples from each training subject by repeating the above procedure.

Testing stage—Given the image of a test subject, multiple samples are generated in a similar way to the process of generating training samples for the combined classifier/regressor C . Each training sample will be assigned by C the class probabilities in the case of classification or a predicted score in the case of regression. For classification, the final output is the class that has the maximum average class probability output by C from the multiple samples. For regression, the final score is the average score predicted by C from the multiple samples.

2.3 Dimensionality Reduction based on Stacked Autoencoder

In this section, we will describe the stacked autoencoder (SAE) for dimensionality reduction in training R_k . A typical autoencoder proposed by Bengio et al. [7] takes an input vector $\mathbf{x} \in [0, 1]^d$, and maps it to a hidden representation $\mathbf{y} \in [0, 1]^d$ through a mapping function $\mathbf{y} = f_\theta$

$(\mathbf{x}) = \text{sigmoid}(\mathbf{W}\mathbf{x} + \mathbf{b})$, parameterized by $\theta = \{\mathbf{W}, \mathbf{b}\}$, the weight matrix and bias, respectively. This latent representation in the hidden layer $\mathbf{y}$ is projected back to reconstruct the input vector $\mathbf{x}$. The reconstructed vector $\mathbf{z} \in [0, 1]^d$ is in the input space, and can be written as $\mathbf{z} = g_{\theta'}(\mathbf{y}) = \text{sigmoid}(\mathbf{W}'\mathbf{y} + \mathbf{c}')$ with $\theta' = \{\mathbf{W}', \mathbf{c}'\}$. Often it is constrained in such a way that $\mathbf{W}' = \mathbf{W}^T$, and are referred to as tied weights. Each training sample $\mathbf{x}^{(i)}$ is mapped to a corresponding $\mathbf{y}^{(i)}$ and has a corresponding reconstruction $\mathbf{z}^{(i)}$. The parameter space is optimized to minimize the average reconstruction error:

$$\min_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^n L(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \min_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^n L(\mathbf{x}^{(i)}, g'_{\theta'}(f_{\theta}(\mathbf{x}^{(i)}))), \quad (1)$$

where

$$L(\mathbf{x}, \mathbf{z}) = - \sum_{k=1}^d [x_k \log z_k + (1 - x_k) \log (1 - z_k)], \quad (2)$$

is a loss function based on the reconstruction cross-entropy.

Autoencoders are used as building blocks to build the deep networks, called stacked autoencoder (SAE), and trained in a greedy layer-wise way [7]. In our work, we stacked two autoencoders to form the SAE, see Fig. 4. The input layer has 1024 nodes, with each node corresponding to a pixel in the image patches. The first and second hidden layer have 500 and 100 nodes respectively. After the auto-encoders are constructed and trained, a classification/regression layer is added on top of SAE, and the entire deep network is fine-tuned to find the optimal parameter for the classification/regression purpose. Given a new image patch S , a feature vector s can be computed by feed-forward propagation through the SAE.

2.4 Classification and regression

Random forests [10] are used for both the weak classifier/regressor R_k and the combining classifier/regressor C . Random forests are an ensemble learning method for classification and regression. It operates by constructing a set of decision trees at training time and outputting the class that is the mode of the classes output by individual trees. It also assigns the probability $f^t(x)$ that an observation x belongs to a particular class t as the mean predicted class probabilities of the trees in the forest. Random forest have been shown to achieve robust and accurate classification while avoiding over-fitting.

3 Experiments

3.1 Cerebellar ataxia classification

In this experiment, we consider the classification problem of classifying four groups: HC, SCA2, SCA6 and AT. We applied a 10-fold cross validation on a dataset of 80 subjects, formed by 31 HC subjects, 4 SCA2 subjects, 27 SCA6 subjects, and 18 AT subjects. The dataset is partitioned into 10 subsets in a way that the proportion of the four groups in each subset are roughly the same. For each trial, one subset is selected for testing and the other nine subsets are used for training. The following classification methods are compared: 1)

ROI volume PCA: using the relative age-adjusted regional volumes as features [1], PCA for dimensionality reduction and random forest for classification; 2) **Image PCA:** using the whole MR image (masked and intensity normalized) as features, PCA for dimensionality reduction and random forest for classification; 3) **Log-Jacobian PCA:** using the log of the jacobian determinant computed from the deformation field from an template as features [3], PCA for dimension reduction and random forest for classification; 4) **Proposed method with PCA:** the proposed method using PCA for dimensionality reduction in R_k ; 5) **Proposed method with SAE:** the proposed method using SAE for dimensionality reduction in R_k .

As shown in Table 1 the proposed method produce the best performance, with error rate being 13.75%. It outperforms the ROI volumetric analyses, indicating that high-dimensional data can reveal structural information that is not captured by the low-dimensional volumetric measurement. It also outperforms direct dimensionality reduction on the whole image or jacobian based features, which verifies the effectiveness of the proposed learning framework, and the strategies of dimensionality reduction and feature selection. Within the proposed learning framework, the performance of using SAE for dimensionality reduction is better than using PCA.

Fig. 5 shows the resulting confusion matrix of the proposed method with SAE. There are two major sources of classification errors: 1) Classifying SCA2 as other classes, due to the limited SCA2 subjects in the dataset; 2) Classifying ataxias as HC, because the cerebellum of people with a short disease duration has mild cerebellar atrophy, and it's difficult to distinguish from healthy subjects. The learning framework can be further optimized to discover features that distinguishes the subjects with mild disease from the healthy.

