International Journal of Neural Systems, Vol. X, No. X (2019) 1–19

DYSLEXIA DIAGNOSIS BY EEG TEMPORAL AND SPECTRAL DESCRIPTORS: AN ANOMALY DETECTION APPROACH

ANDRÉS ORTIZ[1,2], FRANCISCO J. MARTINEZ-MURCIA[1,2], JUAN L. LUQUE[3], ALMUDENA GIMÉNEZ[4], ROBERTO ORTEGA-MORALES[5], JULIO ORTEGA[5]

[1]*Department of Communications Engineering, University of Malaga, Campus de Teatinos s/n, 29071, Malaga, Spain*
E-mail: aortiz@ic.uma.es
http://www.biosip.uma.es/

[2]*Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI)*

[3]*Department of Developmental Psychology and Education.Campus de Teatinos s/n, 29071, Malaga, Spain*

[4]*Department of Basic Psychology.Campus de Teatinos s/n, 29071, Malaga, Spain*

[5]*Department of Computer Architecture and Technology, University of Granada, Periodista Daniel Saucedo Aranda, 18071,Granada, Spain*

Diagnosis of learning difficulties is a challenging goal. There is a high number of factors involved in the evaluation procedure that present high variance among the population with the same difficulty. Diagnosis is usually performed by scoring subjects according to results obtained in different neuropsychological (performance based) tests specifically designed to this end. One of the most frequent disorders is Developmental dyslexia (DD), a specific difficulty in the acquisition of reading skills not related to mental age or inadequate schooling. Its prevalence is estimated between 5% and 12% of the population. Traditional tests for DD diagnosis aim to measure different behavioural variables involved in the reading process. In this paper we propose a diagnostic method not based on behavioural variables but in involuntary neurophysiological responses to different auditory stimuli. The experiments performed use electroencephalography (EEG) signals to analyse the temporal behaviour and the spectral content of the signal acquired from each electrode to extract relevant (temporal and spectral) features. Moreover, the relationship of the features extracted among electrodes allows to infer a connectivity-like model showing brain areas that process auditory stimuli in a synchronized way. Then an anomaly detection system based on the reconstruction residuals of an autoencoder using these features has been proposed. Hence, classification is performed by the proposed system based on the differences in the resulting connectivity models that have demonstrated to be a useful tool for differential diagnosis of DD as well as a method to step towards a better knowledge of the brain processes involved in DD. The results corroborate that non-speech stimulus modulated at specific frequencies related to the sampling processes developed in the brain to capture rhymes, syllabes and phonemes produce effects in specific frequency bands that differentiate between controls and DD subjects. The proposed method showed relatively high sensitivity above 0.6, and up to 0.9 in some of the experiments.

Keywords: EEG, Dyslexia, Deep Learning, Autoencoder, Anomaly detection, Automatic Diagnosis

1. Introduction

Developmental Dyslexia (DD) is a specific difficulty in the acquisition of reading and writing skills not related to mental age or inadequate schooling. DD occur as part of a continuum and it is the low end of a normal distribution of word reading ability.[1,2] A common definition sets the cut-off for reading achievement 1.5 standard deviations below the mean for age. Indeed, its prevalence is estimated between 5% and 12% of the population,[3] depending on the reading performance benchmark. It has an important social impact and may determine school failure. Common symptoms of dyslexia include slow and difficult reading, unreadable handwriting, bad spellings, letter migration or reversals.[4] In addition, it has harmful effects in the self-esteem of affected children. Early diagnosis and prognosis to start an adequate, early and individualized, intervention is decisive in the personal and intellectual development of these children. Usually, DD (as other learning difficulties) are diagnosed by performing tests specifically designed to measure different behavioural variables involved in the reading process. Examples of these variables are reading efficiency, or the ability to split words in their constituent syllables. These tests are individually applied by specialists who need further time to analyse the results. Finally, diagnosis is established by means of cut-off points computed over not very large populations. Thus, the evaluation process is a time-consuming and prone to error task.[5] In addition, most benchmarks are designed for children who are already learning to read and fail, establishing the diagnosis at a late age. research work oriented towards early diagnosis and individualized intervention would have a theoretical and a practical impact.[6] The development of objective diagnosis methods is currently a challenge and an active research field directed to find biomarkers that could be used to improve the diagnosis accuracy by means of neuropsychological responses associated with the brain processes involved in reading or language processing tasks.[7] Moreover, since biological causes and processes of DD are not well known, these biomarkers could reveal unknown aspects of the DD related to its neural basis.[8,9] This can offer valuable information for a better understanding of differences between dyslexic and non-dyslexic subjects, with special application in the design of individualized intervention tasks. At the same time, it provides the arena to design diagnosis tools not only applicable to readers, but also to pre-readers (early diagnosis). In this way, works such as Ref.10 shows that partial learning in feedforward neural networks can produce a behaviour that is reminiscent of that of dyslexic persons. The search for the neurological basis of dyslexia (that remains unknown) has led the researchers to use different techniques to acquire functional brain information while the subject is developing some specifically designed tasks. Current advances in medical imaging acquisition systems provide different alternatives to infer the brain functional activity. These advances include functional neuroimaging techniques such as functional magnetic resonance imaging (fMRI), magneto-encephalography (MEG) or Positron Emission Tomography (PET). On the other hand, electroencephalography (EEG) is a popular technique to explore the cortical brain activity that has been used to study different neurological diseases such as Alzheimer's disease,[11–14] Parkinsonian syndromes,[15,16] epileptic disorders,[17–19] and other psychiatric disorders such as schizophrenia.[20] Nevertheless, it is not straightforward to extract useful information contained in EEG recordings, due to its low Signal-to-Noise ratio (SNR). In other words, transforming EEG information into knowledge requires different preprocessing and processing stages to 1) remove noise and artifacts and 2) extract descriptors. Currently, procedures to fulfill with 1) include SNR improvement through signal averaging and electrooculographic (EOG) artifacts removal using Independent Component Analysys (ICA). Unfortunately, the computation of representative or discriminative descriptors in search for biomarkers depends on the specific disorder, and on the specific stimuli used to trigger differential functional brain patterns. Thus, in Section 2, we show an overview of different processing and classification techniques of EEG signals for DD diagnosis. Nevertheless, since the classification problem tackled here is especially complex due to the high variability of DD subjects, we propose an anomaly detection system to classify DD subjects as those whose patterns are different enough from those computed for controls (CN) subjects. As defined in Refs.21, 22, *anomalies are patterns in data that do not conform to a well defined notion*

of normal behavior that rarely occur by nature. Indeed, anomalies can be detected as deviations from the behaviour considered normal. Specifically, our proposal consists on extracting time and frequency features that describes the EEG signals on each band, acquired during the application of different auditory stimulus explained in Section 4. These features are used to compute a connectivity model based on the similarity among channels in that feature space. Subsequently, an anomaly detection system based on an autoencoder is proposed. This procedure uses the reconstruction residuals to differentiate between controls and DD subjects by means of a SVC classifier. Our approach provides similar performance than previous methods that use MRI imaging or MEG signals with less electrodes and a simpler experimental methodology.

The rest of the paper is structured as follows. Section 3 shows details on the database used and the applied preprocessing. Then, this section describes the auditory stimulus, EEG preprocessing and post-processing (feature extraction) as well as the classification method. Section 4 describes the auditory stimuli used during the EEG acquisitions, along with EEG preprocessing and post-processing (feature extraction) methods. Moreover, this Section details the proposed methodology that includes tools for exploratory analysis based on the similarity between features extracted from different EEG channels in different frequency bands. Section 5 presents and discusses the classification results, and finally, Section 6 draws the main conclusions.

2. Related Work

The risk of DD in children is usually assessed by behavioural tests that measure different variables realated to writting and reading efficiency.[23] This diagnosis procedure is currently standarized (DSM-5) and includes six months of evidence-based intervention. Nevertheless, any arbitrary cut points obtained from behavioral measures might not have biological validity and be compromised by a variety of uncontrolled variables. In addition, heritage information also results relevant due to the genetic component of DD.[24] However, current research directions point to the exploration of the neurological basis of DD through the search for characteristic patterns while performing different tasks. Thus in Ref.25, the authors study the power distribution in each EEG band and the coherence among electrodes during word articulation, phoneme deletion, rapid naming letters and word spelling. In Ref.26 EEG topography is used to obtain a functional map of the brain cortex during a letter writing task finding activation differences between hemispheres. In Ref.4, a pseudo-word reading task is used with a reduced population of adults, while statistical features are extracted from the power spectrum computed for each EEG band and each electrode. These features are eventually classified using a Support Vector Classifier (SVC).[27] Other works are focused on the study of the connectivity patterns inferred during reading tasks. Thus, Ref.28 study disruptions of connectivity within visual and language processing networks during visual word and false font processing tasks. Connectivity patterns in Ref.28 are computed according to the statistical significance of the differences in the PSD observed at each EEG band. Previous works are based on EEG signals recorded while the subject is performing a language-related task. Other works such as Ref.29 use EEG measurements in resting state to study differences in connectivity patterns using graph theory. Specifically, in Ref.29 weighted connectivity matrices are computed for multiple frequency bands using the phase lag index (PLI).[30] Previous commented works are mainly focused on finding discriminant patters in EEG signals acquired during reading or writing related tasks.

The main theory about dyslexia -phonological theory- postulates that dyslexics have a specific impairment in the representation, storage and/or retrieval of speech sounds. While theorists have different views about the nature of the phonological problems, they agree on the central and causal role of phonology in dyslexia. Learning to read requires learning the grapheme-phoneme correspondence, i.e. the correspondence between letters and speech sounds. If these sounds are poorly represented, stored or retrieved, the learning of grapheme-phoneme correspondences will be affected accordingly.[31–34] The phonological theory therefore postulates a straightforward link between a cognitive deficit and the behavioural problem to be explained. Other theories are the rapid auditory processing, which postulates that the deficit lies in the perception of short or rapidly varying sounds.[35] There is indeed also evidence that dyslexics may have

poorer categorical perception of certain contrasts[36] (allophonic perception theory).

