

Automatic classification of dyslexic children by applying machine learning to fMRI images

Yolanda García Chimeno^{a,*}, Begonya García Zapirain^a, Ibone Saralegui Prieto^b and Begonya Fernandez-Ruanova^c

^a*DeustoTech-LIFE Department, University of Deusto, Bilbao, 48007, Spain*

^b*Department of Neuroradiology, Osatek, Galdakao, 48960, Spain*

^c*Research Department, Osatek, Bilbao, 48011, Spain*

Abstract. Functional Magnetic Resonance Imaging (fMRI) and Diffusion Tensor Imaging (DTI) are a source of information to study different pathologies. This tool allows to classify subjects under study, analysing in this case, the functions related to language in young patients with dyslexia. Images are obtained using a scanner and different tests are performed on subjects. After processing the images, the areas that are activated by patients when performing the paradigms or anatomy of the tracts were obtained. The main objective is to ultimately introduce a group of monocular vision subjects, whose brain activation model is unknown. This classification helps to assess whether these subjects are more akin to dyslexic or control subjects. Machine learning techniques study systems that learn how to perform non-linear classifications through supervised or unsupervised training, or a combination of both. Once the machine has been set up, it is validated with the subjects who have not been entered in the training stage. The results are obtained using a user-friendly chart. Finally, a new tool for the classification of subjects with dyslexia and monocular vision was obtained (achieving a success rate of 94.8718% on the Neuronal Network classifier), which can be extended to other further classifications.

Keywords: Classifier, dyslexic, fMRI, PCA, monocular vision

1. Introduction

Today the use of Functional Magnetic Resonance Imaging (fMRI) and Diffusion Tensor Imaging (DTI) [1] is a key part of any research work in neuroscience.

The Machine Learning technique [2] will be used to perform classification as it is an effective tool to classify subjects. Therefore, the combination of fMRI and DTI images, together with subject classification, help to create a tool that allows to classify subjects into the desired groups, thanks to the features provided by those images.

*Corresponding author: Yolanda García Chimeno, DeustoTech-LIFE Department, University of Deusto, Bilbao, 48007, Spain. Tel.: +34 944 13 90 00; Fax: +34 944 45 68 17; E-mail: yolanda.garcia@deusto.es.

Feedback was obtained from neurologists and psychologists, who helped to define the key features to perform a good classification, thus achieving reliable results that could also be useful for them. In fact, these features are brain areas related to language, speech and lexical decision.

Therefore, this new software tool allows users to interact with all the classifiers available. The user can enter the desired features and numbers of subjects, then select the classifier to be applied, to finally obtain the results of such classification.

One of classification phases is training, which consists in creating a group of samples that have been classified as true so that they can be identified. Therefore, the features describing the samples must be discriminatory for subsequent classification.

The learning phase includes supervised [3], unsupervised [4] and semi-supervised [5] learning.

Supervised learning develops a mathematical function that, based on the pre-labelled training samples, deduces which category or groups the set of input samples belong to. Regarding unsupervised learning, there is not a set of training samples available that allow to know the labels of those samples, so it is necessary to use clustering techniques [6] to create these labels. This type is aimed at categorising samples into groups with very similar characteristics. Semi-supervised learning arises from the difficulty of labelling a large set of samples of supervised classifiers. This is why this type resorts to using a limited set of labelled samples and another larger set of non-labelled samples.

To remove redundant features and achieve optimal classification, Principal Component Analysis (PCA) can be applied [7].

2. Material and methods

2.1. Image acquisition

A Philips 3.0-T Achieva scanner with a 32-channel coil was used to obtain fMRI and DTI images. First, high-resolution 3D structural/anatomic image was obtained. The acquisition time is 5'. The DTI sequence consists of a single axial shot, which takes 4'20''.

2.2. Image processing

To obtain the features for DTI and fMRI, two different processing methods were carried out using FSL tools, FMRIB (Functional MRI of the Brain) Software Library, version 4.1.9 (<http://www.fmrib.ox.ac.uk/fsl>).

For DTI processing, the Brain Extraction Tool (BET) [8], FDT (FMRIB's Diffusion Toolbox) and TBSS (Tract-Based Spatial Statistics) [9] functions were used, obtaining FA values [10]. This was registered to the MNI152 standard [11].

