Similarity based classification of ADHD using Singular Value Decomposition

Taban Eslami
taban.eslami@wmich.edu
Department of Computer Science
Western Michigan University

Fahad Saeed
fahad.saeed@wmich.edu **
Department of Computer Science
Western Michigan University

1 Abstract

Attention deficit hyperactivity disorder (ADHD) is one of the most common brain disorders among children. This disorder is considered as a big threat for public health and causes attention, focus and organizing difficulties for children and even adults. Since the cause of ADHD is not known yet, data mining algorithms are being used to help discover patterns which discriminate healthy from ADHD subjects. Numerous efforts are underway with the goal of developing classification tools for ADHD diagnosis based on functional and structural magnetic resonance imaging data of the brain. In this paper, we used Eros, which is a technique for computing similarity between two multivariate time series along with k-Nearest-Neighbor classifier, to classify healthy vs ADHD children. We designed a model selection scheme called *J-Eros* which is able to pick the optimum value of k for k-Nearest-Neighbor from the training data. We applied this technique to the public data provided by ADHD-200 Consortium competition and our results show that J-Eros is capable of discriminating healthy from ADHD children such that we outperformed the best results reported by ADHD-200 competition more than 20 percent for two datasets.

The implemented code is available as GPL license on GitHub portal of our lab at https://github.com/pcdslab/J-Eros.

2 Introduction

In the past two decades, machine learning techniques have been applied as useful tools for classifying healthy vs neurological patients by discovering patterns in the data and using them for predicting new data [1]. Mild Cognitive Impairment and Alzheimer Dementia, temporal lobe epilepsy, Attention Deficit Hyperactivity Disorder (ADHD), Psychosis etc. are examples of brain disorders which are investigated using machine learning algorithms [2, 3, 4, 5]. In this study we focus on ADHD disorder. ADHD causes aggression, excitability, fidgeting, hyperactivity and many other symptoms in children. The cause of ADHD is still unknown and in many cases it continues to adulthood [6, 7]. Clinicians mostly use standard guidelines from the American Academy of Pediatrics or the American Psychiatric Association's Diagnostic and Statistical Manual (DSM) for diagnostic and treatment purposes. The diagnosis involves gathering information from several sources, including schools, caregivers, and parents and is subject to the skill and experience of the caregiver. There is no known quantitative diagnosis for ADHD. To this end, ADHD-200 Global Competition was held in 2011 with the goal of identifying ADHD bio-markers from brain imaging data [8]. The goal of such a competition and the purpose of this paper is to come up with quantitative techniques that will allow non-invasive diagnosis of ADHD. The data consists of structural and functional magnetic resonance imaging (fMRI) data collected from 8 different brain imaging centers around the world containing four different groups of subjects: control (healthy), inattentive, hyperactive or combined. Given the importance of this problem many machine learning and deep learning algorithms have been applied to classify the data into the four groups [9]. Colby et al. [10] used structural and functional data along with demographic information for ADHD classification. Kuang et al. [11] applied fast Fourier transform to convert the fMRI

