Detecting Dyslexia from Audio Records: An AI Approach

Jim Radford, Gilles Richard, Hugo Richard and Mathieu Serrurier
Dystech Ltd, Traralgon, Australia

Keywords: Dyslexia, Screening, Machine Learning.

Abstract: Dyslexia impacts the individual's ability to read, interferes with academic achievements and may also have long term consequences beyond the learning years. Early detection is critical. It is usually done via a lengthy battery of tests: human experts score these tests to decide whether the child requires specific education strategies. This human assessment can also lead to inconsistencies. That is why there is a strong need for earlier, simpler (and cheaper) screening of dyslexia. In this paper, we investigate the potential of modern Artificial Intelligence in automating this screening. With this aim in mind and building upon previous works, we have gathered a dataset of audio recordings, from both non-dyslexic and dyslexic children. After proper preprocessing, we have applied diverse machine learning algorithms in order to check if some hidden patterns are discoverable, making a difference between dyslexic and non-dyslexic readers. Then, we built up our own neural network which outperforms the other tested approaches. Our results suggests the possibility to classify audio records as characteristic of dyslexia, leading to an accurate and inexpensive dyslexia screening via non-invasive methods, potentially reaching a large population for early intervention.

1 INTRODUCTION

Learning disorders such as dyslexia, dyspraxia, dysgraphia, etc., are deeply connected to an individual's outcomes not only during their academic years, but also when it comes to employment, mental health and more. Despite the fact there is no exact figure about the number of dyslexic people on earth, it is widely accepted by the community that dyslexia affects about 5%-10% of school-age children and, if we include adults, then it can go up to 15% (see the report coming from Duke University (Duke, 2016) for instance). The Diagnostic and Statistical Manual of Mental Disorders also known as DSM–5 (American Psychiatric Association, 2013), is often considered as one of the reference documents on this matter. To go to the point, Dyslexia is:

- defined as a basic deficit in learning to read (i.e. decode print) (Vellutino et al., 2004),
- characterized by a significant impairment in the development of reading skills,
- observable by reading performances well below the normal range for given age groups and IQ levels (L. et al., 2016),
- not explained by sensory deficits such as visual, hearing impairment, insufficient scholarship or overall mental development only (L. et al., 2016).

At least, the scientific community agrees on this definition even if some technical details remain debatable (Waesche et al., 2011). Clarifying the causes of dyslexia has been a serious goal of research over the past twenty years. Despite much progress being made across diverse fields, the causes of dyslexia still remain opaque and there is no scientific consensus. Several theories coexist, some of them have already been discredited by empirical observations, others still remain as serious candidates waiting for confirmation. Testing these hypotheses is a difficult task: dyslexic people do not form a homogeneous population and exhibit diverse patterns of errors. That is also why dyslexia is often divided into subtypes (phonological, visual, etc.), possibly originating from deficits at various stages of the comprehension system. Nevertheless, following (Tamboer et al., 2014), *it is possible to identify dyslexia with a high reliability, although the exact nature of dyslexia is still unknown*.

Dyslexia remains throughout a persons lifetime but can be mitigated with appropriate training sessions. Obviously, it is not a matter of Yes or No and the symptoms range from mild to severe.

Without known causes, detecting dyslexia remains a challenging task. The process is complex and per-

Radford, J., Richard, G., Richard, H. and Serrurier, M.
Detecting Dyslexia from Audio Records: An AI Approach.
DOI: 10.5220/0010196000580066
In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 5: HEALTHINF*, pages 58-66
ISBN: 978-989-758-490-9

formed by an accredited professional. This professional looks for many different indicators, intended to detect whether reading, writing (and also calculus skills) are being acquired at a proper rate. Generally a questionnaire to acquire information regarding medical history, social environment, school performances, etc. is filled out, then the final diagnostic is done via a battery of tests (mainly paper-based). The questionnaire is mainly a self-report which, in the case of a child, can be supervised by the tutor. The assessment may be lengthy, costly and emotionally painful. Moreover the limited number of accredited professionals may make this process time consuming. Without mentioning the cost of the assessment which can also be prohibitive.

