Identifying Reading Fluency in Pupils with and without Dyslexia Using a Machine Learning Model on Texts Assessed with a Readability Application

Jure Žabkar[1], Tajda Urankar[1], Karmen Javornik[2] and Milena Košak Babuder[*3]

∽ Measurement of readability is an important tool for assessing reading disorders such as dyslexia. Among the screening procedures for dyslexia is the reading fluency test, which is defined as the ability to read with speed, accuracy and proper expression. The reading fluency test often consists of a sequence of unrelated written texts ranging from simple short sentences to more difficult and longer paragraphs. In psychological testing instruments, subjective text assessment is often replaced by objective readability formulas, e.g., the Automated Readability Index. Readability formulas extract multiple features from a given text and output a score indicating the difficulty of the text. The aim of the present study is to build a machine learning model that discriminates between pupils identified with dyslexia and a control group without dyslexia based on fluency in oral reading of texts assessed with a readability application developed within the project For the Quality of Slovenian Textbooks. We focus on differentiation between both groups of pupils by analysing data obtained from transcriptions of audio recordings of oral reading. The empirical study was conducted with 27 pupils aged 8 and 9 with officially diagnosed dyslexia and a control group without identified dyslexia.

Keywords: dyslexia, readability application, reading fluency, machine learning

---

1   Faculty of Computer and Information Science, University of Ljubljana, Slovenia.
2   Faculty of Education, University of Ljubljana, Slovenia.
3   *Corresponding Author. Faculty of Education, University of Ljubljana, Slovenia; Milena.Kosak-Babuder@pef.uni-lj.si.

Prepoznavanje tekočnosti branja pri učencih z disleksijo in brez nje z uporabo modela strojnega učenja na besedilih, ocenjenih z aplikacijo za berljivost

Jure Žabkar, Tajda Urankar, Karmen Javornik in
Milena Košak Babuder

∽ Merjenje berljivosti je pomembno orodje za ocenjevanje motenj branja, kot je disleksija. Med presejalnimi postopki za disleksijo je tudi preizkus tekočnosti branja. Tekočnost branja je opredeljena kot sposobnost hitrega in natančnega branja ter pravilnega izražanja. Preizkus tekočnosti branja je pogosto sestavljen iz zaporedja nepovezanih zapisanih besedil, od preprostih kratkih povedi do zahtevnejših in daljših odstavkov. V psiholoških testih subjektivno ocenjevanje besedil pogosto nadomeščajo objektivne formule berljivosti, npr. avtomatski indeks berljivosti (Automated Readability Index). Formule za berljivost iz danega besedila izluščijo več značilnosti in izpišejo oceno, ki označuje težavnost besedila. Cilj te raziskave je zgraditi model strojnega učenja, ki bo na podlagi tekočnosti ustnega/glasnega branja besedil, ocenjenih z aplikacijo za ocenjevanje berljivosti, razvite v okviru projekta Za kakovost slovenskih učbenikov (KaUč), razlikoval med učenci, pri katerih je bila ugotovljena disleksija, in kontrolno skupino učencev brez disleksije. Pri tem se osredinjamo na razlikovanje med obema skupinama učencev z analizo podatkov, pridobljenih s transkripcijami zvočnih posnetkov ustnega/glasnega branja. V empirični raziskavi je sodelovalo 27 učencev, starih 8 in 9 let, s potrjeno disleksijo in kontrolna skupina brez ugotovljene disleksije.

Ključne besede: disleksija, aplikacija za berljivost, tekočnost branja, strojno učenje

Introduction

Reading and writing are basic skills that are taken for granted in today's society. They are key elements of literacy, enabling individuals to develop the skills of reflection, critique and empathy, leading to a sense of self-efficacy, identity and full participation in society. Among learning difficulties, it is reading difficulties that have a significant impact on an individual's educational success throughout life. Despite an education system that focuses on literacy development, there are still many pupils who leave primary school without adequately developed literacy skills and who are unable to overcome this deficit even in adulthood (Carpentieri, 2012). Learning to read is one of the most important outcomes of early education, and developing reading and writing skills as two key communicative skills are among the basic goals of teaching Slovenian in the first educational period in primary school (Poznanovič et al., 2018). There are increasing numbers of pupils in schools who have difficulties in learning to read and write due to dyslexia (Snowling et al., 2020). Moreover, difficulties in reading also lead to difficulties in other areas of learning, including writing, spelling, reading fluency and comprehension (Moats & Dakin, 2008; Shaywitz, 2003).

The best-known and most widely researched specific learning difficulty is dyslexia, which is a neurophysiologically conditioned reading disorder originating from a developmental or central nervous peculiarity (Magajna et al., 2015; Raduly Zorgo et al., 2010). It includes a group of diverse but interrelated factors that are part of the individual and affect him/her and his/her functioning throughout life (Magajna et al., 2015; Raduly Zorgo et al., 2010). Dyslexia is characterised by difficulties in accurate and/or fluent word recognition, poor spelling and poor decoding skills, all of which affect reading acquisition, reading comprehension and writing (IDA, 2002). The difficulties are not limited to reading and spelling; there are also difficulties with sustaining attention and automating new knowledge, as well as with gross and fine motor skills (Nicolson & Fawcett, 1990, 2007; Rose, 2009). In addition to neurological differences, dyslexia is also associated with cognitive difficulties that can affect organisational skills, numeracy and other cognitive and emotional abilities (Rose, 2009). People with dyslexia can be extremely talented and original when it comes to solving different types of problems and often have good visual skills (Nijakowska, 2016). Approximately seven percent of children and adolescents in the population have dyslexia (Hulme & Snowling, 2016). It is more common in males and often co-occurs with other developmental disorders, such as specific language disorder, attention-deficit/hyperactivity disorder (ADHD) or developmental coordination disorder (dyspraxia) (Hulme & Snowling, 2016).

Dyslexia affects the ability to decode or transfer phonological skills to spelling. Over the last decade, decoding skills and phonological awareness in pupils with reading difficulties have been identified as serious inhibitors of successful reading (Klingner et al., 2007), as they affect the fluency of reading. Decoding depends primarily on letter knowledge and phonological skills, which include phonological awareness (Hulme & Snowling, 2009). Phonological awareness is the ability to recognise and manipulate phonemes and is a strong predictor of the development of decoding skills or the successful onset of learning to read (Hulme & Snowling, 2009). Inefficiency in performing these skills can lead to reading being a slow and difficult process (Anderson, 1999; Erbeli & Pizorn, 2012; Segalowitz et al., 1991) and may even lead to a decrease in motivation for reading (Erbeli & Pizorn, 2012).

