DSCREEN- A Game To Screen Dyslexia

Prof. Laila V
Computer Science and Engineering
MEA Engineering College
Perinthalmanna
laila@meaec.edu.in

Swetha N
Computer Science and Engineering
MEA Engineering College
Perinthalmanna
19gcs27@meaec.edu.in

Muhammed Aslam P S
Computer Science and Engineering
MEA Engineering College
Perinthalmanna
19gcs41@meaec.edu.in

Thanseeha C
Computer Science and Engineering
MEA Engineering College
Perinthalmanna
19gcs11@meaec.edu.in

***Abstract*—A significant portion of the population, over 10%, has dyslexia and the majority are not recognized until they struggle in school. This project aims to address this issue by using machine learning to screen for dyslexia early on, through a game called DScreen. The game observes a person's behavior to determine if they have dyslexia.Since the majority of them go undiagnosed, children with learning difficulties generally struggle academically and have trouble learning compared to their peers. Even though dyslexia is unrelated to overall intellect, children with this disorder will face psychological issues.Our approach to identifying dyslexia involves using machine learning techniques and game data. The game content was designed based on an analysis of the difficulties experienced by individuals with dyslexia. Our research demonstrates that dyslexia may be evaluated using a machine learning strategy.**

Keywords*—Dyslexia, Random Forest, Screening, Games, Machine learning

I. Introduction

Nearly 2 billion people worldwide or 37.5% of the world's population have disabilities, according to the World Report on Disability, which was released by the World Health Organization (WHO) and the World Bank. According to the World Health Organization, a person with a disability is anyone who has a problem with their body's structure or function, a restriction on their activities, or difficulties carrying out a task or action. Dyslexia is one of the disabilities, a matter that has been questioned for decades and is still up for debate today.

The International Dyslexia Association (IDA) Board of Directors defines dyslexia as: a particular learning problem with neurological roots. It is characterized by issues with precise and/or fluid word recognition, as well as by weak decoding and spelling skills. The challenges faced by people with dyslexia are usually a result of difficulties with language sounds and patterns, which can be surprising considering their other cognitive abilities and academic performance in the classroom.

Individuals with dyslexia can improve their situation by adopting strategies to manage their condition. However, without adequate support, people with dyslexia often struggle in academic settings. Dyslexia is considered a "hidden disability" because it is less noticeable in languages with straightforward writing systems. As a result, students often only become aware of their dyslexia after experiencing difficulties in school, which can be too late for effective intervention.

The current approach to diagnosing and screening for dyslexia is based on an in-person test that is carried out by trained experts. This test is designed to evaluate various aspects of a person's reading and writing performance, such as their speed in reading words, ability to read pseudowords, fluency in reading, number of writing errors, number of reading errors and comprehension of text.This method is very time-consuming and requires a lot of resources, including trained professionals and specialized equipment. It also requires the person being tested to be physically present, making it difficult to reach people in remote or under-resourced areas. This can lead to many people with dyslexia going undiagnosed and struggling in school, which is why alternative methods of early screening, such as the DScreen game, are being developed.

Past studies have looked at various symptoms of dyslexia aside from difficulties with reading and writing, such as visual perception, memory, executive functions and auditory perception. This presents new opportunities for identifying dyslexia in children. Our research builds on these findings by integrating machine learning with data collected from an online game that includes both language-independent and language-specific elements. The game is user-friendly and can serve as a means of informing parents about the possibility of dyslexia, encouraging them to seek help from professionals like doctors or therapists. To predict dyslexia, we utilize basic machine learning classifiers, such as the Random Forest from the Scikit-Learn toolkit.

This study's findings may help dyslexic students improve their academic performance. Furthermore, it has the potential to advance the field of game design by providing novel, imaginative methods for modeling the severity of dyslexia in learning games.

Learning disorders that affect school-age children frequently go undetected due to a lack of understanding. Furthermore, current screening techniques rely on paper-based exams, which are difficult and time-consuming to administer.