Very often, dyslexic children can get government supports in diverse ways (specific teaching lessons, extra tuition, extra time for exams, specific staff helping during the classroom, etc.). To get such a support, the criteria is to provide a certificate coming from an accredited specialist. This situation obviously prevents a lot of people, often among the most vulnerable members of the population, to carry on an assessment. This makes the search of easy, fast, reliable and widely available assessments (or pre-assessments) of primary interest.

From another perspective, there has been an increasing interest to carry over the success stories of Artificial Intelligence (AI) and Machine Learning (ML) in diverse domains, from natural language understanding to cancer screening through autonomous vehicles. That is why some researchers from the AI community have started to investigate ML techniques for dyslexia screening[1].

In this paper, our initial assumption is that a properly trained ML algorithm might be able to distinguish between non dyslexic children and dyslexic children, starting from a set of simple features. In the case of dyslexia, we consider features extracted from audio records of word reading (known words and nonsense words[2]). Our experiences on a restricted set of data coming from speech pathologist partners in Australia demonstrate the soundness of this approach. This paper is organized as follows. In Section 2, we discuss existing related works. In Section 3, we describe the main principles underlying our method. Section 4 is dedicated to our experiments: we describe the dataset, the protocol and the results we get. Before providing future works in our conclusion Section 6, we discuss the limitations of our current approach in Section 5.

[1] Word screening is more appropriate at this stage. But, in the remainder of this paper, we indifferently use the words diagnostic, assessment and screening.

[2] A nonsense word follows phonetic rules, has no meaning and does not belong to the English dictionary but *looks like* a proper English word.

2 RELATED WORKS

We are not alone in our belief that Machine Learning or AI in the large part are there to help provide an effective dyslexia screening. Among the recent (post 2010) AI-based approaches to diagnose dyslexia, we can roughly distinguish 2 categories:

- **Category 1:** Approaches using the results of human-expert scoring to provide a diagnostic. In these cases, the diagnostic process does not change a lot: the user has still to undertake a battery of tests and these tests are human-marked. Since it may be difficult to maintain children's attention throughout the tests, an option is to include all the tests in a serious-game, to better grab the attention of the participant. But only the final diagnostic is done via an AI algorithm.

- **Category 2:** Approaches taking, as an initial assumption, one of the candidate theories explaining dyslexia. In these cases, the AI algorithm is fed with data related to the underlying theory. In the case of neurological explanations for dyslexia, the authors use brain scans or EEG. In case of oculo-motor deficits origin, the authors implement eye-tracking technologies.

2.1 Multi-tests Approaches (Cat. 1)

The approaches considered in this section take as input the results of tests undertaken by the user. These tests can be:

- pre-defined tests generally administered by an accredited professional. In this case, the tests can also be marked by the practitioner. Optionally a computer-assisted marking process is implemented.
- specifically designed computer-based tests capturing information relevant to dyslexia. In this case, a professional is needed to supervise a child participant.

Then the output of these tests are fed into an AI algorithm which ultimately provides a diagnostic as a likelihood of dyslexia (i.e. a number between 0 and 1). We can cite the works of (Palacios et al., 2010; Costa et al., 2013; Al-Barhamtoshy and Motaweh, 2017; Shamsuddin et al., 2017) which are still at a research stage. Only the work of (Rello et al., 2018), implemented via a serious game, led to a commercial system *Dytective*[3]. The authors suggest an accuracy of 85% on their dataset, which is relatively good. Another interesting option has been recently implemented in (Spoon et al., 2019), (which is not category 1, stricto sensu). Because ‘reading is intimately connected to writing, and children who struggle to learn to read often struggle to write’, Spoon et al. provide a proof-of-concept by developing a system that used computer vision to classify handwriting samples as indicative of dyslexia or not. They get 77.6% accuracy in determining whether a patch of handwriting was written by a student with dyslexia or not. Still in early development, these works provide relatively good results using very accessible information with no need for a questionnaire or battery of tests.