**To whom correspondence should be addressed data from time domain to frequency domain and then used Deep Belief Network to discriminate healthy controls from ADHD sub types. Deshpande et al. [12] used fully connected cascade artificial neural network which is a variation of multi layer perceptron networks on functional connectivity between brain regions of subjects to discriminant healthy vs ADHD subjects and in another experiment discriminating ADHD subtypes. In another attempt, Liu et al [13] first reduced the dimensionality of fMRI data by applying deep belief network to detect regions which perform well in classification, then used Bayesian network to extract relationship between well performed brain areas and finally applied SVM classifier to predict subjects as control, combined, inattentive or hyperactive. Qureshi et al. [14] selected 90 subjects (30 from each of ADHD sub-group ADHD-1 and ADHD-3, and 30 from TDC) that match with age and hardness. After calculating the global connectivity maps, they used the average of the connectivity measure as feature and used hierarchical extreme learning machine (H-ELM) classifier. In our proposed study, we look at this problem as its original version by considering four classes and using all available subjects data. We only focus on using fMRI data without demographic information. Our work is based on computing the similarity between brains using Eigenvalues and Eigenvectors of their entire time series covariance matrix. To the best of our knowledge, none of the existing approaches for this problem investigate the use of Eigenvalues and Eigenvectors of the *entire* brain signals to discriminate ADHD from healthy subjects. To compute the similarity between brain images based on Eigenvalue and Eigenvectors, we used a measure called Eros which is weighted cosine similarity of pairs of Eigenvectors based on their corresponding Eigenvalues. Using Eros as the similarity measure between a pair of brain images, k-Nearest Neighbor is used for classification. In order to pick the best value of k, we designed a model selection scheme which considers both sensitivity and specificity of the classification into account. Our experiments on three public datasets show significant improvement in classification accuracy compared to the results released by ADHD-200 competition. This paper is organized as follows: Section 2 describes multivariate time series and Eros similarity measure. Section 3 describes our model selection scheme, section 4 describes the time complexity of the approach, section 5 shows the results of experiments we performed on ADHD data and section 6 describes conclusion of this work and future direction of the research.

3 Multivariate time series and Eros similarity measure

In this study, we used fMRI data in which a time series is extracted from each region of the brain. These time series indicate the functional behavior of different regions over time. Each sample of fMRI data is a multivariate time series (MTS) which can be stored in an $n \times m$ matrix, where n and m correspond to the number of regions (variables) and length of time series respectively. Correlation between different regions/voxels, is a useful measure for analyzing brain functional connectivities [15]. As confirmed by many experiments performed on fMRI data this functional connectivity is an important factor for distinguishing brain disorders [16, 12]. Therefore, it is important to consider algorithms that take the functional correlations among these time series into account. Several distance measures such as Euclidean Distance, Diffusion Time Warping, PCA similarity factor [17] and Extended Ferobenius norm (Eros) [18] have been proposed to compute the similarity of two MTS items. In this study, we investigate the application of Eros similarity measure to analyze the MTS data from different brains.

3.1 Extended Frobenius norm

Extended Frobenius norm (Eros), proposed by Yang and Shahabi [18], is a similarity measure for multivariate time series datasets. Given two MTS items A and B with sizes $n \times m_1$ and $n \times m_2$, where n is the number of time series in each item and m_1 and m_2 are the length of time series, Eros first applies Singular Value Decomposition (SVD) to the covariance matrix of each MTS item. Right eigenvector matrices $V_A = [a_1, a_2, ..., a_n]$ and $V_B = [b_1, b_2, ..., b_n]$ generated by SVD are used for similarity calculation. a_i and b_i correspond to i_{th} Eigenvectors which are column orthonormal vectors. Since SVD is applied to covariance matrix, right Eigenvectors are equal to principal components of the data. Considering V_A and V_B, the Eros

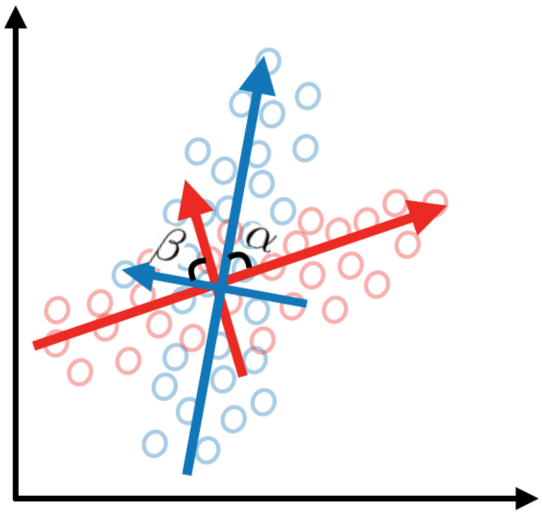

Figure 1: Corresponding Eigenvectors of two MTS item and the angles between them.