The search for the neural basis of DD requires to go beyond experiments based on reading and writing tasks, as it is a more general language processing deficit. In fact, one of the most accepted causes of DD in the scientific community regards to an incorrect phonological processing[5] that causes the abnormal encoding of words in memory. These results on inaccurate or incomplete stored phonological forms that cause unreliable mappings, affecting to the phoneme-grapheme conversion. Contrarily, language processing, in particular, word encoding depends on brain processes based on a sampling process that split the auditory stimuli into their segmental forms: syllable, onset-rime and phoneme.[37] This way, recent works are focused on finding neural patterns related to defective speech encoding or faulty phonological representation.[38,39] Hence, Ref.40 proposes the use of tasks related to the rapid neural learning mechanism for online acquisition of novel speech material, while measuring the neural dynamics through Event Related Potential (ERP) responses. Statistical test performed on the participants showed differences in the early neural response after the word divergence point.

Our experimental setup has been designed according to Ref.41, under the hypothesis that the core deficit in dyslexia is phonological. The Temporal Sampling Framework (TSF)[42] proposes that atypical oscillatory sampling at one or more temporal rates in children with dyslexia could cause phonological difficulties in specifying linguistic units such as syllables or phonemes. Indeed, temporal sampling of synchronous activity of oscillating networks of neurons at different frequency bands (e.g. Delta, 1.5-4 Hz; Theta-Alpha, 4-12 Hz; and Gamma, 30-80 Hz) could explain the perceptual and phonological difficulties with rhymes, syllabes and phonemes found in individuals with dyslexia. Atypical oscillatory entrainment to rhythm and speech in the Delta band have been identified in several studies.[43,44] However, there is evidence of atypical oscillatory differences in Theta and Gamma bands. Interestingly, such evidence has been found using both speech and non-speech stimuli.[41] Moreover, auditory-steady-state responses (ASSRs) to amplitude-modulated white noise stimuli provides a non-bias measure of the synchronization of neural oscillations in the auditory cortex. Although previous studies found inconclusive results in relation to EEG and DD,[45,46] the findings in latest works[43,44] have directed the scientific community to focus on the search for the neural basis of DD. In the same line, Refs.47, 48 have shown differences in readers due to cognitive impairment of the phonological representation of word forms. Speech encoding related to speech prosody and sensorimotor synchronization problems can be revealed by finding patterns in EEG channels at different sub-bands as it provides enough time resolution. Moreover,[44] uses speech-based stimuli to find a match between speech features and neural dynamics through time-frequency descriptors.

3. Database

The EEG data used in this work was provided by the *Leeduca* study at the University of Malaga. The procedure to establish the control and experimental groups was carried out according to both, the standard criteria used in similar studies and the Special Education School Services (SESS) and following the standards.[49] Moreover, the procedure contained specific guarantees from the longitudinal study developed and applied by the Leeduca Project.The Leeduca Project has been implementing a Response to Intervention (RtI) system applied for 20 years in the US and 10 years in Finland. The system applied a dynamic evaluation three times a year from 4 to 8 years to large population samples. Specifically, the control and experimental groups came from a cohort (N = 700) followed from 4 years to the second evaluation of 7 years in 20 different primary schools (Junta de Andalucía). This way, SESS had a longitudinal dynamic evaluation of the subjects, plus ATLAS (A self-report questionnaire on reading-writing difficulties for adults) family risk information, plus a complete report at 7 year old, which included standard assessment tasks. Additionally, SESS services had an official census of other neurodevelopmental disorders, including disorders such as Language Impairment (LI), Speech Sound Disorder (SSD), Attention Deficit Hyperactivity Disorder (ADHD), Autism, and other auditory or visual sensory deficits. Moreover, SESS collected information about other relevant conditions which can affect reading achievement, as immigration or bilingualism. Hence, comorbidities

were taken into account in the screening process.

The equipment used for EEG acquisitions was the Brainvision actiCHamp Plus with a 32 channels amplifier that affords a sampling rate up to 100 KHz. In addition, active electrodes were used (actiCAP, Brain Products GmbH, Germany), as they allowed a wider range of impedances, improving the SNR and making the system more robust to head movement. Hence, the experimental setup was appropiate for working with children. On the other hand, the equipment was powered from li-ion batteries to ensure the isolation with the power line and to reduce the acquisition noise. EEG signals were recorded at a sampling rate of 500 Hz during 15-minute sessions, while participants were presented with auditory stimuli using headphones. As explained in Section 4.1, a session consisted of a sequence of white noise stimuli modulated in amplitude at a rate of 2 Hz, 8 Hz, and 20 Hz presented sequentially for 5 minutes each. Stimulus change was recorded in the EEG acquisition by means of a trigger coming from the computer that generates the stimuli. This methodology facilitates the extraction of EEG recordings corresponding to the response associated to each individual stimulus for further processing.

Acquired signals signals were then pre-processed and analyzed in the frequency domain. Spectral features extracted from the EEG signals were then used to classify the subjects between controls (CN) and DD. The present experiment was carried out with the understanding and written consent of each child's legal guardian and in the presence thereof, and was approved by the Medical Ethical Committee of the Malaga University (ref. 05/02/2020 PND016/2020) and according to the dispositions of the World Medical Association Declaration of Helsinki.

Forty-eight participants took part in the present study, including 32 skilled readers (17 males) and 16 DD readers (7 males) matched in age ($t(1) = -1.4, p > 0.05$, age range: 88-100 months). The mean age of the control group was $94,1 \pm 3.3$ months, and $95,6 \pm 2.9$ months for the DD group. All participants were right-handed Spanish native speakers with no hearing impairments and normal or corrected–to–normal vision. Dyslexic children in this study had all received a formal diagnosis of dyslexia in the school. None of the skilled readers reported reading or spelling difficulties or had received a previous formal diagnosis of dyslexia. The location of 32 electrodes in the 10-20 standarized system used in the experiments is shown in Figure 1.

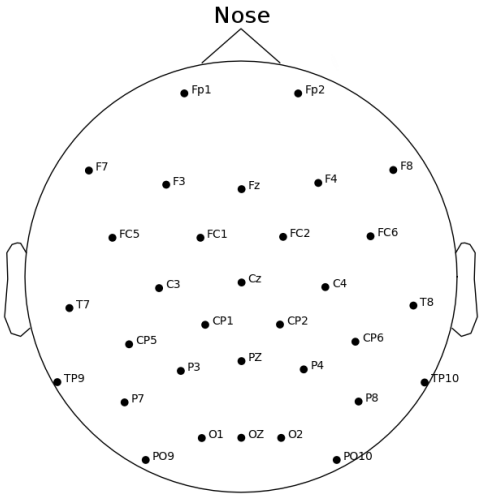

Fig. 1. Electrode montage in the 10-20 system used in the experiments

4. Methods

4.1. *Stimuli*

DD is a reading disorder often characterized by reduced awareness of speech units.[50]

Up to five types of sampling have been identified that might be relevant to explain dyslexia and other language disorders (for a recent review see Ref.51):

(1) On the one hand, Delta oscillations (1-4 Hz; around 2 Hz), located in the right hemisphere, would allow to process the prosodic aspects of speech, that is, the rhythm of speech and its cadence.
(2) Theta frequency oscillations (4-7 Hz; around 5 Hz) are preferably shown in the right hemisphere and allow syllable segmentation.
(3) In the Beta frequency band (around 20 Hz) the internal segmentation of the syllable occurs, initially (onset) and its rhyme (coda nucleus), that item from the vowel that contains all syllable (for example: you; s-ol, pl-an, etc.).
(4) High Beta-Low frequency Gamma waves (20-40 Hz) allow phoneme segmentation and are asymmetrically present in the left hemisphere.
(5) In the Alpha frequency band (8-12 hz) differences have been found that affect attention

control, selective inhibition and audio-visual integration

Other models of neuronal speech coding suggest that dyslexia originates from the atypical dominant neuronal entrainment in the right hemisphere to the slow-rhythmic prosodic (Delta band, 0.5-1 Hz), syllabic (Theta band, 4-8 Hz) or the phoneme (Beta-Gamma band, 12-40 Hz), speech modulations, which are defined by the time of increase in amplitude (i.e., the envelope) generated by the speech rhythm.[43,44]

In our study we approach the first three for their relevance - controversy - in the current studies. Then, have followed the approach described in Ref.49, 52, considering Delta, Theta and Beta bands, at rate of 2Hz, 8Hz and 20Hz to explore stress word patterns, syllable Spanish rate and phoneme segmentation.

Thus, we compared the cortical entrainment to Amplitude Modulated (AM) white-noise at a fixed rates: 2 Hz, 8Hz and 20 Hz. A sample composed of 7 year old children listened to stimuli obtained by rhythmically modulating the amplitude (AM) of white-noise sound either in the Delta, Theta and Gamma band.

Unlike works that use speech-based or word processing-based stimuli that requires the subject interaction during the experiment, we generated different auditory stimuli directed to trigger the functional cortical brain networks involved in language processing at different levels. These regions include Broca's area in the inferior frontal gyrus (IFG), Wernicke's area in the superior temporal gyrus (STG), as well as parts of the middle temporal gyrus (MTG) and the inferior parietal and angular gyrus in the parietal lobe, detailed in Ref.53. However, language processing (in general) is a complex task that implies the interaction among different brain areas regarding auditory, visual and linguistic processing.[54]

In order to reduce interactions, and aiming to explore differences in the auditory cortex, we used auditory stimuli in such a way that no action is required from the subject under test, only to listen the audio signal in a *resting state*-like fashion.

