IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 December 2024, accepted 21 January 2025, date of publication 7 February 2025, date of current version 18 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3539719

APPLIED RESEARCH

Screening Dyslexia Using Visual Auditory Computer Games and Machine Learning

MARIA RAUSCHENBERGER[1], RICARDO BAEZA-YATES[2], (Fellow, IEEE), AND LUZ RELLO[3]
[1]Faculty of Technology, University of Applied Science Emden/Leer, 26723 Emden, Germany
[2]Institute for Experiential AI, Northeastern University at Silicon Valley, San Jose, CA 95113, USA
[3]IE Business School, IE University, 28046 Madrid, Spain

Corresponding author: Luz Rello (luz.rello@ie.edu)

This work was supported in part by the fem:talent Scholarship from the Applied University of Emden/Leer, and in part by the Deutschen Lesepreis 2017 from the Stiftung Lesen and the Commerzbank-Stiftung.
This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ministry of Education, Science and Culture of Schleswig-Holstein (*Ministerium für Bildung, Wissenschaft und Kultur, MBWK)* and by the *Education Authority* of the State of Lower Saxony (*Niedersächsische Landesschulbehörde*).

ABSTRACT Reading acquisition is one the main keys for school success and a crucial component for empowering individuals to participate meaningfully in society. Yet, it is still a challenging skill to acquire for around 10% of children that have dyslexia, a type of neuro-developmental disorder that affects the ability to learn how to read and write. Dyslexia is often under-diagnosed, and normally children with dyslexia are only detected once they fail in school, even though dyslexia is not related to general intelligence. In this work, we present an approach for screening dyslexia using language-independent games in combination with machine learning models. To reach this goal, we designed the content of a computer game, collected data from 137 children playing this game (51 with dyslexia) in different languages -German, Spanish and English- and created a prediction model using different machine learning classifiers. Our method provides a precision of 0.78 and recall of 0.79 for German and a precision of 0.83 and recall of 0.80 for all languages when Extra Trees are used, with an accuracy of 0.67 and 0.75, respectively. Our results open the possibility of inexpensive online early screening of dyslexia for young children using non-linguistic elements.

INDEX TERMS Dyslexia, language disorder, language-independence, machine learning, reading disorder, dyslexia screening, serious games.

I. INTRODUCTION

Achieving reading fluency is considered a critical component for empowering individuals to participate meaningfully in society and is a major contributor to improved livelihoods. However, globally, up to 250 million children are unable to acquire basic literacy skills [1]. While there are many factors which contribute to poor learning outcomes, one widespread challenge is developmental dyslexia, a reading-specific disability, which by some estimates may affect up to 10-15% of the global population [2]. Despite being one of the most common learning disabilities and accounting for up to 80% of diagnosed learning disabilities [3], developmental dyslexia is still under-diagnosed and often goes untreated, leading in many cases to school failure. In fact, nowadays learning difficulties are the primary cause of eventual school failure and school dropout [4]. For instance, 35% of persons with dyslexia drop out of school, and it is estimated that less than 2% of persons with dyslexia will complete a four-year college degree [5]. Moreover, school failure can lead to an excluded population with limited ability to become productive adults who function independently in society. For instance, it has been found that dyslexia negatively impacts workplace performance as well as career progression [6]. So, we need to address this accessibility challenge [7].

Dyslexia has been called a ‘‘hidden disability’’ due to the difficulty of its diagnosis [8]. There are many reasons that make dyslexia detection challenging. A person with

The associate editor coordinating the review of this manuscript and approving it for publication was Nkaepe Olaniyi.

dyslexia demonstrates normal or high levels of intellectual functioning [2] and is able to compensate for these deficits, making dyslexia hard to detect. Also, there is a high comorbidity of dyslexia with other disorders, such as attention-deficit hyperactivity disorder or dyscalculia. Hence, detecting dyslexia normally requires scarce resources such as a professional assessment or even special hardware. For instance, some approaches detect dyslexia through electroencephalography (EEG) and Magnetic Resonance Imaging (MRI) brain scans. Finally, it is even more difficult to diagnose in languages with transparent orthographies, where symptoms of dyslexia are less severe, such in German or Spanish [9].

Deficits in children with dyslexia are ameliorated after remediation [10]. However, first they need to be detected as soon as possible because early intervention is key to avoid its negative effects such as school failure [11].

In this study, we focus on dyslexia detection using games and machine learning. For the game's content we did not use the core indicators of dyslexia, which are linguistic (reading and writing), but decided to focus on other indicators such as visual and auditory factors, also strongly associated with dyslexia. Our motivation for this is threefold: first, by including visual and acoustic content the results from this study could be potentially extended to other languages since the items of the tool are almost language independent. Second, visual and auditory perception are developed by children before they learn how to read; hence, our contributions can potentially be used in future research for early screening of dyslexia in pre-readers. And finally, the social impact, developing a method for online dyslexia screening could contribute to make dyslexia screening widely available for everyone.

The main contributions of this article are the game content, the data sets collected, and the screening results using machine learning prediction techniques that range from 78% to 83% in precision and recall, and from 67% to 75% in accuracy.[1]

The article is organized as follows. After the introduction, we provide the background to this study in Section II. Next, we present how we designed our screening method (Section III) and the study that we performed (Section IV) to collect our data sets that are presented in Section V. Finally, we present the machine learning models used and their setup (Section VI), followed by our results in Section VII. We end with a discussion and our conclusions (Sections VIII and IX).

II. BACKGROUND AND RELATED WORK

A. DYSLEXIA

The *American Psychiatric Organization* defines dyslexia as a *specific learning disorder* which affects from 5% to 15% of the global population [2]. According to the *International Dyslexia Association*, dyslexia is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unrelated to other cognitive abilities and the provision of effective classroom instruction [13].

Even though language acquisition depends on the syllabic complexity and orthographic depth of a language [14], results show that similarities between readers with dyslexia in English and German are far bigger than their differences [15]. Also, similar types of errors were found in texts written by people with dyslexia for English, German [16], and Spanish [17].[2]

The vast majority of the current studies agree on the deficit of the phonological component regarding dyslexia [8]. However, no scientific agreement of the causal origin has been achieved [19]. In fact, there are some studies that consider visual perception a key attribute for the cause of dyslexia [20], while others consider auditory perception to be a causal component [21]. Our screening method focuses specifically on these approaches, that link some visual and auditory manifestations to dyslexia.

B. AUDITORY PERCEPTION IN DYSLEXIA

Various studies found evidence linking basic auditory processing to phonological deficits of dyslexia [22], as well as to prosodic skills and phonemic awareness related to dyslexia [23].

Also, auditory perception of children with dyslexia has been proven to be related to sound structure [24] as well as to the auditory working memory [25]. On this line, the *rise time theory* suggests a connection between dyslexia and slow auditory procession or impaired discrimination of amplitude [26]. For instance, there has been found significant differences in the perceptions of readers with dyslexia on the syllable stress compared to those of the control group at the age of 9 [23]. People with dyslexia also present short-term memory difficulties [27], [28] as well as to recall of information chunks [29]. For example, musicians with dyslexia scored better in auditory perception tests than the general population but, at the same time, these participants score worse in tests addressing auditory working memory, *i.e.,* the ability to keep a sound in mind for seconds [25]. These difficulties can be also linked to the acoustic perception performance. For example, questions like *Which sound did you hear first* or *Which sound is pitched higher?* [24] could determine groups. Huss et al. [24] already showed significant differences on the performance of children with and without dyslexia (8 to 13 years old) using a musical metrical structure in a controlled setting.

Our goal aims to use these auditory indicators as a language-independent approach for dyslexia screening.

[1]The game content is freely available at https://github.com/Rauschii/DGamesContent. The datasets are freely available at https://github.com/Rauschii/DGamesDataSet. Most of these results are part of the Ph.D. thesis of the first author [12] and a demo of the game is available at http://bit.ly/DGamesEN

[2]The dataset for German error words is freely available at [18].

C. VISUAL PERCEPTION IN DYSLEXIA

Another line of research suggests that reading difficulties are due to visual-spatial attention problems and poor coding instead of phonological difficulties [20].

A three year longitudinal study with 96 Italian pre-readers (children in kindergarten) found that visual parietal-attention may explain future reading difficulties and as well as the development during the first and second grade of school. The participants performed a set of visual discrimination search tasks (searching for symbols), which showed significant differences in the error rate for poor readers in first grade compared to their peers. Hence, it is suggested that visual parietal-attention could be a predictor for future difficulties in reading acquisition [30]. More recently, a missing visual left-right asymmetry in adults with dyslexia has been proposed as one of the many possible causes of dyslexia [31].