2.2 Serious Games Approaches (Cat. 1)

When one wants to target a population of very young children (let’s say under 7), it becomes crucial to design a data gathering process which is sufficiently attractive to motivate them. Serious games are therefore natural candidates for this task. An early attempt has been done by (Lyytinen et al., 2007), but it was before the effective emergence of AI. The works of (Van den Audenaeren et al., 2013) within the DYSL-X project and the works of (Gaggi et al., 2012) are targeted to this aim: they have designed serious games, dedicated to young children, available on computers or tablets and allowing the measurement of some parameters characteristic of dyslexia. These works are more focused on the design and implementation of a system than on estimating its accuracy. The work of (Geurts et al., 2015) DIESEL-X was also developed to detect a high risk for having dyslexia in preschoolers. Several theories on the underlying cause of dyslexia are converging on the idea that one fundamental problem derives from abnormal neurological timing, or ’temporal processing’ (Johnson, 1980). As a consequence, the perception of musical elements could differ from children with and without dyslexia. Starting from this assumption, (Rauschenberger et al., 2017) proposes DysMusic, a prototype which aims to predict the risk of having dyslexia before acquiring reading skills. The prototype was designed to observe participants listening to music (via a web app), using the *think aloud protocol*, varying different acoustic parameters such as frequency, duration which relate with perceptual parameters such as pitch and loudness. Among other tasks, the participants (10 in the cohort) have to evaluate how difficult it was to distinguish between the various sounds. (Gaggi et al., 2017) have developed a set of 6 serious games, using a 2D graphic design, and experimented on 24 participants. Both (Rauschenberger et al., 2017) and (Gaggi et al., 2017) have experimented on very small cohorts but their works still demonstrate that serious gaming is an interesting pathway for dyslexia screening.

[3]Dytective is a cross-platform app available on https://www.changedyslexia.org.

2.3 Neuro-based Approaches (Cat. 2)

Neuro-based works start from the widely admitted assumption that dyslexia is linked to a specific brain configuration, either in terms of anatomical shape or in terms of functional organisation. In (Tamboer et al., 2016) three dimensional whole-brain scans are acquired from each participant, each acquisition sequence lasting approximately 6 minutes. Using a standard classifier, they properly classify 80% of the scans between dyslexics and non dyslexics. This accuracy declines to 59% with a larger range of participants. In fact, their algorithm provides a large percentage of false alarms, i.e. many people without dyslexia are labelled with dyslexia (which is a cautious behavior but can lead to serious stress for a child).

The authors of (Frid and Manevitz, 2018) start from the assumption (Asynchrony Theory) that dyslexia could come from a gap in the speed of processing between the different brain entities activated in the word decoding process. This gap may prevent the synchronization of information necessary for an accurate reading process. Starting from this, they monitor a population of 32 children, with a more or less 50/50 percentage of dyslexic/non dyslexic readers. Getting the children to read 96 real words and 96 non-sense words, they record brain activity via electroencephalogram (EEG) and implement a binary classification algorithm (namely Support Vector Machines SVM) to distinguish between dyslexic and non dyslexic readers. They obtain an accuracy in the range of 78.5%. Obviously, this kind of technique cannot be used in a realistic way to screen a large part of the population. And, due to the heavily controlled environment needed to capture the data, it is unlikely these neuro-based approaches will lead to a publicly available tool anytime soon.

2.4 Eye Tracking-based Approaches (Cat. 2)

Eye-tracking can provide serious insight into perceptual/cognitive processes (PH. et al., 2013). Despite the fact that the work of (Hyona et al., 1995) tends to dismiss the oculo-motor dysfunction hypothesis of dyslexia, it is a fact that today most studies agree that there is a link between visual-attention and oculomotor control during reading: see for instance (Bellocchi et al., 2012; Huettig and Brouwer, 2015) for recent publications on this topic. From an AI perspective, it is then natural to monitor the eye movement of a user during reading activity.

For the first time in (Rello and Ballesteros, 2015), an eye tracking technology associated to an SVM classifier was used to predict dyslexia starting from a dataset of 97 subjects, 48 of them with diagnosed dyslexia. The eye tracking technology allows the extraction of information such as *Number of visits* (Total number of visits to the area of interest in the text), *Mean of visit* (Duration of each individual visit within the area of interest in the text), etc. The resulting accuracy is in the range of 80% which is relatively good. (M. et al., 2016) start from the same idea but monitor different parameters such as for a given eye saccade, the duration of the event, the distance spanning the event, the average eye position during the event, etc. Their standard classifier shows a very high accuracy (around 96%). The whole process also takes into account the results from a battery of other common tests, such as rapid automatic naming, reading of nonsense words, etc.: this obviously increases the duration of a screening session[4].