3.2 Functional score regression

In this experiment, we considered the regression of functional staging score for ataxia (FSFA). We applied a 10-fold cross validation on a dataset of 120 subjects which have FSFA evaluated. As in Sec. 3.1, five methods are compared. Table 2 shows that the proposed method with SAE outperforming the other methods in root mean square error (RMSE), and is compatible with the lobule volume PCA for Pearson's correlation coefficient. Fig. 6 shows the predicted FSFA vs. true FSFA for the test subjects from all 10 trials. Example coronal slices of subjects are plotted against their corresponding points in the plot. The slices from the bottom to the top of the plot show a trend from mild to severe atrophy, which indicates that the proposed method is able to capture the correlation between the degree of cerebellar atrophy and the functional loss. There are several sources for prediction error. Some of the ataxia types, such as AT and SCA3, have profounder sensory changes than other types, which could contribute to a worse FSFA than would be predicted by cerebellar changes alone. Many other factors, such as statue, weight, profession and the amount of exercise can affect their functional performance. Also, as discussed in Section 3.1, there's still room for optimizing the learning framework to discover overlooked features that correlate with FSFA.

4 Conclusion

We presented a learning framework for MR image based classification of cerebellar ataxia types and prediction of a disease related functional score. We addressed the difficulty in analyzing high-dimensional image data with limited training subjects by: 1) training weak classifiers/regressors on a set of image subdomains separately, and combining the weak classifier/regressor outputs to make the decision; 2) perturbing the image subdomain to increase the training samples; 3) using stacked auto-encoder (SAE), a deep learning technique, to develop highly representative feature vectors of the input data. Experiments show that our approach can reliably classify between one of four categories (HC and three types of ataxia), and predict the functional staging score for ataxia.

References

1. Jung BC, Choi SI, Du AX, Cuzzocreo JL, Geng ZZ, Ying HS, Perlman SL, Toga AW, Prince JL, Ying SH. Principal component analysis of cerebellar shape on mri separates sca types 2 and 6 into two archetypal modes of degeneration. *The Cerebellum*. 2012; 11(4):887–895. [PubMed: 22258915]
2. Jung BC, Choi SI, Du AX, Cuzzocreo JL, Ying HS, Landman BA, Perlman SL, Baloh RW, Zee DS, Toga AW, et al. Mri shows a region-specific pattern of atrophy in spinocerebellar ataxia type 2. *The Cerebellum*. 2012; 11(1):272–279. [PubMed: 21850525]
3. Leow AD, Klunder AD, Jack CR Jr, Toga AW, Dale AM, Bernstein MA, Britson PJ, Gunter JL, Ward CP, Whitwell JL, et al. Longitudinal stability of mri for mapping brain change using tensor-based morphometry. *Neuroimage*. 2006; 31(2):627–640. [PubMed: 16480900]
4. Ashburner J. Computational anatomy with the spm software. *Magnetic resonance imaging*. 2009; 27(8):1163–1174. [PubMed: 19249168]
5. Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. Compare: classification of morphological patterns using adaptive regional elements. *Medical Imaging, IEEE Transactions on*. 2007; 26(1):93–105.
6. Batmanghelich NK, Taskar B, Davatzikos C. Generative-discriminative basis learning for medical imaging. *Medical Imaging, IEEE Transactions on*. 2012; 31(1):51–69.
7. Bengio Y. Learning deep architectures for AI. *Foundations and trends in Machine Learning*. 2009; 2(1):1–127.
8. Subramony S, May W, Lynch D, Gomez C, Fischbeck K, Hallett M, Taylor P, Wilson R, Ashizawa T, et al. Measuring friedreich ataxia: interrater reliability of a neurologic rating scale. *Neurology*. 2005; 64(7):1261–1262. [PubMed: 15824358]
9. Dale AM, Fischl B, Sereno MI. Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction. *NeuroImage*. 1999; 9(2):179–194. [PubMed: 9931268]
10. Breiman L. Random forests. *Machine learning*. 2001; 45(1):5–32.

Fig. 1.
Example coronal sections of the cerebellum from HC and three ataxia types.

Fig. 2.
 (a) Image subdomains. (b) Method Diagram.

Fig. 3.
Image patches.

Fig. 4.
Stacked auto-encoder.

True \ Pred	HC	SCA2	SCA6	AT
HC	0.97	0.00	0.03	0.00
SCA2	0.25	0.00	0.50	0.25
SCA6	0.11	0.00	0.89	0.00
AT	0.06	0.00	0.11	0.83

Fig. 5.

Average confusion matrix for the proposed method using SAE for dimensionality reduction.

Fig. 6. Predicted FSFA vs. true FSFA, using the proposed method. Typical MR images of the cerebellum are shown.

Table 1

Classification error rate.

Method	Error rate (%)
ROI volume PCA	16.25 ± 8.44
Image PCA	16.25 ± 11.86
Log-Jacobian PCA	22.50 ± 15.37
Proposed method with PCA	15.00 ± 11.49
Proposed method with SAE	13.75 ± 12.43

Table 2

Root mean square error (RMSE) and Pearson correlation coefficient between the measured functional score and predicted functional score.

Method	RMSE	Pearson
Lobule volume PCA	1.187±0.217	0.693± 0.110
Image PCA	1.191±0.274	0.669± 0.153
Log-Jacobian PCA	1.250±0.181	0.635± 0.091
Proposed method with PCA	1.154±0.209	0.687± 0.144
Proposed method with SAE	1.148±0.211	0.685± 0.133