Other evidence comes from the analysis of error words from children with dyslexia, that shows that dyslexic errors are visually motivated, that is, that letters that are more likely to be mistakes are visually similar. For instance, 38.23% of the errors written by children with dyslexia in Spanish has a rotation feature <b, d, p, q>, and 46.91% of the mistaken letters have vertical or horizontal symmetries such as <m, w> and <n, u>. However, this evidence is contested since some letters that are visually similar also present phonological similarities. For instance, the letters <b, d, p, q>, are plosive consonants, that is, they all share the same manner of articulation [32].

Nonetheless, these visual indicators, such as horizontal or vertical symmetry in visual representations, together with search tasks addressing visual-spatial attention, could be used in a language-independent approach to screen dyslexia.

D. SCREENING DYSLEXIA USING MACHINE LEARNING

Historically, the most common way to detect a person with dyslexia was applying some of the widely-used standard assessments that focus on different indicators of reading and writing performance, such as reading speed (words per minute), reading comprehension or spelling errors [33], [34].

More recently, a growing number of computer-based approaches to detect dyslexia have appeared [35]. Here, we only focus on the approaches that include machine learning models for dyslexia predictions.

Successful machine learning approaches to detect dyslexia have used different types of data, such as eye-tracking measures while reading using support vector machine models [36]; content from electroencephalography (EEG) using also support vector machine models [37], [38]; using handwriting with machine learning classifiers [39]; or by human computer interaction measures derived from linguistic games and applying Random Forests [40].

Since collecting health data is costly in terms of time and resources [41], these machine learning approaches to predict dyslexia have used small groups (small data), where precautions of over-fitting need to be addressed [42].

A larger study ($n = 200$) focusing on Chinese handwriting analysis employed machine learning classifiers to predict the risk of dyslexia, achieving a reported accuracy of 81% [43]. While most of the approaches use data derived from linguistic tasks to screen dyslexia, we found one machine learning method that uses auditory content *Lexa* [44]. This model reports an accuracy (89.2%) using features related and not-relevant to phonological processing. These features are collected with extensive tests and the machine learning classification is carried out on a small sample ($n = 56$), with no precautions reported about over-fitting.

1) ORIGINALITY OF OUR APPROACH

To the extent of our knowledge, we present a novel study, since there are no current approaches that apply machine learning methods using visual and auditory content together to screen dyslexia. This approach comprises the advantage of using non linguistic content which makes the approach potentially language independent and applicable to pre-readers, that is, to children who have not yet developed reading skills. Also, it offers the possibility of online screening using games making it easy, inexpensive, and even enjoyable.

This work is an extension of a previous study that explored how auditory elements could aid in dyslexia [45]. The preliminary results from that paper were used in the design of the exercises for this paper.

III. METHOD

Dyslexia has a neurological origin, it does not develop when a child starts learning how to read or write. Some authors have explored other dyslexia manifestations that could appear even before reading acquisition occurs, such as visual and auditory perception.

A. CONTENT DESIGN

To build the content, we considered evidence from the literature on dyslexia indicators related to visual and auditory perception. After that, we incorporated other general features associated with dyslexia, adapting them to be language-independent.

The aim of the game is to collect measures derived from the interaction with the game in order to find differences between children with and without dyslexia. Its duration is less than 15 minutes and has 16 stages. Each stage has 2 rounds. The game follows a Whac-A-Mole interaction where the participant needs to find a visual or an acoustic cue that has been seen or heard before, respectively.

1) AUDITORY CONTENT

To create the acoustic cues we combined the deficits of children with dyslexia in auditory working memory with the results of previous literature related to dyslexia. Auditory working memory helps a person to keep a sound in mind, and our games items focus on remembering different auditory cues. At the same time, we took into consideration the type of

TABLE 1. Description of the auditory attributes used in the method.

Key	Attribute	Description
B	Beginning	70% of the spelling errors are at the third position of a word for German and Spanish [18], [32].
L	Length	The average word length for German and Spanish is just above 7 letters [18], [32].
Si	Simple	For 73.3% of the analyzed words for Spanish the Damerau-Levenshtein distance was one, which means that only single mistakes were made [32]. For German it is 81.3% [18].
Su	Substitution	The error category *Substitution* (exchanging a letter for another one) is frequent for German, English and Spanish [18], [32].
O	Omission	The error category *Omission* (leaving a letter out) is frequent in German [18].
St	Structure	CwD find it more difficult to recall a target item with a similar prosodic structure [29].
Pst	Phonological short-term memory	CwD showed difficulties in the phonological short-term memory [29].
Sip	Short-interval perception	Copying and discrimination tasks are used to predict phonological awareness [46].
Pm	Pitch modulation	CwD have difficulties in processing pitch patterns [47].
Cb	Combinations	Discrimination of rise time is related to language processing [48].
C	Complexity	CwD have difficulties with *the phonological similarity effect* and the *phonological neighborhood* when long memory spans are used [29]. English has a greater percentage of multi-errors compared to Spanish [32].

TABLE 2. Mapping of auditory types and stages of the game.

Type No. Stages	Pho-neme 3	Con-fu sion 8	Com-bina-tion 1	O- -mis -sion 1	Rhy-thm 1	Struc-ture 1	Sub-sti-tution 1
Key							
B	✓		✓	✓		✓	✓
L	✓	✓	✓	✓		✓	✓
Si	✓			✓		✓	✓
Su	✓		✓				✓
O	✓		✓	✓	✓		
St	✓	✓	✓	✓	✓	✓	✓
Pst	✓	✓	✓	✓	✓	✓	✓
Sip		✓	✓		✓		
Pm	✓	✓	✓			✓	✓
Cb			✓		✓		
C	✓		✓		✓		

errors that children with dyslexia make with letters and words found by previous work [16], [32], [49], [50] and applied them into sounds.[3]

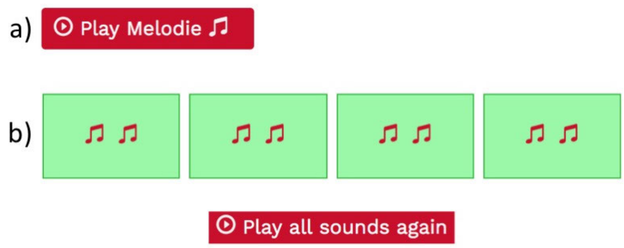

FIGURE 1. Example of the auditory part of the game with the priming of the target cue (a) and then the distractors for each auditory cue (b).

Then, we mapped the attributes explained in Table 1 to the game rounds in Table 2. Additionally, Table 2 shows the relations between our designed auditory types and the literature that provides evidence for distinguishing people with dyslexia.

The auditory part has an auditory type for each of the rounds: *substitution, omission, structure, phoneme* (one Spanish and one German vowel; Spanish consonant), *confusion* (twice Spanish and German; four times English), *combinations*, and *rhythm*. Each auditory type has one auditory cue target and three auditory cue distractors.

Some auditory types are partly related to linguistic features when using the pronunciation of letters or the confusion of words. For example, we created cues from the vowel pronunciation of letters using the Mac OS High Sierra 10.13.6 voice for the different languages, *e.g.,* Spanish and German. We chose vowels because vowel errors are the most frequent in the substitution category [32], [49].

The auditory game round has two phases: (1) remembering the target audio cue; and (2) finding the target audio cue among a collection of audio cues. In the first phase, children click on the *play* button and can listen to the auditory cue target as many times as they like (see Figure 1, a). In the second phase, a row of four buttons is displayed (see Figure 1, b) and, automatically, the assigned auditory cues for each button are played one after another. The buttons are disabled until the auto-play is done to ensure the children listen to all auditory cues. In order to distract the player, the first button/auditory cue is never the auditory cue target. The order of auditory cues is randomly assigned and always starts from left to right. With the *Play all sounds again* button, the children can listen to all cues as many times as they like.

2) VISUAL CONTENT

To design the visual cues and their distractors we also took into account the specific difficulties explained in literature.

Since visual search tasks addressing visual-spatial attention have been successfully used to predict risk of dyslexia in previous lab controlled experiments [20], [30], [31], the visual part of our games followed a search task interaction.

Also, the analyses of mistakes made by people with dyslexia suggest trends in dyslexic errors related to visual features, such as different types of symmetries and

[3] A video showing an example of the visual part of the game is available at https://youtu.be/lVxuNSMZXvE?si=yMwm62vJ4aRGfcHK. A video presenting an example of the auditory part of the game is available at https://youtu.be/TxG6okyh9jg

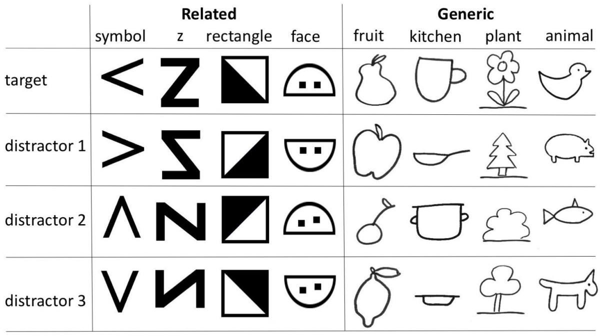

FIGURE 2. The figure shows the target cue (top) and distractor cues (below) for the eight different stages (*symbol, z, rectangle, face, fruit, kitchen, plant, animal*) of the visual part of the game.