For many years, experts in the field of reading disabilities have agreed that phonological deficits are a primary cause of dyslexia, as they directly affect learning to read (Snowling & Hulme, 2012). Such deficits are therefore an early and strong predictor of dyslexia (Mather & Wendling 2012). For pupils with dyslexia, difficulties in learning to read accurately and at an adequate speed (reading fluency) are usually at the forefront (Snowling & Hulme, 2012). Even when a pupil achieves adequate reading accuracy, it is significantly more difficult to achieve adequate reading speed with treatment (Fletcher et al., 2007). Young pupils with dyslexia are characterised by (Rief & Stern, 2010):

- slowness in learning the connection between letters and phonemes,
- letter reversals and inversions,
- lack of a systematic approach to sounding out words,
- difficulty in reading words,
- frustration with reading tasks.

Such pupils have good comprehension of material read to them as opposed to material they attempt to read themselves (Rief & Stern, 2010).

Screening and assessment of dyslexia in Slovenia

In Slovenia, pupils with mild to moderate dyslexia receive adapted methods and forms of teaching and testing under the Primary Education Act (Primary Education Act ZOsn-UPB3, 2006), while pupils with severe dyslexia receive more intensive accommodations and additional professional support under the Act on the Guidance of Children with Special Needs (ZUOPP-1, 2011). The process of identification and diagnostic assessment of dyslexia, which requires a multidisciplinary team of professionals (psychologist, special and rehabilitation teacher, speech therapist), involves several stages, from detection, classification,

support planning and progress monitoring to evaluation (Magajna, 2011). The first stage of identifying pupils with dyslexia (detection) is screening, which aims to identify students in need of diagnostic assessment and inform individuals of the likelihood of dyslexia (Pollak, 2009). Screening tests allow dyslexia to be confirmed in young pupils, thus enabling appropriate treatment to be implemented before they experience a sense of failure (Snowling, 2013). Tests used to detect dyslexia include phonological awareness tests, tests of reading aloud and silently (decoding, spelling, reading fluency – speed and accuracy), reading comprehension, rapid naming, memory, attention, etc.

In Slovenia, there are several tests for dyslexia-like reading and writing difficulties that test different elements of reading and writing (phonological awareness, reading speed and accuracy, reading automation, reading comprehension, dictated writing, written expression): *The Reading and Writing Disability Test* or Šali Test (Šali, 1971) (the test is only partially standardised for the population of children in the second grade); *SNAP – Special Needs Assessment Profile* (SNAP is not a test in the psychometric sense, but an instrument for gathering information about the pupil relevant to identifying potential difficulties in a particular skill) (Weedon & Reid, 2018); *The One-Minute Test of Reading Aloud* (Gradišar & Pečjak, 1991); *The Reading Comprehension Test* (Elley et al., 1995); *The Reading Test* (Pečjak et al., 2012b) (the test is a standardised measurement instrument that assesses general reading ability at the end of the first three years of primary school); *The Reading Ability Assessment Scheme – OSBZ* (Pečjak et al., 2012a) (the test is a standardised measurement instrument and the data collected with the OSBZ provide information about what reading skills the student has already developed); *The Test of Reading Fluency Based on the Curriculum Model for Grades 2, 3 and 4* (Košir, 2011); and *The Phonological Awareness Test* (Magajna, 1994).

Early identification of dyslexia is a key to providing appropriate support and intervention for pupils with dyslexia. Due to the multidimensional nature of the disorder, a variety of tests and test batteries are used to effectively identify dyslexia. Good screening is important in order to distinguish pupils who are at risk of developing reading and writing disorders from those who are not. To identify reading difficulties, pupils are screened for various components of reading, such as phonological awareness, reading fluency (speed and accuracy of decoding), reading automaticity, reading comprehension, etc.

Information and communication technology (ICT) appears to be an increasingly important tool for dyslexia screening and the necessary interventions to address the specific learning difficulties and needs of individual learners (Drigas & Politi-Georgousi, 2019). ICT is an important factor in improving traditional

methods of identifying dyslexia, as well as in exploring new perspectives on identifying individuals with dyslexia (Perera et al., 2016). Rooms (2000) highlights the potential benefits of using ICT for pupils with dyslexia in primary schools, emphasising the fact that it can be accessible and available without making pupils with dyslexia feel different or excluded. Multisensory approaches (auditory, oral, visual, kinaesthetic) and systems are incorporated to mitigate the difficulties of pupils with dyslexia (Rooms, 2000). Diagnostic assessment using ICT allows psychologists and other professionals to easily and quickly assess cognitive abilities and other important skills (Singleton, 2001). Interactive multimedia, virtual environments, neural networks, software, fuzzy logic, game-based techniques and mobile applications improve the effectiveness of traditional dyslexia screening procedures, with each approach offering sophisticated features that facilitate assessment procedures (Menghini et al., 2011).

Research problem and Research question

Dyslexia often manifests itself in young pupils through slow progress in learning to read and write. The difficulties are frequently reflected in poorer academic achievement and, consequently, lower self-esteem. It is therefore important to identify dyslexia as early as possible and treat it appropriately. This helps to prevent the stigmatisation of children and adolescents with dyslexia, to promote their inclusion in society and to reduce difficulties in adulthood.

The use of computer systems to identify pupils with dyslexia is already relatively well established worldwide. A wide range of software is available to teachers, from screening software to more detailed computer-based assessment batteries. Most computer-based dyslexia detection programs rely on assessments of reading and spelling skills as well as cognitive abilities such as phonological awareness and verbal memory, which support literacy development and are generally good predictors of dyslexia (Singleton et al., 2009). Both traditional tests and applications have their advantages and limitations. The advantage of traditional tests is the presence of an expert who administers the test while observing the pupil, checking the pupil's comprehension, adjusting the instructions so that the pupil understands them, and observing the pupil's attention span and possible fatigue. At the same time, the expert can encourage and support the pupil. The main disadvantages of traditional testing are the exposure of the individual and the time-consuming nature of the test. These factors can be eliminated with the help of an application. Moreover, the application can be used by several pupils at the same time, so that many pupils can be assessed in a short time, enabling at-risk pupils to be differentiated from those

who are not at risk. The application also has advantages from a motivational point of view, as it often resembles a computer game rather than an assessment.

Our overall goal was to train a machine learning model to differentiate between pupils with identified dyslexia and a control group of pupils without dyslexia. In this context, our research problem was to identify the important parameters of pupils' oral reading fluency and to investigate whether we can use these parameters as features for our model. In order to determine the parameters of oral reading fluency, we first performed manual transcriptions of audio recordings and defined the types of errors that pupils made most frequently in reading. We defined the parameters of reading based on the defined error types for each word in the six texts obtained from the test battery of the Slovenian National External Assessment of Knowledge for third-grade pupils. Based on the values of these parameters for each pupil, we extracted a subset of the most important parameters and used machine learning methods to build models to classify pupils into one of two groups: 'identified dyslexia' or 'control'.

Method

Participants

The participants of the study were 12 pupils with dyslexia officially diagnosed by experts from the Counselling Centre for Children, Adolescents and Parents Ljubljana and 15 pupils without identified dyslexia. The pupils were from six different primary schools in Ljubljana, from the third (n = 13) and fourth (n = 14) grades. The age of the participants ranged from 8 to 9 years. Five of the pupils in the third grade and seven in the fourth grade were officially diagnosed with dyslexia. We only included pupils who had a signed parental consent form confirming participation and storage of the collected data for further analysis. Participation was anonymous. We did not record the pupils' first and last names; we only recorded their age and whether they had already been diagnosed with dyslexia.