This procedure aims to assess the quality of the oscillatory neural processes measured through AM modulations contributing to the optimal construction of predictions of incoming auditory information (such as linguistic sequences or their simplification through AM modulations). These neurophysiological responses should explain the manifestations of the temporal processing deficits described in dyslexia.

4.2. *Signal preprocessing*

EEG signals were pre-processed in order to remove artifacts related to eye blinking and impedance variations due to movements. Since eye blinking signal is recorded along with EEG signals, these artifact were removed by blind source separation using Independent Component Analysis (ICA).[55,56] Other artifacts related to movement or noise from unknown sources required the removal of EEG segments. Afterwards, the remaining, cleaned signals were segmented into 5 seconds excerpts. As a result, different number of segments were available for different subjects. These segments are normalized to zero mean and unit variance independently. Subsequently, EEG bands defined in Section 2 are derived from preprocessed signals using a Butterworth, 5th order band pass filter, allowing to extract features from each EEG band separately. An initial exploratory analysis of the EEG signals revealed differences between CN and DD subjects in the power distribution across different EEG channels in different frequency bands. Moreover these differences depended on the stimulus. Figures 2, 3 and 4 show the average response computed for all the CN and DD subjects, exposing different activation patterns for the 2Hz, 8 Hz and 20 Hz stimulus focused in specific EEG bands.

Although average differences can be noticed, these exist very high variability among DD subjects, and it is very difficult to group them in the same cluster by the only means of power distribution profiles across different EEG channels. This way, it is neccessary to compute more elaborated features that characterize temporal and spectral aspects of the signals. On the other hand, the classification problem is treated here as an anomaly detection task, aiming to detect significant deviations with respect to the behaviour of the CN features. These features are detailed in the following section.

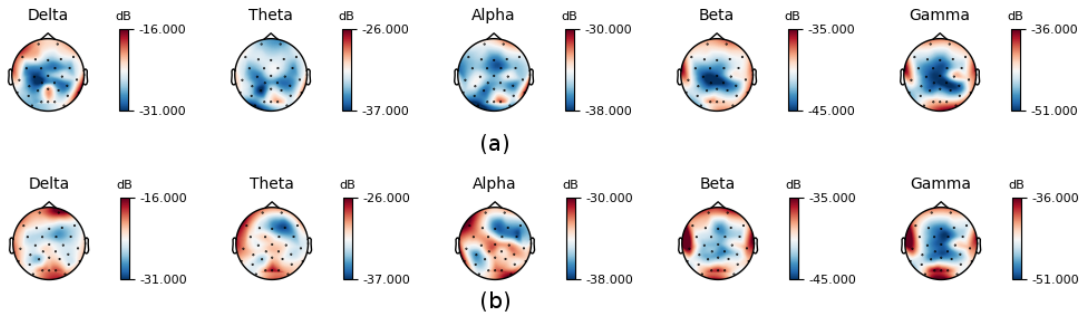

Fig. 2. Power distribution across EEG channels in different bands for (a) CN and (b) DD subjects, when the 2Hz stimulus is used. Average power over all segments over 5 minutes recording is shown.

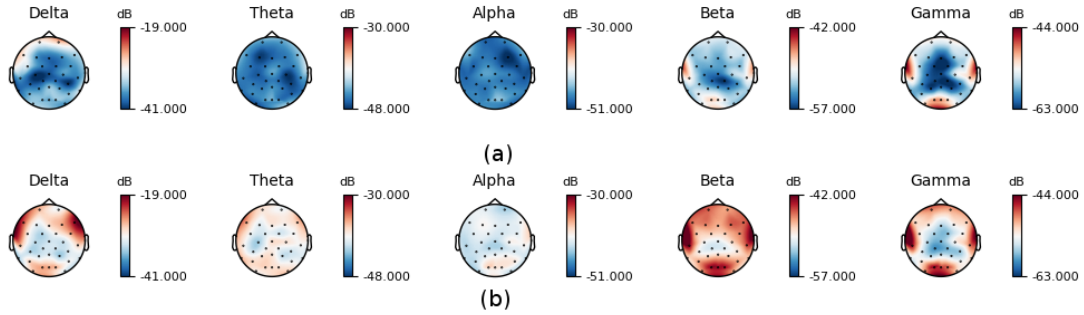

Fig. 3. Power distribution across EEG channels in different bands for (a) CN and (b) DD subjects, when the 8Hz stimulus is used. Average power over all segments over 5 minutes recording is shown.

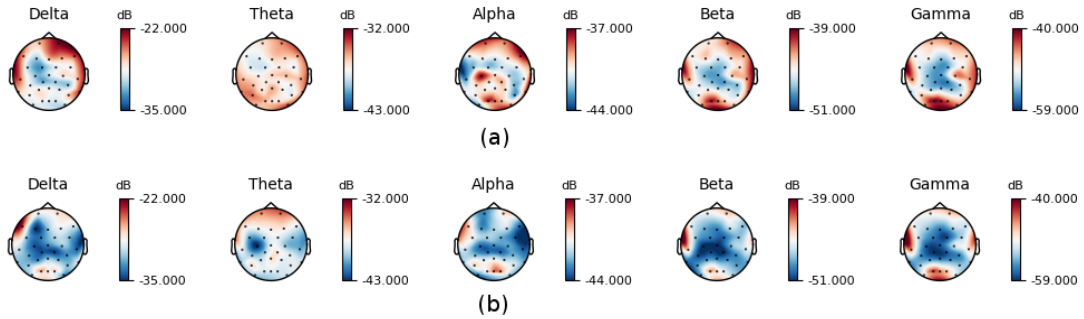

Fig. 4. Power distribution across EEG channels in different bands for (a) CN and (b) DD subjects, when the 20Hz stimulus is used. Average power over all segments over 5 minutes recording is shown.

4.3. *Feature extraction*

Previous works have shown that statistical features extracted from EEG signals contain discriminative information for the diagnosis of different pathologies.[57–59] In this work, time, frequency and fractal descriptors have been extracted from each segment in order to characterize the signal in the temporal domain, along with the shape of the power spectral density distribution. In the following, a mathematical description of the features computed is provided. Let x_i^c be the time signal for a channel c, where $i = \{1, .., N\}$ is the time instant.

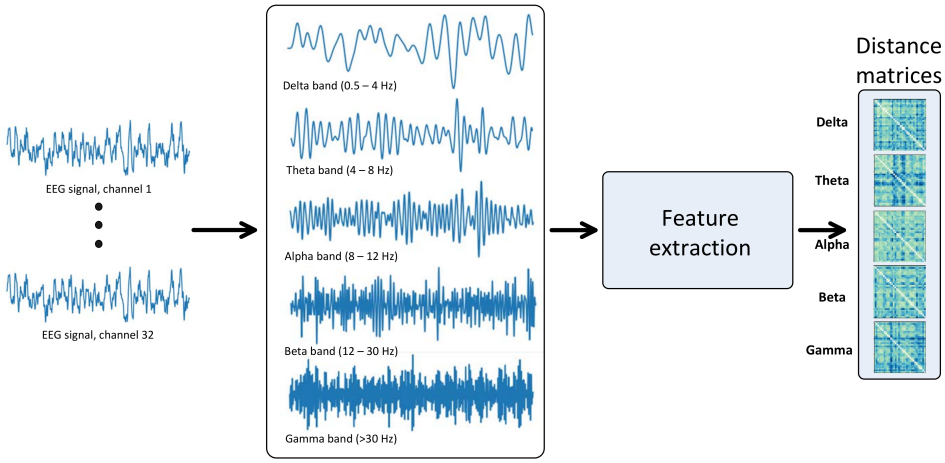

Fig. 5. Feature extraction process.

(1) **Time features**

- Mean amplitude:

$$\mu_t = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

- Amplitude variance:

$$\sigma_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_t)^2} \tag{2}$$

- Temporal skewness:

$$\beta_t = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu_t}{\sigma^t}\right)^3 \tag{3}$$

(2) **Frequency features** Frequency features are based on the computation of the PSD. Due to the non-stationary nature of EEG signals, PSD computation using classical Fourier transform is not reliable, since EEG signals cannot be expressed as a combination of pure tones. Moreover, the accuracy of the PSD computed by this method is reduced by 1) high variance of the estimate, which makes the spectrum noisy, and 2) the bias created by the leakage of energy across frequencies.[60] To overcome the limitations of Fourier transform in the computation of the PSD, more robust techniques such as Wavelet[61] and Welch's periodogram[62] are commonly used in EEG appications. Welch's periodogram is a robust estimator which improves the standard periodogram by reducing the noise, but at the cost of reducing the spectral resolution. In order to take advantage of classical and Welch's periodogram methods while avoiding their drawbacks, the solution proposed by Thomson[60] consists in using windows (also called *tapers*) in the time domain, reducing the leakage produced by multiple side lobes of a window in the frequency domain. This is also achieved by using tapers with low spectral power in the side lobes. Thus PSD can be computed as:

$$PSD(\omega) = \left|\sum_{i=0}^{N-1} x_i a_i e^{-j\omega i}\right|^2 \tag{4}$$

where $x_i, i = 1, ..., N$ is the N-samples time series corresponding to the signal and a_i is the window (taper) in the time domain. The total energy of these tappers is normalized to keep the total power invariant. This approach can be extended to reduce the variance of the estimate at each frequency by using multiple tapers. Especifically, Thomson proposed the use of K orthogonal tapers, providing K orthogonal samples of the data x_i. As a result, we have K spectral estimations $PSD_k(\omega)$ that can be averaged to reduce the variance. Furthermore, the method devised by Thomson includes an optimization step to find the tapers that minimize the leakage by maximizing the energy within a specific bandwidth. This method is also known as *multitapper method*.[63] Once the PSD has been computed, it can be characterized by different descriptors:

- Spectral Centroid. It is the weighted mean of the frequencies present in the signal, and can be seen as the center of mass of the power spectrum. In other words, it represents the frequency at which the most part of the energy is concentrated. It is defined as:

$$SC = \frac{\sum_{\omega} \omega \cdot PSD(\omega)}{\sum_{\omega} PSD(\omega)} \tag{5}$$

where $PSD(\omega)$ is the PSD of $\omega - th$ frequency bin in the spectrum.