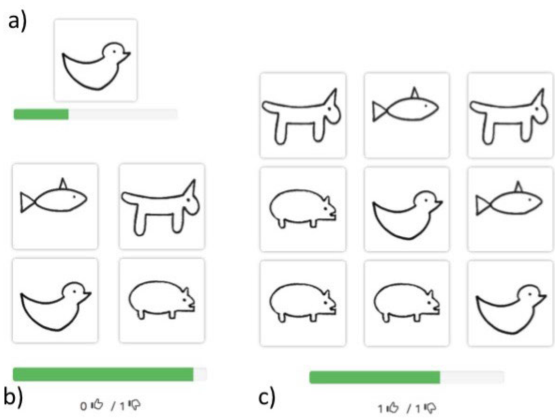

FIGURE 3. Example of the visual part of the game with the priming of the target cue *animal* (a) and then the four-squared (b) and (c) nine-squared design including the distractors for each *animal*.

similarities in letters. For instance, mirror letters such as <b> and <d> appear more frequently in dyslexic errors [32]. Hence, we designed our visual cues using visual features found in the error made by people with dyslexia, including horizontal and vertical symmetries that are known to be difficult for a person with dyslexia (Figure 2, left).

Regarding the interaction, at the beginning participants are shown the target visual cues for three seconds. They are asked to remember this visual cue. After that, the participants are presented with a setting where the target visual cue and distractors are displayed (Figure 2). The participants need to find and click on the target cue as many times as possible within a span of 15 seconds. The arrangement of the target and distractors cues changes randomly after every click.

The visual part of the game has a total of 8 stages. Each stage is assigned to one visual type (*symbol, z, rectangle, face, fruit, kitchen, plant* and *animal*). One visual cue is the target, which the participants need to find and click. The other three visual cues are *distractors* for the participants (Figure 2). Each stage has two rounds (the total number of rounds is 16), the first round presents a 2×2 grid and the second a 3×3 grid design (see Figure 3). The target and all three distractors are displayed in the 4-squared design. In the 9-squared design, the target is displayed twice as well as distractors two and three. Only one distractor is displayed three times.

B. GAMIFICATION

Gamification was proven to be beneficial to improve students' engagement and motivation and, at the same time, computer games are suitable for engaging children with dyslexia in reading and writing tasks [51], [52] as well as for screening [40]. Consequently, we integrated the following gamification mechanics that were identified (in a previous usability study with the game) to increase motivation through emotional engagement and visualize participants' progress: rewards (points), feedback (instant feedback), challenges (time limit), and game components (story for the game design). The reporting of game elements adheres to the guidelines proposed in the 2023 study on game element definitions [53].

C. IMPLEMENTATION

We implemented our games using *PHP, JavaScript, jQuery*, and a *SQL-database*. The visual part is also implemented with *Angular*, which allowed us to perform remote online studies, with the advantage of adapting the application for different devices in future research studies. This study was optimized for desktop and tablet environments [54]. Data pre-processing and analysis was carried out using *RStudio, Python*, and *JupiterLab*.

The layout of the games follow the SO 9241 requirements and recommendations for the ergonomics of human-computer interaction [55].

IV. EXPERIMENTAL STUDIES

Using a within-subject design, we conducted a study comprising a total of 137 participants (51 with dyslexia). Each participant had to play one of our games for approximately 15 minutes. The goal of the study was to collect the data needed to run machine learning experiments to find out if language-independent computer games can predict risk of dyslexia.

We postulated the following ***Hypothesis:*** *Is it possible to screen dyslexia in children by applying machine learning to the data derived from a game that uses auditory and visual language-independent content?*

A. PARTICIPANTS

We conducted experiments with participants with ages ranging from 7 to 12, along with their supervisors: parents, legal guardian, teacher or therapist.

1) RECRUITMENT

Participants with diagnosed dyslexia were recruited via public calls on social media (support groups), learning centers and schools, in collaboration with dyslexia non-profit organizations to reach wider audiences. In contrast, the control group was recruited in collaboration with Spanish and German schools. However, some English speakers contacted

TABLE 3. Participants per data set.

Data set	N	Dyslexia				Control			
		n	$\overline{age}$	f.	m.	n	$\overline{age}$	f.	m.
DE	120	36	9.1	17	19	84	10	46	38
ALL	137	51	9.7	23	28	86	8.8	48	38

us through these calls. Since our games are language independent (only the instructions use native languages) we welcomed the English participants who meet the inclusion criteria to join the study. All participants played the game either with the instructions in English, German or Spanish depending on their native language. Participation in the research was completely voluntary.

2) INCLUSION CRITERIA

The participant call raised attention for parents who either did not know whether their child had dyslexia (18.3%) or suspected their child had dyslexia but did not have an official diagnosis (9.8%), *e.g.,* from a medical doctor. To have a precise data set and a binary classification to simplify the prediction, we only considered participants with an official dyslexia diagnosis or no signs of dyslexia for the control group. So, 28.1% of the initial group of participants with suspected dyslexia but without official diagnoses were not included in the study.

3) EXPERIMENT PARTICIPANTS

A total of 137 participants took part of the study, where most of them were German speakers (DE, $n = 120$). The other data set includes all languages (ALL, $n = 137$) where we also added participants that used the game instructions in English ($n = 2$, 1 with dyslexia and 1 control) and Spanish ($n = 15$, 14 with dyslexia, 1 control). We use the ALL data to explore their influence on the prediction. Due to limited resources, we could not completely balance our groups for dyslexia ($n = 51$) and control ($n = 86$). Hence, we addressed the problem of our unbalanced data sets in our analysis.

4) BILINGUALISM

Participants played the game either in English, German, or Spanish, depending on their native language. However, we had some bilingual participants, $n = 54$. For these cases, we used the language they reported to be more comfortable with, which was used for the instructions of the game. We do not use the native language, but rather the language the game was played in as the criterion to split the data sets for three reasons. The definition of a native language or mother tongue can be made easily when a participant speaks only one language. But this is not the case for bilingual participants because they might not be able to choose, and then we cannot distinguish the mother tongue or native language clearly [56]. Second, this question is a self-reported question and every participant's supervisor might define it differently for each child.

B. DESIGN AND DEPENDENT MEASURES

We used a within-subject experimental design, so all participants contributed to all the conditions of the study *e.g.,* tasks or game rounds. When applying a *within-subject* design, the conditions need to be randomized to avoid *order effects* produced by order of the conditions. Hence, we used *Latin Squares* to counterbalance our conditions and avoid order effects.

To quantify the participant performance, we collected a number of dependent variables that are used as input (features) for the machine learning classifiers, explained in detailed in Section V, where Table 5 shows the auditory measures while Table 6 shows the visual measures.

C. COMPLIANCE AND ETHICS STATEMENTS

Our research is in accordance with the ethical standards of the Universitat Pompeu Fabra, Barcelona. We also address regional ethical requirements such as from the State of Lower Saxony, *e.g.,* additional permit for user studies at each school, and the requirement that no schools, teachers, or pupils are named [57]. The prototypes used for this research are in compliance with the European *General Data Protection Regulation* (GDPR) in regard to the processing and protection of personal data [58], [59]. Personal information of the participant's supervisor such as name or email is not published and it is stored separately from the participant data for communicating results, if given. The name of the child is not collected and all data is stored on a password secured web server in Germany. The data collection for this work has been approved by the Ministry of Education, Science and Culture of Schleswig-Holstein (*Ministerium für Bildung, Wissenschaft und Kultur, MBWK)* and by the *Education Authority* of the State of Lower Saxony (*Niedersächsische Landesschulbehörde*).

D. MATERIALS AND PROCEDURE

The study adhered to the following procedure. First, the parents were informed about the purpose of the voluntary study. Next, only after the parents had given their consent, children were allowed to participate in the user study from home or from school, with either the first author of this work present or always available through digital communication. The communication with the participants was mostly via email or phone.

If the study was conducted in a school or learning center, the parental or legal guardian consent was obtained in advance, and the user study was supervised by the participant's supervisor (*i.e.,* parent, legal guardian, teacher or therapist).

After the online consent form was approved, we collected demographic data using a questionnaire, which was completed by the participant's supervisor. This included the age of the participant, whether they had an existing dyslexia diagnosis (yes/no/maybe), and their native language, among others. Table 4 presents the complete demographic data collected.

This was followed by explaining instructions to the participant's supervisor (*e.g.,* turn up the volume, use headphones, play without interruptions, or explain and help your child

TABLE 4. Description of the demographic and technical features.

