METHODS FOR IMPROVING ACCURACY OF THE DEMENTIA DIAGNOSIS USING FEATURE DIMENSION REDUCTION

MARYAM NADERAN, YURIY ZAYCHENKO

Abstract. In this paper, the problem of choosing the right feature for diagnosing Dementia is discussed. Several features that could affect dementia were reviewed and their importance was evaluated. Random forest algorithm and SVM for the dementia diagnosis have been developed and investigated. Experiments were conducted on the open-source database and compared with the related works' results. The purpose of the paper is to improve the accuracy of diagnosis of dementia using the reduction of features' dimension. This article is devoted to analysis of the main distinguishing features of Alzheimer`s dementia, applicable methods and treatment of Alzheimer`s dementia on early stage that could help to avoid negative consequences connected with progress of the disease. The purpose of the paper is to improve the accuracy of diagnosis of dementia.

Keywords: diagnosis Alzheimer's disease, ensemble learning methods, classification, Convolutional Neural Network.

INTRODUCTION

This article is devoted to analysis of the main distinguishing features of Alzheimer's dementia and applicable methods, main elements of successful diagnostics and treatment of Alzheimer's dementia on early stage, and review of the Alzheimer's association experience, that could help to avoid negative consequences connected with progress of the disease. In addition, it's necessary to find important features which play the big role in Alzheimer's disease diagnostics using by Random Forest Algorithms and Support Vector Machine (SVM) Algorithm. Distinguishing features of Alzheimer's dementia could be found using physical & neurological exam, mental status tests, computerized tests, screening, genetic tests and etc. For this research, Kaggle's data about patients Nondemented and Demented was used.

General features of Alzheimer's disease were described by foreign scientists S. Sarraf, G. Tofighi and J. Anderson in publication devoted to classification of the considered features using convolutional neural networks, that contain results necessary for distinguishing classification criteria and significant features taken into consideration for the patients with the mentioned disease. Authors stated, that "early detection and classification of Alzheimer's disease are critical for proper treatment and preventing brain tissue damage. Alzheimer's disease has a certain progressive pattern of brain tissue damage. It shrinks the hippocampus and cerebral cortex of the brain and enlarges the ventricles [6]. The considered general characteristics could be considered as common for many other mental diseases, but the underlined damage to the brain tissue and peculiarities of negative impact on the ventricles provide the grounds for differentiation at site.

Approaches as to the stages of Alzheimer's disease. Modern Ukrainian and foreign scholars have different views as to the issue of the main stages of Alzheimer's dementia, for example, in Ukrainian practice we found the approach providing differentiation into four main stages: prior dementia, early dementia, moderated dementia and final dementia, according to the classification provided by Ukrainian Clinics of modern neurology "Aximed".

RELATED WORK

Special attention should be paid to modern steps in differentiation of special types of Alzheimer's dementia revealed by L.V. Zhdaneeva in PhD thesis devoted to the issues of "Movement disorders of Alzheimer's disease" [1]. Author substantiated important statistical data, regarding internal differentiation of special features, common to patients with Alzheimer's dementia: "87,7 % of patients with Alzheimer's disease had movement disorders, including: 18,9 % — a kinetic-rigid syndrome; 27,8 % — hyperkinesias syndrome without rigidity; 8,8 % — tremor; 12,2 % — walking and postural stability disorder syndrome; 76,6 — pseudo bulbar syndrome; 45,6 % — stereotypes; 12,2 % — pyramidal syndrome; 6,7 % — epileptic seizures" [1, p. 23]. These data show strong correlation between general and special features, disorders common not only to patients with Alzheimer's dementia, that could be used for further scientific research and practical clinical activities, as the obtained data represent a wide range of patients with Alzheimer's disease and provide solutions for facilitation of differentiation of patient's status.

Modern research results in the field show, that the leading Harvard team was "the first to try to combine fMRI scans and deep learning into a program that could predict an MCI patient's chance of developing Alzheimer's disease. The fMRI scans used in their analysis were taken when patients were at rest. As with any fMRI scans, they reveal where electrical signals are flashing in the brain and how these areas relate to one another". But one of the latest publication in IEEE underlined, that there is no such a method found yet to provide 90 and more percent in recognition of Alzheimer's dementia. Such a conclusion was made by Dinggang Shen, famous scientist from University of North Carolina: "Nobody in the field can get from 80 to 90 percent," he says. "That's impossible, just based on this method" [5].

Negative impact on patient's memory can also be considered as a general feature for differentiation of the status of Alzheimer's patients. Progress of the disease makes it impossible to realize mental function as previously, during active adult or youth periods, even daily activities cannot be fulfilled without toil, patients thus need special assistance and support, understanding of their needs and problems. Scientists consider Alzheimer's disease as a severe disorder, that has neurological nature and affects human brains in a special way described above.

