Dytective: Diagnosing Risk of Dyslexia with a Game

Luz Rello
HCI Institute
Carnegie Mellon University
luzrello@cs.cmu.edu

Miguel Ballesteros
LT Institute
Carnegie Mellon University
NLP Group
Universitat Pompeu Fabra
miguel.ballesteros@upf.edu

Abdullah Ali
University of Maryland
Baltimore County
aali6@umbc.edu

Miquel Serra
Department of Basic Psychology
University of Barcelona
miquel.serra@ub.edu

Daniela Alarcón Sánchez
Change Dyslexia
daniela@changedyslexia.org

Jeffrey P. Bigham
HCI & LTI Institutes
Carnegie Mellon University
jbigham@cs.cmu.edu

ABSTRACT

More than 10% of the population has dyslexia, and most are diagnosed only after they fail in school. This work seeks to change this through scalable early detection via machine learning models that predict reading and writing difficulties by watching how people interact with a linguistic web-based game: *Dytective*. The design of *Dytective* is based on *(i)* the empirical linguistic analysis of the errors that people with dyslexia make, *(ii)* principles of language acquisition, and *(iii)* specific linguistic skills related to dyslexia. Experiments with 243 children and adults (95 with diagnosed dyslexia) revealed differences in how people with dyslexia read and write. We trained a machine learning model that was able to predict dyslexia with 83% accuracy in a held-out test set with 100 participants. Currently, we are working with schools to put our approach into practice at scale to reduce school failure as a primary way dyslexia is diagnosed.

Keywords

Dyslexia; Screening; Diagnosis; Serious Games; Linguistics

Categories and Subject Descriptors

K.3 [**Computers in Education**]: Computer Uses in Education—*Computer-assisted instruction*.

1. INTRODUCTION

More than 10% of the population has dyslexia. For instance, the U.S. Congress reported that from 10 to 17.5% of the population has dyslexia. Dyslexia has a neurological origin, and results in difficulty with reading and writing [13, 33]. If people know they have dyslexia, they can train with effort over time to overcome its negative effects.

Figure 1: *Dytective* is a web-based game designed to detect dyslexia in an affordable and scalable way. Players complete linguistically motivated activities designed to reveal differences between people with and without dyslexia. This photo shows a stage in which players hear a syllable that they should click, *e.g.*, and then click that letter as many times as possible within a time limit.

When people with dyslexia are not diagnosed and provided with appropriate support, they often fail in school. For instance, the Spanish Ministry of Education states that over 40% of the school dropout rate is due to dyslexia [16]. School failure is a primary way that people first learn that they might have dyslexia, which often comes too late for effective intervention. Students are not properly diagnosed because current procedures for diagnosis are expensive [18, 23] and require professional oversight [6, 9, 31]. Our goal is for everyone to know as early as possible if they might have dyslexia. Our approach to achieving this goal is to make it easy, inexpensive, and even enjoyable to find out.

Our method[1] started with a large corpus of errors made by people with dyslexia in reading and writing tasks. We then

[1]This work is protected by a provisional patent application titled "Method to Detect Individuals with or at Risk of Neurodevelopmental Specific Learning Disorders using Human Computer Interaction" filed on March 20, 2015.

PervasiveHealth 2016, May 16-19, Cancun, Mexico

DOI 10.4108/eai.16-5-2016.2263338

created game activities with progressing levels of difficulty by leveraging theory of linguistic and attentional tasks that are challenging for people with dyslexia. Finally, we packaged these activities into a web-based game called *Dytective*.

Dytective records a wide variety of human-computer interaction measures from people playing the game, *e.g.* mouse movements, click times, errors. A machine learning model trained on 243 participants (95 with diagnosed dyslexia) is able to correctly determine if a person has (or has not) dyslexia with 83% accuracy on a held-out test set of 100 participants. A number of challenges make this prediction difficult: *(i)* our activities are based on linguistic exercises, and so may not work well for children still learning to read, *(ii)* many people with dyslexia have already started to work to overcome the effects of dyslexia and so may perform better than we expect, and *(iii)* absolute ground truth is difficult to obtain because even professional diagnoses are imperfect.

2. DYSLEXIA

In the *Diagnostic and Statistical Manual of Mental Disorders (DSM-V)*, dyslexia is listed as a *specific learning disorder* having a neurological origin [1]. Dyslexia is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling. These difficulties typically present as a deficit in the phonological component of language that is unexpected in relation to other cognitive abilities.