General Features	Description	*Number*
Age	It ranges from 7 to 12 years old.	**1**
Gender	A binary feature with two values, *female* or *male*.	**2**
Language	It is either *Spanish*, *German*, or *English*.	**3**
Native Language	It indicates if the language used for the instructions is the first language of the participants, being *Yes* or *No*.	**4**
Number of languages	It describes the number of languages a participant reported knowing, ranging from 1 to 4 languages.	**5**
Class level	It ranges from 0 to 8 and it describes in which year of education the participant is. The integer value corresponds to the main model of primary and lower secondary education in Europe.	**6**
Hearing Limitations	It indicates the hearing limitation the participant reported, being *No limitations, little limitations* or *limitations*.	**7**
Fun	It indicates the expressed "fun" mentioned in the feedback question, being *No fun, little fun, or fun*.	**8**
Difficulty Level	It indicates the expressed level of difficulty for the game mentioned in the feedback question, being *Not challenging, middle challenging* or *challenging*.	**9**
Instrument	It indicates if a participant plays a musical instrument, being *No*, *Yes, less than 6 months* or *Yes, over 6 months*.	**10**
Device	*Computer* or *Tablet*.	**11**
Operating system	*Mac OS, Windows, Android* or *Linux*.	**12**
Browser	*Safari, Chrome, Edge, Firefox, Opera* or *Internet Explorer*.	**13**

TABLE 5. Description of the auditory features.

Auditory Feature	Description	*Number*
Instructions	Number of times a participant listened to the instructions.	**14-29**
Duration per Round	It starts with the beginning of the game round, which is the listening of the target queue, and ends when the participant has made a choice in the second phase, that is, a click on a button (ms.).	**30–45**
Thinking Duration	It starts right after the participant finished hearing the auditory cues of the round and ends when the participant clicks over the selected auditory cue (ms.).	**46–61**
Target Melody Repetitions	Number of times a participant listened to the target auditory cue.	**62–77**
Correct Answers	A binary feature either with *wrong* or *correct* value.	**78–93**
Wrong Answers	A binary feature either with *wrong* or *correct* value.	**94–109**
Total Correct Answers	Number of correct answers for the previous rounds by summing the correct answers from all previous auditory rounds.	**110–125**
Total Wrong Answers	Number of wrong answers for the previous rounds by summing the wrong answers from all previous auditory rounds.	**126–141**
Cue	It indicates the participant's click choice, being *target, distractor 1, distractor 2*, or *distractor 3*.	**142–157**
Button Position	It describes the position of the clicked button, being *left, middle-left, middle-right*, or *right*.	**158–173**

only with the instructions of the games). Then, participants watched the short video with the instructions and played our game for 15 minutes approximately. The dependent variables were collected while playing. Participants could choose to discontinue participation at any time during the study.

V. DATA SETS

Our data sets were derived from the experimental studies presented in the previous section. The data sets from the experiment have 429 features per subject (137), that is, 58,773 data points.

To extract the features we used the (i) demographic questionnaire, (ii) technical metadata collected from the browser (Table 4), and (iii) game measures derived from the participant performance while playing, divided in a auditory features (Table 5) and visual features (Table 6).

VI. MACHINE LEARNING SETUP

A. METHODS

1) RATIONALE

Traditional predictive machine learning methods such as *regression* were designed before big data existed. Since we collected rather small data, it would be obvious to use these traditional methods. However, we did not use them since our data has a high variance which causes a high R-squared error and *multiple-colinearity* (*i.e.*, two or more variables have a high correlation). As dyslexia's origin is not fully decoded yet, more than one cause is assumed and therefore more than one indicator is needed for dyslexia, hence a regression is not the best technique. The data also has complex dependencies, and using causal dependencies (*e.g.* having more correct or incorrect answers), does not give a precise prediction.

For the reasons above we used non-linear methods such as Random Forest (RF) without and with class weights (RFW), Extra Trees (ETC), and Gradient Boosting (GB) from the Scikit-learn library version 0.21.2 [60].

2) AVOIDING RISK OF OVER-FITTING

Since our collected data are considered *small data* [41], [61], [62], we need to analyze them accordingly. To avoid the risk of over-fitting, we used 10-fold cross-validation [42], [63] and the default parameters suggested in the Scikit-learn library to avoid training a model by optimizing the parameters specifically for our data [60]. We also use 10-fold cross-validation because a small test or training set with high variances is not representative, hence a prediction based on it could be misleading. Apart from that we followed the following small data recommendations [64]. Moreover,

TABLE 6. Description of the visual features.

Visual Feature	Description	*Number*
1st Click Interval	Duration between the start of the second phase and the first click of the participant on a visual cue, in milliseconds (ms.).	**174–189**
2nd Click Interval	Duration (ms.) between the first and second click on a visual cue.	**190–205**
3rd Click Interval	Duration (ms.) between the second and third click on a visual cue.	**206–221**
4th Click Interval	Duration (ms.) between the third and fourth click on a visual cue.	**222–237**
5th Click Interval	Duration (ms.) between the fourth and fifth click on a visual cue.	**238–253**
6th Click Interval	Duration (ms.) between the fifth and sixth click on a visual cue.	**254–269**
Time Last Click	Duration (ms.) of the last click within a game round in the second phase.	**270–285**
Total Clicks	Number of total clicks within a game round.	**286–301**
Correct Answers	Number of hits or correct answers within a game round.	**302–317**
Wrong Answers	Number of wrong answers or non-correct answers within a game round.	**318–333**
Distractor 1	Number of times distractor 1 is clicked within a round.	**334–349**
Distractor 2	Number of times distractor 2 is clicked within a round.	**350–365**
Distractor 3	Number of times distractor 3 is clicked within a round.	**366–381**
Efficiency	*Time Last Click* divided by *Correct answers*.	**382–397**
Accuracy	*Correct answers* divided by *Total clicks*.	**398–413**
Effect	*Correct answers* multiplied by *Total clicks*.	**414–429**

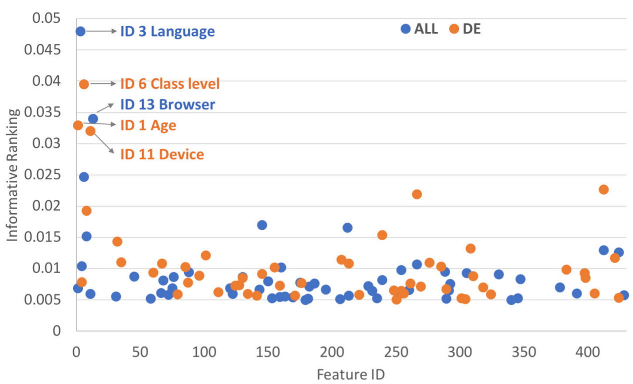

FIGURE 4. Features ranking for the DE and ALL data sets of the experiment with the highest-ranked features highlighted.

we used classifiers designed to avoid over-fitting such as Random Forest with weights and Extra Trees with their default parameters, and metrics for imbalanced data (balanced accuracy, F1-score), as well as no optimization of features within the cross-validation loop.

TABLE 7. DE and ALL have 15 features in common among the highest-ranked informative features (DE $n = 53$, ALL $n = 58$) in the experiment.

Feature Category	No. of Features	Features
Auditory	3	Cue, Total Number of Wrong Answers, Button Position.
Visual	7	Accuracy, Total clicks, Effect, 3rd Click Interval, 5th Click Interval, 6th Click Interval, 6th Click Interval.
Subject	4	Class Level, Age, Fun and Native Language
Technical	1	Device

TABLE 8. Overview of the subsets of features used to compare the quality of the prediction for our method.

Selected Features ID	Description
All features	All the 429 features.
Informative	They are the most informative features with a ranking score over 0.005 for ALL (58 features) and DE (53 features).
Auditory	Only the auditory features.
Auditory related	They are measures taken from the auditory game rounds where the game content was related to language.
Auditory generic	They are measures taken from the auditory game rounds where the game input was generic.
Visual	Only the visual features.
Visual related	They are measures taken from the visual game rounds where the game content was related to language.
Visual generic	They are measures taken from the visual game rounds where the game content was generic.

3) DEALING WITH AN UNBALANCED DATA SET

To compare different predictions we use *balanced accuracy* as main measure to deal with imbalanced data. However, since our aim is to detect a person with dyslexia, if we only consider the balanced accuracy, then we do not mainly focus on the detection of dyslexia, but rather on the overall accuracy of our model. Obtaining both high precision and high recall is unlikely, which is why we also report the F1-score (the weighted average between precision and recall) for dyslexia to compare our model's results. For an operational tool, we need to also look at the ratio of false positives and false negatives, as missing a person with dyslexia is worse than telling a child that may have dyslexia when that is not the case [40].

