

Research article

Eye movement behavior identification for Alzheimer's disease diagnosis

Juan Biondi^{1,2,*}, Gerardo Fernandez¹, Silvia Castro^{1,2}, Osvaldo Agamennoni^{1,3}

¹Laboratorio de Desarrollo en Neurociencia Cognitiva, Instituto de Investigaciones en Ingeniería Eléctrica (IIIE), Departamento de Ingeniería Eléctrica y de Computadoras (DIEC), Universidad Nacional del Sur (UNS) - CONICET, San Andrés 800 – Bahía Blanca, 8000, Buenos Aires, Argentina

²Laboratorio de Visualización y Computación Gráfica (VyGLab), Departamento de Ciencias e Ingeniería de la Computación (DCIC), Universidad Nacional del Sur (UNS), San Andrés 800 – Bahía Blanca, 8000, Buenos Aires, Argentina

³Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC), San Andrés 800 – Bahía Blanca, 8000, Buenos Aires, Argentina

*Correspondence: juan.biondi@uns.edu.ar (Juan Biondi)

<https://doi.org/10.31083/j.jin.2018.04.0416>

Abstract

We develop a deep-learning approach to differentiate between the eye movement behavior of people with neurodegenerative diseases during reading compared to healthy control subjects. The subjects with and without Alzheimer's disease read well-defined and previously validated sentences including high- and low-predictable sentences, and proverbs. From these eye-tracking data trial-wise information is derived consisting of descriptors that capture the reading behavior of the subjects. With this information a set of *denoising sparse-autoencoders* are trained and a deep neural network is built using the trained *autoencoders* and a *softmax* classifier that identifies subjects with Alzheimer's disease with 89.78% accuracy. The results are very encouraging and show that such models promise to be helpful for understanding the dynamics of eye movement behavior and its relation with underlying neuropsychological processes.

Keywords

Eye-tracking; Deep-learning; Alzheimer's disease; neurodegenerative diseases; eye movement behavior; neuropsychological processes

Submitted: June 19, 2017; Accepted: January 9, 2018

1. Introduction

Alzheimer's disease (AD) is a nonreversible neurodegenerative disease characterized by progressive impairment of cognitive and memory functions. It develops over a period of years and is the most prevalent cause of dementia in elderly subjects. Initially, people experience memory loss and confusion, which may be mistaken for the kinds of memory changes that are sometimes associated with normal aging [1]. The subtle changes in behavior and response of the early manifestation of this disease make it difficult to diagnose with classical neuropsychological tests such as the Mini-Mental State Examination. The use of more advanced diagnostic tools such as magnetic resonance imaging (MRI) and positron emission tomography (PET) is critical for early diagnosis. Since AD is nonreversible, early treatment can improve a patient's life by delaying the full manifestation of the disease. In recent years, the study of eye movement during reading, known as *eye-tracking*, has proved helpful for performing this task [2–5].

Reading is a cognitive activity that has received considerable attention from researchers for the evaluation of human cognitive performance. This activity requires the integration of several central cognitive subsystems, from attention and oculomotor control to word identification and language comprehension. Eye movements show reproducible movement patterns during normal reading. Each eye movement ends up at a fixation point, which allows the brain to process incoming information and program the following saccade. Different neuropsychiatric pathologies produce abnormalities in eye movements and disturbances in reading each have a particular pattern

that can be recorded and measured [6–11]. Eye movements can be classified into three groups:

1. Movements for maintaining the image on the fovea (area of the retina with high acuity vision), compensating for head or object movement;
2. Movements for shifting the eyes when attention is changed from one object to another. There are subtypes of shifting movements: saccades (looking for a new center of visual attention), monitoring and vergence (slower than saccades and responsible for moving the image of interest to both foveae, thus enabling stereoscopic vision);
3. Movements of binocular fixation that also prevent fading of the image. These movements have three variations: tremor, drift, and microsaccade.