Instruments

As a research instrument, we used the desktop application PKP Dyslexia[4] to test skills that are typically less well developed in people with dyslexia. The application contains six tests (sequencing concept test, reading comprehension test, phonological awareness test, working memory test, reading aloud

---

4   The study used a desktop version of the PKP – Dyslexia Web Application, previously developed at the Faculty of Computer Science and Informatics as part of the Creative Pathways project. The desktop application was completed as part of a thesis by Kunej (2021).

test and silent reading with an eye tracker), each of which comprises a series of tasks. The tests require the use of cognitive and language skills, which are key to successful reading and writing. In designing the tests, we followed the protocols for developing psychological tests according to international guidelines (e.g., various International Testing Commission guidelines) and the American Standards for Educational and Psychological Instruments (Standards for Educational and Psychological Testing, 2014), as well as guidelines for developers of computer-based psychological tests. Experts from various fields participated in the development. In this study, we present only the results of the Reading Aloud test used to test reading fluency (speed and accuracy/correctness/error).

The reading aloud test included six texts from the test battery of the Slovenian National External Assessment of Knowledge (CEAK) in the mother tongue (Slovenian) for third-grade pupils. This is the first national assessment of knowledge in mother tongue (Slovenian) proficiency in which pupils take part. For the purposes of the present study, the texts were selected from previous years' test batteries. The six texts were all informative and were about topics of general interest to the children (e.g., wild animals and a fairy tale). The level of difficulty of the texts was assessed using an application developed within the project For the Quality of Slovenian Textbooks (KaUč). According to the Automated Readability Index and the Coleman-Liau Index, which take word and sentence length as a criterion, each of the six reading tasks had acceptable reliability indices in the respective years in which they were administered to a national sample of students. The texts have a very similar difficulty level, with the exception of the text entitled The Mountain Gorilla, which is slightly more difficult but still much easier than average (see Table 1).

The texts used in the task vary in length. The shortest text contains 28 words, three of the texts contain about 40 words (36, 40 and 41 words, respectively), one is slightly longer at 57 words, while the longest text has 123 words.

Table 1

Characteristics of the input texts

Text	Title	Automated Readability Index[6]	Coleman-Liau Index[7]	Length in words	Rare words[8]
1	Gorska gorila (*Eng. Mountain Gorilla*)	18.2	23,6	28	6
2	Leopard (*Eng. Leopard*)	2.8	5,4	57	14
3	Šimpanz (*Eng. Chimpanzee*)	2.9	6.5	36	11
4	Koala (*Eng. Koala*)	3.4	7.1	41	10
5	Lev (*Eng. Lion*)	2.8	5.1	40	9
6	Dobra vila v dolini Soče (*Eng. A Good Fairy in the Soča Valley*)	2.0	3.6	123	14

Note. Texts 1 to 5: adapted from National Geographic Junior, issue 124, December 2015. Text 6: Slovenian folk tales about fairies and elves. Published in Zmajček, Vol. 20, No. 1, September 2013.

The difficulty level of the texts used to test reading fluency (speed and accuracy) is important, as they must be simple enough to be appropriate for pupils in the third and fourth grades. At the same time, the texts should contain enough specific features that might cause reading difficulties for pupils with dyslexia.

The texts used in the test battery of CEAK also contain rare words (Table 1) that are considered more difficult to process for pupils with dyslexia (Rüsseler et al., 2003; Suárez-Coalla & Cuetos, 2015). Pupils with dyslexia read the words they encounter frequently in texts faster and more accurately, so they become part of their reading vocabulary. Building a reading vocabulary is challenging for pupils with dyslexia, as they have difficulty learning and recognising new words in print. In pupils with dyslexia, there is often a discrepancy between their spoken vocabulary, which can be very large, and their reading vocabulary (Bailey, 2020).

Below we present the six texts included in our study and a graphical representation of their readability. The graphs included show (1) how the entered text compares with texts from the ccKres[8] corpus in terms of readability, and

---

5   The Automated Readability Index is a simple measure of readability based on two components: word length and sentence length. The higher the number of words with many letters and sentences with many words, the higher the Automatic Readability Index. Higher values indicate lower readability (Škvorc et al., n. d.).

6   The Coleman-Liau Index is similar to the Automated Readability Index and is based on the length of words and sentences. The more words with many letters and sentences with many words a text contains, the higher the Coleman-Liau Index. Higher values indicate lower readability (Škvorc et al., n. d.).
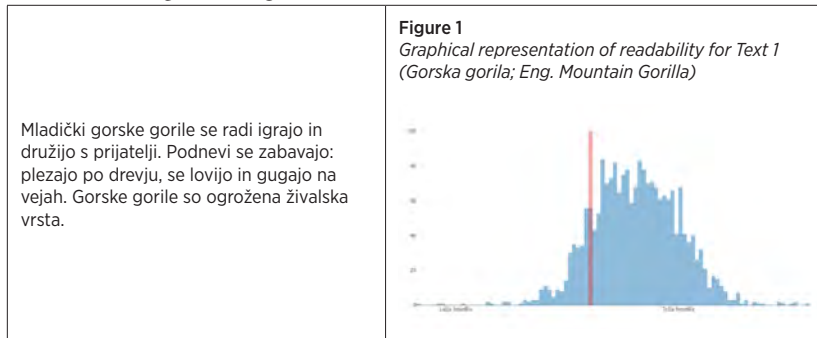
7   Rare words are words not included in the list of common words (Škvorc et al., n. d.).

8   ccKres is a collection of Slovenian texts from fiction, non-fiction, newspapers, magazines and web texts, containing a total of 10 million words (Škvorc et al., n. d.).
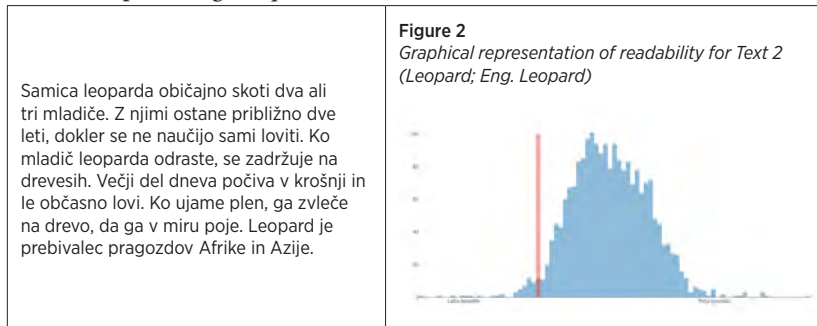
(2) a histogram of the readability measures across texts in the ccKres corpus, where the red line shows where the evaluated text is located compared to all of the texts in the corpus (Škvorc et al., n. d.).