- Spectral Spread. Defines the dispersion of the spectrum around the spectral centroid

$$\sigma_s^2 = \frac{\sum_{\omega} (\omega - SC)^2 PSD(\omega)}{\sum_{\omega} PSD(\omega)} \tag{6}$$

- Spectral Skewness. It measures de degree of symmetry of a distibution. It is defined as:

$$\beta_s = \frac{\sum_{\omega} \left(\frac{\omega - SC}{\sigma}\right)^3 PSD(\omega)}{\sum_{\omega} PSD(\omega)} \tag{7}$$

- Spectral Flatness. It Measures how the power is spreaded across the spectrum.

$$\tau_s = \frac{\left(\prod_{\omega} PSD(\omega)\right)^{1/B}}{(1/B)\sum_{\omega} PSD(\omega)} \tag{8}$$

where B is the number of frequency bins.

(3) **Fractal features**

Fractal dimension aim to measure the complexity and self-similarity of a time series. Specifically, Higuchi fractal dimension (HFD)[64] has been demonstrated to be the most appropiate fractal descriptor for electrophysiological signals,[65] and provides a reasonable estimate of the fractal dimension for short signal segments.[66] Their works[67–70] show the effectiveness of fractal-based features for the analysis of EEG signals in the diagnosis of autism and Alzheimer's disease or to analyse the functional activity of frontal lobes in major depressive disorder.[71] We computed the Higuchi dimension for a maximum scale (k_{max}) of 10.[72] Moreover HFD was computed for time signals and their corresponding *psd*. These two fractal descriptors are noted as hf_t and hf_s.

This way, a feature vector for each electrode and each band b, composed of the descriptors previously defined can be expressed as:

$$F^{b,c} = \left[\mu_t, \sigma_t, \beta_t, hf_t, SC, \sigma_s^2, \beta_s, \tau_s, hf_s\right] \tag{9}$$

4.4. *Feature similarity based connectivity*

Brain connectivity measures and graph models computed from the correlations among EEG channels have been successfully used for the diagnosis of different neurological disorders.[73–75] Morover, connectivity analysis also provides useful information about the brain processes involved during the development of different tasks[76] that can differentiate controls from subjects with a specific disorder.[77,78] According to previous works,[43,44,79,80] useful information to distinguish between CN and DD is present in the relationship among the spectral content of different channels. Thus, we hereafter use the term "*connectivity*" to referring to similarity between the signals captured from different EEG channels. This way, instead of using the feature vectors F^c, we compose a distance matrix for each subject, containing the connectivity between channels based on the correlation distance. Hence, each element d_{ij} of the 32×32 distance matrix for a specific subject, is computed as:

$$cdist(p^i, p^j) = 1 - \left| \frac{(p^i - \overline{p^i}) \cdot (p^j - \overline{p^j})}{\left\|(p^i - \overline{p^i})\right\|_2 \left\|(p^j - \overline{p^j})\right\|_2} \right| \tag{10}$$

where $\overline{p^i}$ is the mean of the elements of p^i and $p^i \cdot p^j$ is the dot product of p^i and p^j. Norms measurements $|\cdot|$ and $\|\cdot\|_2$ refers to the absolute value and $\ell_2 - norm$, respectively. Correlation distance is a similarity measure widely used in gene expression analysis.[81] It is defined by subtracting the Pearson correlation coefficient from 1. The expression in equation 10 defines a positive distance in the range [0,1], obtaining values near to 1 when the two vectors are most similar. The feature extraction process along with the construction of the distance matrices for each band is summarized in Figure 5.

Distance matrices computed from the feature vectors for each band, can be used to determine the relative importance of each electrode pair. The core idea consists in computing the similarity between distance matrices corresponding to CN and DX

subjects for each band. Thus, the more similar distance matrix entries, the less discriminative the corresponding electrode pair is. Since distance matrices are symmetric by definition, only the lower triangle is needed to stablish the comparison.

4.5. *Anomaly based classification*

Early research already noted that the population of dyslexic children is heterogeneous.[82,83] Thus, the literature of subtypes of dyslexia is voluminous, reaching hundreds of studies since 1978. For example, Morris et al.[84] carried out a large-scale study of the performance of normally developing and reading disabled children on a wide set of cognitive variables. Morris's analyses found nine subtypes, including five subtypes with specific reading disability, two with impairments in oral and written language, and just two slightly different categories representing normal readers. As a consequence, one of the most important challenges in Dyslexia detection relies on the high variability among DD subjects, while CN usually present low variances in different measured behavioural variables.[84] Indeed, it is difficult to build accurate models including both (CN and DD) classes.[82] On the other hand, due to the class imbalance problem (usual problem in biomedical engineering) the database contains more controls than experimental subjects, and it is not straightforward to balance the database by obtaining more experimental subjects, due to the distribution of CN and DD subjects in the general population. As a result, two class models generated from unbalanced databases are biased, showing special affinity to the most probable class. There are different methods to mitigate the biasing effect such as using cost sensitive objective functions by assigning different weights to miss-classification of samples from different classes. An alternative method to overcome the biasing effect while taking advantage of it consists of modelling the most probable class and then, identifying whether a new sample belongs to that distribution or not. This is also known as anomaly detection.

In this work, we propose an anomaly detection system in which the core is composed of an autoencoder. An autoencoder is a symmetric neural network (although fully symmetrical architecture is not mandatory) with a bottleneck in the middle, which is trained to reconstruct the input samples by minimizing the reconstruction error. Thus, a trained autoencoder is only able to reconstruct samples similar to the ones in the training set, since dissimilar samples will lie far away from the training samples in the projection (latent) space. In other words, the more dissimilar the input samples with respect to the training samples, the higher the reconstruction error. This fact can be exploited to construct an anomaly detection system: on the one hand, the autoencoder can be trained with the most numerous class (CN in our case). On the other hand, the reconstruction residual (i.e. the difference between the data sample and its reconstructed version generated by the autoencoder) can be used as a multidimensional similarity measurement of the current sample with respect to the class of the training samples (CN). Indeed, we tackle the problem by detecting deviations from normal (CN) behaviour descriptors.

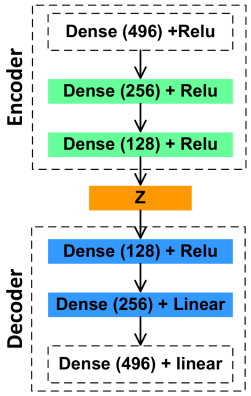

Fig. 6. Architecture of the autoencoder used to compute reconstruction residuals. The reconstruction provided by the Decoder network should reproduce the inputs depending on the subspace learned by the autoencoder. The difference between an input and its reconstruction constitutes the so called residual

The autoencoder can be formally defined as follows. Let $x_i \in X$ be an input sample, f_{enc} the encoder function, $f_{enc} : X \mapsto Z$ and f_{dec} the decoder function, $f_{dec} : Z \mapsto X'$. The residual for a sample x_i can be defined as $\|f_{enc}(x_i) - f_{dec}(x_i')\|_2$.

The architecture of the autoencoder is shown in Figure 6. Encoder and decoder are composed of three fully-connected layers, in which the number of neurons and activation funcion for each layer have been optimized by experimentation. Input and output layers have 496 neurons, corresponding to the

dimension of the feature space. On the other hand, different classification experiments have been carried out varying the dimension of the embedding layer Z to find the value providing the best classification results.

Then we use the residual computed using the autoencoder (once it is trained with CN subjects) to feed a support vector machine classifier. The overall classification system is shown in Figure 7.

In this work, the data samples used to train the autoencoder consist of vectors containing the lower triangle entries of 32×32 distance matrices D (distance matrices are symmetrical by definition), computed as explained in the previous section, resulting in feature vectors of dimension 496.

Summarizing, we follow this pipeline, once the EEG data is preprocessed to remove EOG and movement artifacts and segmented:

(1) Each signal segment is normalized to zero mean and unit variance (standarization). This removes possible DC component present in the signal and limits the range of variation to [-1,1].
(2) All signal segments from a subject are averaged to compose the mean segment.
(3) Time, frequency and fractal features are computed to compose the $F^{c,b}$ feature vector for each channel c and each band b.
(4) Distances among vectors $F^{c,b}$ from different channels are computed, and a distance matrix D is composed for each frequency band, with $d_{ij} = cdist(F^{i,b}, f^{j,b})$ for band b. Since D matrices are symmetrical we only store the lower triangle, D^t.

Training

(5) D^t matrices from CN, namely D^t_{cn} are used to train the autoencoder.
(6) Residuals $r_{cn} = \left\| D^t_{cn} - D^{t'}_{cn} \right\|$ for CN and $r_{dd} = \left\| D^t_{dd} - D^{t'}_{dd} \right\|$ for DD subjects in the training set are computed by using the autoencoder and used to train a linear support vector classifier (SVC).

Test

(7) Compute residuals r_{cn} and r_{dx} for subjects in the test set using the autoencoder and use them to predict the class with the previously trained SVC.

5. Results and Discussion

Classification experiments have been carried out using the methods described in Section 4. The overall procedure has been assessed using k-fold cross-validation (k=5), to determine the generalization ability and ensuring that training data is never used for testing. Particularly, it can be seen that the autoencoder is only trained with the control subjects in the training subset. This is a popular method to estimate the generalization error,[85–87] where such estimation, in practice, will always result in an overestimate of the true prediction error since the entire training set is not used but just $k-1$ folds. This overestimation will depend on the slope of the learning curve of the classifier and will be reduced when k increases.