We do not apply over-sampling to address our unbalanced data because the variances among people with dyslexia are broad; for example, difficulty levels or the individual causes for perception differences vary widely. Similarly, we do not apply under-sampling to address our unbalanced data because our data set is already very small and under-sampling would reduce it to $n < 100$. The smaller the data set, the more likely it is to produce unwanted over-fitting.

TABLE 9. Best results of ALL (on the left) and DE (on the right) data sets for the different classifiers and subsets of features. The best two results for the F1-score and accuracy are highlighted as well as difference in the classifier ranking.

	ALL data set					DE data set				
Selected Features	**Classifier**	**Recall**	**Precis.**	**F1**	**Accuracy**	**Classifier**	**Recall**	**Precis.**	**F1**	**Accuracy**
All features	ETC	0.70	0.67	0.66	0.64	ETC	0.73	0.68	0.68	0.60
All features	GB	0.64	0.64	0.60	0.59	GB	0.70	0.65	0.66	0.59
All features	Baseline	0.63	0.39	0.48	0.50	Baseline	0.70	0.49	0.58	0.50
Informative	**ETC**	**0.77**	**0.81**	**0.75**	**0.73**	**ETC**	**0.79**	**0.78**	**0.74**	**0.65**
Informative	GB	0.75	0.79	0.73	0.70	GB	0.74	0.72	0.71	0.66
Informative	Baseline	0.63	0.39	0.48	0.50	Baseline	0.70	0.49	0.58	0.50
Auditory	*RF*	0.64	0.61	0.61	0.58	*ETC*	0.75	0.72	0.71	0.63
Auditory	*RFW*	0.65	0.64	0.61	0.58	*RF*	0.71	0.62	0.65	0.56
Auditory	Baseline	0.63	0.39	0.48	0.50	Baseline	0.70	0.49	0.58	0.50
Auditory related	ETC	0.72	0.76	0.70	0.68	RF	0.69	0.65	0.65	0.56
Auditory related	GB	0.66	0.67	0.65	0.63	GB	0.69	0.66	0.65	0.56
Auditory related	Baseline	0.63	0.39	0.48	0.50	Baseline	0.70	0.49	0.58	0.50
Auditory generic	**ETC**	**0.80**	**0.83**	**0.77**	**0.75**	**ETC**	**0.77**	**0.77**	**0.74**	**0.67**
Auditory generic	GB	0.66	0.66	0.64	0.62	RFW	0.73	0.71	0.69	0.60
Auditory generic	Baseline	0.63	0.39	0.48	0.50	Baseline	0.70	0.49	0.58	0.50
Visual	GB	0.69	0.69	0.66	0.65	GB	0.74	0.72	0.70	0.64
Visual	ETC	0.67	0.71	0.65	0.62	ETC	0.69	0.62	0.64	0.55
Visual	Baseline	0.63	0.39	0.48	0.50	Baseline	0.70	0.49	0.58	0.50
Visual related	ETC	0.71	0.73	0.68	0.66	RF	0.76	0.74	0.70	0.61
Visual related	GB	0.69	0.69	0.68	0.67	GB	0.70	0.70	0.68	0.61
Visual related	Baseline	0.63	0.39	0.48	0.50	Baseline	0.70	0.49	0.58	0.50
Visual generic	ETC	0.71	0.72	0.69	0.66	ETC	0.68	0.66	0.65	0.57
Visual generic	GB	0.69	0.71	0.66	0.65	GB	0.69	0.66	0.65	0.59
Visual generic	Baseline	0.63	0.39	0.48	0.50	Baseline	0.70	0.49	0.58	0.50

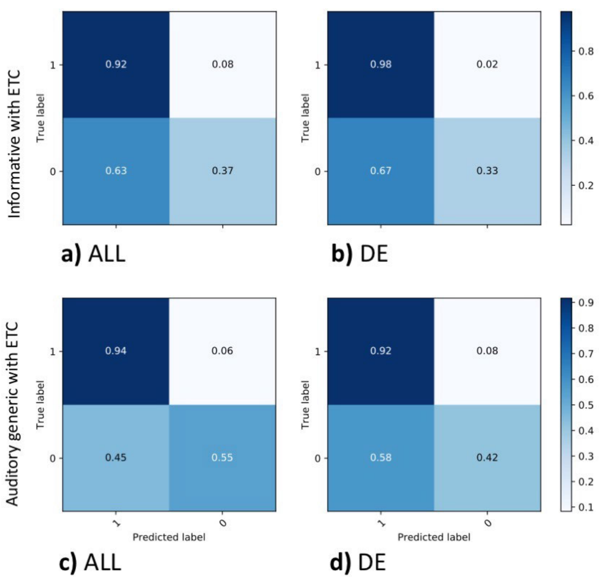

FIGURE 5. Normalized confusion matrix from the two best results (F1-score and accuracy): a) ALL data set, Informative features with ETC; b) DE data set, Informative with ETC; c) ALL, Auditory generic features with ETC; and d) DE, Auditory generic features with ETC.

B. FEATURE SELECTION

To explore the best prediction conditions and gain knowledge about our data, we used feature selection and ranked the most informative features with *Extra Trees*.

1) FEATURE SELECTION

The results of the ranking for the data sets (see Figure 4) show a step at the information score of 0.032: ALL 2 features, and DE 3 features. The two highest-ranked features of the ALL data set (*language* and *browser*) are different from the highest ranked features of the DE data set (*class level, age* and *device*). The comparison of the highest-ranked features (score over 0.005) reveals that the data sets have fewer features in common (15 features out of 58 for ALL and 53 for DE, see Table 7).

Since parameter optimization of predictive models can lead to over-fitting in small data, we use the feature selection to compare screening results. We explore the improvement of our measures for the predictive models with the subsets of features as described in Table 8. We address the danger of selecting the correct features [65] by taking into account knowledge of previous literature about the differences of children with and without dyslexia to avoid finding spurious correlations. All feature subsets include the subject features (1 to 13 in Table 4).

VII. RESULTS

The two best results for the F1-score and accuracy obtained for each data set and feature selection as well as the baseline are presented in Table 9. For the ALL data set, we achieved the best two results for the F1-score and accuracy using the feature selections *auditory generic* $(0.77, 0.75)$; and *informative* $(0.75, 0.73)$ and the ETC model. For the DE data set, we achieved the best two results for the F1-score and accuracy using the feature selections *auditory generic* $(0.74, 0.67)$ and *informative* $(0.74, 0.65)$ and using the ETC model. The ranking of classifiers for each selected feature is nearly the same for ALL and DE except for the *auditory* feature selection.

For the auditory feature selection, ALL predicts best with RF and RFW, while DE has the best results with ETC and RF. The prediction for the subsets of auditory features has a higher accuracy and F1-score than for visual features. Finally, the normalized confusion matrix (see Figure 5) does not show over-fitting for the best results for ALL and DE.

VIII. DISCUSSION

A. HYPOTHESES

After the results of the experiment, we accept the ***Hypothesis:*** *It is possible to screen dyslexia in children by applying machine learning to the data derived from a game that uses language-independent content: auditory and visual and generic content.* The comparison of subsets of features for visual and auditory shows a slightly higher score for accuracy and F1-score when auditory content is used. Auditory subsets of features showed the best results (F1-score and accuracy) with the feature selection *auditory generic* using Extra Trees: ALL data set (0.77, 0.75, $n = 137$) and DE data set (0.74, 0.67, $n = 120$).

Data set ALL ($n = 137$) achieved a better prediction result than DE ($n = 120$), but the unbalanced Spanish participant group ($n = 15$) biased the prediction, something that in retrospect made sense. Due to the limited resources, only a few participants played in English and Spanish. Hence these data sets were not computed separately. English has a balanced group of participants, with only one participant for each group. The Spanish participant group mainly contained participants with dyslexia ($n = 14$), with only one control participant. We assume that the model uses the unbalanced Spanish participant group as the relationship for the prediction and thus achieves higher prediction results compared to DE. The reason that ALL (generic content) performs better than DE is probably due to the unbalanced Spanish participant data, not because of the model. While ALL (related content) does not perform better than the other two data sets (ES or DE) probably due to *e.g.*, bilingualism or features canceling each other. An additional data collection of Spanish participants for the control group is needed to confirm our current results for ALL.

1) VISUAL VS. AUDITORY

A clustering of data might help to ensure the quality of data and a better prediction. For example, it could help to cluster abilities of participants who are better for the visual or auditory game or participants who have a tendency for visual or auditory difficulties. In our case, the subsets with auditory features have slightly better metrics than visual (see Table 9), although it is difficult to compare visual and auditory content (for example, based on the level of difficulty or similarity). The subsets with auditory features probably have higher prediction scores for two reasons: the auditory content is related to more characteristics of dyslexia (see Tables 1 and 2); and the sample size is bigger (in terms of more auditory stages and features) providing more information for the models. These results support the theory that a stronger effect can be measured when content is related to many indicators at the same time. Hence, more indicators combined in game content could create a more prominent effect between groups.

