БІОМЕДИЧНА ІНЖЕНЕРІЯ І ТЕХНОЛОГІЯ

Biomedical Engineering and Technology

Вітчизняний науковий фаховий журнал

№ 1 (1) 2018
Vol. 1 (1) 2018

STATISTICAL FEATURES RANKING FOR ALZHEIMER'S DISEASE DIAGNOSIS

Domashenko D., student, 0000-0003-4937-7151
dmytrodomashenko@gmail.com
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
Manko M., student, 0000-0001-6851-943X
max.manko2595@gmail.com
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
Ciklum
Kyiv, Ukraine
Popov A., PhD, Associate Professor, 0000-0002-1194-4424, popov.kpi@gmail.com
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
Ciklum
Kyiv, Ukraine
Krashenyi I., PhD, 0000-0003-0424-147X, igor.krashenyi@gmail.com
University of Granada, Granada, Spain
Ciklum
Kyiv, Ukraine
Ramirez J., Dr., 0000-0002-6229-2921, javierrp@ugr.es
University of Granada, Granada, Spain

Abstract — This paper deals with the automated Alzheimer's disease diagnosis. In particular, the feature extraction and selection methods for the most significant features of magnetic resonance (MRI) images are considered. The algorithm for extracting statistical features of MRI images using the brain anatomical regions atlas was used for calculating the six statistical features (mean, mean absolute deviation, median, standard deviation, root mean square, skewness) for segmented MRI images of white and gray brain matter of 188 subjects with Alzheimer's disease, 401 subjects with Mild Cognitive Impairment and 229 Normal Controls. The new method for feature ranking using Wilcoxon criterion for binary classification is proposed. As a result, ranked list of features linked to the anatomical regions of the brain for each group by diagnosis was obtained. Among the most descriptive feature for AD diagnosis there are mean values in hippocampus region, mean absolute deviation in cingulum, root mean square in insula. This data indicates the features that have to be used in classification to increase the effectiveness of automated Alzheimer's disease diagnosis.

Key words — Alzheimer's Disease; MRI; feature ranking; diagnosis

I. Introduction

According to the recent findings, there are more than 35 million persons with different forms of dementia all over the world. Between 60 % and 80 % of all cases of dementia are the Alzheimer's disease (AD), affecting onein 20 people over the age of 65 years. In 2040 the number of patients with AD in European Union will be around 13.1 million [1]. Alzheimer's disease refers to diseases that impose the heaviest financial burden on society in developed countries. This disease is incurable, but the diagnosis in the early stages can significantly alleviate the disease and slightly extend the duration of patient's life [2]. Magnetic resonance imaging (MRI) is one of the best methods for diagnosis of Alzheimer's disease, since it is able to accurately measure the three-dimensional (3D) images of the structural components of the brain [3] and identify signs of pathological changes due to AD. There are many methods of MRI image processing to automatically classify images by diagnosis depending on the specific characteristics obtained during the analysis. They are based on statistical feature extraction [4], nonlinear parameters [5], wavelets [6]. Butmost methods require improvement in a sense that thefeatures used as an initial data for recognition of thedisease presence should be as descriptive as possible. One way to improve the efficiency of diagnosis is to determine suitable characteristics which are calculated for the initial MRI image and then used as input for theclassification. The objective of this work is to improve methods for selecting statistical features of MRI imagesof the human brain.

II. Diagnosis of Alzheimer's disease based on MRI images classification

For automated diagnosis of Alzheimer's disease, MRI images of the patient should be classified, i.e. passed through the algorithm that determines the class (in this case the diagnosis), to which the object under investigation refers. The system for automated diagnosis of Alzheimer's disease consists of the following components: a block of image pre-processing, features extraction block, features selection and features reduction block, classification and post-processing block (Fig. 1).
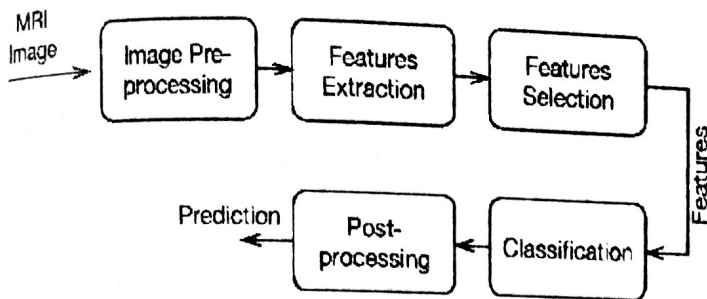

Fig. 1. Block diagram of the system
for automatic diagnosis of Alzheimer's disease

Image pre-processing involves their segmentation, i.e. separation of images into segments. Segmentation of brain MRI images is an important task in many clinical applications because it affects the whole outcome of the analysis. This is because different stages of processing results rely on accurate segmentation of anatomical regions. For example, segmentation of MRI is commonly used to measure and visualize different brain structures, to define affected areas or to analyze brain development [7].