However, to obtain the features of fMRI images, FEAT functions (FMRI Expert Analysis Tool v5.98) and BET (Brain Extraction Tool) were used, and a 3mm FWHM (Full Width at Half Maximum) Gaussian filter was applied. This was also registered to the MNI152 standard space, and it was also applied to functional and structural images [12].

2.3. Databases

The participants' database included 56 children aged between 9-12 years, with their parents' consent.

These subjects were classified by reading ability. These groups included dyslexic children and subjects with monocular vision, and there was also a control group.

To perform this classification, an ophthalmic specialist examined each of these subjects to decide which group each subject should fall into. Apart from this, a neuropsychological study was also carried out using three types of tests. The first was an IQ test, and the other two were reading tests, which helped to assess reading fluency. Depending on the score achieved, children were classified into different groups.

In total, 20 children were classified into the control group (ten boys), 19 children into the dyslexic group (12 boys) and 17 into the monocular vision group (10 boys).

2.4. Technologies

2.4.1. Supervised classifiers

Linear Discrimination Analysis (LDA) classifier is a statistical method for pattern recognition and automatic learning aimed at finding a linear combination of features that characterise or separate two or more classes of samples [13].

Support Vector Machine (SVM) [14], given a set of training samples, classes can be labelled and trained to build a model that classifies a new sample class. When the new samples are entered into that model, based on their proximity, they can be classified into one or another class.

Artificial neural networks [15–17] are based on the operation of the neural system of the human body. It has a learning phase, based on data that are used as patterns. This learning phase is aimed at finding the precise values of the weights of its connections, according to a specific criterion, based on the error of the network itself (supervised training).

Last, K-means classifier [18] is a technique that allows the grouping of cases or variables in a data set based on the similarity between them. This classifier classifies cases based on the number of classes previously known by the user.

2.4.2. Unsupervised classifier

The purpose of this classifier is to assign the most common class among its K nearest neighbors to an invisible point in the training phase. These nearest neighbors are obtained using a distance metric [19]. The second phase of this classifier would be to perform the same process, but this time with the validation data.

2.4.3. Combination: supervised + unsupervised classifier

AdaBoost [20] can be used in conjunction with many other learning algorithms to improve their performance. It is an adaptive classifier in the sense that the classifiers built in each iteration are adjusted to improve the cases that were incorrectly classified by the previous ones.

2.4.4. Selection criteria for classifiers

The tool (in this study) provides several results on classification success rate and assesses which classifiers have the best results. That is why the classifiers described in the sections 2.4.1. to 2.4.3., have been selected.

The selection criteria for these classifiers are the following: on the one hand they are applied most of the times. On the other hand, they are expected to provide greater efficiency. However, the latter criteria depends on the inserted data, in this case features shown by children with dyslexia.

Despite getting the same data, the different classifiers can classify unequally, because each classifier has its own region of space of characteristics where it provides the best results. The Neural Network classifier can even show a diversity of results due to the randomness of the training procedure.

3. System design

3.1. High-level design

Figure 1 shows the block diagram for high-level design.

The first block contains all the resulting images from the tests performed on the subjects scanned, obtaining output activation data. Then, the second block includes a matrix that contains the characteristics used to classify subjects. Regarding the user interface, it is responsible for implementing classifiers and showing the results. Finally, the classifier block is implemented thanks to the user interface, showing the obtained data.

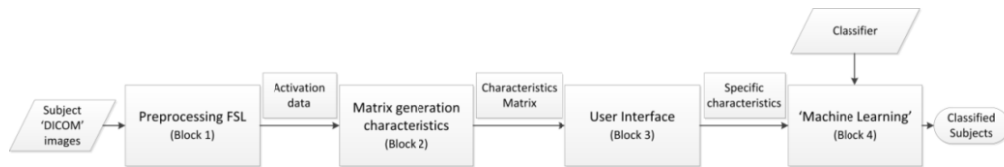


Fig. 1. High-level design chart.

