IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115112

PREDICTION OF DYSLEXIA BASED ON EYE TRACKING

Pranav Pawar[1], Anisha Deochake[2], Bhagyoday Patil[3], Nachiket Mali[4], Dr. Snehal Kamlapur[5]

Student, Computer Engineering, K.K.W.I.E.E.R, Nashik, India[1-4]

Professor, Computer Engineering, K.K.W.I.E.E.R, Nashik, India[5]

Abstract: Dyslexia is a neurodevelopmental reading disability. Dyslexia is thought to be a standard learning disability characterized by a persistent deficit in rapid word recognition and by spelling. Detection of dyslexia is critical, expensive. Early diagnosis of dyslexia is incredibly important. Eye tracking may be a useful approach to detect dyslexic and non-dyslexic people. Eye tracking doesn't rely on the person's verbal response so it provides a natural means to objectively assess the reading process. Although dyslexia may be a learning disorder, eye movements in reading can predict individual reading ability. Eye tracking movements are going to be fed to the model for classification and will predict whether the person is dyslexic or not.

Keywords: Classification, Dyslexia, Eye Tracking

I. INTRODUCTION

Although the reasons for dyslexia are as yet not completely comprehended, and definitions and terminology fluctuate, it is for the most part concurred that kids who neglect to procure perusing expertise at a typical rate need cautious checking and backing during the early long stretches of school. Early identification and proficient help is the best type of mediation for kids with articulated understanding troubles, and it is perilous to delay until kids are officially determined to have dyslexia prior to helping their requirements. Thus we intend to create a system which predicts high risk and low risk of having dyslexia using machine learning and predictive modeling. Dyslexia is a neurodevelopmental perusing inability assessed to influence 5 to 10\% of the population[1]. While there is yet no full comprehension of the reason for dyslexia, or settlement on its exact definition, it is sure that numerous people suffer persistent problems in learning to read for no apparent reason. It is conceivable to recognize 9 to 10 year elderly people in danger of determined perusing troubles by utilizing eye tracking during perusing to test the cycles that underlie understanding capacity[1]. In spite of the fact that dyslexia is in a general sense a language-based learning inability, the outcomes recommend that eye movements in reading can be profoundly prescient of individual understanding capacity and that eye following can be a productive method to recognize youngsters at risk of long-term reading difficulties.

II. LITERATURE SURVEY

In [1] that utilizes eye-tracking while reading to examine the mechanisms that underpin reading ability, we can identify 9–10-year-old persons at risk of chronic reading challenges. Unlike other screening methods that rely on spoken or written responses from the patient, eye tracking requires no overt verbal response from the patient and thus provides a natural way to objectively assess the reading process as it unfolds in real time. The research used a sample of 97 high-risk individuals who had previously reported word decoding difficulty, as well as a control group of 88 low-risk individuals. These topics were picked from a larger sample of 2165 second-graders. We use predictive modeling and statistical re-sampling methodologies to develop classification models utilizing eye-tracking records less than one minute in length, and show that the models can accurately distinguish high-risk participants from low-risk participants. Despite the fact that dyslexia is primarily a language-based learning disorder, the findings suggest that reading eye movements are highly predictive of individual reading ability, and that eye tracking could be a useful tool for identifying children who are at risk of long-term reading difficulties.

In [2] to make it more realistic, come up with a unique technique to display the user. Selecting the dyslexia stage, loading the treatments, and detecting gestures and speech are all topics covered in this study. The end results are majorly governed by a few technologies namely artificial intelligence, virtual reality, image processing and voice recognition. Anyone with dyslexia or who wants to improve their writing, reading, or speaking abilities can use "The Hope" with the help of a parent or guardian and experience major improvements by following the program's daily therapy. To incentivize and encourage users to use the programme on a daily basis, the application includes an appreciation and reward system. Following the therapies on a daily basis will undoubtedly aid the user's abilities to improve. Users can use this tool to swiftly identify their weaknesses in writing, reading, and speaking, and it will help them overcome them in a creative and proper way.