The goal of this project is to determine if machine learning can be used to screen for dyslexia using input data collected from interactions with gamified questions in an online test. This method would make the screening process easier to administer, as it eliminates the need for in-person assessments by trained professionals.

II. Literature Review

In [1], This research report proposes that dyslexia is a phonological processing issue. The best indications of children with dyslexia were found to be trials of oddity and rise time. Quiz and survey are responsible for the creation of the dataset. Language vocab, speed, memory, visual segregation, audio discrimination are parameters. 70% of the

dataset is for model preparation and 30% is used to test the models. The dataset is trained with different machine learning algorithms. Precision and accuracy for each algorithm is obtained. SVM Grid Search CV was discovered to have the highest degree of accuracy, or 97.24

In [2], Dytective is a tool for identifying dyslexia. This game employs linguistic and attentional exercises to highlight the distinctions between those who have dyslexia and those who do not. Varying reading and writing patterns were discovered in experiments involving 243 kids and adults (95 of whom had dyslexia that had been diagnosed). In a held-out test set with 100 individuals, a machine learning model that was developed was able to detect dyslexia with 83% accuracy. In the polynomial Support Vector Machine (SVM) setup, the binary classifier LIBSVM is utilized.

In [3], A study has proposed a new approach for detecting dyslexia through EEG using a non-linear SVM machine learning model. The model uses various attributes to understand brain activity and relies on data from fewer electrodes to reduce complexity. The results showed that the Gaussian (RBF) kernel was the most effective, with a maximum accuracy of 62.4%. To improve accuracy in the future, the study suggests considering other features and using other computational techniques. Although optimizing the channel grouping can make the process more efficient, it may also decrease accuracy. Early diagnosis of dyslexia is important and enhancing this research approach could have a positive impact on society.

In [4], the research demonstrated a newly developed educational game that analyzes game play to determine the degree of dyslexia in Arabic learners. The game's validity was established by comparing its identification results to those of a conventional identification approach. The goal is to achieve phonological awareness at the second level (confusion between letters and sounds). When compared to the conventional method, the game has a high accuracy rate and a reasonable level of agreement, according to the results and programs.

In [5], the study uses a mobile game to identify kids with particular learning problems and offer treatments. These particular learning disorders are screened by using deep learning and machine learning techniques. Outputs from the convolutional neural network are fed into the models used for screening learning disabilities. k-nearest neighbors, random forest and Support Vector Machine the machine learning algorithms used in building models. The screening findings from the models developed in this study have accuracy levels for dyslexia, letter dysgraphia, dyscalculia and numeric dysgraphia of 89%, 90%, 92% and 92%, respectively. The application's results were highly accurate when tested on pre-diagnosed learning disabled kids enrolled in special education classes at Senehasa Education Resource Research and Information Centre (SERRIC).

In [6], machine learning models and data from a language-independent web game were used to screen for dyslexia. This game is designed in German and Spanish to examine dyslexic people's errors in multiple languages as well as other dyslexic factors such as auditory and visual perception. The researchers carried out a user study with 313 children (116 of whom were dyslexic) and used the data gathered to train predictive machine learning models. For German and Spanish, the Random Forests and Extra Trees algorithms are used. A web-based game that uses language-neutral content and machine learning to screen for the likelihood of dyslexia.

In [7], This article presents a method of detecting dyslexia in languages with simple writing systems, like Spanish, using machine learning and data collected from 15 minute online test that uses game elements. However, it is important to note that other factors such as IQ and specific dyslexia conditions must also be evaluated by professionals and therefore the results of this method should only be used as a preliminary screening test. This method of dyslexia screening online is simple to use because no other instruments are required. The preliminary findings of the screener have been published and it has already been used more than 200,000 times as an open access internet tool in Spanish-speaking nations. This strategy has the potential to have a huge social impact because estimates of dyslexia are substantially greater than the population that has been diagnosed with it. Similar techniques might enable earlier dyslexia detection and prevent children from getting a dyslexia diagnosis after experiencing academic failure.