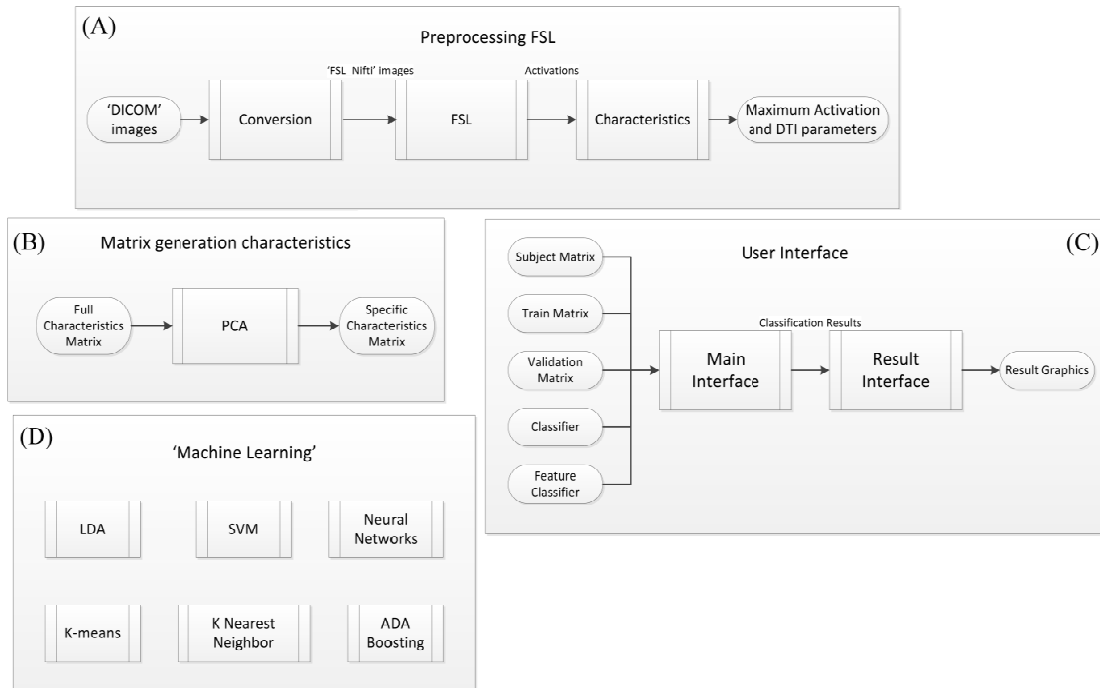


Fig. 2. Low-level diagram. FSL Pre-processing Block (A), Characteristics Matrix Generation block (B), User Interface block (C) and Machine Learning block (D).

3.2. Low-level design

3.2.1. Block A: FSL pre-processing

The FSL pre-processing block shown in Figure 2(A) contains three sub-blocks; one refers to the conversion of images, another to FSL processing, thanks to which the necessary characteristics for matrix generation are obtained. The last refers to obtaining data from FSL processing. These data are the characteristics that are entered as data in the classifiers once they have been selected.

The entrance to the first block are images in DICOM format and their corresponding output are images converted into FSL Nifti format, which are the input to the next block, FSL. This block contains the subjects' activation outputs, which are entered into the characteristic's block, thus obtaining the maximum output activations and DTI parameters.

3.2.2. Block B: matrix generation characteristics

Figure 2(B) shows the second block where the characteristic matrix that will be used for subject classification is generated.

This can be done in two ways, one being that if users have the necessary knowledge, they can discard or include the right features themselves. However, if they do not have this knowledge, the PCA can be applied to the initial matrix to discard redundant characteristics.

3.2.3. Block C; user interface

The user interface shown in Figure 2(C) allows to run different classifiers and displays the results of these classifications.

Inputs can include the subject matrix, comprising all the subjects undergoing classification, which includes the three study groups.

The training matrix includes the subjects from the control groups and dyslexic subjects who will be taken as a reference for the classifier.

The validation matrix comprises the subjects from the control groups and dyslexic subjects who have not been used for the training matrix.

The monocular vision group matrix only contains the subjects under this group. This matrix is used once control and dyslexic groups have been classified.

3.2.4. Block D: machine learning

Figure 2(D) shows all the possible classifiers that can be selected by the user. There are supervised and unsupervised classifiers, and a classifier that is a mixture of both.

4. Results

4.1. Efficiency of classifiers

4.1.1. Without PCA/with PCA

To obtain the results displayed in Table 1, all the classifiers available in this tool were implemented. In addition, for the case of the Neural Networks classifier, four results are presented, one for each hidden layer added up to a total of four hidden layers.