In [6] The study focuses on an E-learning system built with Moodle. Using dedicated courses built on the basis of several elements of an LD student, this method recognises two learner profiles: students with Learning Disability (LD) and students without Learning Disability (Non-LD). This investigation was done in stages to capture the learning parameters for dyslexic kids using an informal assessment approach. Based on the amount of factors employed, there are two techniques to data gathering in the initial step. Machine Learning is utilised to determine whether or not the user has LD (Dyslexia in this case). The findings are provided for both approaches, and a comparison of the datasets generated in the final strategy for capturing parameters involving Natural Language Processing (NLP) demonstrates that the dataset obtained in the final approach is better and more robust. When it comes to doing detection based on the created dataset, the LR algorithm for machine learning outperforms SVM.

In [7] Electrooculography (EOG) signals are frequently employed in biomedical applications such as Human Computer Interaction (HCI), and their usage in the diagnosis of neurodegenerative illnesses is growing. This study presents the use of a one-dimensional convolutional neural network (1D CNN) based on EOG data as a novel technique for dyslexia identification. In the first step of the investigation, EOG signals were captured while healthy and dyslexic youngsters read four different texts. In the second stage, the EOG signals were filtered and divided into frames. In the first step of the investigation, EOG signals were captured while healthy and dyslexic youngsters read four different texts. In the second stage, the EOG signals were filtered and divided into frames. In the final stage, the EOG signals were classified using 1D CNN. According to the results, the healthy and dyslexic groups were classified with a classification accuracy of 73.6128 ± 2.8155 % .

In [8] Augmented reality is a visual technology which combines virtual objects into the real environment, in real time Augmented reality is a visual technology that seamlessly integrates virtual elements into the real world.In this study, a heuristic model of multimedia learning using augmented reality would be developed for a type of neurological illness known as dyslexia. Dyslexia is a complex mental-brain illness that impacts children in a number of ways, including verbal and nonverbal communication, social connections, instruction understanding, reading, writing, learning, and other concerns. The use of an interactive augmented reality based multimedia application to facilitate and provide pedagogy for such extraordinary children would provide a new and different dimension of treating and supporting such young souls in overcoming their challenges in a pleasant and straightforward manner. The study's purpose is to develop a cognitive learning framework for an interactive multimedia learning software that uses augmented reality technology that is suited to autistic children's needs and allows them to interact with the system.

In [9] There is proof that dyslexic people have aberrant eye movements while reading. Machine Learning (ML) classifiers were used in a few recent research to predict dyslexia using eye movement data.These studies' eye movement data sets are limited to reading saccades and fixations, which are compounded by reading challenges; for example, it's unclear if abnormalities are the result or cause of reading difficulty. Ward and Kapoula recently used LED targets (together with the REMOBI AIDEAL approach) to show irregularities of huge saccades and vergence eye movements in depth, demonstrating that dyslexia has intrinsic eye movement problems that are not induced by reading. Another study looked at binocular eye movements while reading two different texts: one with the meaningless "Alouette" text another with a written text. It was revealed that the Alouette text causes dyslexics to have abnormal eye movements. In this article, we assess the accuracy of eye tracking descriptors for dyslexia identification.Using the descriptors generated in the four distinct configurations as input, we compare the generalization performances of numerous classifiers. Our findings show that eye tracking data from the Alouette test positively predicted dyslexia with just an accuracy of 81.25 percent; similarly, data from saccades to LED targets just on Remobi device positively predicted dyslexia with just an accuracy of 81.25 percent, and data from vergence movements to LED targets with an accuracy of 77.3 percent. Eye movement data with meaningful text, for example, had the lowest error (70.2 percent). Following that, utilizing eye tracking descriptors retrieved from meaningful reading, ML algorithms were applied to estimate reading speed, following by Remobi saccade & vergence testing. The meaningful reading test's vergence in Remobi Eye tracking descriptors can predict reading speed much better.