In [8], the study aimed to predict dyslexia by using machine learning algorithms on a data set of 313 participants. The algorithms used were Extra Trees, Random Forest and Gradient Boost. The study found that the best accuracy for German was 74% using Random Forest and for Spanish was 69% using Extra Trees Classifier. The main advantage of the language-independent content approach is that it has the potential to screen pre-readers, which could help to detect dyslexia at an early stage. However, the study acknowledges that this will require more personnel to screen more children and collect more data. The aim is to improve the results by collecting more data from younger children, using different inputs related to dyslexia characteristics and different game designs.

In [9], the study only utilized camera software to assess the reading speed and eye movement (saccade amplitude) of young people. The eye tracking was implemented using Python and OpenCV relying on convolutional neural networks, machine learning and computer vision. The software tracks the user's eye movements, including the trajectory, coordinates and time interval, based on the movement of their pupil. The analysis of the data collected determines the reader's reading fluency, including their reading mode, speed and weight, to assess their risk of having a reading disability. The results are then categorized as high, medium high, medium low, or low. This eye tracking-based dyslexia detection system provides a new and innovative approach to dyslexia identification, reducing the impact of subjective and unstable variables in current testing methods. This can lead to early recognition and intervention at home and in schools, as well as earlier screening.

In [10], suggested method, dyslexia in the brain region is detected and segmented using a machine learning approach. The results are examined for the presence of tumors using evaluation criteria including sensitivity, specificity and accuracy. Clients can explore and build models in an electronic information science environment, find and distribute informational indexes and collaborate with other AI researchers and AI designers using Kaggle. The usefulness of the suggested algorithm is demonstrated by implementing the existing and proposed techniques through simulation results utilizing programming tools like MATLAB and comparing the performance of the proposed approach with the existing method.

III. Methodology

The proposed methodology aims to identify dyslexia in children aged 6 to 12 years old by an Android game and machine learning techniques. The game will be designed to assess cognitive and linguistic abilities that are known to be affected in dyslexia, such as reading and language

processing. The performance data from the game will be collected and combined with demographic information and any available information about the presence or absence of dyslexia. The collected data will then be preprocessed and used to train a Random Forest machine learning algorithm, which will then predict whether a child has dyslexia. The goal of this methodology is to provide a non-invasive and fun method for identifying dyslexia in children.

A. Participants

Children ages 6 to 12 make up the participants. They are supposed to be students. When playing the games, the individual needs to be accompanied by an instructor, teacher, guardian or parent.

B. Procedure

The proposed methodology for identifying dyslexic individuals playing an android game consists of several steps. Firstly, a large dataset of dyslexic individuals playing the game would be collected and preprocessed to remove any missing or irrelevant data. The preprocessed data would then be used to extract relevant features such as average reaction time, accuracy and others which would be used as inputs for the machine learning model. The model of choice would be Random Forest and it would be trained using cross-validation techniques to prevent overfitting. The performance of the Random Forest model would then be evaluated using metrics such as accuracy, precision, recall and F1 score. If the model's performance is not satisfactory, the model would be retrained and improved until a satisfactory level of performance is achieved. Finally, the Random Forest model would be deployed on the android game as an API to identify individuals with dyslexia. The API would analyze the player's performance in real-time and provide a report indicating whether the player is dyslexic or not.

Continuous improvement would be an essential part of this methodology. The Random Forest model would be continuously monitored and improved over time by updating the dataset and adjusting the parameters of the model as needed. Additionally, any feedback received from users of the API would be considered and the model would be adjusted accordingly to improve its accuracy and effectiveness. Overall, this methodology provides a comprehensive approach to identify dyslexic individuals playing an android game. By continuously improving the model, the accuracy of the identification process can be enhanced, which can help to identify dyslexic individuals early on and provide them with the necessary support and assistance.