In order to select the best embedding dimension at the autoencoder Z layer, accuracy, sensitivity and specificity values have been computed. These metrics are defined in equations 11, 12 and 13, respectively. In these equations TP, TN, FP, FN refers to True Positives, True Negatives, False Positives and False Negatives, respectively.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{11}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{12}$$

$$Specificity = \frac{TN}{TN + FP} \tag{13}$$

Thus, results in Table 1 show the classification performance obtained for the best case. Indeed, Figure 8 shows the classification results obtained using all EEG bands for different values of the dimension of the embedding space generated by the autoencoder.

Table 2 shows the best classification results obtained with different methods. It is worth noting that these results use different databases, different features and different classification methods. However, they are included in the table to provide a view of the diagnostic accuracy achieved in previous works using biomedical signals. In the method proposed in this work, all the experiments have been assessed by the k-fold cross validation strategy. As expected, CN subjects are correctly classified in most

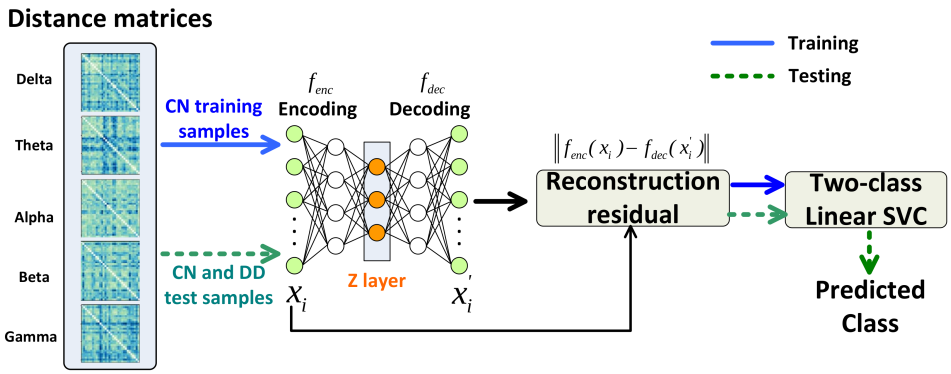

Fig. 7. Classification method based on the distance matrices

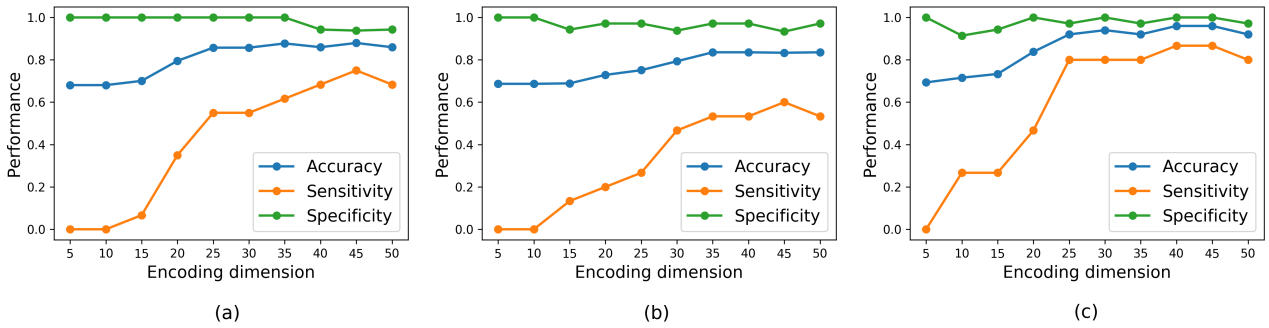

Fig. 8. Classification results obtained by combining all bands for different values of the embedding dimension. Results for (a) 2 Hz, (b) 8Hz and (c) 20 Hz are shown.

experiments, since they are the most homogeneous group and the autoencoder is trained with CN samples. However, sensitivity values are greatly dependant on the band taken into account. Thus, individual results for each band depicted in Table 1 shows that Theta band provides the best Area Under ROC curve (AUC) value with the 2 Hz stimulus, and sensitivity values up to 0.75. AUC is interpreted as the probability that a randomly chosen positive instance to obtain a higher score from the classifier than a randomly chosen negative one.

On the contrary, although features belonging to other EEG bands show reasonable accuracy values, they provide low sensitivity values (i.e. are not acutally detecting the outliers). A similar behavour is observed with the 8 Hz stimulus, where the higher sensitivity value is also obtained for the Theta band. In the case of 20 Hz Stimulus, the best sensitivity value is provided by features extracted from the Beta band.

Previous experiments using each individual band have been carried out to explore the discriminative capabilities of each band for each stimulus. However, experiments combining the five EEG bands have been also performed, concatenating the corresponding feature vectors belonging to the same channel, and eventually composing the all-band descriptor:

$$F^c = \left[F^{c,\Delta}, F^{c,\Theta}, F^{c,\alpha}, F^{c,\beta}, F^{c,\gamma}\right] \qquad (14)$$

Thus, we can notice that the combination of features from all bands increases the classification rates in all cases except for the 8 Hz stimulus due to the very low sensitivity values obtained for Alpha and Beta bands that decreases the overall performance. In fact, the highest gain when combining the features are obtained for the 20 Hz stimulus, where each individual EEG band provides higher sensitivity values.

Receiver Operating Curves (ROC) provides

Table 1. Summary of the best classification results obtained for the different EEG bands. Results combining all the bands are marked in bold. Bands providing the best sensitivity values are highlighted in gray. Error values shown in table are computed from the standard deviation.

Stimulus	Band	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**
2 Hz	Delta	0.85 ± 0.05	0.65 ± 0.17	0.97 ± 0.05	0.82 ± 0.22
	Theta	0.86 ± 0.04	0.75 ± 0.14	0.97 ± 0.05	0.85 ± 0.05
	Alpha	0.77 ± 0.10	0.31 ± 0.30	1.00 ± 0.01	0.61 ± 0.15
	Beta	0.69 ± 0.13	0.20 ± 0.20	0.93 ± 0.07	0.73 ± 0.20
	Gamma	0.78 ± 0.11	0.43 ± 0.32	0.97 ± 0.05	0.74 ± 0.10
	All bands	**0.92 ± 0.11**	**0.75 ± 0.32**	**1.00 ± 0.05**	**0.86 ± 0.10**
8 Hz	Delta	0.85 ± 0.08	0.53 ± 0.26	1.00 ± 0.05	0.84 ± 0.15
	Theta	0.88 ± 0.08	0.63 ± 0.24	1.00 ± 0.01	0.88 ± 0.11
	Alpha	0.68 ± 0.01	0.10 ± 0.10	1.00 ± 0.01	0.62 ± 0.14
	Beta	0.70 ± 0.05	0.06 ± 0.13	1.00 ± 0.01	0.71 ± 0.29
	Gamma	0.85 ± 0.05	0.53 ± 0.16	1.00 ± 0.01	0.73 ± 0.16
	All bands	**0.83 ± 0.05**	**0.60 ± 0.24**	**1.00 ± 0.32**	**0.87 ± 0.10**
20 Hz	Delta	0.83 ± 0.80	0.59 ± 0.38	0.94 ± 0.06	0.82 ± 0.09
	Theta	0.85 ± 0.04	0.66 ± 0.05	0.94 ± 0.06	0.85 ± 0.05
	Alpha	0.77 ± 0.10	0.26 ± 0.32	1.00 ± 0.01	0.72 ± 0.26
	Beta	0.85 ± 0.10	0.70 ± 0.19	0.94 ± 0.06	0.87 ± 0.13
	Gamma	0.85 ± 0.04	0.53 ± 0.16	0.97 ± 0.05	0.86 ± 0.11
	All bands	**0.96 ± 0.11**	**0.86 ± 0.32**	**1.00 ± 0.01**	**0.92 ± 0.10**

Table 2. Comparison of the performance obtained by different methods. GC: Network Global Cost, GE: network Global Efficiency, CI: Network Complexity Index, wIFCG: weighted Integrated Functional Connectivity Graph. (*) Data not provided by the source

Method	**Sensors**	**Acq.Time**	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**
MRI+SVC[88]	T1-MRI	*	0.8 ± *	0.82 ± *	0.78 ± *	*
MEG+SVC+GC[89]	253	3 minutes	0.63 ± 4.13	0.64 ± 4.01	0.65 ± 4.15	*
MEG+SVC+GE[89]	253	3 minutes	0.94 ± 1.78	0.93 ± 1.39	0.93 ± 2.32	*
MEG+SVC+CI[89]	253	3 minutes	0.80 ± 1.14	0.80 ± 1.41	0.79 ± 2.17	*
MEG+SVC+wIFCG[89]	253	3 minutes	0.97 ± 1.89	0.96 ± 1.89	0.95 ± 1.98	*
Proposed	**32**	**5 minutes**	**0.96 ± 0.11**	**0.86 ± 0.32**	**1.00 ± 0.01**	**0.92 ± 0.10**

relevant information regarding the performance of a binary classifier, as they show the trade-off between sensitivity and specificity. Figure 9 shows the ROC curves obtained for all the EEG bands when different stimulus are used. As can be seen, the combination of the features extracted from all EEG bands enhances the classification capabilities of each band separately. According to Ref.41, atypical oscillatory entrainment in Delta and Theta bands are associated to difficulties in rhymes and syllabes while differences in Gamma band are associated to difficulties with phoneme sampling. However specific frequencies depend on the language due to differences in the sampling processes in the brain during language and auditory processing tasks. Then, results for each frequency band provide valuable exploratory information regarding the brain processes involved in DD indicating the level of the sampling processes (rhyme, syllabe or phoneme) at which there are more differences between CN and DD subjects. Thus, although band-wise results contain relevant and interesting information (especially related to the

sensitivity values), all the bands are simultaneously used for the best classification performance. A discussion about the relevance of the exploratory analysis regarding the channels and EEG bands is provided later in this section.