Saccades are large fast eye movements that are of particular importance as cognitive processes have a direct influence on such movements. Each saccade has a particular direction. People, depending on language, read from left to right and most saccadic eye movements are oriented accordingly. These reading movements are called forward saccades. Reading movements going from right to left are called regressions. Saccade movement alternates when a fixation is made as eyes are directed to a particular target (See for a review [12]). As shown in, patients with early Alzheimer disease show changes during the execution of tasks such as reading [5]. These alterations can be related to an impairment in working memory [3, 13]. It has been shown that it is possible to infer a diagnosis from this change in eye-movements [4].

The use of computer-aided diagnosis is a key challenge as the growth of computational power permits the creation of more complex models. These models can be used to create biomarkers that help in disease identification. Since the popularization of *deep-learning* neural networks [14, 15], much effort has been directed to their use in medicine. This technique is commonly used in conjunction with imaging diagnoses provided by PET or MRI mainly because the feature representation that this technology provides may help even when data is incomplete [16]. Specifically, there have been advances in the detection and pattern differentiation of physical brain alterations that neurodegenerative diseases produce, such as AD and mild cognitive impairment (MCI) [17, 18], as well as advances in its early diagnosis [19]. The problem is that, when a physical brain alteration is observable, brain damage is already both irreversible (even though the disease is in an early stage) and may be sufficient to cause deterioration in the quality of life for a patient. Eye-tracking techniques enable the discovery of more subtle changes made by the brain, thus to alleviate small memory deficits for a patient. These changes are not noticed by a patient, but small changes in the way they read our set of test sentences can be identified with the technique described here.

In this work, a deep-learning neural network is trained on reading habits, obtained from controls and patients with probable AD, to identify the patterns they make during the reading process and then separate them into their respective groups. Throughout this work AD patients and patients with probable AD are used interchangeably due to the nature of the AD diagnosis. The hypothesis was that using deep-learning for feature detection of key characteristics of a patient's eye behavior while reading sentences may lead to a classification suited to infer a diagnosis. Employing this type of technology may improve the results obtained (see [4]) as it provides a smaller granularity for detection of the disease and consequently, better performance. Additionally, this technology improves the effectiveness of classification as more "ground truth" subjects are collected.

2. Methods

2.1. Ethics Statement

All patients, their caregivers, and all control subjects signed informed consent forms prior to their inclusion in the investigation. The research adhered to the principles of the Declaration of Helsinki and was approved by the Institutional Bioethics Committee at Hospital Municipal de Agudos (Bahía Blanca, Buenos Aires, Argentina).

2.2. Participants and data

The group of readers consisted of sixty-nine subjects: twenty-six patients with mean age 69 years, standard deviation (SD) = 7.3 years, with the diagnosis of probable AD (recruited at Hospital Municipal of Bahía Blanca, Buenos Aires, Argentina) and forty-three healthy elderly adults, mean age 71 years, SD = 6.1 years, with no known neurological or psychiatric disease according to their medical records, and no evidence of cognitive decline or impairment in daily activities.

Physicians based their diagnoses on the criteria for dementia outlined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), the clinical criteria for diagnosis of early stage Alzheimer's disease is not settled [20]. All AD patients presented an APO E3E4 genotype and underwent a detailed clinical history, physical/neurological examination, thyroid function test, and brain imag-

ing tests (magnetic resonance/computerized tomography scans). Biochemical studies were made so as to discard other common pathologies (hemoglobin, full blood count, erythrocyte sedimentation rate, urea and electrolytes, blood glucose). Patients were excluded if diagnosed with any condition that might interfere with the current study, for example, if they suffered from any other medical conditions that could account for, or might interfere with, their cognitive decline; had evidence of vascular lesions in computed tomography or functional MRI; had evidence for an Axis I diagnosis (e.g. major depression or drug abuse) as defined by the DSM-IV. To be eligible for the study, patients had to have a caregiver providing regular care and support. Patients taking cholinesterase inhibitors were not included. None of the subjects were taking sedative drugs, major tranquilizers or hypnotics.

A one-way ANOVA showed no significant differences between AD and *Control* subject ages. Patients diagnosed with ophthalmologic diseases, such as glaucoma, visually significant cataract or macular degeneration as well as those with visual acuity less than 20/20, were excluded from the study.