2.1 Why is Dyslexia Difficult to Detect?

Even in the UK, a country that effectively treats dyslexia as compared with other countries, only 5% of the individuals with dyslexia are diagnosed and given appropriate help; it is estimated that over 85% of adult illiterates have dyslexia [7]. Even if research agrees in the neurological universality of dyslexia, its manifestations are different across languages, depending on the grade of regularity of the language orthographies.

For instance, English has an opaque –or deep– orthography (the relationships between letters and sounds are inconsistent) and Spanish has a transparent –or shallow– orthography with more consistent mappings between letters and sounds [27]. While dyslexia manifestation in languages with opaque orthographies are related to reading and writing performance, the manifestations of dyslexia in languages with shallow orthographies are not that evident, with reading speed and fluency the main predictors [26]. While in an English speaking country a child that read slower but accurate might not been diagnosed with dyslexia; in a Spanish speaking country it could be diagnosed as dyslexic. In fact, dyslexia has been called a *hidden* disability due to the difficulty of its diagnosis in languages with shallow orthographies.

3. RELATED WORK

Lyytinen *et al.* [15] created the computer game *Literate*, later called *GraphoGame* [14], to identify children at risk of having dyslexia before school age in Finland. Its exercises are aimed towards the connection of graphemes (letters) and phonemes (sounds). They conducted two user studies with 12 and 41 children between 6 and 7 years old with promising results. The authors provide statistical differences between populations but they do not run any machine learning prediction model. In comparison *Dytective* aims at people from all ages starting from 7 years old and cover a wider spectrum of cognitive skills including different levels of language and attentional abilities.

There are three projects on developing games to predict dyslexia in pre-readers that, to the best of our knowledge, did not report any prediction results. First, Gaggi *et al.* [12] tested a game with 24 pre-schoolers in Italy, that aimed at eye-hand coordination, visual spatial attention, rapid speech-sound identification and discrimination as well as visual-to-speech sound. Second, Van den Audenaeren *et al.* [32] performed a user study with 20 pre-schoolers in Flanders and are currently developing the game *DYSL-X* for early risk detection of dyslexia, which includes letter and end-phoneme recognition as well as psycho-acoustical tests.

3.1 Why is Dytective Different?

Dytective differs from prior approaches in its content design and prediction model. First, the content of *Dytective* are exercises based on *(i)* the empirical linguistic analyses of the errors that people with dyslexia make, *(ii)* principles of language acquisition, and *(iii)* specific linguistic skills related to dyslexia. Second, this first game to use human-computer interaction measures to train a machine learning model to predict dyslexia. That is, to the best of our knowledge, *Dytective* is the first game that aims at screening dyslexia in Spanish applying machine learning to measures extracted from linguistic and attentional exercises designed on the basis of generated content by people with dyslexia.

4. METHOD: DYTECTIVE GAME DESIGN

Dytective is designed for dyslexia detection. It targets linguistic and attention abilities associated with having dyslexia. Players proceed through a series of timed stages composed of linguistic exercises of increasing difficulty. Detection is possible when people with dyslexia perform differently than those without dyslexia. To increase the likelihood of this happening, we designed the exercises using a corpus of real errors produced by people with dyslexia that we collected. Therefore, we expect that people with dyslexia will make more errors, and thus our method will be able to differentiate.

The goal for *Dytective* players is to solve as many linguistic problems as possible within a time limit. For instance, in Figure 2 players need to correct as many errors as possible (right), or hear a non-word and then click it as many times as possible within the time limit (left).

4.1 Content Design

Dytective has 17 stages, split into 32 levels consisting of 212 exercises in total. Each stage targets a different linguistic or attentional skills and all together aim to cover the maximum potential indicators of dyslexia that could be caught via a computer-game. The exercises are intended for 7 year old and older.

4.1.1 Errors as Source of Knowledge

Most approaches to diagnosis are based on reading and writing measures such as the the number of words read per minute or comprehension of written material. *Dytective* uses linguistic and attentional exercises designed to distinguish populations with and without dyslexia. Empirical analyses show that errors made by people with dyslexia are different from the errors made by people without dyslexia [29] and,