Feature extraction stage is characterized by calculation of various parameters using image processing methods. These may be statistical parameters, principal component analysis, nonlinear parameters, wavelet analysis, etc. [8]. For example, in paper [4] the first and second order statistical features were calculated, such as Mean, Central Moments, Angular Second Moment, Contrast, Correlation, Homogeneity, and Entropy. In paper [5] the fractal dimension of edges in the Hilbert domain and the skewness and kurtosis of their spectral energy distribution are used as nonlinear parameters of the extracted features. In [6] level-3 decomposition via Harr wavelet was utilized to extract features.

The task of feature selection techniques used in classification pipeline is to learn the entire set of calculated features, and identify precisely the most descriptive and useful for the prediction. There are two main techniques for doing feature selection: principal component analysis (PCA) and independent component analysis (ICA) [9]. The main goal of PCA is to reduce the dimensionality of the data. An important step is to choose the correct number of components to remove the effect of noise and along with it to not delete useful information from the model. The purpose of ICA is not data filtering but identification of the features from all their possible clustered in large quantities on such groups as statistically independent as possible [9].

Features selection block may be omitted because many features extraction methods do not need any additional selection technique for further classification. Also, this block can be combined with classification block (for example, using neural networks) [10]. Feature processing will be carried out if improvements to their further analysis needed and usually it uses standard methods of signal processing, such as smoothing, normalization, and others [10].

Thus, a variety of methods for feature extraction of MRI images exists, that can be used in classification. The particular method to use is a matter of choice and subject to the most accurate diagnosis results achieved. In this paper, we study the alternative approach [8][11][12], when the feature selection uses the fact that different anatomical and functional areas of the brain are exposed to affection during Alzheimer's disease in varying degrees. Generally, the most significant changes are taking place in areas that are responsible for memory, understanding, language and other cognitive function. So promising way can be the consideration of image attributes primarily in those regions.

III. Statistical parameters of MRI images of brain regions

In this paper the following statistical parameters in anatomical regions of the brain are used:

A. Mean

If we have a data set containing the values $a_1, a_2, ..., a_n$, then the arithmetic mean of A is calculated using the formula:

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n}. \quad (1)$$

B. Mean absolute deviation

For a data set containing the values

59

$a_1, a_2, \ldots, a_n$, the mean absolute deviation A is calculated using the formula:

$$A = \frac{1}{n}\sum_{i=1}^{n} |a_i - a|, \qquad (2)$$

where $\bar{a}$ — mean.

C. Median

Median of a finite set of numbers can be found by placing all the observations from lowest value to the maximum value and choose the number that is in the middle. If the number of observations is odd $(n = 2k + 1)$, the median will be equal to:

$$\tilde{a} = a_{k+1}. \qquad (3)$$

If a number of observations are even $(n = 2k)$, then the middle will not be in any number, while the median is usually defined as the average of the two values that are in the middle of:

$$\tilde{a} = \frac{a_k + a_{k+1}}{2}. \qquad (4)$$

D. Standard deviation

For a data set containing the values $a_1, a_2, \ldots, a_n$, the standard deviation is calculated using the formula:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(a_i - \frac{1}{n}\sum_{i=1}^{n}a_i\right)^2}. \qquad (5)$$

E. Root mean square

Root mean square is defined as the square root of the arithmetic mean of the squares of numbers. For a data set containing the values $a_1, a_2, \ldots, a_n$, the root mean square is calculated using the formula:

$$A = \sqrt{\frac{1}{n}\left(a_1^2 + a_2^2 + \cdots + a_n^2\right)}. \qquad (6)$$

F. Skewness

Asymmetry γ of theoretical probability distribution of a random variable is the ratio of the central point of order to the cube standard deviation:

$$\gamma = \frac{\mu_3}{\sigma^3}. \qquad (7)$$

In probability theory and mathematical statistics, the central point of k-th order of random variable X with real values is:

$$\mu_k = M(X - MK)^k, \qquad (8)$$

where M — mathematical expectation.

After calculating these features in all brain regions one needs to choose only those that are the most different for MRI image of a normal and sick brain, that are the most significant. In this paper it is proposed to rank features using independent evaluation criterion for binary classification. As a criterion for accessing the relevance of each feature to separate the two groups it is proposed to use the absolute value of two sample t-test with pooled estimate of variance or absolute value of standardized U-statistic for two sample Wilcoxon test, also known as Mann-Whitney U-test. The required criterion is determined by the results of Kolmogorov-Smirnov test for normality.