The mean scores of *Controls* and AD subjects in the Mini-Mental State Examination (MMSE) [21] were 27.8 (SD = 1.0) and 24.2 (SD = 0.8), respectively, the latter suggesting early mental impairment. A one-way ANOVA suggested significant differences between AD subjects and *Controls* for MMSE ($p < 0.001$). The mean score of AD subjects in the Adenbrook's Cognitive Examination - Revised [22] was 84.4, (SD = 1.1), the cut-off being 86. The mean duration of school education for AD patients was 15.2 years (SD = 1.3) and for the *Control* group 15.1 years (SD = 1.0), a one-way ANOVA showed no significant differences. For a description of the sentence corpus (see [23]). For Apparatus technical specifications and procedures (see [7]).

Eye movement data from 69 participants reading 184 sentences resulted in a total of 48,716 fixations – 13,002 for *Control* and 35,714 for AD subjects. This data was deblinked and tracking errors were removed.

2.3. Information used

The information used for this study was a *trial-wise* compaction of the original data that comprised individual subject descriptors of reading behavior for each sentence read. Data measures included: the mean and SD of saccade amplitude, fixation duration and fixation duration of single words by the subject during the reading of each sentence. Additionally, the total number of fixations were recorded and classified as first pass fixations, refixations, unique fixations, and total fixations:

1. First pass fixations: The first fixation on a specific word of a sentence.
2. Unique fixations: Fixations that occurred once in a word that was skipped during the first pass.
3. Multiple fixations: Multiple fixations on a word during the first pass.
4. Refixations: Fixations that occurred once a word already had a *first pass fixation* or a *unique fixation*, implying a regression.

Categorical diagnostic data (used as training labels) were replaced by numerical values to enable data type pooling to improve the classification procedure. Two integer values were employed for construction of diagnostic information: 0 for "*Control*" and 1 for "*AD*". Subject identification and diagnostic information were sepa-

rated from the data. Variables used as input for model construction are given in Table 1. As the tag (AD or *Control*) was associated with a patient and not each sentence, and, since a per-trial classification approach was employed, the subject's tag was applied to all the sentences they read. However, this procedure may introduce noise at the classifying stage, as in a per-trial classification approach, a *Control* subject could, for example, be distracted during the reading of a specific sentence, leading to data being misclassified as coming from a non-healthy person. Nevertheless, it was assumed the system should detect and ignore such artifacts due to the number of samples used at the training stage.

Table 1. Variables employed for model construction

Name	Description
nw	Number of words in the sentence.
gaze	Global (sentence) mean of the sum of fixation durations on the same word.
sd_gaze	Standard deviation of gaze.
as	Mean saccade amplitude in the sentence.
sd_as	Standard deviation of as
ntf	Count of the total number of fixations on the sentence.
ntm	Count of the number of multifixations on the sentence.
dfp	Mean duration of the first pass fixations on the sentence.
sd_dfp	Standard deviation of dfp.
fpp	Count of the number of first pass fixations on the sentence.
rf	Count of refixations on the sentence.
nfu	Count of unique fixations on the sentence.
dfu	Mean duration of unique fixations on the sentence.
sd_dfu	Standard deviation of dfu.

All data were outlier-checked by a dropout policy to reduce dataset noise. Outlier checking proceeded by finding the group mean and SD of each condition and checking the two groups separately. All *trials* with a SD greater than twice the SD of the group were considered outliers and were dropped from the analysis. This eliminated 10% of the samples. The sentence identification, order, and type were kept separate from the training information. This was done because the SDs in the data from the AD patients for proverbs and high predictability sentences appeared to be particularly high after the data were outlier-checked. Thus generating highly unbalanced datasets.

The resulting dataset consisted of 3, 235 *trials* with a mean of 46.88 (SD = 1.76) *trials* per subject. The dataset was divided into two groups: one for network training composed of data from 61 subjects and other for data testing with eight randomly selected subjects. Finally, the training dataset consisted of 2, 922 *trials* for 61 subjects, i.e. 39 *Control* and 22 AD subjects each with a mean number of 47.9 (SD = 10.47) *trials*; the test dataset consisted of 313 *trials* of 8 subjects, i.e. 4 *Control* and 4 AD subjects each with a mean number of 39.12 (SD = 10.37) *trials*.