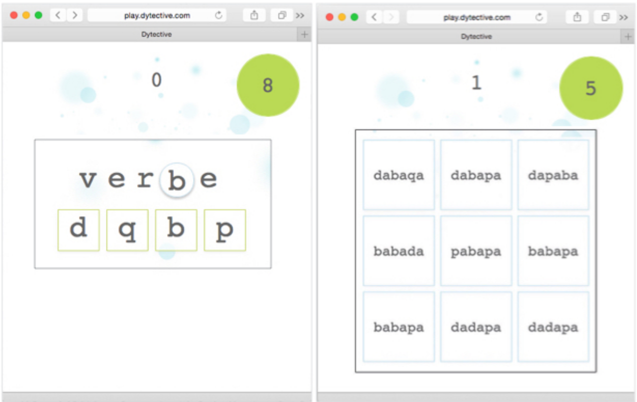

Figure 2: *Dytective* screenshots of first (left) and last (right) exercises of Stage 1, illustrating one way that difficulty is increased progressively.

have been successfully used for dyslexia intervention [24]. Hence, the criteria for the linguistic exercises were built on the basis of an analysis of an existing resource of errors made by people with dyslexia [22].

We analysed the errors from a *visual* point of view (shapes and visual features shared by the letters of other linguistic segments involved in the errors) and from a *linguistic* point of view, taking into consideration all the language levels that were involved in the errors, mainly, phonetic, phonologically, morphologically and syntactic levels. The most frequent linguistic and visual features shared in the errors were incorporated into the exercises, as described below.

4.1.2 Targeting Linguistic and Attentional Skills

We manually created different kinds of exercises to cover the maximum number of linguistic and attentional abilities related to dyslexia and whose performance can be catched via a computer-based game. In Table 1 we present the cognitive skills that each exercise type targets (all the exercises can be grouped in 17 types of exercises or stages). Most of the exercises address phonological awareness because that meta-cognitive skill is the primary factor in solving reading and spelling problems, also in the case of dyslexia [3, 13]. Indirectly, all the exercises target visual attention skills which have been found to have a causal relationship with reading acquisition in the case of dyslexia [8].

4.1.3 Linguistic Content and Ranking Criteria

The *Dytective* exercises get progressively more difficult both in later stages and within each stage [10]. Each stage is composed of a number of exercises, ranging from four to twenty-six exercises and they are ranked by their difficulty. We selected the linguistic input of the exercises, *e.g.*, letters, syllables, etc., using both linguistic patterns extracted from our error analysis, and the order in which the linguistic elements and structures are naturally acquired [19]. Thus, exercises that appear earlier should be those that are easiest for people with dyslexia to complete, and should also be those that are easiest for younger children to complete.

In higher difficulty levels, the target letter, syllable, or word(s) tends to be less frequent, longer, has a more complex morphology, and has a higher phonetic and orthographic similarity with other words. Both, error analyses and previous literature show that phonetic and orthographic similarity makes it more challenging for people with dyslexia as together with frequency, length, morphological complexity.

- At **Stages 1** to **5** the player hears a letter and needs to map it to its visual representation (Stage 1), recognize letters by sound, not letter names (Stage 2), map the syllable they hear with what they see (Stage 3) or recognize words and non-words, respectively (Stages 4 and 5). Exercise complexity is added gradually, higher levels gradually include distractors [2] that share more phonetic features or visual features with the target.

- In **Stage 6** players are presented with a number of letters and need to spot the one that is different *(Visual Attention)*. Distractors gradually become more phonetically and orthographically similar (Table 1).

- In **Stages 7** to **12** players must produce correct words by fixing errors based on the real errors that people with dyslexia make. These exercises target *Phonological Awareness* at a lexical (word) level. They were designed based on the type of errors that appear in texts written by people with dyslexia, such as *addition* of letters, **arround (around)* (Stage 7); *omission* of letters, **emty (empty)*; *substitution* of letters, **scholl (school)*; *transposition* of letters, **littel (little)*; and *word boundary errors* such as split words, **mis understanding (misunderstanding)*, and run-ons, **alot (a lot)*. Depending on the fixing operation the exercises are grouped in following stages. Stage 7 (Insertion of a letter); Stage (Substitution of a letter); Stage 9 (Reordering letters; Stage 10 (Reordering syllables); Stage 11 (split a string of characters into words) and Stage 12 (Deletion of a letter). The exercises of these stages were selected from the game *Piruletras* used to train the spelling of children with dyslexia [24].

- At **Stages 13** and **14** the player needs to spot written errors in sentences. The reason why we included these exercises is because one of the main challenges that people with dyslexia face is that they do not consciously detect errors while reading [2]. We used real-word errors (correctly spelled words that are not the one the user intended to write, *i.e. a letter *form you* instead of *i.e. a letter from you*). In Stage 13 the errors occur in lexical words[3] and in Stage 14 in function words.[4] This way each of the group of exercises aim at different comprehension language levels and linguistic skills, *Syntaxic Awareness* and *Semantic Awareness*, respectively. We also made this differentiation because lexical and function words are processed differently [19].