The proposed method of determining a set of the most significant features of MRI images for computer-aided diagnosis consists of the following steps. First, the database of MRI images is formed. Since the features extraction is performed for each region pooled from the the brain atlas, it is necessary that all the images were scaled and spatially normalized according to the atlas. This assures the selection of necessary voxels of the image in the desired anatomical region. Considering the voxel values intensity as random values, the statistical parameters of the set of voxels in each region of the MRI image are calculated. The next step is forming a set of calculated parameters for future ranking. It uses statistical characteristics described above for each region of the brain.

Previously received data set should be divided according to diagnosis (classes) of patients, with the presence of Alzheimer's disease and healthy. Then two statistical hypotheses are formulated: H0 — that the values of characteristic in the relevant region for MRI images of two classes are samples from same general population, and alternative H1 — that the values of characteristic in the region of MRI images of two classes come from two different general populations.

The data should be checked for normality for selecting independent evaluation criterion for binary classification. Depending on the obtained results we should use Student t-test if the data is normally distributed or Wilcoxon test if not.

IV. Experimental research of feature selection method

The data used in this study were obtained from ADNI database (adni.loni.usc.edu). ADNI database includes 1.5T and 3.0T MRI images. This database consists of images from 818 subjects (229 healthy patients (NOR), 401 patients with mild cognitive impairment (MCI) and 188 patients with Alzheimer's disease (AD)).

The statistical atlas IBASPM116 [13] is used in this work. Atlas label combines anatomical regions of the normalized spatial pattern, which corresponds to the dimension of the MRI images. Each voxel is characterized by an integer value in the range from 1 to 116, corresponding to one of the 116 regions, thus linking every voxel to the brain region of normalized MRI image.

All MRI images are three-dimensional, they were preliminary scaled, normalized to the area and divided into gray and white matter that allows to select exactly desired area according to the coordinates obtained using the atlas regions.

Using a set of SPM (Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk/spm/) tools, image intensity values were recorded in the three-dimensional matrix, according to the standard practice in MRI image preprocessing [8][14][15]. The next step for each MRI image was calculating six statistical parameters in each of the 116 areas of the cerebral cortex according to the atlas.

Three groups of data were created for the ranking during binary classification: 1) AD and MCI; 2) AD and NOR; 3) MCI and NOR. In each group, the initial matrices with calculated statistical parameters were combined into one. The distribution of data in each group is found to be not normal, therefore it is decided to use the U-Mann-Whitney test as an independent evaluation criterion for binary classification.

As a result, ranked statistical parameters linked to the anatomical region of the cerebral cortex were received in each group. Ten most important parameters for segmented images of gray and white matter are presented in Таблица 3 and Таблица 3 respectively.

V. Conclusion

The new approach to determining a set of the most significant features of MRI images for automated diagnosis of Alzheimer's disease based on the intensity of the image voxels is proposed in this paper. Using this method, one can get ranked statistical parameters linked to the anatomical region of gray and white matters of the brain. In the experimental study using ADNI database, ten most important parameters and regions for segmented images are identified, most of which are located in cingulum, hippocampus and temporal lobe. Obtained results can be used as input parameters of classification.

Table 1. **Ranged statistical parameters and regions of gray matter**

Wilcoxon absolute value	Parameter	Region	Wilcoxon absolute value	Parameter	Region	Wilcoxon absolute value	Parameter	Region
0.1752	Mean absolute deviation	Temporal Pole	0.1706	mean	Hippocampus	0.1432	root mean square	Insula
0.3836	standard deviation	Temporal Pole	0.0852	median	Cingulum	0.2308	skewness	Rectus
0.4098	root mean square	Temporal Sup	0.3669	root mean square	Insula	0.9359	mean	Hippocampus
0.2131	skewness	Heschl	0.4956	skewness	Rectus	1.0198	median	Cingulum
0.5730	median	Temporal Sup	0.0249	median	Cingulum	0.3723	mean	Hippocampus
0.9766	standard deviation	Temporal Pole	0.2736	mean	Hippocampus	0.4810	root mean square	Insula
0.2036	Mean absolute deviation	Temporal Pole	0.7594	mean	Amygdala	1.1273	skewness	Rectus
0.0728	mean	Temporal Inf	0.5046	root mean square	Cingulum	0.9648	median	Cingulum
0.2492	mean	Temporal Mid	0.0314	skewness	Cingulum	0.1622	mean	Amygdala
0.9212	median	Cerebelum	0.2625	mean	Amygdala	0.3849	root mean square	Cingulum

AD and MCI — spans first three data columns; **AD and NOR** — spans middle three; **MCI and NOR** — spans last three.