Graphical representation of readability for Text 6 (Dobra vila v dolini Soče; Eng. A Good Fairy in the Soča Valley)
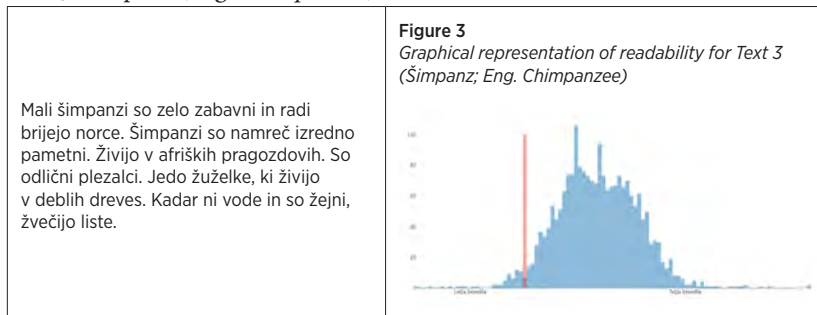
Text 1: Gorska gorila (*Eng. Mountain Gorilla*)

Mladiči gorske gorile se radi igrajo in družijo s prijatelji. Podnevi se zabavajo: plezajo po drevju, se lovijo in gugajo na vejah. Gorske gorile so ogrožena živalska vrsta.

Figure 1
Graphical representation of readability for Text 1 (Gorska gorila; Eng. Mountain Gorilla)

Text 2: Leopard (*Eng. Leopard*)

Samica leoparda običajno skoti dva ali tri mladiče. Z njimi ostane približno dve leti, dokler se ne naučijo sami loviti. Ko mladič leoparda odraste, se zadržuje na drevesih. Večji del dneva počiva v krošnji in le občasno lovi. Ko ujame plen, ga zvleče na drevo, da ga v miru poje. Leopard je prebivalec pragozdov Afrike in Azije.

Figure 2
Graphical representation of readability for Text 2 (Leopard; Eng. Leopard)

Text 3: Šimpanz (*Eng. Chimpanzee*)

Mali šimpanzi so zelo zabavni in radi brijejo norce. Šimpanzi so namreč izredno pametni. Živijo v afriških pragozdovih. So odlični plezalci. Jedo žuželke, ki živijo v deblih dreves. Kadar ni vode in so žejni, žvečijo liste.


Figure 3
Graphical representation of readability for Text 3 (Šimpanz; Eng. Chimpanzee)

Text 4: Koala (*Eng. Koala*)

Ko koale pridejo na svet, so velike kot bonbon. Približno pol leta preživijo kar v materini vreči. Zato malim avstralskim koalam ni treba hoditi v vrtec. Koal ne smemo zamenjati z medvedi, čeprav so jim podobne.
So vrečarji, tako kot kenguruji.

Figure 4
Graphical representation of readability for Text 4 (Koala; Eng. Koala)
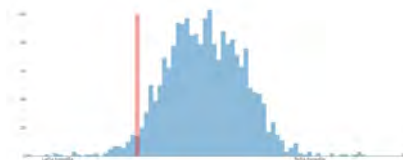
Text 5: Lev (*Eng. Lion*)

Mladi levi bi radi čim prej odrasli.
Takrat pri večerji ne bodo več čakali, da se najprej najedo starejši samci in samice iz krdela.
Pri levih po navadi lovijo odrasle samice.
Naloga samcev pa je, da stražijo in branijo ozemlje.

Figure 5
Graphical representation of readability for Text 5 (Lev; Eng. Lion)
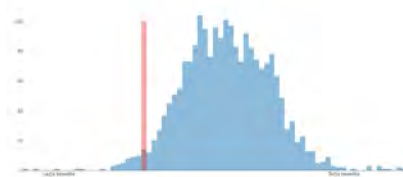
Text 6: Dobra vila v dolini Soče (*Eng. A Good Fairy in the Soča Valley*)

Na bregu Soče je nekdaj stala koča, v kateri je živel reven kmet s sinom, ki je pasel ovce. Nekega dne je fantič zašel v gozd in ni našel poti domov. Prišel je do studenca in v travi ob vodi zagledal ribico, ki se je nemočno premetavala. Hitro jo je položil v vodo, a v tistem trenutku se je spremenila v prelepo vilo.
»Hvala ti. Rešil si mi življenje. Kako naj ti to poplačam?« je spregovorila vila.
»Prosim, pokaži mi pot domov,« jo je zaprosil pastir. Vila je vodila dečka skozi gozd in ga pripeljala do njegovega doma. Pastir bi se ji rad zahvalil, a dobra vila je nenadoma izginila, njegova pastirska palica pa se je v tistem trenutku spremenila v zlato palico.

Figure 6
Graphical representation of readability for Text 6 (Dobra vila v dolini Soče (Eng. A Good Fairy in the Soča Valley))

The position of the red lines in the histograms above for texts 1 through 6 indicates where the scored text is placed compared to all of the texts in the corpus, thus showing that all six texts are relatively easy texts.

The user interface is designed to attract pupils while ensuring that a single display does not contain unnecessary and distracting stimuli or too many elements at once. Information is displayed sequentially and in small sections.

The colour contrast between the text and the background is specifically designed to suit the visual processing characteristics of pupils with dyslexia. The text is left-aligned to make it easier and faster for pupils to find the beginning of the text on a new line. When the text is displayed, the program begins to time and record the voice. The time is stopped when the pupil reads the text and clicks the 'NEXT TEXT' button. The purpose of this task was to obtain audio recordings of the pupils reading aloud.

Research Design

In our experiments, we used the desktop application *PKP – Disleksija*. In the Reading Aloud Test, the pupils were asked to read six texts each, which were displayed on a 15-inch laptop screen. The test contains written and auditory instructions that are carefully prepared in such a way that it is assumed that pupils will understand them. However, it is also accepted that parents will help pupils to understand the instructions. The instructions are followed by a brief demonstration that gives the pupil a clear visual idea of how to approach the test. After the initial instructions, the pupil is given a series of exercises to check that he or she has understood the instructions (verifying that the pupil has understood how to complete the task). These preliminary exercises are not scored and the pupil has the opportunity to review the instructions again while performing them. This is followed by six reading aloud tasks that are recorded and then scored. The Zoom H4n Pro handheld digital recorder was used to collect the audio data. The read-aloud test data was collected between 9 June and 18 June 2021.

Results

Our experimental work focused on using machine learning methods for the classification of pupils into one of two groups: those with 'identified dyslexia' and a control group 'without identified dyslexia'. Due to the small sample of pupils, we were limited to using machine learning methods that require a great amount of pre-processing; we could not use raw audio recordings for input, but instead had to extract the features from them. We struggled to automate the feature extraction process, but managed to construct the features manually. This limits the applicability of our models to the six texts that were used in this study.