Partitioning data in this way ensures the network can not infer a condition in another way, avoids over-fitting, and ensures that the testing data is totally unknown by the network. All data was normalized between 0 and 1 for processing by the neural network.

2.4. Deep learning with denoising sparse-autoencoders

In this study, *sparse-autoencoders* were used at the codification stage. The *sparse-autoencoders* function as regular *autoencoders*, i.e. they are neural networks under supervised learning with their

targets set equal to the input (identity) values. In the case of *sparse-autoencoders*, an average number of activations per neuron restriction was applied in the hidden layer by penalizing the average number of activations different from the desired number (known as sparsity proportion) by adding a penalty term to the cost function. This restriction is introduced so that each neuron specializes on a particular feature. The lower the sparsity proportion, the more specific the feature. The resulting trained neural network can be thought of as: an encoder, involving the input and the hidden layer, and a decoder, involving the hidden and the output layer. In this study activation restriction was set to 10%.

In a *denoising-autoencoder*, the idea is to force the hidden layer to discover more robust features and to prevent it from simply learning the identity, by training the *autoencoder* to reconstruct the input from a corrupt version. The altered version of the input is generated by introducing noise, which is obtained by clamping some fields to zero. The corrupt data is used as the *sparse-autoencoder* input, and the clean (unaltered) data as the target. Using this type of data corruption mechanism forces the network to learn a way of reconstructing a given field based on other fields. When combined with the sparsity restriction, this generates more robust features.

The deep-learning neural network was built using two stages of these *denoising sparse-autoencoders*. At each stage, the *autoencoder* was trained by corrupting clean encoded data obtained from the previous stage, and then employing it as input to the next stage. Following these two stages, a *softmax* layer was set as a classifier and trained with uncorrupted data and the corresponding tag. As a *per-trial* classification approach was used, the subject diagnosis was extended to all sentences they read, the classifier was trained with this data as the target. The *softmax* layer is a non-linear, multiclass generalization of binary *Logistic Regression*, and its output is the “probability” of each class (the word “probability” is quoted as its shape depends on the regularization used at the training stage, it can either be more diffuse or peaky).

3. Results

Several configurations were generated by varying the sparsity proportion, the number of units and layers, and the shape of the network (same vs. decreasing number of units between layers). The one that produced the best results was adopted. It consisted of two layers of *denoising sparse-autoencoders* with 16 and 4 hidden units using a sparsity proportion of 10% in each hidden layer. After the training of the network, a series of tests were performed with data not included in the training dataset. This test dataset consisted of 313 sentences from 8 subjects - 4 *Control* and 4 AD - with a mean number of 39.12 (SD = 10.37) *trials* for each subject. We used a *softmax* layer for classification training by using the condition translation of 0 for *Control* subjects and 1 for AD subjects. This meant there was a single AD class and, since the output of the classifier was a real number between 0 and 1, the read sentences classified by the network with values close to 0 had a low “probability” of being read by an AD patient (i.e. a high probability of being read by a *Control* patient) and vice-versa. The “ground truth” values were known, so the output could be split into groups and the number of misclassified sentences observed by the network. Consequently, the output of the network with values lower than 0.5 could be classified as *Control*, and higher values classified as AD. As can be seen, Fig. 1 shows the output of the network was consistent with the expected values.

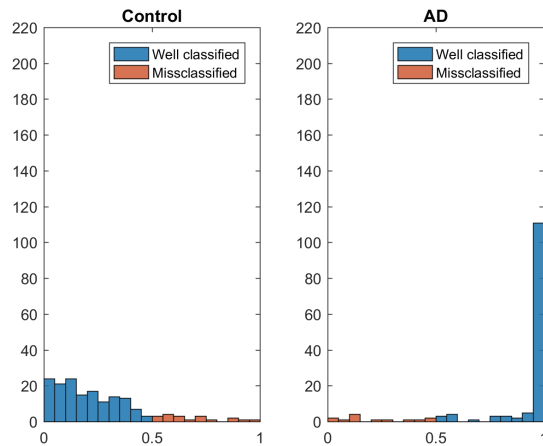


Fig. 1. Classification result histogram giving the number of sentences split by “ground truth” values. Values below 0.5 are classified as *Control*, and higher values are classified as AD.