Table 2. Ranged statistical parameters and regions of white matter

AD and MCI			AD and NOR			MCI and NOR		
Wilcoxon absolute value	Parameter	Region	Wilcoxon absolute value	Parameter	Region	Wilcoxon absolute value	Parameter	Region
0.0973	Mean absolute deviation	Cingulum	0.5150	mean	Hippocampus	1.0558	mean	Hippocampus
0.0976	mean	Hippocampus	0.4826	mean absolute deviation	Cingulum	0.7120	standard deviation	Cingulum
0.4114	root mean square	Insula	0.3949	root mean square	Insula	1.0593	mean absolute deviation	Cingulum
0.7937	skewness	Rectus	0.3262	skewness	Rectus	1.3029	mean	Hippocampus
0.1446	standard deviation	Cingulum	0.0850	mean	Hippocampus	0.3112	mean absolute deviation	Hippocampus
0.0462	mean	Hippocampus	0.2229	mean absolute deviation	Cingulum	0.3089	root mean square	Cingulum
0.1640	Mean absolute deviation	Cingulum	0.5852	Standard deviation	Cingulum	1.3138	skewness	Insula
0.3517	root mean square	Insula	0.2226	root mean square	Insula	0.7459	standard deviation	Cingulum
0.3271	skewness	Rectus	0.1751	skewness	Rectus	0.0742	mean absolute deviation	Cingulum
0.3366	standard deviation	Cingulum	0.0485	Standard deviation	Cingulum	0.1591	root mean square	Cingulum

REFERENCES

[1] H. Tomaskova, "Prediction Of Population With Alzheimer's Disease In The European Union Using A System Dynamics Model," *Neuropsychiatric Disease And Treatment*, Vol. 12, P. 1589–1598, 2016.

[2] A. Burns, "Alzheimer's Disease," *Bmj*. No. 338, 2009.

[3] Ramachandran T.,"Alzheimer Disease Imaging," *Medscape*, 2016.

[4] N. Aggarwa And R. Agrawal, "Second Order Statistics Features For Classification Of Magnetic Resonance Brain Images," *Journal Of Signal And Information Processing*, Vol. 3, No. 2, Pp. 146-153, 2012

[5] S. Lahmiri And M. Boukadoum, "Automatic Brain Mr Images Diagnosis Based On Edge Fractal Dimension And Spectral Energy Signature," In *Annual International Conference Of The Ieee Engineering In Medicine And Biology Society*, 2012.

[6] Y. Zhang, Z. Dong, L. Wu And S. Wang, "A Hybrid Method For Mri Brain Image Classification," *Expert Systems With Applications*, Vol 38, P. 10049-10053, 2011.

[7] I. Despotović, B. Goossens And W. Philips, "Mri Segmentation Of The Human Brain: Challenges, Methods, And Applications," *Computational And Mathematical Methods In Medicine*, P. 23, 2015.

[8] I. Krashenyi, J. Ramírez, A. Popov And J. M. Górriz, "Fuzzy Computer-Aided Alzheimer's Disease Diagnosis Based On Mri Data," *Current Alzheimer Research*, Vol. 13, No. 5, Pp. 545-556, 2016.

[9] A. H. Vedmid, S. V. Mashtalir And E. S. Sakalo, "Restoration Of Images Using Analysis Of Principal And Independent Component," *Information Processing Systems*, Vol. 6, Pp. 66-72, 2010.

[10] O. Panichev, "Methods Of Eeg Analysis For Prediction Of Epileptic Seizures," *Electronics And Communications: Scientific And Technical Journal*, Vol. 20, No. 3, P. 68–77, 2015.

[11] D. Domashenko, M. Manko, A. Popov, I. Krashenyi, J. Ramírez And J. M. Górriz, "Feature Ranking For Mild Cognitive Impairment And Alzheimer's Disease Diagnosis," In *Signal Processing Symposium (Spsympo)*, Jachranka, 2017.

[12] I. Krashenyi, A. Popov, J. Ramírez And J. M. Górriz, "Fuzzy Computer-Aided Diagnosis Of Alzheimer's Disease Using Mri And Pet Statistical Features," In *Ieee 36th International Conference On Electronics Andnanotechnology (Elnano)*, 2016.

[13] Y. Alemán-Gómez, L. Melie-García And P. Valdes-Hernandez, "Ibaspm: Toolbox For Automatic Parcellation Of Brain Structures," In *12th Annual Meeting Of The Organization For Human Brain Mapping*, 2006.