The mentioned approach differs from the one provided by J. Islam and Y. Zhang in the above-mentioned publication, providing classification into three main stages of Alzheimer's disease: very mild, mild and moderate [4]. The distinction between the considered approaches lies in the attitude to the prior stage of the disease that often shows very close features to other common mental, psychological, general diseases, in this respect it is important to discover modern methods of early diagnostics and treatment of Alzheimer's dementia, as it could provide the best results for evaluation of patient's health.

COMPARATIVE ANALYSIS

In [2] was carried out comparative analysis about methods which are using in medical diagnosis. It was emphasized that Convolutional Neural Network's diagnosis ability is greater than endoscopists in general. In [3], authors have compared CNN and endoscopists. The results obtained show that the sensitivity, specificity, accuracy, and diagnostic time were 88,9%, 87,4%, 87,7%, and 194 s, respectively, for Convolutional Neural Network. The same indicators obtained for the 23 endoscopists were 79,0%, 83,2% and 82,4% respectively.

Among foreign scholars and researchers we should address to modern scientific articles of Jyoti Islam, Yanqing Zhang "An Ensemble of Deep Convolutional Neural Networks for Alzheimer's Disease Detection and Classification" authors substantiated "a novel Alzheimer's Disease detection and classification model using brain MRI data analysis, developed an ensemble of deep convolutional neural networks and demonstrated superior performance on the Open Access Series of Imaging Studies dataset" [4]. This aspect has special importance for domestic research and clinical practice as provides new knowledge and methods for development of early diagnostics of Alzheimer's dementia.

Modern computerized systems provide experts with detailed information on the nature of dementia, it's stage, or precisely define, a contrario, Nondemented status. For the past ten years significant development was achieved in using MRI, but modern foreign researchers insist on automated brain MRI results, providing at site segmentation and classification, comparison with other groups of patients, i.e. control groups collected in computerized memory. This could be an example of conclusions and recommendations of the 31-st Conference held in 2017 on Neural Information Processing Systems in USA, where it was proposed to use "handcrafted feature generation and extraction from the MRI data, improvement of machine learning models such as Support Vector Machine, Logistic regression model etc" [4, p. 2].

RESULTS

In related works, authors hadn't mentioned which attributes are more informative for classification. In this paper, three more important features for dementia classification are detected.

Table 1 illustrates some of the data which were used in this paper [7]. The origin of the data set consists several visits of each patients. We consider only those patients who visited for the first time, since during second and third visit, patients could use medicine to prevent degradation of the dementia.

T a b l e 1. Attributes with their definition

Attrib-utes	Subject ID	MRI ID	Group	Visit	MR Delay	M/F	Hand
Definition	Subject identification	MRI identification	Class	Visit followup	MR contrast	Gender	Dominant hand
Attributes	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
Definition	Education level	SES cognition test	MMSE cognition test	CDR cognition test	Estimated intracranial volume	Standardized brain volume	Atlas factor scaling

T a b l e 2. Data of some patients during medical examination

Group	Visit	MR Delay	M/F	Hand	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
Nondemented	1	0	M	R	87	14	2	27	0	1987	0,696	0,883
Demented	1	0	M	R	75	12		23	0,5	1678	0,736	1,046
Nondemented	1	0	F	R	88	18	3	28	0	1215	0,71	1,444
Nondemented	1	0	M	R	80	12	4	28	0	1689	0,712	1,039
Demented	1	0	M	R	71	16		28	0,5	1357	0,748	1,293
Nondemented	1	0	F	R	93	14	2	30	0	1272	0,698	1,380
Demented	1	0	M	R	68	12	2	27	0,5	1457	0,806	1,205
Demented	1	0	F	R	66	12	3	30	0,5	1447	0,769	1,213
Nondemented	1	0	F	R	78	16	2	29	0	1333	0,748	1,316
Nondemented	1	0	F	R	81	12	4	30	0	1230	0,715	1,427
Demented	1	0	M	R	76	16	3	21	0,5	1602	0,697	1,096
Demented	1	0	M	R	88	8	4	25	0,5	1651	0,66	1,063
Nondemented	1	0	M	R	80	12	3	29	0	1783	0,752	0,985
Converted	1	0	F	R	87	14	1	30	0	1406	0,715	1,248
Converted	1	0	M	R	80	20	1	29	0	1587	0,693	1,106

Alzheimer's disease is diagnosed based on the value of some key indicators. Data analysis in table 3 illustrates some features with the average of each patients (Demented and Non-Demented).