Rounded values can be used to plot a confusion matrix and approximate the number of misclassified sentences so as to measure network performance. Fig. 2 shows a confusion matrix of the output. The abscissa represents the expected output values and the ordinate the rounded output of the network. Overall performance of the network was good with 89.8% of sentences well-classified. The performance of the network using sentences read by *Control* subjects (88.7% correct) was slightly less than the performance of AD subjects (91.0% correct).

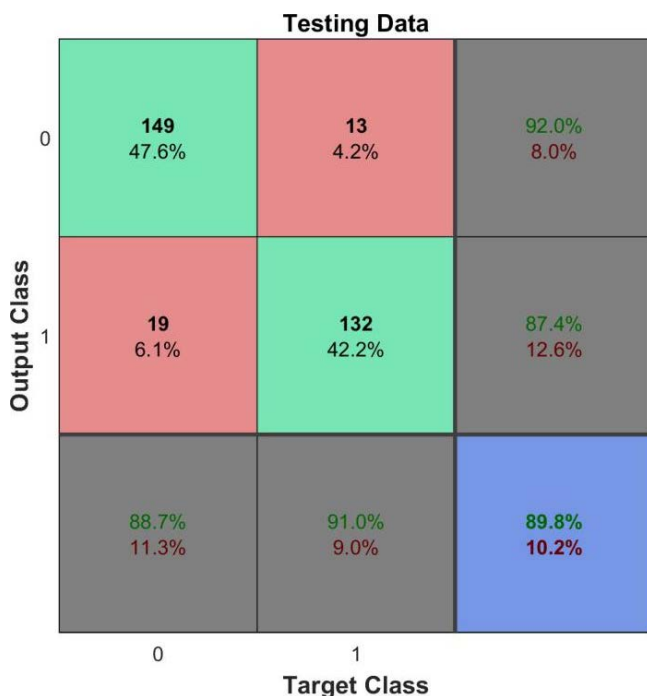


Fig. 2. Classification results. Values below 0.5 are considered classified as *Control* (class 0), and higher values are considered classified as AD (class 1).

Alternatively, Fig. 3 shows that no particular type of sentence

was preferentially misclassified. It gives the original concentration of sentences types in the testing dataset and the correctness of the classification following the given method.

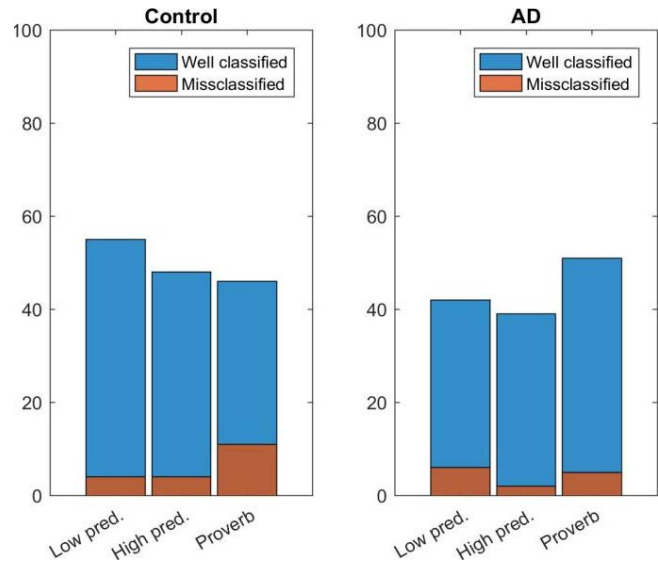


Fig. 3. Number of misclassified sentences by type, split by “ground truth” label.

This result, when combined with the fact that neither was a consequence of misclassification in a particular sentence, suggests that most of the misclassification was likely stochastic. Trained networks were additionally evaluated with a spread result test to determine the softness of the model. These tests checked if similar information is encoded in a similar way in subsequent stages of the network. A significant differentiation in later stages of encoding may show over-fitting either by the network and/or by different stages.