- At **Stage 15** the user needs to memorize sequences of letters with increasing difficulty *(Visual Memory &*

[2]Distractors are incorrect options in a multiple-choice question, designed to resemble the correct answer [17].
[3]Lexical words are content words, *i.e.* nouns, verbs, adjectives and most adverbs. They have a lexical meaning in contrast with the grammatical meanings expressed by function words, such as prepositions or conjunctions.
[4]Function words are words that have little lexical meaning, but instead serve to express grammatical relationships with other words within a sentence, such prepositions, pronouns, or conjunctions.

Exercise Type	Cognitive Skill	Example	Exercise Type	Cognitive Skill	Example
Stage 1 Letter recognition by name	*Orthographic Processing*	u i e i e u e o i	**Stage 2** Letter recognition by sound	*Phonological Awareness*	c g c d a e g g d c b d e d b c
Stage 3 Syllable recognition	*Phonological Awareness*	qu nu ba an da ad qu na ba up ab eb de ap du an an ne eb un be pu ba pa ad	**Stage 4** Word recognition	*Word Recognition*	pala mala sala gala mala dala pala mala dala
Stage 5 Non-word recognition	*Phonological Memory*	padama madapa padama dapama bamapa bamapa damapa pamaba madapa	**Stage 6** Letter differentiation	*Visual Attention*	n n n u n n n n n n n n n n n n
Stage 7 Insertion of letter	*Phonological Awareness*	lu_es s c m z u	**Stage 8** Substitution of letter	*Error Correction & Phonological Awareness*	h a z e r b t d l
Stage 9 Letter ordering	*Phonemic Segmentation & Phonological Awareness*	s e i t e	**Stage 10** Syllable ordering	*Syllabic Segmentation & Phonological Awareness*	be da bi
Stage 11 Sentence segmentation	*Word Recognition*	L a p a l o m a	**Stage 12** Deletion of letter	*Phonological Awareness*	b e b d e r
Stage 13 Error detection	*Syntaxic Awareness*	Está la final de la sala.			
Stage 14 Error detection	*Semantic Awareness*	Voy a la pastelería a comparar un pastel.			
Stage 15 Letter sequence memorization	*Visual Memory & Working Memory*	pgdj q w e r t y u i o p a s d f g h j k l z x c v b n m			
Stage 16 Word dictation	*Word Writing*				
Stage 17 Non-word dictation	*Non-Word Writing & Phonological Memory*				

Table 1: Linguistic and attentional abilities targeted by each stage in *Dytective*. The 17 stages are split into 32 levels consisting of 212 exercises in total.

Working Memory). That is, increasingly the sequence to remember contain letters that are less frequent, orthographically less transparent,[5] and share visual and phonetic features among each other.

- Finally, at **Stages 16** and **17** measure the writing performance and the *Phonological Memory* via Word and Non-Word Dictation. The criteria to select the words and the non-words are same for the rest of the Stages. For instance, the first exercises start with lexically simpler non-words, *i.e. tada* while in the higher levels the player is asked to write *mabadana* whose letters are more likely to be mistaken by people with dyslexia according to our empirical analyses of errors, because they share phonological and visual features.

4.2 Interface Design

Since text presentation significantly impacts the text readability of people with dyslexia we used black text on a white background, a large font size (minimum 18 points) and the monospaced *Courier* font face, which benefits both populations with and without dyslexia [21].

4.3 Implementation

Dytective is a web based game written in HTML5, CSS and Javascript with a backend PHP server and a database. By using these web technologies it is possible to play *Dytective* on different devices such as desktops, tablets, and mobile phones. It was implemented with a high level abstraction to make it easily portable to native iOS or Android application for future implementations.

5. EXPERIMENTAL STUDY

We conducted a study with 243 participants (95 with dyslexia diagnoses) using a within-subject design. The goal of the study was to collect data needed to run a machine learning experiment to find out if error-based linguistic problems can predict dyslexia in Spanish. All of the participants played all stages of *Dytective* over 15 minutes, but may not have advanced through all of the exercises in each stage.

We conducted a parallel study with 100 participants that is only use as held-out test set for the evaluation and the results are reported in Section 7.2.

