NeuroImage: Clinical 11 (2016) 508–514

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia

P. Tamboer*, H.C.M. Vorst, S. Ghebreab, H.S. Scholte

University of Amsterdam, Faculty of Social and Behavioural Sciences, Weesperplein 4, 1018XA Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 28 December 2015
Received in revised form 8 March 2016
Accepted 17 March 2016
Available online 29 March 2016

Keywords:
Dyslexia
MRI
SVM classification
Gray matter
VWFA

ABSTRACT

Meta-analytic studies suggest that dyslexia is characterized by subtle and spatially distributed variations in brain anatomy, although many variations failed to be significant after corrections of multiple comparisons. To circumvent issues of significance which are characteristic for conventional analysis techniques, and to provide predictive value, we applied a machine learning technique – support vector machine – to differentiate between subjects with and without dyslexia.

In a sample of 22 students with dyslexia (20 women) and 27 students without dyslexia (25 women) (18–21 years), a classification performance of 80% ($p < 0.001$; d-prime = 1.67) was achieved on the basis of differences in gray matter (sensitivity 82%, specificity 78%). The voxels that were most reliable for classification were found in the left occipital fusiform gyrus (LOFG), in the right occipital fusiform gyrus (ROFG), and in the left inferior parietal lobule (LIPL). Additionally, we found that classification certainty (e.g. the percentage of times a subject was correctly classified) correlated with severity of dyslexia (r = 0.47). Furthermore, various significant correlations were found between the three anatomical regions and behavioural measures of spelling, phonology and whole-word-reading. No correlations were found with behavioural measures of short-term memory and visual/attentional confusion. These data indicate that the LOFG, ROFG and the LIPL are neuro-endophenotype and potentially biomarkers for types of dyslexia related to reading, spelling and phonology.

In a second and independent sample of 876 young adults of a general population, the trained classifier of the first sample was tested, resulting in a classification performance of 59% ($p = 0.07$; d-prime = 0.65). This decline in classification performance resulted from a large percentage of false alarms. This study provided support for the use of machine learning in anatomical brain imaging.

1. Introduction

Dyslexia is usually defined as a specific reading disorder characterized by a specific and significant impairment in the development of reading skills that are unrelated to problems with visual acuity, schooling or overall mental development (World Health Organisation, 2010). For (sub-) groups of dyslexics, reading difficulties have been related to various symptoms of which the most frequently reported are related to phonological difficulties although, in recent years, visual/attentional deficits are reported frequently as well (e.g. Ramus and Ahissar, 2012). Generally, it is assumed that early learning delays cannot be overcome completely despite remedial teaching programs, and that these learning delays interfere with academic achievement into adulthood for most of the dyslexics which are estimated to represent 5% to 15% of the population.

Reliable diagnoses can currently only be determined behaviourally and after some years of education, when the discrepancy between normal cognitive and reading abilities becomes visible. Alternatively, researchers have been searching for biomarkers of dyslexia using MRI or fMRI. Meta-analyses showed that these differences do exist (Richlan et al., 2011, 2012; Vandermosten et al., 2012), although many findings failed to be significant after corrections of multiple comparisons.

A potentially more powerful technique than the univariate voxel-wise evaluation and correction of multiple comparisons are multivariate classification techniques from machine learning. This technique has recently successfully been applied in several clinical neuroimaging studies. For instance, a high accuracy rate of 90% has been reported for discriminating major depressive disorder and controls (Mwangi et al., 2012). An accuracy rate of 81% has been found for autism (Ecker et al., 2010).

Also with regard to dyslexia, this classification approach has been applied. In a study of Hoeft et al. (2011), a multivariate pattern analysis of brain activation during a reading task over the whole brain using linear

Abbreviations: SVM, support vector machine; VBM, voxel-based morphometry; GM, gray matter; VWFA, visual word form area; LIPL, left inferior parietal lobule; LOFG, left occipital fusiform gyrus; ROFG, right occipital fusiform gyrus.

* Corresponding author at: Overtoom 247B, 1054HW Amsterdam, The Netherlands.
E-mail addresses: petertamboer@kpnmail.nl (P. Tamboer), H.C.M.Vorst@uva.nl (H.C.M. Vorst), S.Ghebreab@uva.nl (S. Ghebreab), H.S.Scholte@uva.nl (H.S. Scholte).

http://dx.doi.org/10.1016/j.nicl.2016.03.014
2213-1582/

support vector machine and cross-validation, showed that reading gains over a 2.5 year period in children with dyslexia can be predicted with >90% classification accuracy. A study of Tanaka et al. (2011) showed that in two samples of typical and poor reading children 79% and 80% were classified correctly using leave-one-out linear SVM analyses of brain activation during phonological processing. In a study of Pernet et al. (2009), classification of dyslexic readers brains resulted in dyslexics falling outside the 95% confidence boundaries of the controls in two areas (the right cerebellar declive and the right lentiform nucleus).

The aim of this study is to investigate whether young adults with and without dyslexia can reliably be classified based on anatomical differences. We examined neuro-anatomical networks involved in dyslexia using a whole-brain classification employing SVM and cross-validation. We used the T_1-weighted magnetic resonance images of GM structure of a sample of 22 students with dyslexia and 27 students without dyslexia for acquiring a trained classifier. Next, we determined which voxels were involved with the correct classification. Furthermore, we explored to what degree these results can be used to investigate the relation between different cognitive aspects of dyslexia and neural substrates. We also tested the reliability of the trained classifier in an independent sample of 876 young adults.