Two subsets of trials (one composed of AD subjects and the other of *Control* subjects) that have similar values for the input in each field are shown in Fig. 4. As seen, similar input values map to similar encoded values at each stage of the *autoencoders*. This is attributed to the smoothness of the modeled function. Furthermore, as data is processed through subsequent stages of codification, it tends to group. These results show that output information such as the encoding in the different stages of the network are reliable. Alternatively, they show that at later stages certain neurons tend to specialize in the detection of specific AD or *Control* input features.

4. Discussion

Results showed that using a deep-learning architecture for identifying the characteristic eye movement patterns associated with neurodegenerative diseases such as Alzheimer’s disease provides a good approach as this technology is focused on pattern finding and is suited to this work. Moreover, the high performance of a per-trial classification approach, leads to the conclusion that, since a single patient reads many sentences, the success rate per patient is higher than the 89.8% accuracy reported here. Assuming that network outputs higher than 0.5 can be classified as AD and lower outputs as *Control*, if each subject is tagged using “majority voting” over all sentences read, the network reaches a 100% classification accuracy for the testing set - 8 well classified subjects from 8 total. This was expected as, for this test set, the total number of misclassified sentences is 32

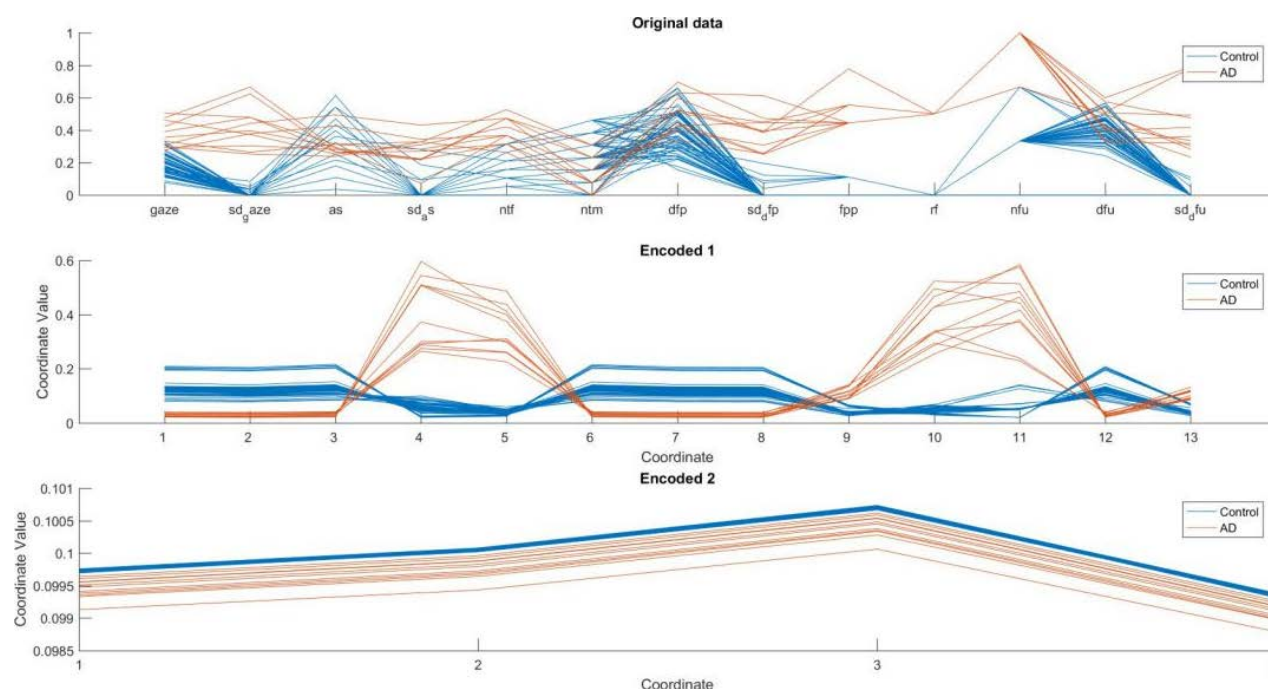


Fig. 4. Parallel coordinates plot with two subsets of *trials* (one composed of AD subjects and the other of *Control* subjects) that have similar values for the input in each field and its codification within the different stages of the network. As expected, similar values encoded “together”. *Control* subjects encoded closer than AD subjects; this could be attributed to the high “within group” variability of the AD group.

and every subject, after the data optimization, reads on average 39.12 (SD = 10.37) sentences.