B. LIMITATIONS

1) SMALL DATA

Other machine learning approaches to screen dyslexia presented higher accuracy rates. One possible explanation of our results is the high heterogeneity of our participants. Nonetheless, dyslexia itself presents a wide spectrum [66] and there is also a high variance in the measures of current diagnostic tools [34], [67]. Another possible reason is that we took the precautions needed when dealing with unbalanced small data; so, our model is not over-fitted at the cost of lower accuracy rates. Also, having small data is a limitation itself. However, small data was expected because collecting health-related data is challenging in many ways: dyslexia is under-diagnosed and privacy issues, among others. As a matter of fact, our study has a good number of participants in contrast to some previous studies. Moreover, in some circumstances, only small data is available, and it can lead to more reliable data, lower costs, and faster results [41], [61].

2) GROUND TRUTH

Another possible limitation is the heterogeneity of our ground truth. Dyslexia is diagnosed in different languages using different diagnostic assessments that are connected to a specific language. These tests do not necessarily use the same measures to diagnose dyslexia, such as reading errors, reading speed, different types of reading comprehension, among others [33], [34], [68]. Having an extra online screening test as a control test in our experiments would have further ensured the quality of the data. However, implementing or accessing such a test would have taken even more resources. Besides, it would have been more challenging to find the required number of participants because of having a longer study that includes diagnostic assessments instead of just a game.

3) LANGUAGE-INDEPENDENT CONTENT

Another challenge was to design new language-independent content that could show measurable differences between children with and without dyslexia, when most manifestations of dyslexia occur in reading and writing. Designing new language-independent content was probably the greatest challenge (it is also the case for the *National Center on Improving Literacy* [69]) because our indicators, though related to the reading and writing difficulties, are probably not the main causes of dyslexia.

4) ONLINE EXPERIMENT

Furthermore, our new language-independent content needed to be integrated an online experiment as a game. Most previous approaches using auditory and visual content conducted their experiment in a laboratory setting [23], [30], which means that these indicators were tested in controlled environments. That is not the case for online experiments. We controlled as many variables as possible; however, the control of an online setting is not comparable to a lab controlled environment. This study has been conducted using data extracted via a computer. However, this does not imply that the results can be extended to data collected through a mobile device or tablet. In fact, other approaches using similar data and machine learning to screen dyslexia, have required the development of distinct models for different devices [40], as the interaction differences between them are significant enough to impact the statistical model.

IX. CONCLUSION AND FUTURE WORK

We presented a method to screen risk of dyslexia that applies machine learning to data extracted from a computer game composed of language-independent items. In contrast to previous work, we worked with visual and auditory elements (not linguistic) as dyslexia indicators. Also, we conducted an international user study that included native speakers from three different languages: German, Spanish and English. Our results show that our method is able to screen dyslexia in children from 7 to 12 years old, specially in the larger data sets including groups that share the same native language.

This is the first screening method that integrates in the same tool visual and auditory content and which was evaluated for different languages. Since this approach uses language-independent content, our contributions has the potential to be extended to other languages, as well as to be applied for early prediction of dyslexia in pre-readers, that is, to screen risk of dyslexia in very young children, even before they develop linguistic skills. This approach is not a medical diagnosing tool; however, it offers an easily accessible and cost-efficient way of helping children to be screened in order to avoid the negative consequences of dyslexia. The demo of the games, the content used, and data sets are freely available online.

We consider our results with language-independent *auditory generic content* for DE (highest balanced accuracy of 0.67 and highest F1-scores of 0.74) using Extra Trees (Table 9) as a promising way to screen for dyslexia using language-independent content related to various dyslexia indicators. In addition, it is possible to screen dyslexia with the measures extracted from the games, however models are trained separately for each language, as recommended by previous studies [70]. Hence, predicting dyslexia for different languages with the same prediction model remains a challenge.

For future work we aim to conduct a longitudinal study and collect more data from younger children to find out if this approach with visual and auditory elements can be used for early prediction of dyslexia in pre-readers, without using any linguistic elements, working towards a universal screening tool for dyslexia.

ACKNOWLEDGMENT AND DISCLAIMERS

Part of this work was done while Maria Rauschenberger was doing the Ph.D. degree from the Universitat Pompeu Fabra, Barcelona, Spain. The authors thank ChangeDyslexia.org for their support in the recruitment of the participants of the study. They would also like to thank all teachers, students, and parents for their participation and time and supporters to distribute their participation calls. They also thank all parents and children for playing their games. Finally, thanks to H. Witzel for his advice during the development and to M. Blanca and M. Herrera for the translation of the Spanish version. They have no conflict of interests to disclose.

REFERENCES

[1] UNESCO. (2021). *Literacy*. Accessed: Apr. 14, 2021. [Online]. Available: https://en.unesco.org/themes/literacy/

[2] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*. London, U.K.: American Psychiatric Association, May 2013, doi: 10.1176/appi.books.9780890425596.

[3] S. E. Shaywitz, "Dyslexia," *New England J. Med.*, vol. 338, no. 5, pp. 307–312, Jan. 1998.

[4] C. Cortiella and S. H. Horowitz, *The State of Learning Disabilities: Facts, Trends and Emerging Issues*, vol. 25. New York, NY, USA: National Center for Learning Disabilities, 2014, pp. 2–45.

[5] L. Al-Lamki, "Dyslexia: Its impact on the individual, parents and society," *Sultan Qaboos Univ. Med. J.*, vol. 12, no. 3, p. 269, 2012.

[6] D. Morris and P. Turnbull, "A survey-based exploration of the impact of dyslexia on career progression of UK registered nurses," *J. Nursing Manage.*, vol. 15, no. 1, pp. 97–106, Jan. 2007.

[7] K. Gerling, M. Rauschenberger, B. Tannert, and G. Weber, "The next decade in accessibility research," *I-COM*, vol. 23, no. 2, pp. 231–237, Jul. 2024, doi: 10.1515/icom-2024-0015.

[8] F. R. Vellutino, J. M. Fletcher, M. J. Snowling, and D. M. Scanlon, "Specific reading disability (dyslexia): What have we learned in the past four decades?" *J. Child Psychol. Psychiatry*, vol. 45, no. 1, pp. 2–40, Jan. 2004, doi: 10.1046/j.0021-9630.2003.00305.x.

[9] F. Serrano and S. Defior, "Dyslexia speed problems in a transparent orthography," *Ann. Dyslexia*, vol. 58, no. 1, pp. 81–95, Jun. 2008.

[10] E. Temple, G. K. Deutsch, R. A. Poldrack, S. L. Miller, P. Tallal, M. M. Merzenich, and J. D. E. Gabrieli, "Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2860–2865, Mar. 2003.

[11] G. Schulte-Körne, "Diagnostik und therapie der lese-rechtscheib-störung (The prevention, diagnosis, and treatment of dyslexia)," *Deutsches Ärzteblatt Int.*, vol. 107, no. 41, pp. 718–727, 2010.

[12] M. Rauschenberger, "Early screening of dyslexia using a language independent content game and machine learning," Ph.D. dissertation, Dept. Inf. Commun. Technol., Universitat Pompeu Fabra, Spain, Oct. 2019.

[13] International Dyslexia Association. (2020). *Definition of Dyslexia*. Accessed: Apr. 14, 2021. [Online]. Available: https://dyslexiaida.org/definition-of-dyslexia/

[14] P. H. K. Seymour, M. Aro, and J. Erskine, "Foundation literacy acquisition in European orthographies," *Brit. J. Psychol.*, vol. 94, no. 2, pp. 143–174, May 2003. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12803812

[15] J. C. Ziegler, C. Perry, A. Ma-Wyatt, D. Ladner, and G. Schulte-Körne, "Developmental dyslexia in different languages: Language-specific or universal?" *J. Experim. Child Psychol.*, vol. 86, no. 3, pp. 169–193, Nov. 2003. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/14559203

[16] M. Rauschenberger, L. Rello, S. Füchsel, and J. Thomaschewski, "A language resource of German errors written by children with dyslexia," in *Proc. 10th Int. Conf. Lang. Resour. Eval. (LREC)*, Paris, France, May 2016, pp. 83–87.

[17] L. Rello, "DysWebxia: A text accessibility model for people with dyslexia," Ph.D. dissertation, Dept. Inf. Commun. Technol., Universitat Pompeu Fabra, Spain, Jun. 2014.

[18] M. Rauschenberger, S. Füchsel, L. Rello, and J. Thomaschewski, (2017). *DysList German Resource: A Language Resource of German Errors Written by Children With Dyslexia*. Accessed: Jun. 6, 2019. [Online]. Available: https://zenodo.org/record/809801 and http://rauschii.github.io/DysListGerman/

[19] E. Borleffs, B. A. M. Maassen, H. Lyytinen, and F. Zwarts, "Cracking the code: The impact of orthographic transparency and morphological-syllabic complexity on reading and developmental dyslexia," *Frontiers Psychol.*, vol. 9, pp. 1–19, Jan. 2019, doi: 10.3389/fpsyg.2018.02534.