Audio transcription and feature construction

The audio recordings were manually transcribed using Audacity software (Audacity® software is copyright © 1999-2021). Four attributes were defined for each transcribed word: **start**, **word**, **end**, **error_type**. Each line of the

transcription file refers to a single word in the text. The **start** feature indicates the time when the reader started reading the **word** aloud, and **end** indicates the end time of reading the **word**. Both values are written in the format *{MM:SS. mmm}*, where MM denotes minutes, SS seconds and mmm milliseconds. The **word** feature indicates the word that was read: all of the vowels that the reader read aloud when reading each particular word we written down. The **error_ type** denotes the type of error that occurred while reading the word. From the audio transcriptions of all six texts, seven most common error_types were identified, which were labelled with numbers from 1 to 7:

1.	Misread word (e.g., **balon** instead of **bonbon**),
2.	Word read n-times (e.g., **ko ko**, marked as 2:2),
3.	Word sequence read n-times (e.g., **ki živijo v ki živijo v**, marked as 3:2 at each word of the sequence),
4.	Character elongating,
5.	Reading stutter (e.g., **zazabavni** instead of **zabavni**),
6.	Incorrectly stressing the word,
7.	Omitting the word.

For each pupil, a separate transcription file was created for each text, giving a total of 27 * 6 = 162 transcription files. In order to use this data in our Orange (Demsar et al., 2013) machine learning setting, all transcribed features for each pupil were combined into a single learning example, resulting in 27 learning examples and 618 features (the features from transcriptions, i.e., error types, silence before and reading time). The dataset is well balanced: 12 examples belong to a positive target class (identified dyslexia) and 15 to a negative class (without identified dyslexia).

The attributes were standardised so that they all have $\mu = 0$ and $\sigma^2 = 1$. Despite the small dataset and the large number of features, the goal was to learn a model that predicts the target outcome (identified dyslexia). The leave-one-out method was used in all of our experiments in order to evaluate the models.

The goal was to see how well an ensemble method performed on our data. Ensemble methods are machine learning techniques that combine a set of base models, such as decision trees. Each base model contributes to the ensemble model with its own prediction; ultimately, the ensemble model predicts the outcome based on the votes of all of the base models. We tried extreme gradient boosting of random forest (xgboost), which consists of 100 trees and limits the depth of each tree to 3, but allows all of the attributes in each tree, level and split. The confusion matrix in Table 2 shows the results of the leave-one-out test for the xgboost model, which indicates three misclassified pupils from our dataset.

Table 2

Results of the leave-one-out test for the xgboost model

		Predicted		Σ
		Pupils without identified dyslexia	Pupils with identified dyslexia	
Actual	Pupils without identified dyslexia	13	2	**15**
	Pupils with identified dyslexia	1	11	**12**
	Σ	**14**	**13**	**27**

Ensemble models usually provide good predictions but are difficult or impossible for humans to understand. In order to gain insights, we focused on simple methods that can provide models humans can understand: a naive Bayesian classifier, a decision tree and Freeviz (Demsar et al., 2013). Before learning, feature subset selection was performed using ReliefF (Kononenko, 1994), which selected the following top ten features from all six texts:

- igrajo_silence_before
- gorile_silence_before
- se_reading_time.1.2
- in_silence_before.1.2.1
- do_reading_time
- je_reading_time.4
- se_reading_time
- nenadoma_silence_before
- običajno_reading_time
- običajno_silence_before

These features were used for learning a naive Bayesian classifier, a classification tree and the Freeviz visualisation. The feature names are combinations of the word that was read and the type of feature it describes. Two types of features were chosen:

- silence_before describes how much time was needed before the word was read aloud (example: igrajo_silence_before is a feature that describes the silence needed before the word 'igrajo' was read aloud),
- reading_time describes how much time was needed to read the word aloud (example: običajno_reading_time is a feature that describes the time needed to read the word 'običajno' aloud).

The nomogram in Figure 7 serves as a visual representation of the naive Bayesian classifier. The contribution of each feature is measured as a score and

the individual scores are summed and converted into the probability of the target class (pupils with identified dyslexia). The features are ranked by importance: the strongest influence on the target class (pupils with identified dyslexia) are the features običajno_silence_before and nenadoma_silence_before.

The confusion matrix of the naïve Bayesian classifier in Table 3 shows that only one child was misclassified in the leave-one-out test.

Figure 7

The nomogram of the Naive Bayesian classifier
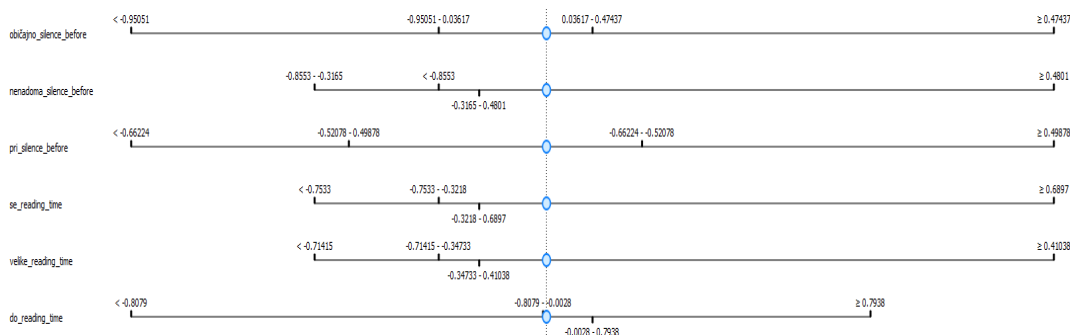
Table 3

Results of the leave-one-out test for the Naive Bayesian model

		Predicted		Σ
		Pupils without identified dyslexia	Pupils with identified dyslexia	
Actual	Pupils without identified dyslexia	14	1	**15**
	Pupils with identified dyslexia	0	12	**12**
	Σ	**14**	**13**	**27**

The classification tree learned with the above features is shown in Figure 8. Again, the same two features turn out to be the most important: the words 'običajno' and 'nenadoma' seem to be the most difficult in the six texts. The values of the splits should be interpreted in the context of feature standardisation ($\mu = 0$ and $\sigma^2 = 1$):

- običajno_silence_before takes the values in the interval [-0.99, 2.35]; the divided value of 0.371 is slightly above the mean and indicates that the pupils with a longer than average pause before this word are classified as pupils with dyslexia. The rest of the pupils – those who make shorter

pauses before reading the word 'običajno' – are further checked in the classification tree for the time of silence before the word 'nenadoma'.

- nenadoma_silence_before takes the values in the interval [-0.88, 2.5]. The divided value of 0.38 is again about one-third of the length of the interval. Those who took less time before reading the word 'nenadoma' are classified as pupils without dyslexia, while the rest are predicted as pupils with dyslexia.