5.1 Participants

Participants were recruited through a public call that dyslexia associations distributed to their members; 95 participants had a confirmed diagnosis of dyslexia including the date the the place where they were diagnosed; 31 were at risk of having dyslexia or suspected that have dyslexia.[6]

The participants with dyslexia or at risk of having dyslexia consisted of 126 people (68 female, 58 male). Their ages ranged from 7 to 68 ($M = 20.09, SD = 14.47$). The group of participants without dyslexia was composed of 117 people (74 female, 43 male). Their ages ranged from 7 to 70 ($M = 18.62, SD = 11.94$).

[5] Letter whose sound correspondence is not straightforward, that is, letters that can correspond to different sounds depending on the context, for instance, letter *c* can be pronounced as /k/ in *casa, 'house'* or as /θ̟/ in *cereza, 'cherry'.*

[6] Except for 17 adults, all were under observation by professionals, the step before having an official diagnosis.

All the participants' first language was Spanish, although 18 with dyslexia and 10 without dyslexia were fully bilingual in Spanish and another language spoken in the area where they live, such as Basque, Catalan, and Galician. 74 participants with dyslexia had previously failed Spanish subject at school, compared to only 30 participants without dyslexia.

5.2 Dependent Measures

To quantify performance, we used the following *dependent measures* extracted for each group of exercises: (i) Number of *Clicks* per stage; (ii) *Hits*, *i.e.* number correct answers; (iii) *Misses*, *i.e.* number in correct answers; (iv) *Score i.e.* sum of correct answers per group of exercises; (v) *Accuracy* defined as the number of *Clicks* divided by the number of *Hits*; (vi) *Missrate* defined as the number of *Clicks* divided by the number of *Misses*.

5.3 Materials and Procedure

We sent an announcement of the study to the main associations of dyslexia of hispanic countries and countries with large Spanish speaking populations, mainly Argentina, Chile, Mexico, Spain and the USA. We also sent the call to specialized centers that support people with dyslexia. Interested potential participants contacted us, and after we checked the participation requirements (age, mother languages and technical requirements) we set up a date to supervise the study. We met with the participants (and their parents in case the participant was underage) online or by telephone. After they signed the online consent and/or parental consent we gave them specific instructions and they completed the study. Parents were specifically warned that they could not helped their children to play *Dytective* and were asked again afterwards to double check. The data of two participants with dyslexia had to be deleted because their mothers explained to us that helped their children. One school and one specialized center collaborated in the study. For these cases the parental consent was obtained in advance and the study was supervised by the school counselor and the therapist respectively. We deliberately carried out the study in three different settings (home, school and a specialized centre) so our results are independent of the these settings.

6. DATASET

Our dataset was derived from the experimental study presented in the previous section with 243 participants (95 with diagnosed dyslexia) and a test set with 100 participants (10 with diagnosed dyslexia) that is representative of the population.[7] The dataset is composed of 197 features per participant, that is, 47,871 data points. From the dataset we extracted the following features, marked as D if the participant has dyslexia, N if not, and M (maybe) if the participant suspects that he or she has dyslexia but is not diagnosed.

[1] Age of the participant ranging from 7 to 70 years old.

[2] Gender of the participant, a binary feature with two values *female* and *male*.

[3] Second mother language in case of bilingualism; all the participants had Spanish as mother language.

[7] Around 10% of the population has dyslexia.

[4] Spanish subject. This is a binary feature with two values, *yes* when the participant has ever failed Spanish subject at school and *no* when the participant have never failed that subject among all the school history.

[5-197] Performance measures. These features correspond with the six dependent measures (*Clicks*, *Hits*, *Misses*, *Score*, *Accuracy*, and *Missrate*) we gather per level played (32 levels), that is, 192 features corresponding to different cognitive skills (Table 1).

Some of the features have numeric (real or integer) values, so we established ranges for each of them to discretize the data by the population median.

7. RESULTS

In order to find out whether it is feasible to detect people with dyslexia after playing *Dytective*, we set up a machine learning experiment. Machine learning is the scientific discipline that studies algorithms that can learn from data and make predictions. The output of a machine learning algorithm is called a model, which is capable of making predictions given unseen data. In this case, the goal is to predict whether someone has dyslexia or not based on the data collected while participants played *Dytective*.

We used the binary classifier LIBSVM [5] in the polynomial Support Vector Machine (SVM) set-up. A SVM is a method for supervised learning that analyzes data and recognizes patterns for classification. Given a set of training examples, each marked as belonging to a category (in our case either having dyslexia or not), an SVM training algorithm builds a model that assigns new examples into the categories. When there is an input for the classifier it tries to assign a category to the input and then this is the classification output. Our SVM is trained on datasets like the one described in the Dataset Section, and it is able to perform predictions on new participants that may play *Dytective*.