This android game aims to detect dyslexia through interactive and entertaining games that are designed to identify specific symptoms related to the learning disorder. The game includes a range of linguistic and non-linguistic games including word matching, visual memory, word association and letter recognition. The games are designed to collect data on the player's performance. This data can then be analyzed to make predictions of dyslexia and provide educational resources for further evaluation. The tests used in the game are as follows:

1. Letter Match: A memory-style game that tests players' ability to remember letters and sequences of letters. 2. Reading Race: A reading-speed game that tests players' ability to quickly identify words and read them accurately.
3. Visual Match: A visual-matching game that tests players' ability to match images to words (Fig. 1). 4. Sound Match: A sound matching game that tests players' ability to match sounds to words.
5. Shape Match: A shape-matching game that tests players' ability to match shapes to correct shape. 6. Picture Match: Difficulty recognizing and associating words with pictures.
7. Identify direction: Show an arrow and ask if it is right or left or up or down.
8. Word Scramble: Players are given a scrambled word and must unscramble it to form the original word. 9. Memory Match: Match pairs of objects.
10. Letter Replacement: A game where players must replace a set of letters in a word to make a new word (Fig. 2).

Fig. 1. Visual Match

Fig. 2. Letter Replacement

Each exercise included in the program targets three or more of the dyslexia-related indicators listed in Table 1, which encompass various language skills, working memory and perceptual processes.

TABLE I. GAMES AND COGNITIVE INDICATORS USED IN TEST DEVELOPMENT

Game	**Symptom screened**	**Cognitive indicators used in test development**
Letter Match	Working memory deficits	1-Alphabetic Awareness 2-visual
Reading Race	Decoding deficits	1-phonological awareness 2-auditory
Visual Match	Visual processing deficits	2-visual

Shape Match	Difficulty with visual discrimination	2-visual
Picture Match	Semantic processing deficits	1-Alphabetic Awareness,phonol ogical awareness 2-visual
Identify direction	difficulties with spatial orientation.	2-visual 3-Visual Discrimination and Categorization
Word Scramble	Phonological manipulation difficulties.	1-Orthographic Awareness
Memory Match	Difficulties in working memory	2-visual,auditory
Letter Replacement	Alphabetic Awareness	1-Alphabetic Awareness

1=language skills,2=working memory,3=Perceptual Processes

C. Algorithm

The Random Forest algorithm is a popular machine learning method that is used in this methodology to identify dyslexia in children. This algorithm is well-suited for this problem due to its ability to handle high-dimensional data and multiple features. The Random Forest model is trained on the collected and preprocessed data to predict whether a child has dyslexia based on their performance in the Android game. The model considers various factors, such as the child's demographic information and their performance on different cognitive and linguistic tasks within the game.

Random Forest is a powerful machine learning algorithm that can be used to make predictions and classifications. In the case of dyslexia prediction, Random Forest can be used to analyze the performance metrics of players playing an android game and determine if they have dyslexia or not. The algorithm works by creating a large number of decision trees and combining the results to make a final prediction. Each decision tree in the forest makes a prediction based on a random subset of the features and training data. The final prediction is made by combining the results from all the trees in the forest, typically through a majority vote. Random Forest is a robust algorithm and can handle a large number of features, including both continuous and categorical features. It is also less prone to overfitting than other algorithms, which makes it a good choice for this application. Once the Random Forest model is trained and evaluated, it can be deployed as an API in the android game to identify dyslexia in real-time. The model can also be continuously improved over time by updating the dataset and adjusting the parameters of the model as needed.

Furthermore, the Random Forest algorithm is able to capture complex non-linear relationships between the features and the target variable, making it well-suited for this application. Additionally, Random Forest is able to handle class imbalances, which can be a problem in datasets where one class is much more prevalent than another. This is important because dyslexia is a relatively rare condition and it is essential to avoid over-representing one class in the training data. Another advantage of Random Forest is that it provides feature importance information, which can be useful in understanding what factors are driving the predictions. This information can be used to improve the accuracy of the model by focusing on the most important features and ignoring less relevant ones. Moreover, Random Forest is computationally efficient, which makes it well-suited for deployment on mobile devices. The algorithm can run quickly even on low-end hardware, making it a good choice for use in an android game environment.