In [10] Although Dyslexia is not an oculomotor disorder, readers with dyslexia have showed distinct eye movements during text reading than typically developing students. In comparison to normal readers, dyslexic readers produce frequent and longer fixations, have lower saccade length, and have more backward fixations. Dyslexic readers are also reported to have trouble with reading long words, skip fewer smaller words, and also have a longer gaze duration on several words. RADAR is a one-of-a-kind, automated, quick, and accurate tool for detecting kids at significant risk in dyslexia that can be applied widely. This work introduces Rapid Evaluation of Difficulties & Abnormalities in Reading (RADAR), a unique, rapid, objective, noninvasive tool for screening for traits linked with the abnormal visual scan of reading text shown in dyslexia. During silent text reading, eye tracking parameter measurements are obtained that are stable under retest and also have high discriminative power, as evidenced by the ROC (receiver operating characteristic) curves. These factors were combined to create a portion score (TRS) that really can reliably distinguish dyslexic readers from non-dyslexic readers. TRS was investigated in a sample of school-aged youngsters aged 8.5 to 12.5 years. To reduce head

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115112

movements, participants were instructed to place their head on a chin rest. A total reading score is based on certain criteria and is used to make the RADAR screening decision. Given the recorded eye-tracking parameter vector, TRS is a multivariate Gaussian function that estimates the probability that reader is non-dyslexic. Under circular validation, the TRS has a specificity and sensitivity of 93.8 % and 94.6 percent, respectively. RADAR correctly detected 30 out of 32 reading comprehension (sensitivity 93.8 percent) in a circular validation circumstance where the individual being evaluated was not part of the test construction group. Finally, examining eye movement characteristics recorded with RADAR while reading will almost certainly be valuable for developing individualized treatment regimens and objectively measuring the success of various interventions. The long-term goal is to be able to use RADAR as an accurate, objective, and quantitative first-pass screening tool for people with reading impairments like dyslexia, which are characterized by faulty oculomotor reading processes.

In [11] Reading eye-movement models that are currently available are difficult and require a significant number of hand-crafted elements. To address these issues, this study proposes a reading eye-movement fixated sequence labeling strategy that improves on existing reading eye-movement models' fixation fineness processing mode and regression processing mode. The suggested model is based on the multi deep learning rnn that employs deep learning to cut down on the number of hand-crafted features necessary while also incorporating cognitive psychology data to boost accuracy. The act of marking the eye's fixation just on words is viewed as the read eye movement process, and the challenging modeling effort of read eye movements is simplified to a simpler multi label problem.The results of the basic LSTMCRF model are utilized as a benchmark, and the theory is improv affect reading eye movement. In addition, this study provides a data enrichment method for detection by incorporating linguistic parameters including such part-of-speech & word length that ng eye-movement to match the data size requirements of deep-learning models. Using this method to add data from eye-movement tissues considerably improves the model's accuracy. Furthermore, a contrast of the experimental data to current reading eye-movement models reveals that the proposed model can achieve equal accuracy with fewer hand-crafted features.

III. METHODOLOGY

Eye descriptors of dyslexic and non-dyslexic subjects recorded while reading are taken as input. Missing values are dropped in order to remove redundancy and noise. Features containing numeric and categorical values are considered. Furthermore, categorical data is converted into numeric form using label encoding technique. These features are used as input for training a classification model to discriminate between dyslexic and non-dyslexic subjects. The test data will be given to the system for testing purpose. At last, the binary classifier will classify dyslexic and non-dyslexic subjects.

Dataset Description

There are three datasets:

- Pseudoword and Meaningful Passage Reading Dataset[12] contains 46 dyslexic adolescents (18 female, 28 male; mean age 15.52, SD 2.45) and 41 nondyslexic adolescents (20 female, 21 male; mean age 14.78 +/ 2.44). Dyslexic adolescents were diagnosed in specialized medical centers and were admitted to their schools based on their dyslexia diagnosis. Typically diagnosis by these centers include multiple testing which confers a diagnosis of visual, phonologic dyslexia, mixed dyslexia, etc. We therefore did not have children self-identify their condition. Both dyslexic and non-dyslexic adolescents had no known neurologic or psychiatric abnormalities.