[20] T. R. Vidyasagar and K. Pammer, "Dyslexia: A deficit in visuo-spatial attention, not in phonological processing," *Trends Cognit. Sci.*, vol. 14, no. 2, pp. 57–63, Feb. 2010.

[21] U. Goswami, "A temporal sampling framework for developmental dyslexia," *Trends Cognit. Sci.*, vol. 15, no. 1, pp. 3–10, Jan. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1364661310002354
[22] J. A. Hämäläinen, H. K. Salminen, and P. H. T. Leppänen, "Basic auditory processing deficits in dyslexia: Systematic review of the behavioral and event-related potential/field evidence," *J. Learn. Disabilities*, vol. 46, no. 5, pp. 413–427, Sep. 2013, doi: 10.1177/0022219411436213.
[23] U. Goswami, N. Mead, T. Fosker, M. Huss, L. Barnes, and V. Leong, "Impaired perception of syllable stress in children with dyslexia: A longitudinal study," *J. Memory Lang.*, vol. 69, no. 1, pp. 1–17, Jul. 2013.
[24] M. Huss, J. P. Verney, T. Fosker, N. Mead, and U. Goswami, "Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology," *Cortex*, vol. 47, no. 6, pp. 674–689, Jun. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S001094521000198X
[25] C. Männel, G. Schaadt, F. K. Illner, E. van der Meer, and A. D. Friederici, "Phonological abilities in literacy-impaired children: Brain potentials reveal deficient phoneme discrimination, but intact prosodic processing," *Develop. Cognit. Neurosci.*, vol. 23, pp. 14–25, Feb. 2017, doi: 10.1016/j.dcn.2016.11.007.
[26] G. De Zubicaray and N. O. Schiller, *The Oxford handbook of neurolinguistics*. New York, NY, USA: Oxford Univ. Press, 2018. https://www.worldcat.org/title/oxford-handbook-of-neurolinguistics/oclc/1043957419&referer=brief_results
[27] D. J. Johnson, "Persistent auditory disorders in young dyslexic adults," *Bull. Orton Soc.*, vol. 30, no. 1, pp. 268–276, Jan. 1980. [Online]. Available: http://link.springer.com/10.1007/BF02653723
[28] K. Overy, "Dyslexia, temporal processing and music: The potential of music as an early learning aid for dyslexic children," *Psychol. Music*, vol. 28, no. 2, pp. 218–229, Oct. 2000, doi: 10.1177/0305735600282010.
[29] U. Goswami, L. Barnes, N. Mead, A. J. Power, and V. Leong, "Prosodic similarity effects in short-term memory in developmental dyslexia," *Dyslexia*, vol. 22, no. 4, pp. 287–304, Nov. 2016, doi: 10.1002/dys.1535.
[30] S. Franceschini, S. Gori, M. Ruffino, K. Pedrolli, and A. Facoetti, "A causal link between visual spatial attention and reading acquisition," *Current Biol.*, vol. 22, no. 9, pp. 814–819, May 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0960982212002709
[31] A. Le Floch and G. Ropars, "Left-right asymmetry of the Maxwell spot centroids in adults without and with dyslexia," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 284, no. 1865, Oct. 2017, Art. no. 20171380, doi: 10.1098/rspb.2017.1380.
[32] L. Rello, R. Baeza-Yates, and J. Llisterri, "A resource of errors written in Spanish by people with dyslexia and its linguistic, phonetic and visual analysis," *Lang. Resour. Eval.*, vol. 51, no. 2, pp. 379–408, Feb. 2016, doi: 10.1007/s10579-015-9329-0.
[33] A. Fawcett and R. Nicolson, *The Dyslexia Screening Test: Junior (DST-J)*. U.K.: Pearsons Assessments, 2004.
[34] M. Grund, C. L. Naumann, and G. Haug, *Diagnostischer Rechtschreibtest Für 5. Klassen: DRT 5 (Diagnostic Spelling Test for Fifth Grade: DRT 5)* (Deutsche Schultests), 2nd ed., Göttingen, Germany: Beltz Test, 2004. [Online]. Available: https://www.testzentrale.de/shop/diagnostischer-rechtschreibtest-fuer-5-klassen.html
[35] M. Rauschenberger, L. Rello, and R. Baeza-Yates, "Technologies for dyslexia," in *Web Accessibility Book*, vol. 1, 2nd ed., London, U.K.: Springer-Verlag, 2019, pp. 603–627. [Online]. Available: https://www.springer.com/us/book/9781447174394
[36] M. Nilsson Benfatto, G. Öqvist Seimyr, J. Ygge, T. Pansell, A. Rydberg, and C. Jacobson, "Screening for dyslexia using eye tracking during reading," *PLoS ONE*, vol. 11, no. 12, Dec. 2016, Art. no. e0165508.
[37] A. Frid and Z. Breznitz, "An SVM based algorithm for analysis and discrimination of dyslexic readers from regular readers using ERPs," in *Proc. IEEE 27th Conv. Electr. Electron. Engineers Isr.*, Nov. 2012, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/document/6377068/
[38] H. Perera, M. F. Shiratuddin, and K. W. Wong, "A review of electroencephalogram-based analysis and classification frameworks for dyslexia," in *Proc. Int. Conf. Neural Inf. Process.* Springer, 2016, pp. 626–635.
[39] H. A. Rashid, T. Malik, I. Siddiqui, N. Bhatti, and A. Samad, "DYSIGN: Towards computational screening of dyslexia and dysgraphia based on handwriting quality," in *Proc. 22nd Annu. ACM Interact. Design Children Conf.*, Jun. 2023, pp. 532–536.
[40] L. Rello, R. Baeza-Yates, A. Ali, J. P. Bigham, and M. Serra, "Predicting risk of dyslexia with an online gamified test," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0241687.
[41] J. J. Faraway and N. H. Augustin, "When small data beats big data," *Statist. Probab. Lett.*, vol. 136, pp. 142–145, May 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167715218300762
[42] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 326–327, Sep. 1995. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.2069
[43] S. Zhong, S. Song, T. Tang, F. Nie, X. Zhou, Y. Zhao, Y. Zhao, K. F. Sin, and S.-H.-G. Chan, "DYPA: A machine learning dyslexia prescreening mobile application for Chinese children," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 3, pp. 1–21, Sep. 2023. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3610908
[44] A. Poole, F. Zulkernine, and C. Aylward, "Lexa: A tool for detecting dyslexia through auditory processing," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–5.
[45] M. Rauschenberger, R. Baeza-Yates, and L. Rello, "A universal screening tool for dyslexia by a web-game and machine learning," *Frontiers Comput. Sci.*, vol. 3, p. 111, Jan. 2022, doi: 10.3389/fcomp.2021.628634.
[46] C. Moritz, S. Yampolsky, G. Papadelis, J. Thomson, and M. Wolf, "Links between early rhythm skills, musical training, and phonological awareness," *Reading Writing*, vol. 26, no. 5, pp. 739–769, May 2013.
[47] E. J. Rolka and M. J. Silverman, "A systematic review of music and dyslexia," *Arts Psychotherapy*, vol. 46, pp. 24–32, Nov. 2015, doi: 10.1016/j.aip.2015.09.002.
[48] U. Goswami, R. Cumming, and A. Wilson, "Rhythmic perception, music and language: A new theoretical framework for understanding and remediating specific language impairment background to the project," Nuffield Found., U.K., Tech. Rep., 2016. [Online]. Available: https://www.cne.psychol.cam.ac.uk/files/nuffield_briefing_report.pdf
[49] J. Pedler, "Computer correction of real-word spelling errors in dyslexic text," Ph.D. dissertation, Birkbeck College, London Univ., London, U.K., 2007. [Online]. Available: http://www.learninglink.bbk.ac.uk/research/recentphds/pedler.pdf
[50] C. Coleman, N. Gregg, L. McLain, and L. W. Bellair, "A comparison of spelling performance across young adults with and without dyslexia," *Assessment Effective Intervent*, vol. 34, no. 2, pp. 94–105, Mar. 2009, doi: 10.1177/1534508408318808.
[51] O. Gaggi, C. E. Palazzi, M. Ciman, G. Galiazzo, S. Franceschini, M. Ruffino, S. Gori, and A. Facoetti, "Serious games for early identification of developmental dyslexia," *Comput. Entertainment*, vol. 15, no. 2, pp. 1–24, Apr. 2017, doi: 10.1145/2629558.
[52] F. Kyle, J. Kujala, U. Richardson, H. Lyytinen, and U. Goswami, "Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG rime and GG phoneme," *Reading Res. Quart.*, vol. 48, no. 1, pp. 61–76, Jan. 2013, doi: 10.1002/rrq.038.
[53] S. Hallifax, M. Altmeyer, K. Kölln, M. Rauschenberger, and L. E. Nacke, "From points to progression: A scoping review of game elements in gamification research with a content analysis of 280 research papers," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, pp. 748–768, Sep. 2023.
[54] M. Rauschenberger, L. Rello, and R. Baeza-Yates, "A tablet game to target dyslexia screening in pre-readers," in *MobileHCI*. Barcelona, Spain: ACM Press, 2018, pp. 306–312. [Online]. Available: http://mobilehci.acm.org/2018/
[55] ISO/TC 159/SC 4 Ergonomics of Human-System Interaction, "Part 210: Human-centred design for interactive systems," in *Ergonomics of Human-System Interaction*, vol. 1. Brussels, Belgium: International Organization for Standardization, 2010, p. 32. [Online]. Available: https://www.iso.org/standard/52075.html
[56] I. Kecskes and T. Papp, *Foreign Language and Mother Tongue*, 1st ed., New York, NY, USA: Psychology Press, Jun. 2000. [Online]. Available: https://www.taylorfrancis.com/books/9781410606464
[57] Niedersächsisches Vorschrifteninformationssystem. (2015). *VORIS Ministry of Education and Cultural Affairs | 25b-81402 | Administrative Regulations (Lower Saxony) | Polls and Surveys in Schools*. Accessed: Jun. 13, 2019. [Online]. Available: http://www.nds-voris.de/jportal/?quelle=jlink&query=VVND-224100-MK-20140101-SF&psml=bsvorisprod.psml&max=true
[58] European Union. (2016). *General Data Protection Regulation*. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, Repealing Directive 95/46/EC. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679