Table 1

Classifiers Percentages. The result without applying PCA is displayed on the left and with PCA on the right. (NN, Neural Network; Con, Control; Dys, Dyslexic; MV, Monocular Vision)

Classifier	%Success rate	% Success rate Con.	% Success rate Dys.	MV = Con	MV = Dys
LDA	74.359/79.4872	80/85	68.42/73.68	13/14	4/3
SVM	87.1795/84.6154	90/95	84.21/73.68	13/14	4/3
NN (1 layer)	87.1795/89.7436	90/100	84.21/78.94	13/12	4/5
NN (2 layers)	87.1795/89.7436	90/100	84.21/78.94	13/15	4/2
NN (3 layers)	94.8718/92.3077	95/95	94.73/89.47	16/13	1/4
NN (4 layers)	89.7436/92.3077	95/90	84.21/94.73	10/11	7/6
K-Means	56.4103/66.6667	75/66.6667	36.84/52.63	9/14	8/3
K Nearest Neighbor	76.9231/71.7949	80/80	73.68/68.42	14/12	2/5
AdaBoost	84.6154/89.7436	90/95	78.94/84.21	15/14	2/3

Table 2

Sensitivity and Specificity Results of the percentages applying PCA and not applying it

Classifier	Sensitivity	Specificity	Sensitivity (PCA)	Specificity (PCA)
LDA	68.42	80	73.68	85
SVM	84.21	90	73.68	95
Neural Network (1 layer)	84.21	90	78.94	100
Neural Network (2 layers)	84.21	90	78.94	100
Neural Network (3 layers)	94.73	95	89.47	95
Neural Network (4 layers)	84.21	95	94.73	90
K-Means	36.84	75	52.63	80
K Nearest Neighbor	73.68	80	68.42	80
AdaBoost	78.94	90	84.21	95

Results were obtained both without applying the PCA component reduction method and applying it. In addition, it shows the results of the classification of a group of monocular vision subjects, taking into account whether the characteristics are more similar to a control group or to those of the dyslexic group. The best success rate corresponds to Neural Network classifier (94.8718%), specifically three layers and without applying PCA.

4.1.2. Sensitivity and specificity

Table 2 shows the values of Sensitivity and Specificity.

The best sensitivity rate is 94.73% and specificity is 95%, those results correspond to Neural Network (three layers and no PCA applied).

5. Discussion and conclusion

The tool has been successfully created, which is useful to classify patients with any type of disease, in this case, dyslexia. In addition, there is a wide range of classifiers since there is at least one of each type: supervised, unsupervised and a combination of both [3–5]. At present, this application is an early version, so it is very similar to the basic version, and it can be used as a guide to prevent users from having problems when using this tool. Obviously, this initial version is mainly aimed at users who only want to see the results of the classification without the need to have an in-depth knowledge of all the classifiers available in the tool.

Furthermore, the best way to optimise the tool for future versions is to create an advanced version where users are free to use the tool in any way they wish, that is, characterising each classifier as necessary.

It is evident that this new advanced version would be focused on users with a good knowledge of Machine Learning [2] so that they could modify each classifier and adapt them to their own group of subjects.

As mentioned above, possible further developments of this application would be to allow users more interaction with the tool thanks to the advanced version, thus having more freedom to characterise each classifier.

As it can be observed in the results, generally high classifiers efficiency has been obtained. Thus the classifiers have been chosen successfully.

Although this study shows positive results, some limitations have been faced. Recruiting suitable participants for the study has shown to be the main difficulty for the following reason: firstly all, children had to meet all the inclusion criteria. Secondly potential participants fitting into at least one exclusion criteria where determined not to be suitable.

On the other hand, getting into a scanner can be uncomfortable and even overwhelming for children. In the process of recording images, the movements of the participants inside the scanner cause distortion in the images needing to reject them.

If greater number of subjects were obtained, it could lead to a more robust classification and verify the efficiency of classifiers.

Another is to enable users not only to classify subjects with dyslexia, as the basic version does not allow to modify classifiers, but to give them the possibility of classifying other types of subjects like. For example, those suffering from migraines and children with Attention Deficit Hyperactivity Disorder (ADHD).