2. Materials and methods

2.1. Subjects & procedure

The first sample – used to find a trained classifier – consisted of 22 students with dyslexia (20 women; 4 left-handed; mean age 20.7 years, SD 1.8 years) and 27 students without dyslexia (25 women; 4 left-handed; mean age 20.3 years, SD 0.9 years). All participating subjects were first-year psychology students, native Dutch speakers, had at least twelve years of school education, were free from medical or psychiatric diseases and had no history of sensory deficits or head trauma. None of the participants had a diagnosis of ADHD. Handedness was assessed with a short self-report questionnaire, which included questions about writing hand, general hand preference, as well as 20 specific questions. There were no students with inconsistent reports which could indicate being ambidextrous.

The 49 students of the first sample were invited to participate in the present study by mail and telephone. The students gave informed written consent and were debriefed afterwards. All participants had the option to choose between acquiring participation points required for the first year of study, or a financial reward. This study was approved by the ethics committee at the University of Amsterdam.

The second sample – used to test the trained classifier – consisted of young adults of a general population. Brain data of this sample were available for various studies at the University of Amsterdam. We excluded participants with a serious medical condition, with a diagnosis of autism spectrum disorder and participants using psychiatric drugs or psychiatric medicine. The remaining sample consisted of 876 subjects who were native Dutch speakers and who had at least twelve years of school education. Of this sample, 60 (7%) subjects (27 women; mean age 22.5 years, SD 1.6 years) were diagnosed with dyslexia whilst attending school, and 816 subjects (433 women; mean age 22.9 years, SD 1.7 years) had no reported history of dyslexia.

2.2. Neuropsychological Assessment

The first sample was acquired from a sample of 480 students who participated in a previous study (Tamboer et al., 2014a). In that study, dyslexia and non-dyslexia was assessed using three sources of information: (1) a history of language difficulties, (2) a self-report of language difficulties, and (3) a test-battery measuring numerous abilities such as spelling, reading, pseudoword reading, phonology, attention, and short-term memory. Severity of dyslexia was determined with a regression formula which consisted of 13 test items and 10 self-report questions, and which classified all subjects with and without dyslexia correctly. In a follow-up study (Tamboer et al., 2014b), five behavioural factors accompanying dyslexia were determined using exploratory and confirmatory factor analyses. On the basis of these analyses we acquired five Z-transformed sum scores: spelling, phonology, short-term memory, visual/attentional confusion, and whole-word reading.

We assumed that intelligence of all participants was within the normal range because all had finished the highest level of secondary school education in the Netherlands. Group differences of intelligence were analysed as follows. In the original sample of 480 students, we performed factor analyses over six subtests of a cognitive battery that was based on the Structure of Intellect Model of Guilford and Raven Progressive Matrices for a better interpretation of various aspects of intelligence. Three factors (non-verbal intelligence, speed of numeric processing, vocabulary) were extracted and factor scores were acquired with a mean of zero and standard deviation of one. The smaller sample of the present study shows small deviations from the mean and SD of 1 because this was a selection of the original sample. In the present sample, the groups did not differ on the three aspects of general intelligence. Furthermore, no differences were found on school grades of English language, mathematics, and other courses. However, the dyslexic group had compared to the non-dyslexic group lower final school grades of Dutch language and other languages such as French or German. We conclude that the groups did not differ in terms of general intelligence. Specific details can be found under *Supplementary Information*.

The data of the second sample were collected to be used in various studies regarding brain correlates accompanying various developmental disorders. The subjects of this sample were not tested for dyslexia, because the present study was performed after the collection of data. Available was a large self-report questionnaire which included two questions about dyslexia. One question was whether the subjects had an official certificate of dyslexia and a second question was whether a subject was tested for dyslexia whilst attending school.

2.3. Image acquisition and preprocessing

For both samples, we used the standard population acquisition protocol of the Spinoza Centre for NeuroImaging in Amsterdam. We acquired three 3DT1 whole-brain scans for each subject (3D T1, Turbo Field Echo sequences, voxel size = 1 mm^3, FOV = 256^2 mm, 160 slices, FA = 8°, TE = 3.81 ms, TR = 8.24 ms), using a 3 T Philips Achieva scanner with a 32 channel headcoil. Each sequence lasted approximately 6 min to acquire. The three T1 scans were aligned to the 2nd recorded T1 scan and subsequently averaged. Each averaged brain was manually inspected and subsequently placed in a common space using VBM (Good et al., 2001) as implemented in FSL (Smith et al., 2004).

First, structural images were brain-extracted. Next, tissue-type segmentation was carried out using FAST4 (Zhang et al., 2001). The resulting GM partial volume images were then aligned to MNI152 standard space using the affine registration. The resulting images were averaged to create a study-specific template, to which the original GM images were then non-linearly re-registered with a method that uses a B-spline representation of the registration warp field (Rueckert et al., 1999). The registered partial volume images were then modulated (to correct for local expansion or contraction) by dividing by the Jacobian of the warp field. The modulated segmented images were then smoothed with an isotropic Gaussian kernel with a kernel of 4 mm.

2.4. Pattern classification

We used SVM to train a classifier to distinguish between subjects with and without dyslexia of the first sample (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). The SVM classifier was trained on using 21 randomly selected subjects with dyslexia (of 22) and 21 randomly selected subjects without dyslexia (of 27). The voxels used during the training stage were determined by subtracting the average