[14] I. A. Illan, J. M. Górriz, J. Ramírez And A. Meyer-Base, "Spatial Component Analysis Of Mri Data For Alzheimer's Disease Diagnosis: A Bayesian Network Approach," *Frontiers In Computational Neuroscience*, Vol. 8:156, 2014.

[15] A. R. Hidalgo-Muñoz, J. Ramirez, J. M. Górriz And P. Padilla, "Regions Of Interest Computed By Svm Wrapped Method For Alzheimer's Disease Examination From Segmented Mri," *Frontiers In Aging Neuroscience*, Vol. 6, 2014.

РАНЖУВАННЯ СТАТИСТИЧНИХ ОЗНАК ДЛЯ ДІАГНОСТИКИ ХВОРОБИ АЛЬЦГЕЙМЕРА

Домашенко Д. В., студент, 0000-0003-4937-7151
dmytrodomashenko@gmail.com
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Київ, Україна
Манько М., студент, 0000-0001-6851-943X
max.manko2595@gmail.com
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Київ, Україна
Ciklum, Київ, Україна
Попов А., к.т.н., доцент, 0000-0002-1194-4424
popov.kpi@gmail.com
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Київ, Україна
Ciklum, Київ, Україна
Крашений І., к.т.н, 0000-0003-0424-147X
igor.krashenyi@gmail.com
Ciklum, Київ, Україна
Університет Гранади, Гранада, Іспанія
Рамірез Х., доктор, 0000-0002-6229-2921
javierrp@ugr.es
Університет Гранади, Гранада
Іспанія

Реферат — Стаття присвячена автоматичному прогнозуванню хвороби Альцгеймера та методам вивчення та вибору найбільш значущих ознак зображень МРТ. Використовуючи алгоритм вивчення статистичних характеристик зображень МРТ за допомогою атласу анатомічних областей головного мозку, було розраховано шість статистичних ознак (середнє, середнє абсолютне відхилення, медіана, стандартне відхилення, середнє квадратичне, коефіцієнт асиметрії) для сегментованих зображень білої та сірої речовини мозку. Запропоновано новий підхід до ранжування ознак за критерієм Вілкоксона для бінарної класифікації. В результаті отриманий ранжований список ознак, пов'язаних з анатомічними областями головного мозку для кожної групи за діагнозам. Серед найбільш описових особливостей для діагностики хвороби Альцгеймера є значення середнього арифметичного в гіпокампі, середнє абсолютне відхилення в зоні поясу, середньоквадратичне в острівцевій корі.

Ключові слова — хвороба Альцгеймера, МРТ, діагностика

РАНЖИРОВАНИЕ СТАТИСТИЧЕСКИХ ПРИЗНАКОВ ДЛЯ ДИАГНОСТИКИ БОЛЕЗНИ АЛЬЦГЕЙМЕРА

Домашенко Д. В., студент, 0000-0003-4937-7151
dmytrodomashenko@gmail.com
Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского»
Киев, Украина

Манько М., студент, 0000-0001-6851-943X
max.manko2595@gmail.com
Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского»
Киев, Украина
Ciklum
Киев, Украина

Попов А., к.т.н., доцент, 0000-0002-1194-4424
popov.kpi@gmail.com
Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского»
Киев, Украина
Ciklum
Киев, Украина

Крашеный И., к.т.н, 0000-0003-0424-147X
igor.krashenyi@gmail.com
Ciklum
Киев, Украина
Университет Гранады
Гранада, Испания

Рамирез Х., доктор, 0000-0002-6229-2921
javierrp@ugr.es
Университет Гранады
Гранада, Испания

Реферат — Эта статья посвящена автоматическому прогнозированию болезни Альцгеймера и методам извлечения и отбора наиболее значимых признаков изображений МРТ. Используя алгоритм извлечения статистических характеристик изображений МРТ с помощью атласа анатомических областей головного мозга, были рассчитаны шесть статистических признаков (среднее, среднее абсолютное отклонение, медиана, стандартное отклонение, среднее квадратическое, коэффициент асимметрии) для сегментированных изображений белого и серого вещества мозга. Предложен новый подход к ранжирование признаков по критерию Уилкоксона для бинарной классификации. В результате был получен ранжированный список признаков, связанных с анатомическими областями головного мозга для каждой группы по диагнозу. Среди наиболее описательных особенностей для диагностики болезни Альцгеймера является значение среднего арифметического в гиппокампе, среднее абсолютное отклонение в зоне пояса, среднеквадратическое в островковой коре.

Ключевые слова – болезнь Альцгеймера, МРТ, диагностика.