similarity between A and B is defined as follows:

$$Eros(A,B,w) = \sum_{i=1}^{n} w_i| < a_i, b_i > | = \sum_{i=1}^{n} w_i|cos\theta_i| \tag{1}$$

In this equation, $< a_i, b_i >$ is the inner product of i_{th} Eigenvectors of A and B, and $W = \{w_1, w_2, ..., w_n\}$ is a weight vector computed based on aggregating Eigenvalues from the entire MTS samples. Indeed, Eros computes a weighted sum of the cosine of the angles between corresponding Eigenvectors. Fig 1. shows an example of two corresponding Eigenvectors and angles between them. In order to compute the weights, first, Eigenvalues for each MTS item are normalized by dividing each Eigenvalue to the sum of all Eigenvalues of that MTS item. Then, w_i is computed by applying an aggregate function (min, max or mean) to i_{th} Eigenvalues of all MTS items. w_is are then normalized by dividing their values to the sum of all w_is. Algorithm 1 describes this process. In this paper we used algorithm 1 for computing the weigh vector and used mean as aggregating function f.

Algorithm 1 function computeWeightRaw, Computing a weight vector w based on the distribution of raw eigenvalues [18]

Input: an $n \times N$ matrix S, where n is the number of variables for the dataset and N is the number of MTS items in the dataset. Each column vector s_i in S represents all the eigenvalues for i_{th} MTS item in the dataset. s_{ij} is a value at column i and row j in S. s_{*i} is i_{th} row in S. s_{i*} is i_{th} column

```
for $i = 1$ to $N$ do
    $s_i \leftarrow s_i / \sum_{j=1}^{n} s_{ij}$
end for
for $i = 1$ to $n$ do
    $w_i \leftarrow f(s_{*i})$
end for
for $i = 1$ to $n$ do
    $w_i \leftarrow w_i / \sum_{j=1}^{n} w_j$
end for
```

Algorithm 2 J-Eros

Input: Training set D_train, Test set D_test

```
 1: for $i = 1$ to 10 do
 2:     J-stats[$i$] = 0
 3: end for
 4: for $k = 1$ to 10 do
 5:     sensitivity = 0
 6:     specificity = 0
 7:     for $iter = 1$ to 10 do
 8:         [sens,spef] = 4-fold-cross-validation(D_train)
 9:         sensitivity = sensitivity + sens
10:         specificity = specificity + spef
11:         if sens = 0 or spef = 0 then
12:             J-stats[$k$] = $-\infty$
13:             break
14:         end if
15:     end for
16:     sensitivity = sensitivity/10
17:     specificity = specificity/10
18:     J-stats[$k$] = sensitivity + specificity - 1
19: end for
20: k-optimal = argmax$_k$(J-stats)
21: Perform k-Nearest Neighbors classification with k-optimal on D_test
```

4 Model selection for finding optimum k in k-Nearest-Neighbor

As suggested by [18], we used the k-Nearest-Neighbor (k-NN) for classification of fMRI samples based on Eros as the similarity metric between two MTS items. k-NN classifies a test sample based on majority votes of k neighbors having the smallest distance with it [19]. The only parameter of k-NN that should be taken into account is k or the number of nearest neighbors. Optimum value of k is not known beforehand, so we performed 4-fold cross validation for each value of k in range 1 to 10. Our experiments showed that values of k greater than 10 do not improve the accuracy. In order to evaluate each value of k, we used a measure called J-statistics. J-statistics is defined as follows:

$$J = sensitivity + specificity - 1 \tag{2}$$

In this equation, sensitivity is defined as the proportion of ADHD subjects which are correctly classified as ADHD regardless of subtype (True Positive rate) and specificity is defined as the proportion of subjects which are correctly classified as healthy subjects (True Negatives rate). We used J-statistics because it considers both sensitivity and specificity into account. Since some of the ADHD datasets are imbalanced and contain more healthy subjects than ADHD subjects, considering other evaluation metrics like accuracy instead of J-statistics results in selecting a value for k in which k-NN classifies most or all of the test subjects as healthy. Even though this increases the overall accuracy, sensitivity of classification will be very poor. Algorithm 2 shows our model selection scheme. We repeated the 4-fold-cross-validation on training set 10 times for each value of k between 1 to 10 and measured the J-statistic for each k (line 4 to 19). If k results in sensitivity or specificity equal to 0, that value for k is ignored (line 11 to 14). Finally, the value of k with the best J-statistic is selected and used for performing k-NN on the test set (line 20, 21).