Exploring inter-channel relationships

As explained above, classification has been addressed by linear SVC using the feature vectors consisting of distances among channels. A linear SVC classifier aims to classify a m-dimensional labeled set of n data samples $\{s_i, l_i\}, i = \{1, .., n\}$, where $s_i \in \mathbb{R}^m$ and the output label is $l = \{-1, +1\}$. This is addressed by computing a linear decision hyperplane by maximizing its margin. In the standard SVC formulation, ℓ_2-norm is used in this optimization problem.[27] However, the use of $\ell 1$-norm produces sparser w coefficient sets,[90] which are the weights assigned to each feature during the minimization process. The SVC version using the ℓ_1 norm, called LP-SVC, enhances the feature selection abilities of the SVC, and can be formulated as

$$\min_{w,b} \|w\|_1 = \sum_{j=1}^{p} |w_j| \quad subject\ to \quad l_i(w^T s_i + c) \geq 1 \tag{15}$$

Since the SVC (now, LP-SVC) is trained with vectors containing the entries of the distance matrix, the features selected during the minimization process will indicate the most relevant relationships between EEG channels. Figure 10 shows the most relevant connections according to the mean coeffients of the LP-SVC along the cross-validation loop. Additionally, these figures depict the stimulus/band combinations providing the best sensitivity values according to the classification experiments.

This experiment directed to compute the most relevant EEG channels for the classification task aims to reveal not only the specific electrodes, but also the intra-hemisphere and inter-hemisphere *connections*. As a result, we noticed 5 inter-hemisphere and 7 intra-hemisphere connections for the 2Hz stimulus in Theta band, 5 inter-hemisphere and 10 intra-hemisphere connections for the 8Hz stimulus in the Theta band, and 5 inter-hemisphere and 10 intra-hemisphere connections for the 20Hz stimulus in the Beta band. As a conclusion, different stimuli not only activate different brain areas, but also promote different relationships among channels. Thus, while the most relevant relationships with the 2 Hz stimulus / Theta band are focused between occipital and temporal regions, 8Hz/Theta generates a more symmetric connectivity in between occipital hemispheres and frontal-central channels. Previous fMRI and EEG studies have found abnormal activity in prefrontal and occipital areas,[28,91,92] although these studies also refer to different subcortical areas. On the other hand, differences in channels located at auditory cortex areas with respect to those at frontal and occipital areas have also been reported. Differences in these areas have been previously explained during phonetic processing,[93] although these studies were not developed for Spanish speakers. This could imply differences in the frequency band involved (due to differences in the sampling processes in the brain).[41] In the case of 20Hz Stimulus, connections appears in Beta band, being more noticeable in the frontal area, slightly biased to the right hemisphere. This effect is also linked to phonetic processing tasks.[94]

Table 1 shows the classification performance achieved by different methods in the literature. Specifically, Gray matter distribution obtained by Magnetic Resonance Imaging (MRI) segmentation is used in Ref.88 along with non-linear SVC classifier to detect differences in the brain anatomy. On the other hand, functional measures using sensor-specific and network-level measures are used in Ref.89. Our proposal outperforms the classification results obtained in Ref.88 although MEG-derived features provide best results, the number of sensors is noticeably higher (253 in MEG vs. 32 in our case) and the experimental setup used in our case is considerably simpler, while providing high sensitivity and AUC values.

6. Conclusions and Future Work

In this paper, we tackled the problem of dyslexia diagnosis using EEG signals. This has been addressed by a simple experimental setup using non-interactive, non-speech auditory stimulus consisting on bandwidth limited white noise modulated in amplitude with 2 Hz, 8Hz and 20Hz signals, corresponding to the sampling processes developed in the brain during language processing. Indeed, we proposed the use of temporal and spectral features to describe EEG signals and the

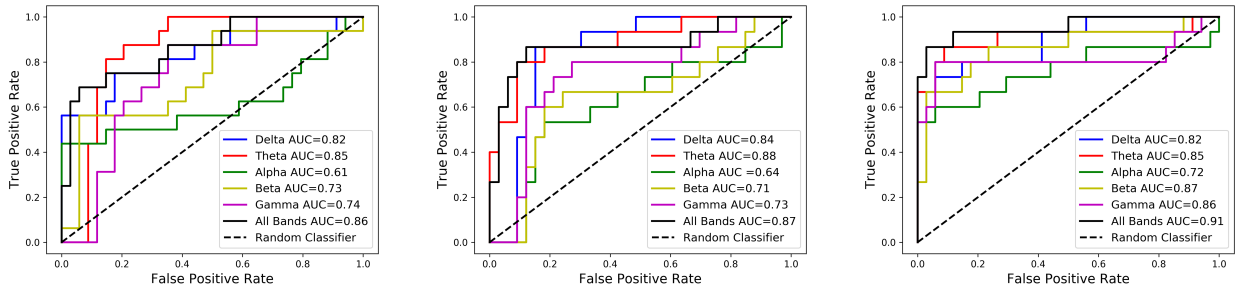

Fig. 9. ROC curves and their corresponding AUC values obtained for different EEG bands for (a) 2 Hz Stimulus, (b) 8 Hz Stimulus and (c) 20 Hz Stimulus. Curves are assessed by cross-validation. Mean and standard deviation AUC values are provided in Table 1.

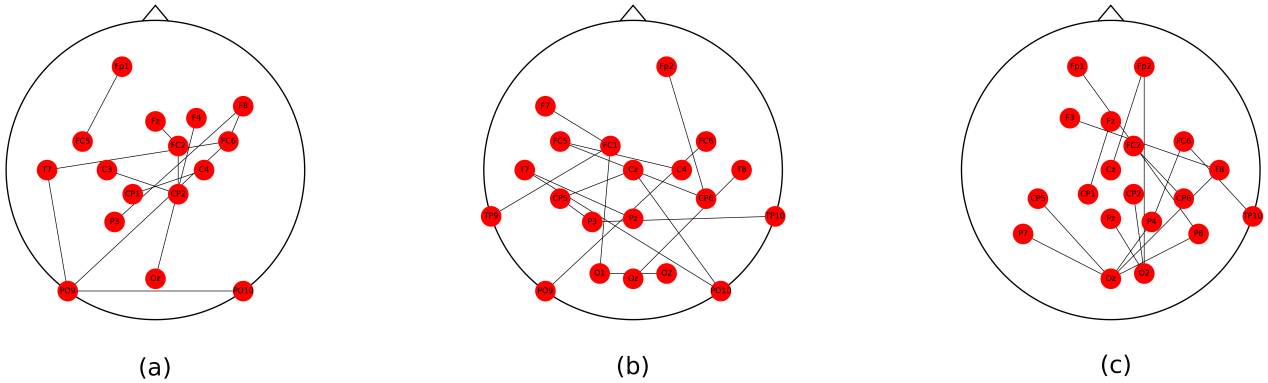

Fig. 10. Most relevant relationships between EEG channels, according to the features selected by LP-SVC. Figures shown here correspond to the Stimulus and band pairs providing the best sensitivity. (a) 2 Hz auditory stimulus / EEG Theta band, (b) 8 Hz auditory stimulus / EEG Theta band, (c) 20 Hz auditory stimulus/ EEG Beta band.

relationships among different acquired channels. These features were used to compose distance matrices by means of the correlation distance, in order to characterize the connectivity among EEG channels (i.e. electrodes showing similar features) with exploratory and discriminant purposes. On the one hand, they helped us to identify the bands that generated the most dissimilar features for each stimulus. On the other hand, the distance matrices were used to classify the subjects. Due to the high variability present in DD subjects, the classification methodology proposed here resembles an anomaly detection system, with the aim to identify DD subjects as outliers. This has been implemented by an autoencoder trained using only the CN, to compute the reconstruction residual for both CN and DD subjects. Subsequently, these reconstruction residuals were classified using a support vector classifier. Distance matrices aimed to find patterns in the relationship among EEG channels, searching for those channels that share similar temporal and spectral features in the different bands. Indeed, although differences in the power distribution were not enough to correctly classify the subjects, the similarity among feature vectors corresponding to different EEG channels provide discriminant information to obtain reasonable AUC values. The experiments carried out also corroborate that non-speech stimulus modulated at specific frequencies related to the sampling processes developed in the brain to capture rhymes, syllabes and phonemes produce differences in specific frequency bands. Specifically, 2 Hz and 8 Hz stimulus focuses the activation in the Theta band, 20 Hz stimuli produces. Moreover, the exploratory analysis performed by computing the power distribution

at different electrodes showed differences between CN and DD subjects. The proposed classification method, based on detecting anomalies in the *connection* between EEG channels has demonstrated a relatively high sensitivity (above > 0.6 and up to 0.9 in some experiments). In addition, the feature selection of the Sparse SVC (LP-SVC) to obtain sparse sets of coefficients to determine the most discriminative connections between EEG channels provide a new and effective tool to reveal brain areas with similar temporal and spectral responses under a specific stimulus.

In the near future, we plan to extend the current study, including a larger database and increasing the number of DD subjects. Moreover, since the *Leeduca* research group owns extensive neuropsychological data for all the subjects in the database, we plan to use information fusion techniques to combine EEG data with behavioural descriptors, searching for brain areas with different responses in CN and DD subjects, during the performance of different tasks.

Acknowledgments

This work was partly supported by the MINECO/FEDER under PGC2018-098813-B-C31, PGC2018-098813-B-C32 and PSI2015-65848-R projects. We gratefully acknowledge the support of NVIDIA Corporation with the donation of one of the GPUs used for this research. Work by F.J.M.M. was supported by the MICINN "Juan de la Cierva - Formación" Fellowship. We also thank the *Leeduca* research group and Junta de Andalucía for the data supplied and the support.