Additionally, the psychiatrists managing other AD patients (that were not included in training process) were asked to score the overall severity of the disease of each patient with traditional tests using a scale from 0 to 1, not knowing the results given by our network. The process of creating this score required physicians to have a deep knowledge of both the psychiatric history and results of all neuropsychological tests for each patient. Table 2 shows the scores given by the psychiatrists compared with the mean value and their SDs obtained for the network model of all the sentences read by the subjects. As seen, for most of the patients, the values obtained were very similar to the scores given by the psychiatrists (mean value 0.19, SD = 0.15). These results show that the created marker was reasonably close to the psychiatric score but required a much simpler process.

Finding a better way to interpret the output of the classifier is left for future investigation. An improvement is required as values near 0.5 are identified as neither AD nor *Control* (equal “probabilities”). However, the strategy reported here is an initial approximation that doesn’t reflect the actual power of the network. Using a fuzzy-logic encoder to obtain the overall diagnosis of a patient might lead to more accurate results. Determination of whether the number given by the classifier is related to the severity of the disease is left for future improvement.

This task is particularly difficult as there are no “ground truth” measurements suited to corroborate the information obtained from current psychological testing methods. Although a per-trial classification approach was adopted, it is possible that overall diagnosis may be related to a measurement extracted from an entire test and not from a single *trial*. As reported here, even with the strategy used

Table 2. Comparison of mean diagnostic value given by the network and a “severity of disease” score given by psychiatrists.

ID Pat	Mean	SD	Score	Difference
58	0, 97	0, 17	0, 9	0, 07
57	0, 95	0, 17	0, 5	0, 45
66	0, 49	0, 32	0, 5	0, 01
60	0, 95	0, 16	0, 6	0, 35
56	0, 96	0, 16	0, 8	0, 16
55	0, 94	0, 17	0, 7	0, 24
63	0, 87	0, 25	0, 8	0, 07
64	0, 51	0, 34	0, 5	0, 01
70	0, 90	0, 24	0, 5	0, 40
69	0, 84	0, 25	0, 5	0, 34
65	0, 91	0, 22	0, 6	0, 31
59	0, 58	0, 36	0, 6	0, 02
71	0, 75	0, 33	0, 6	0, 15
62	0, 76	0, 32	0, 6	0, 16
67	0, 47	0, 35	0, 5	0, 03
68	0, 40	0, 31	0, 8	0, 40
53	0, 84	0, 31	0, 8	0, 04
			Mean	0, 19
			SD	0, 15

in this study, simply using the mean of the scores or the “controlling” label was sufficient for this network to behave as expected.

5. Conclusions and future work

In this paper we showed that the Deep Learning approach is a good alternative for the identification of eye movement patterns because this technology is specifically aimed at finding rich features in com-

plex data and using them later. We have also seen that this technology needs large amounts of tagged data, and in some cases, this could be a limitation. However, we hope that this work will serve as a stimulus for neuroscientists to start thinking about the use of Deep Learning for this type of tasks.

An important and very interesting challenge that remains as future work is to obtain a better way of interpreting the result of the classifier. This would probably make it possible to determine whether the number given by the classifier is related to the severity of the disease.

Acknowledgments

This work was supported by the Agencia Nacional de Promoción Científica y Tecnológica from Argentina under grant PICT 2013 0403, by the Universidad Nacional del Sur under grants PGI 24/K062 and PGI 24/N037 and by Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires.

Conflict of Interest

All authors declare no conflict of interest.