In conclusion, Random Forest is an excellent algorithm for dyslexia prediction in an android game environment. The algorithm's ability to handle large datasets, handle a variety of features and produce accurate predictions, combined with its computational efficiency, makes it a good choice for this application.

IV. Conclusion

This android game aims to detect dyslexia through interactive and entertaining games that are designed to identify specific symptoms related to the learning disorder. The game includes a range of linguistic and non-linguistic games, including word matching, visual memory, word association and letter recognition. The games are designed to collect data on the player's performance. This data can then be analyzed to make predictions of dyslexia and provide educational resources for further evaluation.

The proposed methodology provides a new and innovative approach to the detection of dyslexia in children aged 6 to 12 years old. By utilizing an Android game and the Random Forest algorithm, the methodology aims to assess the cognitive and linguistic abilities affected by dyslexia in a non-invasive and fun way. The game tasks and challenges are specifically designed to target the skills typically impacted by dyslexia such as reading and language processing. The use of machine learning techniques enables the system to learn from a large dataset and make accurate predictions about dyslexia, making it a powerful tool for early detection and intervention. Moreover, the methodology highlights the potential of technology and machine learning to tackle important societal problems and opens new avenues for further research and development in the field. By utilizing a fun and engaging platform, the methodology has the potential to reach a wider audience and provide valuable insights into the abilities of children with dyslexia, ultimately leading to a better understanding of the condition and improved support for those affected. In conclusion, the

proposed methodology offers a unique and promising solution for the detection of dyslexia in children and has the potential to make a significant impact in the lives of children with dyslexia and their families.

REFERENCES

[1] Pralhad, Ghadekar Premanand, et al. "Dyslexia Prediction Using Machine Learning." *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*, 2021, https://doi.org/10.1109/aimv53313.2021.9671004. .

[2] Rello, L. et al. (2016) "Dytective: Diagnosing risk of dyslexia with a game," *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare* [Preprint]. Available at: https://doi.org/10.4108/eai.16-5-2016.2263338.

[3] Parmar, S.K., Ramwala, O.A. and Paunwala, C.N. (2021) "Performance evaluation of SVM with non-linear kernels for EEG-based dyslexia detection," *2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)* [Preprint]. Available at: https://doi.org/10.1109/r10-htc53172.2021.9641696.

[4] Tlili, A. et al. (2020) *"*Unobtrusive monitoring of learners' game interactions to identify their dyslexia level*," 2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)* [Preprint]. Available at: https://doi.org/10.1109/icalt49669.2020.00040.

[5] Kariyawasam, R. et al. (2019) "A gamified approach for screening and intervention of dyslexia, dysgraphia and dyscalculia," *2019 International Conference on Advancements in Computing (ICAC)* [Preprint]. Available at: https://doi.org/10.1109/icac49085.2019.9103336. International Conference on Advancements in Computing (ICAC), 2019

[6] Rauschenberger, M., Baeza-Yates, R. and Rello, L. (2020) "Screening risk of dyslexia through a web-game using language-independent content and machine learning,"*Proceedings of the 17th International Web for All Conference* [Preprint]. Available at: https://doi.org/10.1145/3371300.3383342.

[7] Rello, L. et al. (2020) "Predicting risk of dyslexia with an online gamified test," *PLOS ONE*, 15(12). Available at: https://doi.org/10.1371/journal.pone.0241687. .

[8] Rauschenberger, M., Baeza-Yates, R. and Rello, L. (2022) "A universal screening tool for dyslexia by a web-game and machine learning," *Frontiers in Computer Science*, 3. Available at: https://doi.org/10.3389/fcomp.2021.628634.

[9] Zhang, Y. and Zhou, W. (2022) "Research on dyslexia detection based on eye tracking," 2022 *IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)* [Preprint]. Available at: https://doi.org/10.1109/icetci55101.2022.9832243.

[10] S, S., R, U. and R, V. (2022) "Dyslexia Biomarker Detection Using Machine Learning," 2022 *International Conference on Communication, Computing and Internet of Things (IC3IoT)* [Preprint]. Available at:

https://doi.org/10.1109/ic3iot53935.2022.9767973.