T a b l e 3. Average of each features that could cause the Alzheimer's disease

AVE	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
Female Demented	76,10714	12,89286	3	25,46429	0,625	1373,929	0,728786	1,281036
Female Non-Demented	75,68	15,22	2,38	29,34	0	1402,56	0,75072	1,26152
Male Demented	74,33333	14,30556	2,5625	25,22222	0,583333	1555,278	0,720972	1,135361
Male Non-Demented	74,86364	15,04545	2,5	28,86364	0	1656,364	0,735682	1,070909

Based on the table 3, it could be concluded that the average of education level for Demented people is less than Non-Demented. While average of the age for Demented patients is higher than Non-Demented (for both female and male).

All features were compared separately during training Random Forest Algorithm. After some experiments, Gender, Age and Education level were computed that are the most valuable attributes for classification. The coefficient of importance each features for forest was multiplied by 100 which is illustrated in fig. 1.

There were used 70% of data for training and 30% of them for test. Accuracy of the classification using by Random Forest Algorithms was 89,5%.

Whereas, accuracy of the classification with Support Vector Machine was 88%. The table 4 compares two algorithms result for classification.
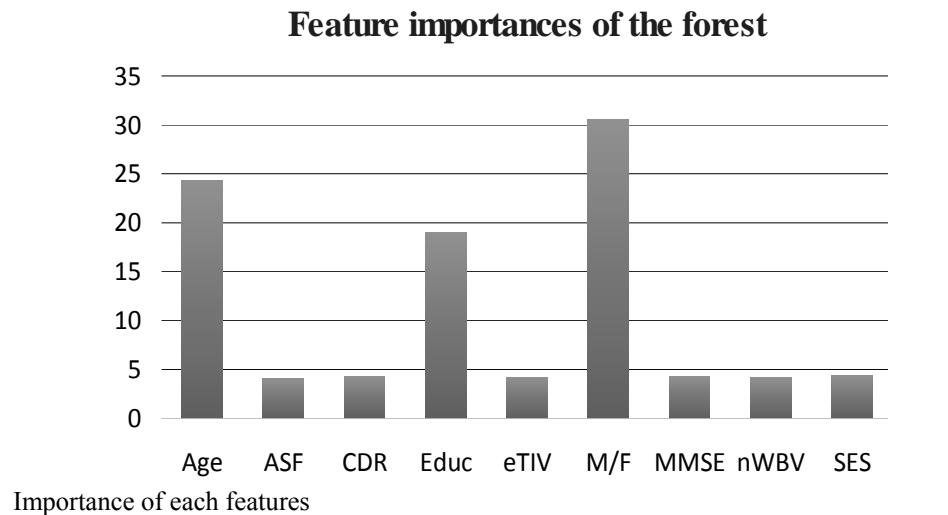
Feature importances of the forest

Importance of each features

T a b l e 4. Average accuracy of each algorithms

Classifier	Average Recall, %	Average Precision, %	Average F-Score, %
SVM	89	88	88
Random Forest	89	89	89

CONCLUSION

This article presents the results of known works analysis on diagnosis of Alzheimer's disease. The following conclusions should be made after analysis of considered works.

By using Random Forest Algorithms (RFA), we determine the most informative features like gender, age and education level which could mostly affect Alzheimer's disease.

In addition, Support vector machine and RF algorithms were compared with accuracy 88% and 89% respectively.

In the future research, it's planned to improve Convolutional Neural Network to reach the higher accuracy. Based on previous works, CNN provides high results in classification and has various benefits: like it's very good feature extractors. In addition, the results of CNN classification will be compared with the algorithms which were consider in the current paper.

REFERENCES

1. *Convolutional* Neural Network. 3 things you need to know. — P. 1–4. — Available at: https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html
2. *Альцгеймера* болезнь. Диагностика. Анализы и инструментальные исследования. — Режим доступа: http://demenciya.com

3. *Islam J.* An Ensemble of Deep Convolutional Neural Networks for Alzheimer's Disease Detection and Classification, Computer Vision and Pattern Recognition / J. Islam, Y. Zhang. — Available at: https://arxiv.org/pdf/1712.01675.pdf

4. *Sarraf S.* Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri / S. Sarraf, J. Anderson, G. Tofighi. — bioRxiv, p. 070441, 2016.

5. *Naderan M.* Diagnosing Lung Cancer Based on Deep Learning Algorithms: Review / M. Naderan, Y.P. Zaychenko // 20-th International conference on System Analysis and Information Technology SAIT 2018, May 21–24, 2018. — P. 111–112.

6. *Satoki S.* Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images / S. Satoki, N. Shuhei, A. Kazuharu, N. Yoshitaka et al. // EBioMedicine. — Vol. 25, November 2017. — P. 106–111.

7. *Boysen Jacob.* Magnetic Resonance Imaging Comparisons of Demented and Non-demented Adults / Jacob Boysen. — Available at: https://www.kaggle.com/jboysen/mri-and-alzheimers

From the Editorial Board: the article corresponds completely to submitted manuscript.