References

- [1] Waldemar G, Dubois B, Emre M, Georges J, McKeith I, Rossor M, Scheltens P, Tariska P, Winblad B (2007) Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. *European Journal of Neurology* **14**(1), e1-e26.
- [2] Fernández G, Schumacher M, Castro L, Orozco D, Agamennoni O (2015) Patients with mild Alzheimer's disease produced shorter outgoing saccades when reading sentences. *Psychiatry Research* **229**(1, 2), 470-478.
- [3] Fernández G, Manes F, Politi LE, Orozco D, Schumacher M, Castro L, Agamennoni O, Rotstein NP (2016) Patients with mild Alzheimer's disease fail when using their working memory: evidence from the eye tracking technique. *Journal of Alzheimer's Disease* **50**(3), 827-838.
- [4] Fernández G, Castro LR, Schumacher M, Agamennoni OE (2015) Diagnosis of mild Alzheimer disease through the analysis of eye movements during reading. *Journal of Integrative Neuroscience* **14**(1), 121-133.
- [5] Fernández G, Mandolesi P, Rotstein NP, Colombo O, Agamennoni O, Politi LE (2013) Eye movement alterations during reading in patients with early Alzheimer disease. *Investigative Ophthalmology & Visual Science* **54**(13), 8345-8352.
- [6] Fernández G, Sapognikoff M, Guinjoan S, Orozco D, Agamennoni O (2016) Word processing during reading sentences in patients with schizophrenia: evidences from the eyetracking technique. *Comprehensive Psychiatry* **68**, 193-200.
- [7] Fernández G, Guinjoan S, Sapognikoff M, Orozco D, Agamennoni O (2016) Contextual predictability enhances reading performance in patients with schizophrenia. *Psychiatry Research* **241**, 333-339.
- [8] Holzman PS, Proctor LR, Levy DL, Yasillo NJ, Meltzer HY, Hurt SW (1974) Eye-tracking dysfunctions in schizophrenic patients and their relatives. *Archives of General Psychiatry* **31**(2), 143-151.
- [9] Iacono WG, Moreau M, Beiser M, Fleming JA, Lin TY (1992) Smooth-pursuit eye tracking in first-episode psychotic patients and their relatives. *Journal of Abnormal Psychology* **101**(1), 104-116.
- [10] Riby D, Hancock PJ (2009) Looking at movies and cartoons: eye-tracking evidence from Williams syndrome and autism. *Journal of Intellectual Disability Research* **53**(2), 169-181.
- [11] Kellough JL, Beevers CG, Ellis AJ, Wells TT (2008) Time course of selective attention in clinically depressed young adults: an eye tracking study. *Behaviour Research and Therapy* **46**(11), 1238-1243.
- [12] Rayner K (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* **124**(3), 372-422.
- [13] Fernández G, Laubrock J, Mandolesi P, Colombo O, Agamennoni O (2014) Registering eye movements during reading in Alzheimer's disease: difficulties in predicting upcoming words. *Journal of Clinical and Experimental Neuropsychology* **36**(3), 302-316.
- [14] Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Networks* **61**, 85-117.
- [15] Deng L, Yu D (2014) Deep learning: methods and applications. *Foundations and Trends in Signal Processing* **7**(3, 4), 197-387.
- [16] Li RJ, Zhang WL, Suk HI, Wang L, Li J, Shen DG, Ji SW (2014) Deep learning based imaging data completion for improved brain disease diagnosis. In: Li RJ et al. (Eds.) *Medical Image Computing and Computer-Assisted Intervention* (pp. 305-312). Boston, Springer.
- [17] Suk HI, Shen D (2013) Deep learning-based feature representation for AD/MCI classification. *Medical Image Computing and Computer-assisted Intervention* **16**(02), 583-590.
- [18] Suk HI, Lee SW, Shen D, Initiative AsDN (2014) Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* **101**, 569-582.
- [19] Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D (2014) Early diagnosis of Alzheimer's disease with deep learning. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). Beijing, China.
- [20] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**(7), 939-944.
- [21] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* **12**(3), 189-198.
- [22] Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR (2006) The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences* **21**(11), 1078-1085.
- [23] Fernández G, Shalom DE, Kliegl R, Sigman M (2014) Eye movements during reading proverbs and regular sentences: the incoming word predictability effect. *Language, Cognition and Neuroscience* **29**(3), 260-273.