5 Time complexity of J-Eros

Now we analyze the time complexity of the J-Eros step by step based on the following the setting:
Given N MTS items, assume n is the number of time series (Number of regions of the brain) and t is the

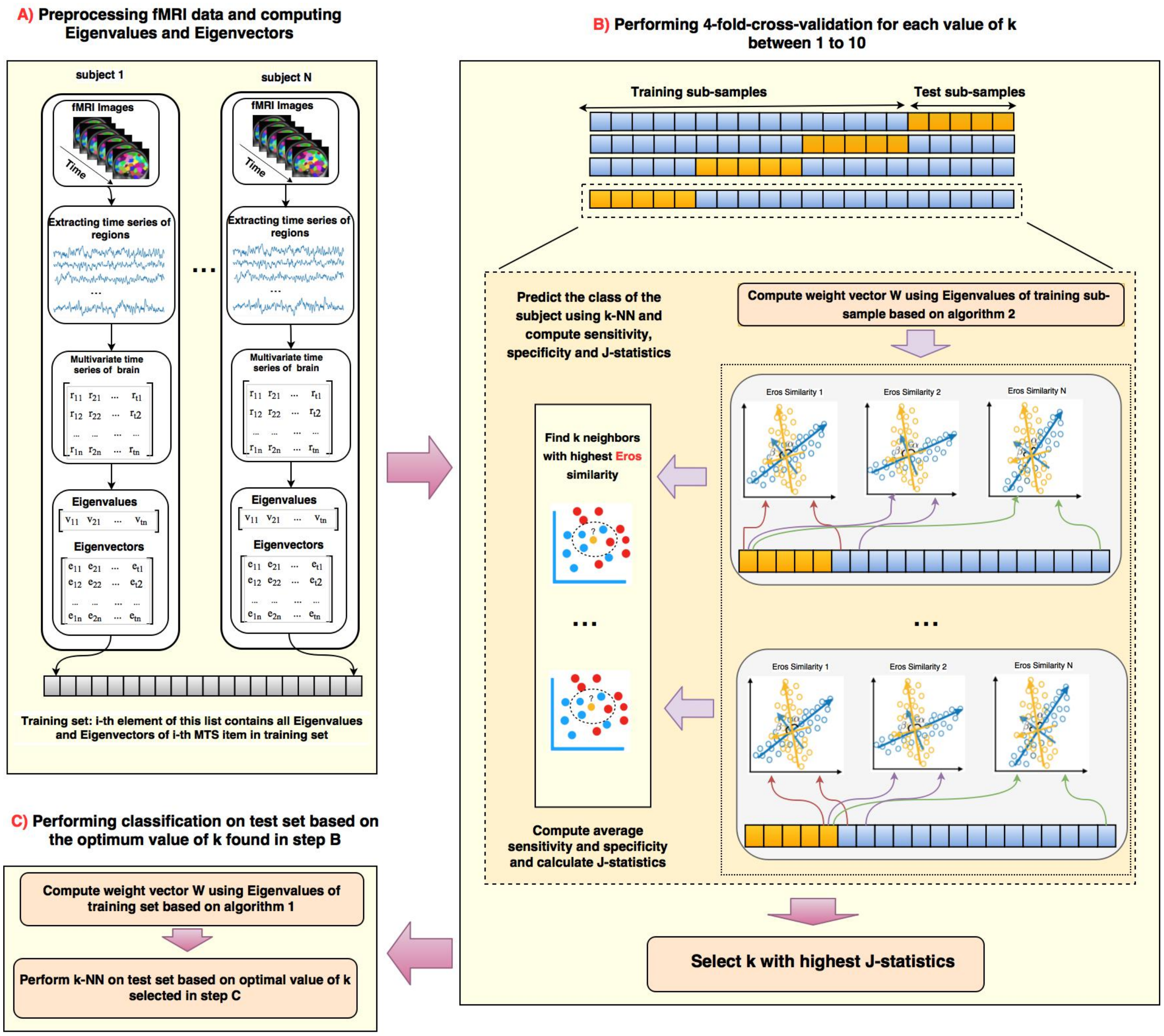

Figure 2: Work flow of J-Eros for classifying ADHD disorder. In part A, time series are extracted from brain regions, stored in n × m matrices and Eigenvalues and Eigenvectors of each matrix is calculated. In part B, 4-fold-cross-validation is performed in which k-NN using Eros as similarity measure is performed and average sensitivity and specificity of classifying test sub-samples are used for computing J-statistics. The k which results in highest J-statistics is picked as optimum k. In part C, using the optimum value of k computed in part B, k-NN is performed on test set.

maximum length of time series among all MTS items (length of time series could be different among different MTS items)

5.1 Eigendecomposition of covariance matrix

Time complexities of computing covariance matrix and applying Eigendecomposition to it are $O(n^2t)$ and $O(n^3)$ respectively. Performing these two steps for all MTS items takes $O(Nn^2t + Nn^3)$ time.

5.2 Computing weight vector W

The first step of algorithm 2 is normalizing Eigenvalues of all MTS items which has time complexity $O(nN)$. Algorithm 1 contains normalizing i_{th} Eigenvalue over all MTS items and then normalizing resulting n Eigenvalues which gives us time complexity of $O(nN)$. So the total time complexity of computing weight vector W is $O(nN)$.

5.3 Computing Eros similarity between two MTS items

The cosine of the angles between two Eigenvectors a and b of size n can be calculated by computing inner product of a and b which has time complexity of $O(n)$. Since Eros computes the weighted some of $\cos(\theta)$ for each pair of corresponding Eigenvectors, its time complexity is equal to $O(n^2)$.