Typical readers had no history of reading difficulty, no visual impairment, or difficulty with near vision. Of dyslexics, 34.0% (16/47) identified that their primary issue was visual/reading based, 4.3% (2/47) auditory, 2.1% (1/47) writing, and 59.6 (28/47) were mixed or unknown. Unfortunately, our small sample size limited further analysis by category.

- S1 Dataset[10] contains 32 dyslexic (15 girls, 17 boys) and 37 non-dyslexic (22 girls, 15 boys).

- Provo Corpus-Eyetracking Dataset[13] contains 10 participants. The Provo corpus is an English-language eye-movement corpus published in 2018. Reading materials include 55 short essays, covering online news, popular science magazines, and popular novels. The eye-movement data used in the experiments in this paper came from 84 of the corpus participants

This dataset consists of various language features like Participant_ID, Word_Unique_ ID,Text_ ID, Word_Number,Sentence_Number,Word_In_Sentence_Number,Word_Length,Total_Response_Count,Unique_Count,ModalResponseCount,POS_CLAWS,,Word_POS,POSMatch,POSMatchModel,,LSA_Response_Match_Score

The main eye movement features include Interest area (IA),saccades, fixations, left, right, top, bottom coordinates, regressions, delay time, fixation start and end time. fixation duration, saccade amplitude and saccade angle.

The algorithm for developing the system follows the following:

- Data cleaning: Cleaning the data by replacing the null missing values with mean or dropping the values wherever necessary.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022
DOI: 10.17148/IJARCCE.2022.115112

- Data transforming: Performing Label Encoding on the data so the data is ready for training in numeric form.
- Train Test split: Splitting the processed data into training and testing datasets with several test sizes(for e.g. 0.2,0.3 resp.).
- Training the data on classification models: Using several classification models such as SVM, Logistic Regression, Random Forest and Naive Bayes.
- Testing the models on the test data: The trained model is then tested on the test data and the values are predicted by the model.
- Comparing various performance metrics: The predicted values are compared to the actual values and several performance metrics are evaluated.