We performed a cross validation experiment by dividing the dataset in 243 different subsets having only one participant each. We then iteratively trained a statistical model on all the data but one participant (242 participants) and tested the one held out. At the end, we had all the data tested independently. The participants marked as M (maybe) are used for training the models as if they are D (participants with dyslexia) but they are not used for evaluation. This means that we have 212 participants to test, and we train each model, a total of 212 models, with 242 experiments performed by participants.

The initial results suggest that the model predicts people having or not having dyslexia quite accurately, with a final result of 81.60% in the cross-validation experiment by using all features (151 performance features extracted from the game, plus age, gender, mother language and school performance). This means that the model correctly classifies 173 of the 212 participants.

7.1 Optimization

In order to improve performance, we carried out a feature selection experiment following a backward algorithm. We start testing a model with all features, and we iteratively removed features one by one by training new models; if the performance was better or equal than before we permanently remove the feature from the feature set, seeking more informative features.

After we have a partially reduced the feature set, we carried out a redundancy selection experiment, in which we removed features in pairs, by testing all possible combinations in a double loop, meaning that we fix a particular feature and we start removing all feature plus the particular feature that we have fixed before. If the performance is better or equal than before we remove the pair of features from the feature set.

After the optimization round we obtained an improved result of 85.85%, which increases the previous score substantially and reduces the number of features, from 198 to 150. Some of the dependent measures from some Stages were left out, such as the number of *Clicks* in Stage 4. The model selected features from all the stages. The model is now capable of correctly predicting the condition of 182 of the 212 participants.

	Score
Accuracy	83%
Precision – Class Dyslexia	36.0%
Recall – Class Dyslexia	90.0%
Precision – Class Not-Dyslexia	98.7%
Recall – Class Not-Dyslexia	82.2%

Table 2: Classifier accuracy in the cross validation experiment, using the optimized feature set and the ablated conditions in which all the features from a particular Stage are removed. The last row shows the result in which all the stages are included. Every feature was necessary for the highest accuracy rate.

7.2 Results in the Held-out Test-Set

We compiled a held-out test with 100 participants, 10 of them with dyslexia, that is similar to the percentage of people with dyslexia that we have in our society. We train a model in the entire training set and we report the results on the held-out test. The results are shown in Table 2.

The first conclusion to extract is that the model is capable of detecting the risk of having dyslexia regardless of the distribution of people we have in our dataset.

Of course, better results are expected if we increase the size of the training set. However, 83% means that we are able to predict the risk of having dyslexia.

8. DISCUSSION

Our machine learning model trained on data from *Dytective* is able to classify people as having dyslexia or not with high accuracy. In this section, we explore the utility of the different features used and discuss the model's errors.

8.1 Feature Analysis

Table 3 depicts the information gain of each of the features for each stage. The left hand side table shows the result of the model when we only use features from a single stage, and the right hand side table shows the results of the model when we use all the selected features except for the features of a particular stage. The main conclusion is that all the stages are needed; the tasks and the cognitive skills that target

Only	**Accuracy**
1	72.17%
2	70.75%
3	65.09%
4	67.92%
5	69.34%
6	61.79%
7	71.70%
8	74.06%
9	74.53%
10	72.17%
11	72.17%
12	64.62%
13	68.40%
14	63.68%
15	73.11%
16	75.94%
17	73.11%

Without	**Accuracy**
1	75.94%
2	74.06%
3	76.42%
4	76.42%
5	78.30%
6	79.25%
7	76.42%
8	76.42%
9	76.89%
10	79.72%
11	77.36%
12	77.83%
13	78.77%
14	81.13%
15	81.60%
16	79.72%
17	83.49%

Table 3: Classifier accuracy in the cross validation experiment, using the optimized feature set and only the features from the given stage (left) and using all features except those from the given stage (right).

each stage complement each other, capturing linguistic and attentional difficulties that people with dyslexia have.

The most informative stage by itself is Stage 16 (Word dictation), which achieves an accuracy of 75.94 using only performance measures extracted from that stage. However, when using the rest of the features, Stage 16 is more redundant than others being the Stage 2 (Letter recognition by sound) the most relevant in order to achieve a higher performance when all the stages are taking into account. When we remove the features from the Stage 2, the model only achieves accuracy of 74.06.