Since in k-NN similarities between one test item and all training items are computed, finding k nearest neighbors of a test sample takes $O(Nn^2)$.
Based on A, B and C, time complexity of finding k nearest neighbors of a test subject is $O(Nn^2t + Nn^3 + nN + Nn^2) = O(Nn^2t + Nn^3)$.

Algorithm 2 performs 4-fold cross validation by applying k-NN repeatedly for each training and testing subsamples. Since the size of training and testing subsamples is $3N/4$ and $N/4$, time complexity of entire k-fold cross validation will be $O(N^2n^2t + N^2n^3)$. Values 10 and 4 regarding repetition number and fold number are considered as constant factors and are omitted from time complexity.

6 Experiments and results

6.1 System specification

We ran our proposed algorithm on a Linux server with Ubuntu Operating System version 14.04. This server includes two Intel Xeon E5 2620 processors with clock speed 2.4 GHz, 48 GBs RAM. We used Python 3.5 to conduct our experiments.

6.2 Datasets

We used preprocessed fMRI data provided by ADHD-200 global competition database for our experiments [1]. The preprocessing steps includes head motion correction, spatial smoothing and normalization, frame-wise displacement adjustment etc., performed by Neuro Bureau using the ”athena” pipeline. All subjects in dataset are scanned on three Tesla scanners using standard resting T2-weighted echo-planar imaging with TR = 2000 ms, flip angle = 90 degrees, echo time (TE) = 30 ms, and in-plane resolution = 64×64 mm^2, filtered using a bandpass filter (0.009 Hz ¡ f ¡ 0.08 Hz). The data is acquired from eight sites worldwide. We picked three datasets for our experiments: KKI, NYU and OHSU. The details about these three datasets are shown in table 1 and table 2.