Figure 8

While the numbers in the splits with their absolute values do not explain much (due to standardisation), both splits show that long silences before the two difficult words 'običajno' and 'nenadoma' predict a positive target class (identified dyslexia). Note that the same two attributes have the largest positive influence on the target class in the above NB nomogram.
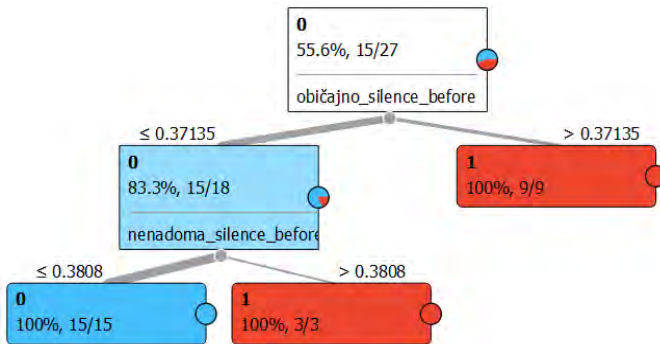

Finally, in Figure 9 we present a FreeViz projection that visually confirms the observations from the nomogram and the decision tree. FreeViz (Demšar, 2007) is a method that optimises a linear projection of data with a discrete class variable (in our case it has two values: 'identified dyslexia' and 'control group') and displays the projected data in a two-dimensional scatter plot. FreeViz can reveal interesting relationships between classes and features; in our domain, the explanation for the FreeViz projection is as follows.

The blue area, concentrated in the middle, represents the pupils from the control group; they have shorter reading times and even pause before the more difficult words. In contrast, the red area, which represents our target class (pupils identified with dyslexia), extends around the blue area and shows higher scores on all observed variables.

Figure 9
FreeViz projection.
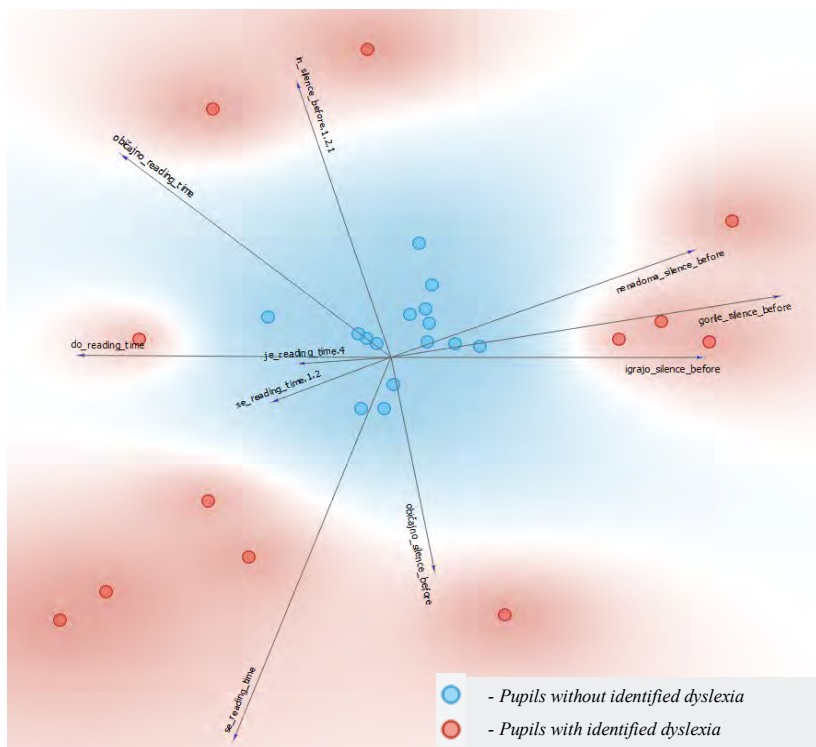
Discussion

In order to become a good reader, pupils need to develop two basic skills: decoding and reading comprehension (Nation, 2006). With practice, decoding soon becomes quick, flexible and efficient in pupils who have no difficulty in this area (Nation, 2006). The reading test in the present study included six texts from the Slovenian National External Assessment of Knowledge (CEAK) battery in the mother tongue (Slovenian), all of which deal with topics of general interest to children (e.g., wild animals and fairy tales).

In order to select texts that are easy enough to be suitable for 8- and 9-year-olds, but at the same time contain enough features (e.g., rare words) that might create reading difficulties for pupils with dyslexia, the difficulty of all six texts was assessed using the KaUč readability application, which is used to evaluate Slovenian textbooks. Since the texts belong to the CEAK battery,

the KaUč readability application was found to be an appropriate tool for assessment. We used the Automated Readability Index, which assesses the difficulty of a text based on the length of words and sentences. All six texts were rated as very easy in the ccKres text corpus, which contains Slovenian texts from various sources and has more than 10 million words. Although the texts were simple, they discriminated between the two groups of learners: those with dyslexia and those without.

Machine learning methods have been used to predict pupil reading disabilities. Based on a small but balanced sample, our models clearly distinguished between pupils with reading difficulties, e.g., dyslexia, and a control group of pupils without dyslexia. Although the six selected texts were classified as easy by the KaUč readability application, we were able to determine that they were suitable for detecting a reading difficulty in pupils in the third grade. Other authors have also emphasised the importance of assessing reading fluency as one of the distinguishing characteristics of pupils with dyslexia. In a meta-analysis, Carioti et al. (2021) explained that reading fluency can be meaningfully considered as the most important parameter for diagnosing developmental dyslexia, as deficits in reading speed, lexical recognition and phonological recoding have been identified as universal manifestations of reading deficits, regardless of age and orthographic depth of language. This suggests that the use of time-limited approaches in reading tasks does not provide contradictory or less robust evidence for the presence of developmental dyslexia (Carioti et al., 2021). In particular, in the context of transparent orthographic systems, where there is a high degree of correspondence between graphemes and phonemes, the authors suggest that the main feature of developmental dyslexia is poor reading fluency (Martínez-García et al., 2019).

The models presented in the aforementioned study are highly relevant to the six selected texts from our study, but are not generally applicable. The data set used in our experiments is very small, consisting of only 27 pupils, which is understandable due to the nature of the task. Although the conclusions are promising, a bigger sample size would be needed to determine whether the results are significant and how well they generalise to a larger population. Our methodology shows that different machine learning methods on audio transcripts can clearly distinguish between pupils with reading disabilities, e.g., dyslexia, and a control group without dyslexia, even for short and simple texts. The latter suggests that basic screening tests could be short and effective. Carioti et al. (2021) make a similar point: it is important to be aware that the reading process can be stressful for those with developmental dyslexia. Therefore, it is useful to use time-limited reading tasks and not to overwhelm pupils with long and complex reading tasks

whose reliability and clinical validity may be questionable. In this context, several authors have pointed out that it is not optimal to adopt an assessment of reading skills based solely on accuracy; although accuracy is an important parameter, it is not the only one, especially when assessing cross-linguistic differences in reading skills, when orthographic transparency or deficit compensation (at least for this parameter) can easily lead to inaccurate results in adulthood (Carioti et al., 2021; Sprenger-Charolles et al., 2011).