More recently, we can cite the works of (Asvestopoulou et al., 2019), still using eye tracking associated to an SVM classifier and getting an excellent 97% accuracy rate over a set of 69 native Greek speaking children, 32 being dyslexic. Their system *DysLexML*, still a work in progress, could ultimately be the basis of another screening tool. Nevertheless, eye tracking-based methods still need an external device to be connected in one way or another to a computer. This could be considered as a serious drawback.

Our work departs from all the previous ones in terms of the data we use, in terms of simplicity and in terms of duration of a session. In the following section, we describe how we proceed to tackle the issue of dyslexia screening.

3 PREDICTION PRINCIPLE

Because one of the main symptoms of dyslexia is difficulty in reading, we have decided to only gather reading audio recordings, from both dyslexic and non dyslexic readers, then to apply machine learning algorithms. Instead of analysing images, brain signals or eyes movements, we directly analyse audio signals. We agree that poor reading performance is not an ultimate marker of dyslexia, but our results demonstrate that a dedicated machine learning algorithm associated with proper audio signal processing can extract patterns that are not accessible to a human expert. Let us start with what a user is supposed to provide.

3.1 Word Selection and Generation

Our process is to have every individual to read 32 words (no sentences, only words). It is well-known that dyslexic children struggle when it comes to reading words they have never seen or heard. They have also difficulties with some letters, or combinations of letters (p and q for instance) and certain syllables. Our initial corpus comes from a set of 82 children's books extracted from the Gutenberg Project (Hart, 1971). We clean the texts and remove proper nouns. We obtain a list of around 100 000 words. Then we produce two lists : one with words from 4 to 6 letters, one with words from 7 to 9 letters. In each list, we consider only words with a high frequency of occurrence to guarantee the words are known by children. After filtering, each of the two lists contains around 2000 words.

In a second step, we create two lists of nonsense words. We also need to guarantee that the nonsense is pronounceable. In order to achieve that, we build a Long-Short Term Memory neural networks (LSTM) (Hochreiter S, 1997) that learn to build such nonsense words. We are then able to generate an infinite list of nonsense words[5]. As for the real words, we build two lists of nonsense words with different size (1000 nonsense words with 4 to 6 letters, 2500 nonsense words with 7 to 9 letters) and we only keep nonsense words that fit within the following constraints :

- Every subset of 4 consecutive letters exists in an English word (to guarantee the word is pronounceable)
- It contains difficult letters or a difficult combination of letters for dyslexic people.

The final list of 32 words to be read by a child is obtained by choosing 16 words in the list of real words and 16 words in the list of nonsense words. We change the length of the words according to the age of the user performing the assessments:

- List 1: From 6 to 8 years old (included) the list is ordered this way:

[4]Starting from this approach, Lexplore (https://www.lexplore.com/) has been founded in 2016 in Sweden which then expanded to the USA in 2017.

[5]We also use the expression 'generated words' because they are the output of an AI process.

  - 2 easy real words
  - $\frac{2}{3}$ of words from 4 to 6 letters (50% real, 50% nonsense):
  - $\frac{1}{3}$ of words from 7 to 9 letters (50% real, 50% nonsense):
- List 2: From 9 to 13 years old (included) the list is ordered this way:
  - 2 easy real words
  - $\frac{1}{3}$ of words from 4 to 6 letters (50% real, 50% nonsense):
  - $\frac{2}{3}$ of words from 7 to 9 letters (50% real, 50% nonsense):
- List 3: 14 years old and over the list is ordered this way:
  - 2 easy real words
  - 100% of words from 7 to 9 letters (50% real, 50% nonsense):

These constrained lists of words are randomly generated and are age-related: short words with simple syllables for children from 7 to 8, more difficult for children from 9 to 13, then difficult for children over 14. It is then very unlikely that 2 sessions lead to the same list of 32 words[6]. Note that 50% of the words are displayed with the times new roman font and 50% are displayed with the Open Dyslexic font. We tune the size of the font to ensure that the vowels appears with exactly the same size on the screen.

3.2 Input Parameters for Dyslexia Classifier

For every audio record, we consider 2 parameters:

- The Reading Reaction Time (RRT) is the interval between the initial display of the word and the start of the reading.
- The Reading Time (RT) is the time it takes for the user to read the corresponding word.