[59] Bildungsministerium. (2018). *Rundschreiben des Bildungsministeriums: Schule Und Datenschutz-Grundverordnung der EU (DSGVO) (Circular letter of the Ministry of Education: School and General Data Protection Regulation of the EU (GDPR))*. Accessed: Jun. 13, 2019. [Online]. Available: https://schulrecht-sh.de/texte/d/dgsvo_und_schule.pdf

[60] Scikit-Learn Developers. (2019). *Scikit-Learn Documentation*. Accessed: Jun. 20, 2019. [Online]. Available: https://scikit-learn.org/stable/documentation.html

[61] R. Baeza-Yates. (2018). *Big, Small or Right Data: Which is the Proper Focus?* Accessed: Jul. 22, 2019. [Online]. Available: https://www.kdnuggets.com/2018/10/big-small-right-data.html

[62] A. C. Weigand and M. Rauschenberger, "Exploring the definition of small data collected with HCI methods and used for ML," in *Mensch und Computer 2023—Workshopband*, P. Fröhlich and V. Cobus, Eds. Rapperswil, Switzerland: Veröffentlicht durch die Gesellschaft für Informatik e.V., Sep. 2023.

[63] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinf.*, vol. 7, no. 1, p. 91, Feb. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16504092

[64] M. Rauschenberger and R. Baeza-Yates, "How to handle health-related small imbalanced data in machine learning?" *i-com*, vol. 19, no. 3, pp. 215–226, Jan. 2021. [Online]. Available: https://www.degruyter.com/view/journals/icom/19/3/article-p215.xml

[65] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997. [Online]. Available: https://ieeexplore.ieee.org/document/574797

[66] D. W. Black and J. E. Grant, *DSM-5 Guidebook: The Essential companion to the Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. USA: American Psychiatric Association, 2014. [Online]. Available: https://www.appi.org/dsm-5_guidebook

[67] C. Steinbrink and T. Lachmann, *Lese-Rechtschreibstörung (Dyslexia)*. Berlin, Germany: Springer, 2014, doi: 10.1007/978-3-642-41842-6.

[68] F. Cuetos, B. Rodríguez, E. Ruano, and D. Arribas, *PROLEC-R: Batería de Evaluación de Los Procesos Lectores, Revisada (Battery of Reading Processes Assessment-Revised)*. Madrid, Spain: TEA Ediciones, 2007.

[69] Y. Petscher, H. Fien, C. Stanley, B. Gearin, J. M. Fletcher, Y. Petscher, C. Stanley, B. Gearin, N. Gaab, and J. M. Fletcher, "Screening for dyslexia," US Dept. Educ., Nat. Center Improving Literacy, Tech. Rep., 2019. [Online]. Available: https://megalodon-glockenspiel-jywd.squarespace.com/s/ScreeningforDyslexia_2019.pdf

[70] A. Bandhyopadhyay, D. Dey, and R. K. Pal, *Prediction of Dyslexia using Machine Learning—A Research Travelogue*, vol. 24. Singapore: Springer, 2018, doi: 10.1007/978-981-10-6890-4.

MARIA RAUSCHENBERGER was born in Germany, in 1986. She received the B.S. degree in computer media and the M.A. degree in online computer science and media application from the University of Applied Science Emden/Leer, Germany, and the Ph.D. degree in computer science from the Universitat Pompeu Fabra, Barcelona, Spain, in 2019.

She was a Project Manager with MSP Medien-Systempartner, Oldenburg, Germany. After the master's study, she was a Research Assistant with OFFIS e.V., Institute for Informatics, Oldenburg, a Research Visitor with the HCI Institute, Carnegie Mellon University, and an internship with the non-profit organization Change Dyslexia. After a postdoctoral position with the Max-Planck Institute for Software Systems, Saarbrücken, Germany, with Krishna P. Gummadi, she became a Full Professor of digital media with the University of Applied Sciences Emden/Leer, in 2020. She specializes in human–computer interaction, machine learning, accessibility, and gamification, focusing on user-centric digital systems to address social issues.

Dr. Rauschenberger has received several accolades, including the German Reading Award 2017 and multiple fem-scholarships during her Ph.D. Her work, which significantly influences digital media and user experience design, has made her a key contributor to the academic and professional communities, where she delivers international keynote speeches and served on editorial boards.

RICARDO BAEZA-YATES (Fellow, IEEE) was born in Chile, in 1961. He received the B.Eng. degree in electrical engineering (EE) and the B.Sc. degree in computer science (CS), the M.Sc. degree in CS, and the M.Eng. degree in EE from the University of Chile, in 1983, 1985, and 1986, respectively, and the Ph.D. degree in CS from the University of Waterloo, Canada, in 1989.

Currently, he is the Director of research with the Institute for Experiential AI, Northeastern University. He was also a part-time Professor with the Universitat Pompeu Fabra, Barcelona, Spain, and the Universidad de Chile, Santiago, in 2004 and 1989, respectively. Before that, he was the CTO of NTENT, a semantic search technology company, California. Prior to these roles, he was a VP of Research with Yahoo Laboratories, Barcelona, and later in Sunnyvale, CA, USA, from 2006 to 2016. His areas of expertise are responsible AI, web search and data mining, information retrieval, and data science and algorithms in general.

Dr. Baeza-Yates is a member of the IEEE's AI Technology Policy Committee. He is the co-author of the best-seller *Modern Information Retrieval* [Addison-Wesley in 1999 and 2011 (second edition)], which won the ASIS&T 2012 Book of the Year Award. From 2002 to 2004, he was elected to the Board of Governors of the IEEE Computer Society, and from 2012 to 2016, he was elected to the ACM Council. Since 2010, he has been a Founding Member of the Chilean Academy of Engineering. In 2009 and 2011, he was named ACM Fellow and IEEE Fellow, respectively, among other distinctions, including national awards in Chile and Spain.

LUZ RELLO was born in Spain, in 1984. She received the B.S. degree in linguistics from the Universidad Complutense de Madrid, Spain, in 2008, the M.A. (international master's) degree in natural language processing and human language technology from the University of Wolverhampton, U.K., in 2010, and the Ph.D. degree in computer science from the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Spain, in 2014. She was a first Postdoctoral Researcher and a Systems Scientist with the Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, USA, from 2014 to 2018. Currently, she is an Assistant Professor with the Department of Information Systems and Technology, IE Business School, IE University, Madrid, Spain. She was the first Spanish person to receive the European Young Researchers' Award. Her research focused on machine learning for dyslexia, has been recognized by several awards including the MIT Technology Review Innovators Under 35 Award, Princesa de Girona Award, or the Forbes 30 Under 30 list. She is also the founder of the social venture Change Dyslexia to help children with dyslexia using machine learning, which was recently awarded with the UNESCO King Hamad Bin Isa Al-Khalifa Prize for the Use of ICT in Education. She was part of the Spanish Artificial Intelligence Advisory Council of the Ministry of Economic Affairs and Digital Transformation. She is a consultant for Google, and has published two books, one of them "Overcoming Dyslexia" which already has eight editions.

• • •