IV. ARCHITECTURAL DESIGN(BLOCK DIAGRAM).

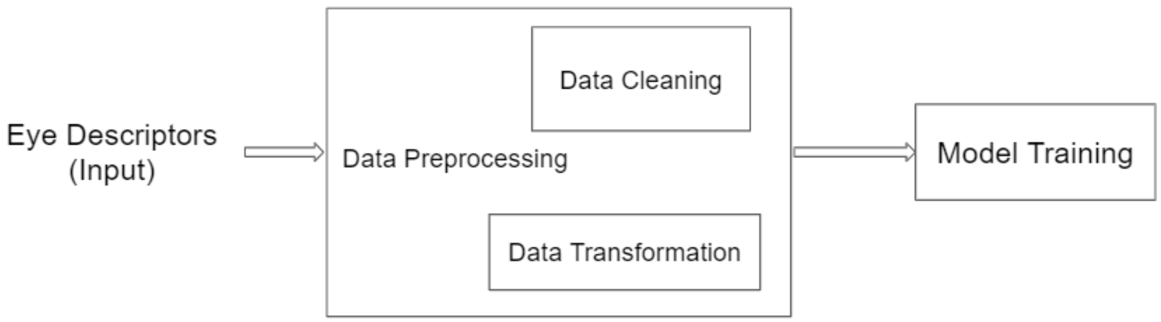


FIG. 1 OFFLINE PROCESSING

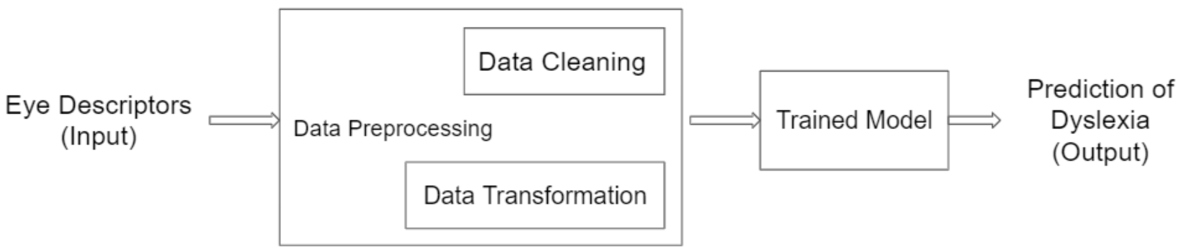


FIG. 2 ONLINE PROCESSING

- Input :

Eye movement features recorded during reading and non reading tasks are used as input for the system.

- Data Cleaning :

Some values were missing for non dyslexic people. Those missing values are either dropped or replaced with the population mean.

- Data Transformation:

All the text data is converted into numeric form so that it can fit into the model while training phase.

- Model Training :

Model is trained using some binary classification algorithms such as Logistic Regression, SVM, K-Nearest Neighbours. Among those classifiers, the classifier with highest accuracy will be selected.

- Trained Model :

The trained model is used to predict whether the person is dyslexic or not.

V. PROJECT MODULES

Data Cleaning: Data cleansing or data cleaning is the process of detecting and correcting corrupt records from a record set, table, or database and refers to identifying incorrect, inaccurate or irrelevant parts of the info then replacing,

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 5, May 2022
DOI: 10.17148/IJARCCE.2022.115112

modifying, or deleting the dirty or coarse data. In the data cleaning process ,the data values which are null are : i)dropped ,because the missing value are vague and can't get replaced with the other value because it will make the info inconsistent(Pseudoword and Meaningful Passage Reading Dataset[12]) ii)replaced with population mean of their respective columns to form the info consistent (S1 Dataset[10]) iii)dropped and further performed label encoding on text data (Provo Corpus-Eye Tracking Dataset[13]:).

Data Preprocessing: Data preprocessing can seek advice from manipulation or dropping of data before it's utilized in order to make sure or enhance performance, and is an important step within the data processing process. We performed data transformation in Provo Corpus-Eye Tracking Dataset[13] via label encoding which is converting the text data to a numeric form in order that it's code and may be further used for training the model.