In both RRT and RT, the time unit is millisecond (ms) and their evaluation is done via a computer (no human in the loop). Consequently, from a session of 32 audio records, we extract 6 numbers which will be used in our ML experiments:

- average *RRT* for 32 words, average *RRT* for 16 real words, average *RRT* for 16 nonsense words
- average RT for 32 words, average RT for 16 real words, average RT for 16 nonsense words

[6]In fact, recently, we have decided to put at the end of the lists, 2 easy real words: it is better to finish a session on a positive!

Note: In an ideal word, $RRT_{real} + RRT_{gene} = 2 \times RRT$, but it often happens that some words are not read by the user. Then this simple linear relation is not valid anymore.

For instance, a simple analysis of the reaction time (RRT) reading time (RT) on our dataset is given in Figure 1. This demonstrates that dyslexic readers need more time than non dyslexic ones, whatever the type of words to be read.

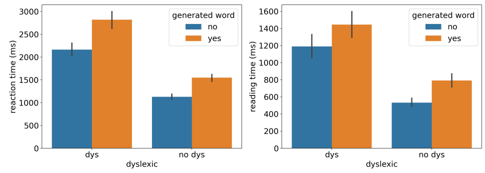

Figure 1: Reaction time (left table) - reading time (right table) for non-dyslexic/dyslexic readers w.r.t. the type of word.

We assessed 93 users, among them 43 are dyslexic and 50 are non dyslexic readers. All in all, we achieved 93 x 32= 2976 audio records. Due to this small cohort, it is not realistic to consider a simple threshold approach such as $RRT > ThresRRT$ and $RT > ThresRT$ would be enough to characterize dyslexic readers. Our final neural network computes a more sophisticated formula.

4 EXPERIMENTS

We provide in the following subsections the general context of screening dyslexia in a global population, the precise metrics we use to be in a position to rigorously compare different algorithms and the results we get from each of these algorithms. Everything has been implemented in Python, with Tensorflow and SciKit Learn libraries.

4.1 Context

Machine learning is a data-driven technology. Apart from designing an algorithm, we have to gather proper data, with their own labels. Not only the quality, but also the quantity is important since training machine learning algorithms may require a lot of data in order to be accurate. In (Wagner, 2018), the author points out that the difficulty to diagnose dyslexia is mainly coming from the unbalanced population. Only a relatively low percentage of the population has dyslexia or dysgraphia, and so to train a mathematical model with such an unbalanced population is always a challenge. When it comes to measuring the performance of the algorithm, standard accuracy can then be a misleading metric. Assuming we have 10% of the population having dyslexia or dysgraphia, a baseline algorithm declaring everybody as non dyslexic will ensure a 90% accuracy.

4.2 Metrics

As a consequence, other metrics are needed such as precision, recall and F1-score. Let us recall below some standard definitions. For a binary classifier, we have at our disposal a set of positive examples (dyslexic children) and a set of negative examples (non dyslexic children). We note tp the number of positive examples predicted as positive, tn the number of negative examples predicted as negative, fp the number of negative examples predicted as positive (false positive), and fn the number of positive examples predicted as negative (false negative). The metrics are defined as follows:

$$accuracy = \frac{tp+tn}{tp+tn+fp+fn} \tag{1}$$

The accuracy measures the probability that the class predicted by the model is the right one. In the latter we will use percentage for this accuracy.

$$precision = \frac{tp}{tp+fp} \tag{2}$$

The precision is the probability of being positive if the example is predicted as positive. In some sense, this measures the correctness of the predictor when it predicts an example as positive. The bigger this number, the better the predictor is.

$$recall = \frac{tp}{tp+fn} \tag{3}$$

The recall is the probability of a positive example to be predicted as positive. In some sense, this measures the ability of the predictor to predict all positive examples as positive. Still, the bigger this number, the better the predictor is.

$$F1\text{-}score = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

The $F1$-score is a balance between precision and recall. Thus, accuracy focus on the performances of the model in general when precision, recall and $F1$-score focus on performances of the model on positive examples only.

4.3 Results

First of all, our baseline is what we call the Dummy classifier that chooses classes randomly with the a priori probability (their frequency) of classes computed on the training set: this provide an accuracy of 0.46 which is quite poor. Then, we compare with the performances of state of the art classifiers: Logistic Regression (LR), KNN, SVM with polynomial kernel (SVC), SVM with linear kernel (LSVC), Naive Bayes NB, Random Forest (RFC) and Decision Tree (DTR). As we just want an estimation of their related performance, we use the default parameters for each of these classifiers, using *scikit-learn* Python library. We are aware that spending time to tune these parameters could lead to better results.