Acknowledgement

This publication has been founding by eVIDA research group grant from Education and Research department of the Basque Country, Deiker from University of Deusto.

The author would like to thank all those who took part in the study, especial department of neuroradiology and radiology of Galdakao Hospital and Osatek.

References

- [1] F. Javad, J.D. Warren, C. Micallef, J.S. Thornton, X. Golay, T. Yousry et al., Auditory tracts identified with combined fMRI and diffusion tractography, *Neuroimage* **84** (2014), 562–574.
- [2] K. Yoshida and A. Sakurai, Machine learning, in: *Encyclopedia of Information Systems*, H. Bidgoli, Ed., Elsevier, New York, 2003, pp. 103–114.
- [3] S.B. Kotsiantis, I. Zaharakis and P. Pintelas, Supervised machine learning: A review of classification techniques, *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI System with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, 2007, 3–24.
- [4] J.A. Richards, Clustering and unsupervised classification, *Remote Sensing Digital Image Analysis*, Springer, 2013, 319–341.
- [5] B. Krishnapuram, D. Williams, Y. Xue, L. Carin, M. Figueiredo and A.J. Hartemink, On semi-supervised classification, *Advances Neural Information Processing Systems* (2004), 721–728.
- [6] T. Næs, P.B. Brockhoff and O. Tomic, Cluster analysis: Unsupervised classification, statistics for sensory and consumer science, in: *Front Matter*, John Wiley & Sons, Ltd, 2010, pp. 249–261.

- [7] H. Abdi and L.J. Williams, Principal component analysis: Wiley interdisciplinary reviews, Computational Statistics **2** (2010), 433–459.
- [8] S.M. Smith, Fast robust automated brain extraction, Hum. Brain Mapp. **17** (2002), 143–155.
- [9] S.M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T.E. Nichols, C.E. Mackay et al., Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data, Neuroimage **31** (2006), 1487–1505
- [10] T. Beppu, T. Inoue, Y. Shibata, A. Kurose, H. Arai, K. Ogasawara et al., Measurement of fractional anisotropy using diffusion tensor MRI in supratentorial astrocytic tumors, J. Neurooncol. **63** (2003), 109–116.
- [11] M. Brett, K. Christoff, R. Cusack and J. Lancaster, Using the Talairach atlas with the MNI template, Neuroimage **13** (2001), 85–85.
- [12] M.S. Asghar, A.E. Hansen, H.B. Larsson, J. Olesen and M. Ashina, Effect of CGRP and sumatriptan on the BOLD response in visual cortex, The Journal of Headache and Pain **13** (2012), 159–166.
- [13] J. Ye, R. Janardan and Q. Li, Two-dimensional linear discriminant analysis, Advances in Neural Information Processing Systems (2004), 1569–1576.
- [14] S. Li, T. Fevens, A. Krzyżak and S. Li, Automatic clinical image segmentation using pathological modeling, PCA and SVM, Eng. Appl. Artif. Intell. **19** (2006), 403–410.
- [15] J.R.H. González and V.J.M. Hernando, Redes neuronales artificiales: Fundamentos, modelos y aplicaciones, Ra-ma , Artificial Neural Networks: Basis, models and applications, Ra-ma, 1995.
- [16] B.J. Andrews, R. Davoodi, R. Kamnik and T. Bajd, Control of FES in paraplegia: Modeling voluntary arm forces, Biomed. Mater. Eng. **8** (1998), 241–251.
- [17] Z. Che, T. Chiang and Z. Che, Feed-forward neural networks training: A comparison between genetic algorithm and back-propagation learning algorithm, Int. J. Innov. Comput. Inf. **7** (2011), 5839–5850.
- [18] T. Niknam, E. Taherian Fard, N. Pourjafarian and A. Roustaa, An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering, Eng. Appl. Artif. Intell. **24** (2011), 306–317.
- [19] K. Hechenbichler and K. Schliep, Weighted k-nearest-neighbor techniques and ordinal classification, Discussion paper 399, Ludwig-Maximilians University, 2004.
- [20] W. Cheng and D. Jhan, A self-constructing cascade classifier with AdaBoost and SVM for pedestrian detection, Eng. Appl. Artif. Intell. **26** (2012), 1016–1028.