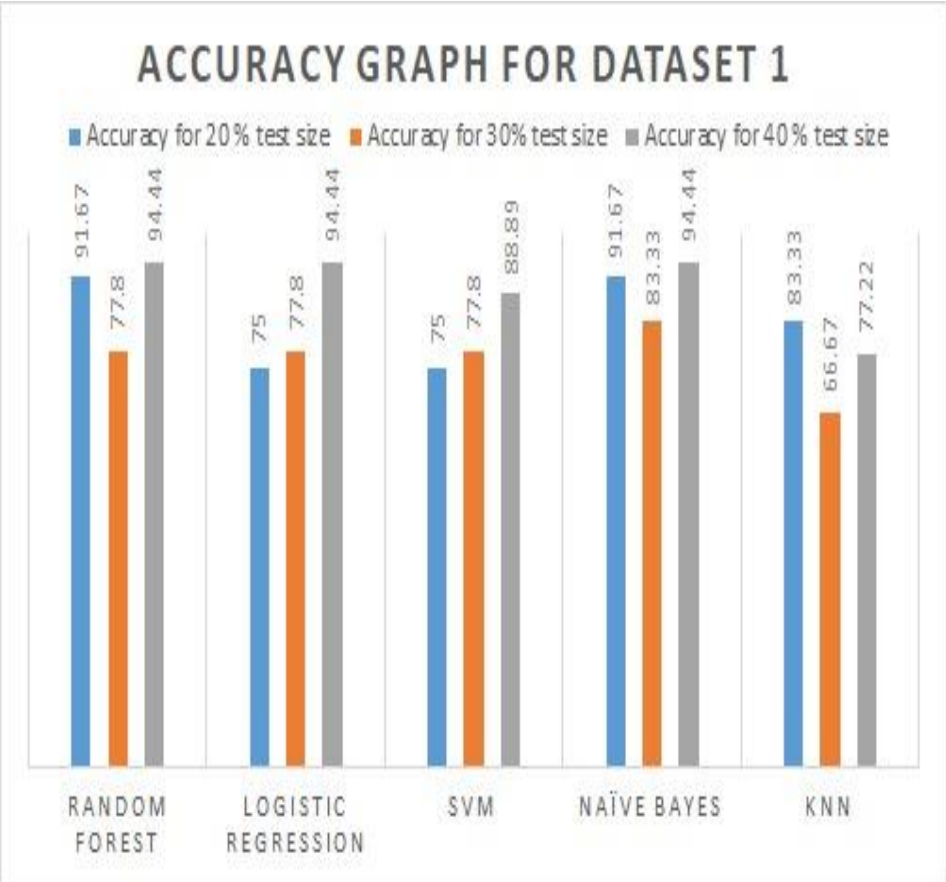
Build Classification model: we've built the classification model using binary classification algorithms like SVM, KNN, Naive Bayes and Random Forest. Training dataset is given as input to the classification model so as to train the model. After this process, the model goes under testing phase.

Evaluate Classification model: to judge the model, we've used performance parameters.We have calculated the classification accuracy, precision and recall for every algorithm.

VI. RESULT

The results of the system are discussed in this section. We trained our model by five binary classification algorithms. The accuracy, precision and recall of each model has been calculated for 20\%, 30\% and 40\% test size respectively. The results of each dataset are as follows:

- Pseudoword and Meaningful Passage Reading Dataset contains 46 dyslexic adolescents (18 female, 28 male; mean age 15.52, SD 2.45) and 41 non-dyslexic adolescents (20 female, 21 male; mean age 14.78 +/ 2.44).



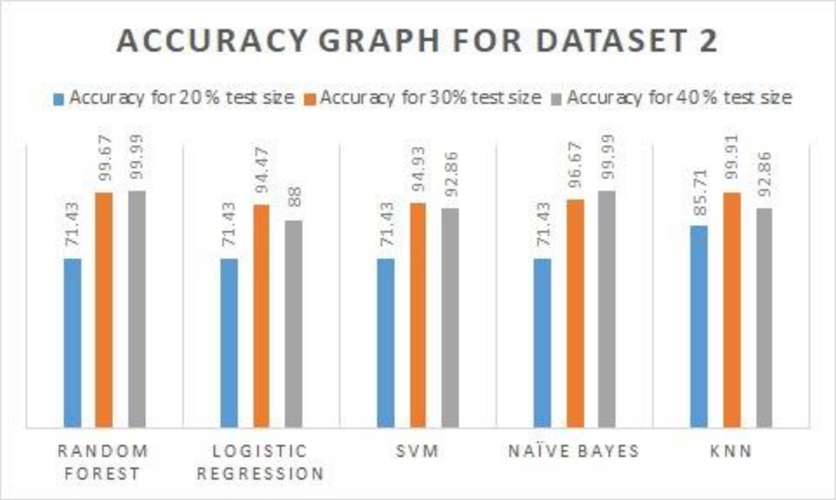
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