In a second step, we build up a neural network NN with 4 dense layers (activation function Relu) then a final dense with sigmoid activation function as it is usual for binary classification.

In order to have an average estimation of the metrics previously described:

1. We perform for each classifier a 10-fold cross-validation scheme and we get the above 4 metrics.
2. We repeat the experiences 10000 times and our table provides the average values of the 4 metrics on these 10000 runs.

Table 1: Performance of different machine learning algorithms for detecting dyslexia from audio signals.

Algorithm	Accuracy	Precision	Recall	F1
Logistic reg.	69.00	0.72	0.63	0.64
KNN	69.01	0.73	0.65	0.66
Naïve Bayes	69.02	0.73	0.58	0.61
Poly. SVM	68.10	0.76	0.33	0.43
Linear SVM	69.00	0.71	0.62	0.63
Random Forest	68.04	0.75	0.65	0.67
Decision Tree	68.01	0.64	0.66	0.62
Neural Network	81.72	0.86	0.72	0.78

The average values of metrics are shown in Table 1 (in all the tables, the subtitle 'dys' means dyslexic reader, 'no dys' means non dyslexic reader). The best results are obtained by the neural network and the other classifiers (when they are not tuned) are more or less equivalent in terms of performance.

The network, providing an accuracy of more than 80% could probably be tuned to get better performances. But due to the relatively small number of data that we have, this could lead to over fitting without giving a clear picture of the accuracy in a real environment. This could be partially overcome by considering more data (provided they are not biased). Getting more data is part of our future task. However, these experiences clearly demonstrate the power of an ML approach to distinguish dyslexic from non dyslexic readers.

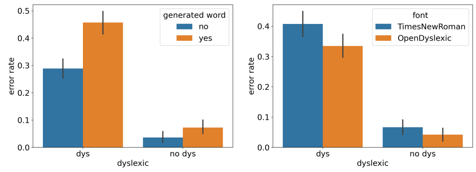

Figure 2: Error rate for non-dyslexic/dyslexic readers w.r.t. type of word (left table) and type of font (right table).

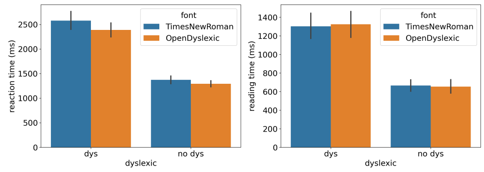

Figure 3: Reaction time (left table) and reading time (right table) for non-dyslexic/dyslexic readers w.r.t. the font.

4.3.1 Additional Experiments

At this stage of our work, we have introduced other features that we tag manually. For instance, for each spoken word in the data set, we manually indicate if the reading was right or not. Figure 2 highlights that the error rate is significantly higher for dyslexic people (which was expected). Also, the nonsense words are more difficult to read for dyslexic than for non-dyslexic people, consequently generating more errors: the gap in error rate between dyslexic and non dyslexic is bigger for nonsense words than for real words. The error rate seems to be a good indicator of dyslexia: we use it in our machine learning approach.

By considering the error rate, we achieve an accuracy of 90% on our dataset (in the same conditions than the previous experiments). This clearly establishes the interest of the error rate for improving the quality of the screening. Because this error rate is not automatically detected, it is not taken into account in Table 1. We currently work on machine learning approaches to compute this error rate automatically.

We also consider the type of font we use to display the word on the screen. Figure 2 shows that the use of the Open Dyslexic font decreases the error rate for both dyslexic and non dyslexic. Nevertheless, we do not observe from Figure 3 a similar positive effects for reaction time and reading time. The improvement of accuracy with Open Dyslexic font has already been reported in (de Leeuw, 2010) but this is still a debatable issue (Wery and Diliberto, 2017). The two last studies also report no effect on the reaction and reading time, which is in accordance with what we got. Finally, we have tested that including the reaction time with respect to the font as a new input feature for the neural network does not improve its accuracy.

5 LIMITATIONS AND OTHER OPTIONS

Despite our promising results, we are well aware that we have to be cautious.