The fMRI data we used for this study is divided into 190 regions using spatially constraint spectral clustering algorithm [20]. Each region contains voxels which are functionally homogeneous and average time series of all voxels in each region is considered as time series of that region. We removed the first ten volume of each fMRI subject and did not apply further preprocessing to the data. Since data centers used different scanners, scanning and data acquisition parameters (Some of them collected brain imaging data while subject’s eyes were closed, and subject’s eyes were opened in another centers) [21], we decided to

---

[1] https://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline

Table 1: Class membership information of training data

	control	combined	inattentive	hyperactive
KKI	61	16	5	1
OHSU	42	23	12	2
NYU	98	73	43	2

Table 2: Class membership information of test data

	control	combined	inattentive	hyperactive
KKI	8	3	0	0
OHSU	28	4	1	1
NYU	12	22	7	0

perform classification on training set and test set of each data center separately as opposed to combining all training sets.

We compared our proposed approach with approach proposed by Colby et al. [10]. and results released by ADHD-200 competition. Algorithm 2 is applied to each training set and prediction accuracy on test set is computed. Prediction accuracy of classification is the fraction of subjects that are correctly classified as control, combined, inattentive and hyperactive. Fig 3. shows the achieved accuracies for all three datasets considered in this study. Our proposed technique significantly improved prediction accuracy on KKI dataset. It also showed better performance for NYU (its worth mentioning that accuracy of NYU dataset achieved lowest accuracy among all other datasets in ADHD-200 competition). Prediction of OHSU is almost similar to the competition result. Table 3 shows the sensitivity and specificity of test sets. For all three datasets, we achieved higher sensitivity than Colby's approach. Prediction accuracy of Colby's approach is higher than ours for OHSU, but as can be seen in table 3, their sensitivity is equal to 0.

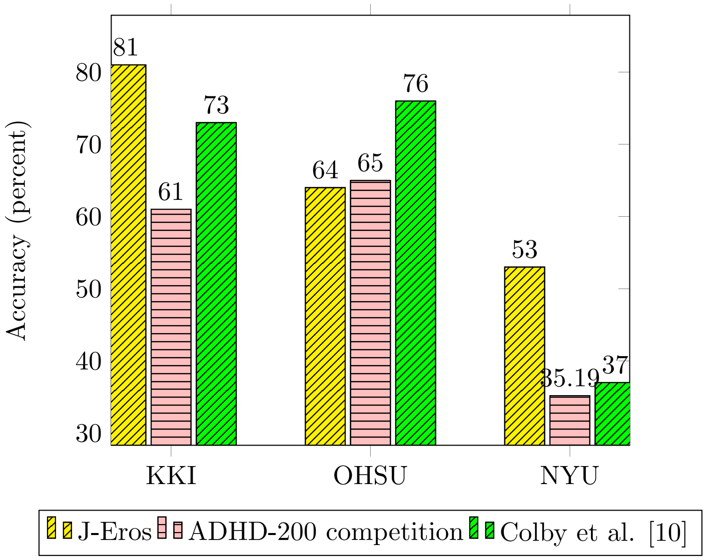

Figure 3: J-Eros achieved more accuracy than the best result in ADHD-200 competition and Colby et al for KKI and NYU. It has almost similar accuracy to competition result and less accuracy than Colby et al for OHSU. Although Colby's approach has more accuracy, its sensitivity is equal to zero.

Table 3: Comparison of sensitivity and specificity on test set.

	Sensitivity		Specificity	
Approach	proposed	Colby et al [10]	proposed	Colby et al [10]
KKI	66	0	87	100
NYU	55	34	83	58
OHSU	66	0	71	93

We also performed another experiment to see the effect of J-statistics for selecting the optimum value of k. In this experiment we compared prediction accuracy obtained by using J-statistics with other metrics like sensitivity and specificity. For each metric algorithm 2 is run and line 18 is changed based on each metric in use. Prediction accuracies corresponding to each metric is shown in Fig 4. Accuracy of KKI is the same when each of these metrics are used. NYU has lower accuracy when either sensitivity or specificity is used. OHSU has same accuracy when sensitivity is used. Regarding specificity, it has higher accuracy but in this case sensitivity of test set is equal to 0. This shows that using J-statistic which considers both sensitivity and specificity into account works better than considering each of them separately.