Assessing pupils' reading fluency is important not only to identify problems, but also to monitor progress in this area. Based on research findings, Kairaluoma et al. (2007) suggest that students with reading difficulties benefit from reading fluency intervention. They add that the intervention should be long term and initially based on emphasising syllables as sublexical reading units and then gradually progressing to larger reading units. It is also worth noting that prior phonological and semantic training facilitates the formation of orthographic representations, as evidenced by a reduction in the length effect (Martínez-García et al., 2019). When comparing 8- to 9-year-old pupils with and without dyslexia before the implementation of a training programme based on letter-sound associations, with a particular focus on increasing reading fluency, González et al. (2015) found that the group of dyslexic pupils showed more severe impairments on measures of word reading speed than on measures of accuracy (González et al., 2015). When evaluating the impact of the training programme comparing 8 to 9-year-old pupils with and without dyslexia, they found that the pupils with dyslexia improved significantly in the main measures of word reading and spelling after the training, progressing at a faster rate than both the group of pupils without dyslexia and the group of pupils with dyslexia in the control group who were waiting for the programme (González et al., 2015).

Conclusion

In our study, we trained different machine learning models to predict pupil reading disabilities. Despite the small sample, all of the models clearly distinguished between pupils with reading disorders and a control group. It was demonstrated that fluency in oral reading can be measured objectively even in short and simple texts. The machine learning methodology used is based on transcription data, which was constructed manually from audio recordings of oral reading. Manual construction of such data is tedious and subjective work, and is therefore impractical for larger datasets of audio recordings. Our future work will focus on automating audio transcriptions and feature construction from automatically obtained transcripts. We will also explore the possibility of

working directly with audio signals and include methods for incorporating the knowledge of domain experts into our learning dataset.

Acknowledgement

References

Anderson, N. J. (1999). *Exploring second language reading: Issues and strategies*. Heinle & Heinle.

Audacity® software is copyright © 1999-2021 Audacity Team. The name Audacity® is a registered trademark.

Bailey, E. (2020). *Tips for teaching vocabulary to students with dyslexia*. ThoughtCo. https://thoughtco.com/teaching-vocabulary-to-students-with-dyslexia-3111207

Boardman, A. G., Roberts, G., Vaughn, S., Wexler, J., Murray, C. S., & Kosanovich, M. (2008). *Effective instruction for adolescent struggling readers: A practice brief.* RMC Research Corporation, Center of Instruction.

Carioti, D., Masia, M. F., Travellini, S., & Berlingeri, M. (2021). Orthographic depth and developmental dyslexia: A meta-analytic study. *Annals of Dyslexia, 71*(3), 399–438. https://doi.org/10.1007/s11881-021-00226-0

Carpentieri, J. D. (2012). *Act now: The EU high level group of experts report on literacy*. https://discovery.ucl.ac.uk/id/eprint/10061875/1/HLGL-final-report_en.pdf

Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbontar, J., Zitnik, M., & Zupan, B. (2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research, 14*(1), 2349–2353.

Demšar, J., Leban, G., & Zupan, B. (2007). FreeViz—An intelligent multivariate visualization approach to explorative analysis of biomedical data. *Journal of Biomedical Informatics, 40*(6), 661–671.

Drigas, A. S., & Politi-Georgousi, S. (2019). ICTs as a distinct detection approach for dyslexia screening: A contemporary view. *International Journal of Online and Biomedical Engineering, 15*(13), 46–60. https://doi.org/10.3991/ijoe.v15i13.11011

Elley, W. B., Gradišar, A., & Lapajne, Z. (1995). Kako berejo učenci po svetu in pri nas [How pupils read in Slovenia and abroad]. *Education, 3.*

Erbeli, F., & Pizorn, K. (2012). Reading ability, reading fluency and orthographic skills: The case of L1 Slovene English as a foreign language students. *Center for Educational Policy Studies Journal, 2*(3), 119–139.

Fletcher, J., Lyon, G., Fuchs, L., & Barnes, M. (2007). *Learning disabilities: From identification to*

intervention. The Guilford Press.

González, G. F., Žarić, G., Tijms, J., Bonte, M., Blomert, L., & van der Molen, M. W. (2015). A randomized controlled trial on the beneficial effects of training letter-speech sound integration on reading fluency in children with dyslexia. *PLoS ONE, 10*(12). https://doi.org/10.1371/journal.pone.0143914

Gradišar, A., & Pečjak, S. (1991). Enominutni test glasnega branja [The One-Minute Test of Reading Aloud].

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language, learning and cognition.* Wiley-Blackwell.

Hulme, C., & Snowling, M. J. (2016). Reading disorders and dyslexia. *Current Opinion in Pediatrics*, *28*(6), 731.

IDA (2002). *Definition of dyslexia.* International Dyslexia Association, Board of Directors: 12 November 2002. http://eida.org/definition-of-dyslexia/

Kairaluoma, L., Ahonen, T., Aro, M., & Holopainen, L. (2007). Boosting reading fluency: An intervention case study at subword level, *Scandinavian Journal of Educational Research, 51*(3), 253–274. https://doi.org/10.1080/00313830701356117

Klingner, J. K., Vaughn, S., & Boardman, A. (2007). *Teaching reading comprehension to students with learning difficulties.* The Guilford Press.

Kononenko, I. (1994). Estimating attributes: Analysis and extensions of Relief. In L. De Raedt and F. Bergadano (Eds.), *Machine learning: ECML-94* (pp. 171–182). Springer Verlag.

Košir, J. (2011). Formativno ocenjevanje s preizkusom tekočnosti branja, ki temelji na kurikulu [Formative assessment with a curriculum-based test of reading fluency]. In L. Magajna & M. Velikonja (Eds.), *Učenci z učnimi težavami: Prepoznavanje in diagnostično ocenjevanje* (pp. 105–123). Univerza v Ljubljani, Pedagoška fakulteta.

Magajna, L. (1994). Razvoj bralnih strategij – vloga kognitivnega in fonološkega razvoja ter fonološke strukture jezika [Development of reading strategies: The role of cognitive and phonological development and the phonological structure of language]. [Doctoral dissertation]. Univerza v Ljubljani, Filozofska fakulteta.

Magajna, L., Kavkler, M., Košak Babuder, M., Zupančič Danko, A., Seršen Fras, A., & Rošer Obretan, A. (2015). Otroci s primanjkljaji na posameznih področjih učenja. [Children with deficits in individual areas of learning]. In N. Vovk Ornik (Eds.), *Kriteriji za opredelitev vrste in stopnje primanjkljajev, ovir oz. motenj otrok s posebnimi potrebami* [Criteria for defining the type and degree of deficits, obstacles and disorders of children with special needs] (pp. 23–31). Zavod RS za šolstvo.

Martínez-García, C., Suárez-Coalla, P., & Cuetos, F. (2019). Development of orthographic representations in Spanish children with dyslexia: The influence of previous semantic and phonological knowledge. *Annals of Dyslexia, 69*, 186–203. https://doi.org/10.1007/s11881-019-00178-6

Mather, N., & Wendling, J. B. (2012). *Essentials of dyslexia assessment and intervention.* John Wiley & Sons.