- First of all, despite our dataset being generally bigger than the ones of our competitors, it is too small to provide a definitive conclusion. The sizes of dataset used in modern machine learning technologies are far bigger than a hundred or so examples.
- If more data are needed, it is quite clear that we need data coming from a controlled environment to be sure of their accuracy. Having noisy data will degrade our prediction.
- Spending time in tuning the standard algorithms (SVM, random forests, etc.) could lead to better results that the ones we got. One thing is for sure: building a basic neural network such as the one we use is effortless and immediately brings better results than non-tuned standard algorithms.
- We are currently focusing on getting an automatic pronunciation error detection which would definitely improve our accuracy. But why not using other parameters such as total reading time (instead of an average value) which is widely used in the Gray Oral Reading Test aka GORT 5 for instance ((Wiederholt and Bryant, 2012)).

6 FUTURE WORKS AND CONCLUSION

Analysing data (verbal/written test results, brain images, eyes movement, etc.) via Machine Learning technologies to detect dyslexia is not a new idea. In this paper, we show that it is also possible to predict dyslexia by analysing audio signals using ML. Our method satisfies the requirements needed to build a mass market screening tool:

- We focus on the human observable symptoms of dyslexia,
- We do not use any other data than the audio records,
- We do not use any external device to gather data,
- A screening session is between 10 to 15 minutes long.

It has been recently proven that properly trained ML-based predictors can be more accurate than human experts on specific tasks. Based on these facts and our encouraging results, we think there is a huge potential for ML-based technologies to help people with dyslexia. As usual with ML, accuracy can still be improved by gathering more data. ML-based technologies could definitely avoid the need of manual analysis and global performances may be improved. In the future, a better understanding of the correlation between the different disorders could also help in providing more informed predictions. For instance adding to audio records, a picture of a handwritten text could help to make the prediction still more accurate. As far as we know, this is the first time audio signals analysis is used to detect dyslexia, leading the path to a non invasive, fast and cost effective screening tool.

REFERENCES

Al-Barhamtoshy, H. M. and Motaweh, D. M. (2017). Diagnosis of dyslexia using computation analysis. In *2017 International Conference on Informatics, Health Technology (ICIHT)*, pages 1–7.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Autor, Washington, DC, 5th ed. edition.

Asvestopoulou, T., Manousaki, V., Psistakis, A., Smyrnakis, I., Andreadakis, V., Aslanides, I. M., and Papadopouli, M. (2019). Dyslexml: Screening tool for dyslexia using machine learning. *CoRR*, abs/1903.06274.

Bellocchi, S., Mathilde, M., Bastien-Toniazzo, M., and Ducrot, S. (2012). I can read it in your eyes: What eye movements tell us about visuo-attentional processes in developmental dyslexia. *Research in developmental disabilities*, 34:452–460.

Costa, M., Zavaleta, J., Cruz, S., Manhaes, L., Cerceau, R., Carvalho, L., and Mousinho, R. (2013). A computational approach for screening dyslexia. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 565–566.

de Leeuw, R. (2010). Special font for dyslexia?

Duke, U. (2016). Dyslexia international: Better training, better teaching. https://www.dyslexia-international.org/wp-content/uploads/2016/04/DI-Duke-Report-final-4-29-14.pdf.

Frid, A. and Manevitz, L. M. (2018). Features and machine learning for correlating and classifying between brain areas and dyslexia. *CoRR*, abs/1812.10622.

Gaggi, O., Galiazzo, G., Palazzi, C., Facoetti, A., and Franceschini, S. (2012). A serious game for predicting the risk of developmental dyslexia in pre-readers children. In *21st International Conference on Computer Communications and Networks (ICCCN)*, pages 1–5.

Gaggi, O., Palazzi, C. E., Ciman, M., Galiazzo, G., Franceschini, S., Ruffino, M., Gori, S., and Facoetti, A. (2017). Serious games for early identification of developmental dyslexia. *Comput. Entertain.*, 15(2):4:1–4:24.

Geurts, L., Vanden Abeele, V., Celis, V., Husson, J., Audenaeren, L., Loyez, L., Goeleven, A., Wouters, J., and Ghesquière, P. (2015). *DIESEL-X: A game-based tool for early risk detection of dyslexia in preschoolers*, pages 93–114. Springer International Publishing.

Hart, M. (1971). Project gutenberg. www.gutenberg.org.