A possible explanation for why all the stages are informative could be that all indirectly –or directly in Stage 6 (Letter differentiation)– aim at *visual attention*. Previous studies have found that visual-spatial attention deficits are an indicator of dyslexia [8, 11]. It is also worth noting that this fact indicates that the method is more general and would not incorrectly classify people according to their age.

We do not directly measure reading speed even though that measure has been pointed out as the strongest indicator of dyslexia in languages with shallow orthographies such as Spanish [26]. However, some of our exercises require a little bit of reading (specially Stages 12, 13 and 14), so indirectly reading speed is considered.

8.2 Error Analysis

We analyzed the cases that our algorithm classifies participants wrong, the could be divided into three groups.

First, we found errors in in classifying young children without dyslexia (7-15 years old). A possible explanation of this is that at that age children are still acquiring basic reading skills and writing skills and the time of acquisition of those reading skills varies across subjects [34]. Therefore it is possible that some classified with risk of having dyslexia are just acquiring linguistic skills in a slower pace. This may be solved with more training data.

Second, we have a group of errors coming from teenagers and college students with dyslexia that are classified and people without dyslexia. These cases mainly correspond to people who was diagnosed a long time ago (more than five years) and that have received intervention since them. This is probably because these group have already learn how to overcome dyslexia and have developed the proper compensating strategies. In fact, previous studies show that with proper intervention the skills of people with dyslexia can be comparable to the rest [28], and there was even found evidence of amelioriation in brain regions associated with phonological processing [30].

Third, we observed that the algorithm fails with older adults (specially starting from 45 years old). For this group there are errors in both directions, people without dyslexia are classified as the risk group and *vice versa*. The possible explanation for this is the frequency of exposure to reading that the older participants have because the frequency in with people are exposed to words have a direct effect on their reading speed [4, 20]. While during school years that frequency is moderately homogenous due the schooling when it comes to adults it really depends on the personal choices and on the type of job that the participant has.

9. FUTURE WORK

Dytective presents a number of opportunities for future research. The current version only works for people with some reading ability, which means it cannot be used effectively for very young children. We plan to design a version of *Dytective* for pre-readers, starting from 3 to 6 years old.

Currently we are carrying out a large scale study with 10,000 participants to improve our prediction algorithm accuracy and being able to apply other machine learning approaches that require more data such as neural networks. We will use *Dytective* to detect dyslexia with students who have not yet been diagnosed. We will then follow them through their schooling to gauge the effectiveness of *Dytective* in its intended context.

Finally, we are extending *Dytective* to other languages [25], such as English and German, and to integrate the game into a free web-based platform that it can be easily shared. If we attract enough people to play the game, then we might be able to estimate the prevalence of dyslexia among web users, which is a long-standing challenge.

10. CONCLUSION

We have presented a game to detect dyslexia called *Dytective*. This game uses linguistic and attentional exercises that we designed in order to illustrate differences between people with and without dyslexia. A machine learning model trained on data collected from 243 *Dytective* players was able to predict dyslexia with 83% accuracy in a test set consisting of 100 new participants. Because *Dytective* is easy to access on the Web and does not require special equipment, it is easy to scale. Our hope is that it will lead to earlier detection of dyslexia and prevent children from being diagnosed with dyslexia only after they fail in school.

Acknowledgments

We thank CREIX Barcelona (Spain), Américo Vespucio school in Santiago (Chile), Colegios de la Hijas de María Auxiliadora (Salesianas, Spain), *Associació Catalana de Dislèxia*, *Associación DISFAM*, and *Madrid con la Dislexia*. Special thanks to Dr. María Teresa Moral and Mariana Moya for their help in the field work, and Nikki García for recording her voice for the game. This paper was developed under a grant from the US Department of Education, NIDRR grant number H133A130057.

11. REFERENCES

[1] A. P. Association. *Diagnostic and statistical manual of mental disorders, (DSM-V)*. American Psychiatric Publishing, Arlington, VA, 2013.

[2] M. Bruck. The word recognition and spelling of dyslexic children. *Reading Research Quarterly*, pages 51–69, 1988.

[3] M. Bruck. Persistence of dyslexics' phonological awareness deficits. *Developmental psychology*, 28(5):874, 1992.

[4] M. Carreiras, A. Mechelli, and C. J. Price. Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human brain mapping*, 27(12):963–972, 2006.

[5] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Trans. on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[6] F. Cuetos, B. Rodríguez, E. Ruano, and D. Arribas. *PROLEC-R. Batería de evaluación de los procesos lectores, revisada.* TEA, Madrid.

[7] Dyslexia Research Institute. Dyslexia, identification, January 2015. www.dyslexia-add.org.