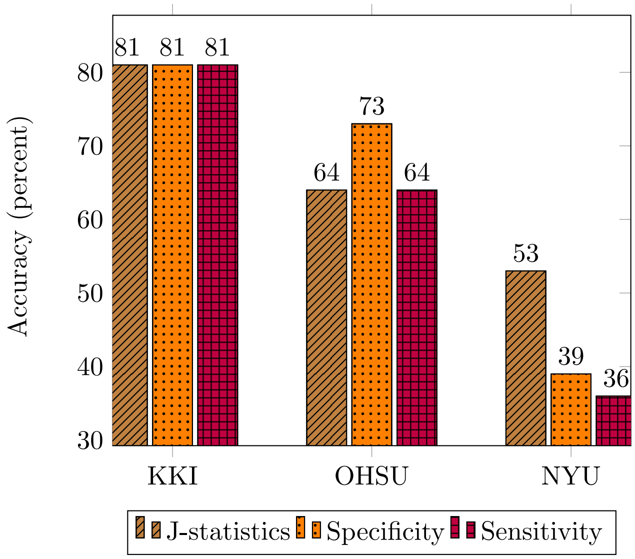

Figure 4: Comparison of prediction accuracy considering J-statistics, sensitivity and specificity for picking the best value of k.

We also measured the running time of our proposed technique. KKI and OHSU run in about 5 minutes and NYU needs 30 minutes. Larger time needed by NYU is due to the fact that it has more subjects rather than KKI and OHSU, and as explained in section IV, number of subjects in training set has a quadratic effect on time complexity. No other methods have reported the wall clock run time for their proposed methods and the source codes are not available (limiting our ability to run the experiments using these algorithms).

7 Conclusion and future works

ADHD classification is one the problems that still needs attention. Although ADHD public dataset is one the most valuable resources and many approaches have been proposed for this problem, existing methods have shown poor performance. In this paper we investigated this problem based on perspective of multivariate time series. We considered each brain as a multivariate time series and applied k-NN approach for classifying test set. We used a similarity measure called Eros for finding nearest neighbors. Eros measures how similar two multivariate time series are based on their corresponding Eigenvector and Eigenvalues of their covariance matrix. Our approach implicitly works with functional connectivity of regions since it computes the covariance matrix from the time series data of different regions. For the classification task, we use k Nearest Neighbor technique. Since the optimum value of k is unknown, we designed a model selection scheme which finds the optimum value of k that performs better than all other k's on training set. In this scheme, 4-fold-cross-validation is repeatedly performed on training data and the k which achieves highest J-statistics is selected. We compared effects of other metrics like sensitivity and specificity instead of J-statistics. Results showed J-statistics could be potentially a better measure for selecting k. We achieved higher accuracies for two datasets and almost similar accuracy for one dataset compared to the best result of ADHD-200 competition. Our proposed approach shows a robustness in terms of sensitivity as we achieved more than fifty percent sensitivity for all three datasets we used.

Acknowledgment

This research was supported by the National Institute Of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under Award Number R15GM120820 and in part by National Science Foundation grants NSF CAREER ACI-1651724. The content is solely the responsibility of the authors and does not necessarily represent the official views of the stated US Federal Agencies.

References

[1] S. Vieira, W. H. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neuroscience & Biobehavioral Reviews*, 2017.

[2] E. Hosseini-Asl, G. Gimel'farb, and A. El-Baz, "Alzheimer's disease diagnostics by a deeply supervised adaptable 3d convolutional network," *arXiv preprint arXiv:1607.00556*, 2016.

[3] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, and V. D. Calhoun, "Deep learning for neuroimaging: a validation study," *Frontiers in neuroscience*, vol. 8, 2014.