Hochreiter S, S. J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Huettig, F. and Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia-evidence from eye-tracking: Word reading and predictive language processing. *Dyslexia*, 21.

Hyona, J., Olson, R., Defries, J., Fulker, D., Pennington, B., and Smith, S. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21:1430–1440.

Johnson, D. J. (1980). Persistent auditory disorders in young dyslexic adults. *Bulletin of the Orton Society*, 30:268–276.

L., V., F., M., G., T., and M., H. (2016). Direct viewing of dyslexics' compensatory strategies in speech in noise using auditory classification images. *PLoS ONE*, 11(4).

Lyytinen, H., Ronimus, M., Alanko, A., Poikkeus, A.-M., and Taanila, M. (2007). Early identification of dyslexia and the use of computer game-based practice to support reading acquisition. *Nordic Psychology*, 59(2):109–126.

M., N. B., G., O. S., J, Y., A., P. T. R., and C, J. (2016). Screening for dyslexia using eye tracking during reading. *PLoS ONE*, 11(12).

Palacios, A. M., Sánchez, L., and Couso, I. (2010). Diagnosis of dyslexia with low quality data with genetic fuzzy systems. *International Journal of Approximate Reasoning*, 51(8):993 – 1009.

PH., T., IG., C., G., P., JN., R., DP., M., and L., I. (2013). High-throughput classification of clinical populations from natural viewing eye movements. *J Neurol.*, 1(260):275–284.

Rauschenberger, M., Rello, L., Baeza-Yates, R., Gomez, E., and Bigham, J. P. (2017). Towards the prediction of dyslexia by a web-based game with musical elements. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work*, W4A '17. Association for Computing Machinery.

Rello, L. and Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. *Proceedings of the 12th Web for All Conference W4A '15*, pages 1–8.

Rello, L., Romero, E., Rauschenberger, M., Ali, A., Williams, K., Bigham, J. P., and White, N. C. (2018). Screening dyslexia for english using HCI measures and machine learning. In Kostkova, P., Grasso, F., Castillo, C., Mejova, Y., Bosman, A., and Edelstein, M., editors, *Proceedings of the 2018 International Conference on Digital Health, DH 2018, Lyon, France, April 23-26, 2018*, pages 80–84. ACM.

Shamsuddin, S. N. W., Mat, N. S. F. N., Makhtar, M., and Isa, W. M. W. (2017). Classification techniques for early detection of dyslexia using computer-based screening test. *World Applied Sciences Journal*, 35(10).

Spoon, K., Crandall, D., and Siek, K. (2019). Towards detecting dyslexia in children's handwriting using neural networks. In *ICML Workshop on AI for Social Good*.

Tamboer, P., Vorst, H. C. M., Ghebreab, S., and Scholte, H. S. (2016). Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *NeuroImage. Clinical*, 11:508–514.

Tamboer, P., Vorst, H. C. M., and Oort, F. J. (2014). Identifying dyslexia in adults: an iterative method using the predictive value of item scores and self-report questions. *Annals of Dyslexia*, 64(1):34–56.

Van den Audenaeren, L., Celis, V., Vanden Abeele, V., Geurts, L., Husson, J., Ghesquière, P., Wouters, J., Loyez, L., and Goeleven, A. (2013). Dysl-x: Design of a tablet game for early risk detection of dyslexia in preschoolers. In *Games for Health*, pages 257–266. Springer Fachmedien Wiesbaden.

Vellutino, F., Fletcher, J., Snowling, M., and Scanlon, D. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of child psychology and psychiatry, and allied disciplines*, 45 1:2–40.

Waesche, J., Schatschneider, C., Maner, J., Ahmed, Y., and Wagner, R. (2011). Examining agreement and longitudinal stability among traditional and rti-based definitions of reading disability using the affected-status agreement statistic. *Journal of learning disabilities*, 44:296–307.

Wagner, R. K. (2018). Why is it so difficult to diagnose dyslexia and how can we do it better? https://dyslexiaida.org/.

Wery, J. J. and Diliberto, J. A. (2017). The effect of a specialized dyslexia font, opendyslexic, on reading rate and accuracy. *Annals of dyslexia*, 67(2):114–127.

Wiederholt, J. L. and Bryant, B. R. (2012). *(GORT-5) Gray Oral Reading Test, Fifth Edition*. Pro-Ed.