References

1. G. Lyon, J. Fletcher, S. Shaywitz, B. Shaywitz, J. Torgesen, F. Wood and R. Olson, Rethinking learning disabilities, *Rethinking special education for a new century* (01 2001) 259–287.
2. G. R. Lyon, S. E. Shaywitz and B. A. Shaywitz, A definition of dyslexia, *Annals of Dyslexia* **53**(1) (2003) 1–14.
3. R. Peterson and B. Pennington, Developmental dyslexia, *Lancet* **379** (2012) 1997–2007.
4. H. Perera, M. F. Shiratuddin, K. W. Wong and K. Fullarton, EEG signal analysis of real-word reading and nonsense-word reading between adults with dyslexia and without dyslexia, *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, June 2017, pp. 73–78.
5. S. Shaywitz, R. Morris and B. Shaywitz, The education of dyslexic children from childhood to young adulthood, *Annual review of psychology* **59** (2008) 451–75.
6. P. A. Thompson, C. Hulme, H. M. Nash, D. Gooch, E. Hayiou-Thomas and M. J. Snowling, Developmental dyslexia: predicting individual risk, *Journal of Child Psychology and Psychiatry* **56**(9) (2015) 976–987.
7. M. Caravolas, *The Nature and Causes of Dyslexia in Different Languages*, *The Science of Reading: A Handbook* (John Wiley & Sons, Ltd, 2008), ch. 18, pp. 336–355.
8. A. Galaburda, J. Loturco, F. Ramus, R. Fitch and G. Rosen, From genes to behavior in developmental dyslexia, *Nature neuroscience* **9** (11 2006) 1213–7.
9. I. Lieder, V. Adam, O. Frenkel, S. Jaffe-Dax, M. Sahani and M. Ahissar, Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia, *Nature Neuroscience* **22** (02 2019).
10. E. A. Ferrán and R. P. J. Perazzo, Dyslexic behaviour of feedforward neural networks, *International Journal of Neural Systems* **01**(03) (1990) 237–245.
11. N. Mammone, L. Bonanno, S. D. Salvo, S. Marino, P. Bramanti, A. Bramanti and F. C. Morabito, Permutation disalignment index as an indirect, eeg-based, measure of brain connectivity in MCI and AD patients, *International Journal of Neural Systems* **27**(05) (2017) p. 1750020.
12. G. Mirzaei, A. Adeli and H. Adeli, Imaging and machine learning techniques for diagnosis of Alzheimer's disease, *Reviews in the Neurosciences* **27**(8) (2016) 857–870.
13. George, Sarah and Adeli, Hojjat, EEG-MEG and imaging-based diagnosis of Alzheimer's disease, *Reviews in the neurosciences* **24** (2013) 563–576.
14. F. C. Morabito, M. Campolo, D. Labate, G. Morabito, L. Bonanno, A. Bramanti, S. D. Salvo, A. Marra and P. Bramanti, A longitudinal EEG study of alzheimer's disease progression based on a complex network approach, *International journal of neural systems* **25**(2) (2015) p. 1550005.
15. G. Gálvez, M. Recuero, L. Canuet and F. Del-Pozo, Short-term effects of binaural beats on eeg power, functional connectivity, cognition, gait and anxiety in parkinson's disease, *International Journal of Neural Systems* **28**(05) (2018) p. 1750055, PMID: 29297265.
16. N. Jackson, S. R. Cole, B. Voytek and N. C. Swann, Characteristics of waveform shape in parkinson's disease detected with scalp electroencephalography, *eNeuro* **6**(3) (2019).
17. H. Adeli, Z. Zhou and N. Dadmehr, Analysis of EEG records in an epileptic patient using wavelet transform, *Journal of Neuroscience Methods* **123**(1) (2003) 69 – 87.
18. S. J. M. Smith, Eeg in the diagnosis, classification,

and management of patients with epilepsy, *Journal of Neurology, Neurosurgery & Psychiatry* **76**(suppl 2) (2005) ii2–ii7.
19. Q. Yuan, W. Zhou, S. Yuan, X. Li, J. Wang and G. Jia, Epileptic EEG classification based on kernel sparse representation, *International Journal of Neural Systems* **24**(04) (2014) p. 1450015, PMID: 24694170.
20. Y. Li, S. Tong, D. Liu, Y. Gai, X. Wang, J. Wang, Y. Qiu and Y. Zhu, Abnormal EEG complexity in patients with schizophrenia and depression, *Clinical Neurophysiology* **119**(6) (2008) 1232 – 1241.
21. V. Chandola, A. Banerjee and V. Kumar, Anomaly detection: A survey, *ACM Comput. Surv.* **41** (07 2009) 447–465.
22. V. Kotu and B. Deshpande, Chapter 13 - anomaly detection, *Data Science (Second Edition)*, eds. V. Kotu and B. Deshpande (Morgan Kaufmann, 2019), pp. 447 – 465, second edition edn.
23. M. Laasonen, J. Väre, H. Oksanen-Hennah, S. Leppämäki, P. Tani, H. Harno, L. Hokkanen, E. Pothos and A. Cleeremans, Project dyadd: Implicit learning in adult dyslexia and adhd, *Annals of Dyslexia* **64** (Apr 2014) 1–33.
24. A. Giménez, A. Ortiz, M. López-Zamora, A. Sánchez and J. L. Luque, Parents' reading history as an indicator of risk for reading difficulties, *Annals of Dyslexia* **67** (Oct 2017) 259–280.
25. M. Arns, S. Peters, R. Breteler and L. Verhoeven, Different brain activation patterns in dyslexic children: Evidence from EEG power and coherence patterns for the double-deficit theory of dyslexia, *Journal of integrative neuroscience* **6** (04 2007) 175–90.
26. N. B. Mohamad, K. Y. Lee, W. Mansor, Z. Mahmoodin and S. Amirin, Normal and dyslexic children: EEG topography versus fMRI brain images during letters writing, *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Dec 2016, pp. 291–295.
27. V. N. Vapnik, *Statistical Learning Theory* (Wiley-Interscience, 1998).
28. G. Žarić, J. M. Correia, G. F. González, J. Tijms, M. W. van der Molen, L. Blomert and M. Bonte, Altered patterns of directed connectivity within the reading network of dyslexic children and their relation to reading dysfluency, *Developmental Cognitive Neuroscience* **23** (2017) 1 – 13.
29. G. F. González, M. V. der Molen, G. Žarić, M. Bonte, J. Tijms, L. Blomert, C. Stam and M. V. der Molen, Graph analysis of EEG resting state functional networks in dyslexic readers, *Clinical Neurophysiology* **127**(9) (2016) 3165 – 3175.
30. C. J Stam, G. Nolte and A. Daffertshofer, Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources, *Human brain mapping* **28** (11 2007) 1178–93.
31. L. Bradley and P. Bryant, Difficulties in auditory organisation as a possible cause of reading backwardness, *Nature* **271** (1978) 746–747.
32. S. Brady, D. Shankweiler and V. Mann, Speech perception and memory coding in relation to reading ability, *Journal of Experimental Child Psychology* **35** (1983) 345–367.
33. M. Snowling, Phonemic deficits in developmental dyslexia, *Psychological research* **43(2)** (1981) 219–234.
34. F. R. Vellutino, Dyslexia: Theory and research, *Applied Psycholinguistics* **4(1)** (1983) 69–79.
35. P. Tallal, S. Miller and R. Fitch, Neurobiological basis of speech: A case for the preeminence of temporal processing, *Annals of the New York Academy of Sciences* **682** (07 1993) 27–47.
36. W. Serniclaes, L. Sprenger-Charolles, R. Carré and J. Demonet, Perceptual discrimination of speech sounds in developmental dyslexia, *Journal of speech, language, and hearing research : JSLHR* **44** (05 2001) 384–99.
37. D. Swan and U. Goswami, Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis, *Journal of Experimental Child Psychology* **66**(1) (1997) 18 – 41.
38. U. Goswami, Speech rhythm and language acquisition an amplitude modulation phase hierarchy perspective, *Annals of the New York Academy of Sciences* **0**(0).
39. U. Goswami, A neural oscillations perspective on phonological development and phonological processing in developmental dyslexia, *Language and Linguistics Compass* **13**(5) (2019) p. e12328, e12328 LNCO-0755.R2.
40. K. Lilli, S. Yury, E. Partanen and T. Kujala, Impaired neural mechanism for online novel word acquisition in dyslexic children, *Scientific Reports* **8** (08 2018) 1–21.
41. U. Goswami, Temporal sampling framework for developmental dyslexia, *Trends in cognitive sciences* **15** (01 2011) 3–10.
42. U. Goswami, A Neural Basis for Phonological Awareness? An Oscillatory Temporal-Sampling Perspective, *Current Directions in Psychological Science* **27** (2017) 56–63.
43. S. Flanagan and U. Goswami, The role of phase synchronisation between low frequency amplitude modulations in child phonology and morphology speech tasks, *The Journal of the Acoustical Society of America* **143** (03 2018) 1366–1375.
44. Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia, *NeuroImage* **175**(1) (2018) 70 – 79.
45. J. Boddy, Evoked potentials and the dynamics of language processing, *Biological Psychology* **13** (1981) 125 – 140.
46. A. L. Benton and J. W. Bird, The eeg and reading disability, *American Journal of Orthopsychiatry*