Menghini, D., Finzi, A., Carlesimo, G. A., & Vicari, S. (2011). Working memory impairment

in children with developmental dyslexia: Is it just a phonological deficity? *Developmental Neuropsychology, 36*(2), 199–213.

Moats, L. C., & Dakin, K. E. (2008). *Basic facts about dyslexia and other reading problems.* The International Dyslexia Association.

Nation, K. (2006). Assessing children's reading comprehension. In M. J. Snowling and J. Stackhouse (Eds.), *Dyslexia, speech and language: A practitioner's handbook* (pp. 128–142). Whurr publishers.

Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research? *Cognition, 35*(2), 159–182. https://doi.org/10.1016/0010-0277(90)90013-A

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: Reuniting the developmental disorders? *Trends in neurosciences, 30*(4), 135–141. https://doi.org/10.1016/j.tins.2007.02.003

Nijakowska, J. (2016). Grasping dyslexia: Bridging the gap between research and practice. *Selected Papers on Theoretical and Applied Linguistics, 21*, 43–58.

Pečjak, S., Magajna, L., & Potočnik, N. (2012a). *Ocenjevalna shema bralnih zmožnosti učencev 1.–3. razreda: OSBZ* [An evaluation scheme for pupils' reading abilities in grades 1–3]. Znanstvena založba Filozofske fakultete, Univerze v Ljubljani.

Pečjak, S., Magajna, L., Potočnik, N., & Podlesek, A. (2012b). Bralni test [The reading test]. Znanstvena založba Filozofske fakultete, Univerze v Ljubljani.

Perera, H., Shiratuddin, M. F., & Wong, K. W. (2016). Review of the role of modern computational technologies in the detection of dyslexia. In *Information Science and Applications (ICISA) 2016* (pp. 1465–1475). Springer. https://doi.org/10.1007/978-981-10-0557-2_141

Pollak, D. (Ed.). (2009). *Neurodiversity in higher education: Positive responses to specific learning differences.* John Wiley & Sons.

Poznanovič, M., Cestnik, M., Čuden, M., Gomivnik Thuma, V., Honzak, M., Križaj, M., Rosc-Leskovec, D., Žveglič, M., & Ahačič, K. (2018). *Učni načrt. Program osnovna šola. Slovenščina.* [Curriculum. Primary school curriculum. Slovenian]. Revised edition. Ministrstvo za izobraževanje, znanost in šport: Zavod Republike Slovenije za šolstvo.

Raduly Zorgo, E., Smythe, I., Gyarmathy, É., Košak Babuder, M., Kavkler, M., & Magajna, L. (2010). *Disleksija – vodnik za tutorje* [Dyslexia – A guide for tutors]. Univerza v Ljubljani, Pedagoška fakulteta.

Rief, S. F., & Stern, J. M. (2010). *The dyslexia checklist: A practical reference for parents and teachers.* Jossey-Bass.

Rooms, M. (2000). Information and communication technology and dyslexia. In *Dyslexia in Practice* (pp. 263–272). Springer, MA.

Rose, J. (2009) *Identifying and teaching children and young people with dyslexia and literacy difficulties: An independent report from Sir Jim Rose to the Secretary of State for Children, Schools and Families.* DCSF Publications.

Rüsseler, J., Probst, S., Johannes, S., & Münte, T. F. (2003). Recognition memory for high- and low-frequency words in adult normal and dyslexic readers: An event-related brain potential study. *Journal of Clinical and Experimental Neuropsychology, 25*(6), 815–829. https://doi.org/10.1076/jcen.25.6.815.16469.

Segalowitz, N., Poulsen, C., & Komoda, M. (1991). Lower level components of reading skill in higher level bilinguals: Implications for reading instruction. *AILA review, 8*(1), 15–30.

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

Singleton, C. (2001). Computer-based assessment in education. *Educational and Child Psychology*.

Snowling, M. J. (2013). Early identification and interventions for dyslexia: A contemporary view. *Journal of Research in Special Educational Needs, 13*(1), 7–14.

Snowling, M. J., & Hulme, C. (2012). Annual research review: The nature and classification of reading disorders – a commentary on proposals for DSM-5. *Journal of Child Psychology and Psychiatry, 53*(3), 593–607.

Sprenger-Charolles, L., Siegel, L. S., Jiménez, J. E., & Ziegler, J. C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: A review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading, 15*(6), 498–521. https://doi.org/10.1080/10888438.2010.524463.

Suárez-Coalla, P., & Cuetos, F. (2015). Reading difficulties in Spanish adults with dyslexia. *Annals of Dyslexia, 65*(1), 33–51. https://doi.org/10.1007/s11881-015-0101-3

Šali, B. (1971). *Test motenosti v branju in pisanju (T – MBP)* [The reading and writing disability test]. Zavod SR Slovenije za rehabilitacijo invalidov.

Škvorc, T., Robnik Šikonja, M., Žagar, A., Arhar Holdt, Š., Pollak, S., Čibej, J., Pori, E., Kosem, I., Krek, S., & Torkar, G. (n. d.). *Za kakovost slovenskih učbenikov (KaUč) – Aplikacija berljivosti besedila* [For the quality of Slovenian textbooks (KaUč) – Text readability application]. http://www.kauc.si/aplikacija-berljivosti-besedila/

Weedon, C., & Reid, G. (2018). *SNAP-3: Profil ocene posebnih potreb, 3. izdaja: računalniško podprto ocenjevanje in izdelava profila specifičnih učnih težav (5–14 let): priročnik* [SNAP-a 3: Special needs assessment profile, 3rd Edition: Computer-assisted assessment and profiling of specific learning disabilities (5–14 years): Manual.] Center za psihodiagnostična sredstva.

Biographical note

Jure Žabkar, PhD, is an Assistant Professor and researcher at the Artificial Intelligence Laboratory at the Faculty of Computer and Information Science, University of Ljubljana, Slovenia. He conducts research in machine learning and data mining, qualitative reasoning, cognitive robotics and systems for decision support, with applications in robotics and healthcare.

Tajda Urankar, BSc, is pursuing a Master of Applied Data Science degree at Frankfurt School of Finance and Management, Frankfurt am Main, Germany. Her main areas of interest are deep learning topics such as natural language processing, quantitative trading and pricing models with the current focus on the growing digital lending market.

Karmen Javornik is a teaching assistant of Special and Rehabilitation Education at the Faculty of Education, University of Ljubljana, Slovenia. Her research interests include inclusion of people with special needs in the context of education, with a focus on general and specific learning difficulties and the development of strategies and models of support and treatment in these areas, which she links to research on executive functioning.

Milena Košak Babuder, PhD, is an Assistant Professor of Special and Rehabilitation Education at the Faculty of Education, University of Ljubljana, Slovenia. Her research interests include the inclusion of people with special educational needs, the impact of general and specific learning difficulties on the academic performance of pupils and students, and the development of strategies and models of support and treatment in these areas, and in particular the impact of dyslexia on learning English as a foreign language.