[8] A. Facoetti, P. Paganoni, M. Turatto, V. Marzola, and G. G. Mascetti. Visual-spatial attention in developmental dyslexia. *Cortex*, 36(1):109–123, 2000.

[9] A. J. Fawcett and R. L. Nicolson. *Test para la detección de la dislexia en niños (DST-J)*. TEA, Madrid.

[10] C. C. Gabe Zichermann. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. O'Reilly, 2011.

[11] J. d. Gabrieli and E. S. Norton. Reading abilities: importance of visual-spatial attention. *Current Biology*, 22(9):R298–R299, 2012.

[12] O. Gaggi, G. Galiazzo, C. Palazzi, A. Facoetti, and S. Franceschini. A serious game for predicting the risk of developmental dyslexia in pre-readers children. In *Proc. ICCCN'12*, pages 1–5. IEEE, 2012.

[13] International Dyslexia Association. Frequently Asked Questions About Dyslexia, 2011. www.interdys.org.

[14] H. Lyytinen, J. Erskine, J. Kujala, E. Ojanen, and U. Richardson. In search of a science-based application: A learning tool for reading acquisition. *Scandinavian journal of psychology*, 50(6):668–675, 2009.

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

[16] Ministerio de Educación, Cultura y Deporte. *Informe Redie. La atención al alumnado con dislexia en el sistema educativo*. Gobierno de España, 2012.

[17] R. Mitkov, L. A. Ha, A. Varga, and L. Rello. Semantic similarity of distractors in multiple-choice tests: extrinsic evaluation. In *Proc. EACL Workshop GeMS '09*, pages 49–56, 2009.

[18] D. L. Molfese. Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain and language*, 72(3):238–245, 2000.

[19] S. Pinker. *Language Learnability and Language Development*. Harvard University Press, 2009.

[20] L. Rello. *DysWebxia. A Text Accessibility Model for People with Dyslexia*. PhD thesis, Universitat Pompeu Fabra, 2014.

[21] L. Rello and R. Baeza-Yates. Good fonts for dyslexia. In *Proc. ASSETS'13*, Bellevue, Washington, USA, 2013. ACM Press.

[22] L. Rello, R. Baeza-Yates, and J. Llisterri. DysList: An annotated resource of dyslexic errors. In *Proc. LREC'14*, Reykjavik, Iceland, May 2014.

[23] L. Rello and M. Ballesteros. Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proc. W4A '15*, Florence, Italy, 2015.

[24] L. Rello, C. Bayarri, Y. Otal, and P. Pielot. A computer-based method to improve the spelling of children with dyslexia using errors. In *Proc. ASSETS'14*, Rochester, USA, October 2014.

[25] L. Rello, K. Williams, A. Ali, N. Cushen White, and J. P. Bigham. Dytective: Towards detecting dyslexia across languagesusing an online game. In *Proc. W4A'16*, Montreal, Canada, 2016. ACM Press.

[26] F. Serrano and S. Defior. Dyslexia speed problems in a transparent orthography. *Annals of Dyslexia*, 58(1):81–95, 2008.

[27] P. H. K. Seymour, M. Aro, and J. M. Erskine. Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2):143–174, 2003.

[28] S. E. Shaywitz. *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. Knopf, 2003.

[29] C. Sterling, M. Farmer, B. Riddick, S. Morgan, and C. Matthews. Adult dyslexic writing. *Dyslexia*, 4(1):1–15, 1998.

[30] E. Temple, G. K. Deutsch, R. A. Poldrack, S. L. Miller, P. Tallal, M. M. Merzenich, and J. d. Gabrieli. Neural deficits in children with dyslexia ameliorated by behavioral remediation: evidence from functional mri. *PNAS*, 100(5):2860–2865, 2003.

[31] J. Toro and M. Cervera. *TALE: Test de Análisis de Lectoescritura (TALE: Literacy Analysis Test)*. Visor, Madrid, 1984.

[32] L. Van den Audenaeren, V. Celis, V. V. Abeele, L. Geurts, J. Husson, P. Ghesquière, J. Wouters, L. Loyez, and A. Goeleven. DYSL-X: Design of a tablet game for early risk detection of dyslexia in preschoolers. In *Games for Health*. Springer, 2013.

[33] World Health Organization. *International statistical classification of diseases, injuries and causes of death (ICD-10)*. World Health Organization, Geneva, 1993.

[34] J. C. Ziegler and U. Goswami. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, 131(1):3, 2005.