33(3) (1963) 529–531.
47. A. J. Power, L. J. Colling, N. Mead, L. Barnes and U. Goswami, Neural encoding of the speech envelope by children with developmental dyslexia, *Brain and Language* **160** (2016) 1 – 10.
48. S. Cutini, D. Szücs, N. Mead, M. Huss and U. Goswami, Atypical right hemisphere response to slow temporal modulations in children with developmental dyslexia, *NeuroImage* **143** (08 2016).
49. A. De Vos, S. Vanvooren, J. Vanderauwera, P. Ghesquière and J. Wouters, A longitudinal study investigating neural processing of speech envelope modulation rates in children with (a family risk for) dyslexia, *Cortex* **93** (05 2017) 206–219.
50. Molinaro, Nicola and Lizarazu, Mikel and Lallier, Marie and Bourguignon, Mathieu and Carreiras, Manuel, Out-of-synchrony speech entrainment in developmental dyslexia, *Human Brain Mapping* **37** (08 2016) 2767–2783.
51. Jimenez-Bravo, Miguel and Marrero, Victoria and Benítez-Burraco, Antonio, An oscillopathic approach to developmental dyslexia: from genes to speech processing, *Behavioural Brain Research* **329** (2017) 84–95.
52. A. D. Vos, S. Vanvooren, J. Vanderauwera, P. Ghesquière and J. Wouters, Atypical neural synchronization to speech envelope modulations in dyslexia, *Brain and Language* **164** (2017) 106 – 117.
53. A. Friederici, The brain basis of language processing: From structure to function, *Physiological reviews* **91** (2011) 1357–1392.
54. B. Horwitz and A. Braun, Brain network interactions in auditory, visual and linguistic processing, *Brain and language* **89** (2004) 377–84.
55. R. Li and J. C. Principe, Blinking artifact removal in cognitive eeg data using ica, *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (2006) 5273–5276.
56. A. Delorme and S. Makeig, Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis, *Journal of Neuroscience Methods* **134**(1) (2004) 9 – 21.
57. R. W. Wang, Y.-C. Chen, I.-N. Liu and S.-W. Chuang, Temporal and spectral eeg dynamics can be indicators of stealth placement, *Scientific Reports* **8** (12 2018) 1–17.
58. F. Riaz, A. Hassan, S. Rehman, I. K. Niazi and K. Dremstrup, Emd-based temporal and spectral features for the classification of EEG signals using supervised learning, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **24** (Jan 2016) 28–35.
59. S. M. S. Alam and M. I. H. Bhuiyan, Detection of seizure and epilepsy using higher order statistics in the emd domain, *IEEE Journal of Biomedical and Health Informatics* **17** (March 2013) 312–318.
60. D. Thomson, Spectrum estimation and harmonic analysis, *Proceedings of the IEEE* **70** (10 1982) 1055 – 1096.
61. Z. Sankari, H. Adeli and A. Adeli, Wavelet coherence model for diagnosis of alzheimer disease, *Clinical EEG and neuroscience : official journal of the EEG and Clinical Neuroscience Society (ENCS)* **43(3)** (2012) 268–278.
62. P. Welch, The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, *IEEE Transactions on Audio and Electroacoustics* **15** (June 1967) 70–73.
63. D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications* (Cambridge University Press, 1993).
64. T. Higuchi, Approach to an irregular time series on the basis of the fractal theory, *Physica D: Nonlinear Phenomena* **31**(2) (1988) 277 – 283.
65. P. Castiglioni, What is wrong in katz's method? comments on: "a note on fractal dimensions of biomedical waveforms", *Computers in biology and medicine* **40** (10 2010) 950–2.
66. A. Eke, P. Herman, L. Kocsis and L. Kozák, Fractal characterization of complexity in temporal physiological signal, *Physiological measurement* **23** (03 2002) R1–38.
67. M. Ahmadlou, H. Adeli and A. Adeli, Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder, *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* **27** (10 2010) 328–33.
68. M. Ahmadlou, H. Adeli and A. Adeli, Fractality and a wavelet-chao methodology for EEG-based diagnosis of alzheimer's disease, *Alzheimer disease and associated disorders* **25** (01 2010) 85–92.
69. M. Ahmadlou and H. Adeli, Visibility graph similarity: A new measure of generalized synchronization in coupled dynamic systems, *Physica D: Nonlinear Phenomena* **241**(4) (2012) 326–332.
70. M. Ahmadlou, H. Adeli and A. Adeli, Improved visibility graph fractality with application for the diagnosis of autism spectrum disorder, *Physica A: Statistical Mechanics and its Applications* **391**(20) (2012) 4720 – 4726.
71. M. Ahmadlou, H. Adeli and A. Adeli, Fractality analysis of frontal brain in major depressive disorder, *International journal of psychophysiology : official journal of the International Organization of Psychophysiology* **85** (05 2012) 206–11.
72. S. Spasic, A. Kalauzi, G. Grbic, L. Martac and M. Culic, Fractal analysis of rat brain activity after injury, *Medical and Biological Engineering and Computing* **43** (Jun 2005) 345–348.
73. M. Ahmadlou, H. Adeli and A. Adeli, New diagnostic EEG markers of the alzheimer's disease using visibility graph, *Journal of neural transmission (Vienna, Austria : 1996)* **117** (09 2010) 1099–109.

74. J. del Etoile and H. Adeli, Graph theory and brain connectivity in alzheimer's disease, *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* **23** (04 2017) 616–626.
75. Y. Rajamanickam, M. M, U. R. Acharya, H. Adeli, N. Mohamed Ibrahim and E. Mesquita, Brain functional connectivity patterns for emotional state classification in parkinson's disease patients without dementia, *Behavioural Brain Research* **298** (10 2015) 248–260.
76. M. Ahmadlou, A. Adeli, R. Bajo and H. Adeli, Complexity of functional connectivity networks in mild cognitive impairment subjects during a working memory task, *Clinical Neurophysiology* **125**(4) (2014) 694–702.
77. M. Ahmadlou and H. Adeli, Complexity of weighted graph: A new technique to investigate structural complexity of brain activities with applications to aging and autism, *Neuroscience Letters* **650** (04 2017) 103–108.
78. M. Ahmadlou and H. Adeli, Fuzzy synchronization likelihood with application to attention-deficit/hyperactivity disorder, *Clinical EEG and neuroscience : official journal of the EEG and Clinical Neuroscience Society (ENCS)* **42** (01 2011) 6–13.
79. F. J. Martinez-Murcia, A. Ortiz, R. Morales-Ortega, P. J. López, J. L. Luque, D. Castillo-Barnes, F. Segovia, I. A. Illan, J. Ortega, J. Ramirez and J. M. Gorriz, Periodogram connectivity of EEG signals for the detection of dyslexia, *Understanding the Brain Function and Emotions*, 2019, pp. 350–359.
80. A. Ortiz, P. López, J. Luque, F. J. Martínez-Murcia, D. Aquino-Britez and J. Ortega, An anomaly detection approach for dyslexia diagnosis using EEG signals, *Understanding the Brain Function and Emotions*, 2019, pp. 369–378.
81. P. A Jaskowiak, R. Campello and I. Costa, On the selection of appropriate distances for gene expression data clustering, *BMC bioinformatics* **15 Suppl 2** (01 2014) p. S2.
82. G. Lyon, J. Fletcher and M. Barnes, *Learning disabilities. From Identification to Intervention*Child Psychopathology, Child Psychopathology, 2 edn. (Cambridge University Press, 2002).
83. A. Ellis, The cognitive neuropsychology of developmental (and acquired) dyslexia: A critical survey, *Cognitive Neuropsychology* **2** (1985) 169–205.
84. R. Morris, K. Stuebing, J. Fletcher, S. Shaywitz, G. Lyon, D. Shankweiler, L. Katz, D. Francis and B. Shaywitz, Subtypes of reading disability: Variability around a phonological core, *Journal of Educational Psychology* **90** (1998) 347–373.
85. T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning* (Springer, New York, NY, USA, 2001).
86. C. Sammut and G. I. Webb, *Statistical Learning Theory* (Springer, 2010).
87. Q. Dai, A competitive ensemble pruning approach based on cross-validation technique, *Knowledge-Based Systems* **37** (2013) 394 – 414.
88. P. Tamboer, H. Vorst, S. Ghebreab and H. Scholte, Machine learning and dyslexia: Classification of individual structural neuroimaging scans of students with and without dyslexia, *NeuroImage: Clinical* **11** (03 2016) 508–514.
89. S. I. Dimitriadis, P. G. Simos, J. Fletcher and A. C. Papanicolaou, Aberrant resting-state functional brain networks in dyslexia: Symbolic mutual information analysis of neuromagnetic signals, *International Journal of Psychophysiology* **126** (2018) 20 – 29.
90. M. Tan, L. Wang and I. W. Tsang, Learning sparse SVM for feature selection on very high dimensional datasets, *Proceedings of the 27th International Conference on International Conference on Machine Learning*, *ICML'10*2010, pp. 1047–1054.
91. T. L. Richards and V. W. Berninger, Abnormal fmri connectivity in children with dyslexia during a phoneme task: Before but not after treatment, *Journal of Neurolinguistics* **21**(4) (2008) 294 – 304, Use of Electrophysiological Measures in Reading Research.
92. L. Stanberry, T. Richards, V. Berninger, R. Nandy, E. Aylward, K. Maravilla, P. Stock and D. Cordes, Low-frequency signal changes reflect differences in functional connectivity between good readers and dyslexics during continuous phoneme mapping, *Magnetic resonance imaging* **24** (05 2006) 217–29.
93. T. Richards, E. Aylward, K. Field, A. Grimme, W. Raskind, A. Richards, W. Nagy, M. Eckert, C. Leonard, R. Abbott and V. Berninger, Converging evidence for triple word form theory in children with dyslexia, *Developmental neuropsychology* **30** (2006) 547–89.
94. M. Eckert, C. Leonard, T. Richards, E. Aylward, J. Thomson and V. Berninger, Anatomical correlates of dyslexia: frontal and cerebellar findings, *Brain : a journal of neurology* **126** (03 2003) 482–494.