[4] Z. Yang, S. Zhong, A. Carass, S. H. Ying, and J. L. Prince, "Deep learning for cerebellar ataxia classification and functional score regression," in *International Workshop on Machine Learning in Medical Imaging.* Springer, 2014, pp. 68–76.

[5] B. C. Munsell, C.-Y. Wee, S. S. Keller, B. Weber, C. Elger, L. A. T. da Silva, T. Nesland, M. Styner, D. Shen, and L. Bonilha, "Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data," *Neuroimage*, vol. 118, pp. 219–230, 2015.

[6] S. J. Kooij, S. Bejerot, A. Blackwell, H. Caci, M. Casas-Brugué, P. J. Carpentier, D. Edvinsson, J. Fayyad, K. Foeken, M. Fitzgerald *et al.*, "European consensus statement on diagnosis and treatment of adult adhd: The european network adult adhd," *BMC psychiatry*, vol. 10, no. 1, p. 67, 2010.

[7] A. I. Frank-Briggs, "Attention deficit hyperactivity disorder (adhd)," *Journal of Pediatric Neurology*, vol. 9, no. 03, pp. 291–298, 2011.

[8] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, "The neuro bureau adhd-200 preprocessed repository," *Neuroimage*, vol. 144, pp. 275–286, 2017.

[9] A. Eloyan, J. Muschelli, M. B. Nebel, H. Liu, F. Han, T. Zhao, A. D. Barber, S. Joel, J. J. Pekar, S. H. Mostofsky *et al.*, "Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging," *Frontiers in systems neuroscience*, vol. 6, 2012.

[10] J. B. Colby, J. D. Rudie, J. A. Brown, P. K. Douglas, M. S. Cohen, and Z. Shehzad, "Insights into multimodal imaging classification of adhd," *Frontiers in systems neuroscience*, vol. 6, 2012.

[11] D. Kuang and L. He, "Classification on adhd with deep learning," in *Cloud Computing and Big Data (CCBD), 2014 International Conference on.* IEEE, 2014, pp. 27–32.

[12] G. Deshpande, P. Wang, D. Rangaprakash, and B. Wilamowski, "Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data," *IEEE transactions on cybernetics*, vol. 45, no. 12, pp. 2668–2679, 2015.

[13] A. J. Hao, B. L. He, and C. H. Yin, "Discrimination of adhd children based on deep bayesian network," 2015.

[14] M. N. I. Qureshi, H. J. Jo, and B. Lee, "Adhd subgroup discrimination with global connectivity features using hierarchical extreme learning machine: resting-state fmri study," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on.* IEEE, 2017, pp. 529–532.

[15] T. Eslami, M. G. Awan, and F. Saeed, "Gpu-pcc: A gpu based technique to compute pairwise pearson's correlation coefficients for big fmri data," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics.* ACM, 2017, pp. 723–728.

[16] A. F. LEUCHTER, T. F. NEWTON, I. A. COOK, D. O. WALTER, S. ROSENBERG-THOMPSON, and P. A. LACHENBRUCH, "Changes in brain functional connectivity in alzheimer-type and multi-infarct dementia," *Brain*, vol. 115, no. 5, pp. 1543–1561, 1992.

[17] W. Krzanowski, "Between-groups comparison of principal components," *Journal of the American Statistical Association*, vol. 74, no. 367, pp. 703–707, 1979.

[18] K. Yang and C. Shahabi, "A pca-based similarity measure for multivariate time series," in *Proceedings of the 2nd ACM international workshop on Multimedia databases.* ACM, 2004, pp. 65–74.

[19] S. Raschka and V. Mirjalili, *Python Machine Learning, 2nd Ed.*, 2nd ed. Packt Publishing, 2017.

[20] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "A whole brain fmri atlas generated via spatially constrained spectral clustering," *Human brain mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.

[21] S. Dey, A. R. Rao, and M. Shah, "Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects," *Frontiers in neural circuits*, vol. 8, 2014.