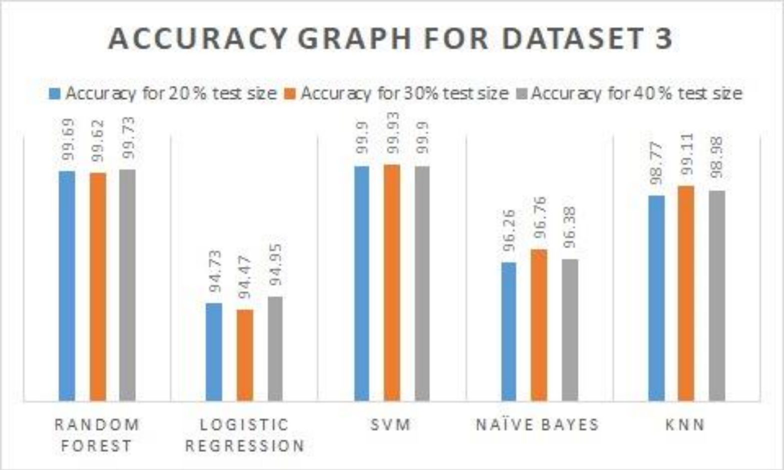
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115112

- The S1 Dataset contains 32 dyslexic (15 girls, 17 boys) and 37 non-dyslexic (22 girls, 15 boys).
-



- Provo Corpus-Eye Tracking Dataset contains 10 participants.
-



VII. CONCLUSION

Children with dyslexia often spend a few years struggling at school before receiving appropriate professional support. Efficient screening methods which will be easily deployed at school settings are important instruments to counter this circumstance. Here, it's investigated the employment of eye-tracking during reading as a screening method and demonstrated that it can produce individual-level predictions with high sensitivity and specificity in but a second of tracking time.Rather than existing screening tests which depend upon paper-and-pencil conventions, this method requires no composed or verbal reaction and negative manual evaluation or reviewing in the customary sense. The main reaction estimated is that the eye development signal and that itself is level headed; it's neither right nor wrong as indicated by some predefined models. additionally, it appears to be plausible that a screening test dependent on eye-tracking may decrease the measure of pressure that more customary test techniques force, since subjects may well be certain to encounter that they're occupied with an undertaking without anyone else instead of unequivocally playing out an errand for another person. Eye-tracking is helpful for screening dyslexia, note that the methodology isn't driven by the presumption that dyslexia is caused by an inherent deficiency in visual insight or oculomotor control. The assumption made is only the benefit or difficulty with which words are processed by the language the system has an essentially immediate influence on eye movements during reading. Finally, it's important to fret that not all children who experience persistent difficulties in learning to read fit the identical neuropsychological profile. it's well established, for example, that there's considerable symptom overlap and a high rate of comorbidity between dyslexia, attention-deficit hyperactivity disorder (ADHD) and language impairment[3-4]. Moreover, it's additionally normal to acknowledge distinctive subforms of dyslexia. Hence, symptomatic development of a positive screening result is consistently important to assemble a more far-reaching comprehension of a person's intellectual profile, so intercession methodologies will be tuned to individual

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022
DOI: 10.17148/IJARCCE.2022.115112

necessities. Nevertheless, early identification of Screening for Dyslexia Using Eye Tracking individuals in need of support is the first important step during this process.

VIII. FUTURE WORK

The main objective of the paper was to predict whether the person is dyslexic or not via reading only. This system can further be improved by more accurate diagnosis of the disease via reading, writing and speaking. Availability of the datasets of eye movement tracking is indispensable for detection of dyslexia, so making health care related datasets available for researchers should be given importance. Also more focus should be given to improve the performance and accuracy of the classification model. Another scope of the paper can be taking live data from the subject while he/she is reading using EOG signals and the data can be fed to the model to give accurate results in real-time.

In healthcare systems, the diagnosis of dyslexia is done manually. It requires manual efforts to make prediction of dyslexia. This study proposes a new approach based on machine learning techniques that can predict whether a person is dyslexic or non dyslexic with high accuracy. The system will help to reduce human efforts and errors.

REFERENCES

1. Nilsson Benfatto M, O¨qvist Seimyr G,Ygge J, Pansell T, Rydberg A, Jacobson C (2016)Screening for Dyslexia Using Eye Tracking duringReading. PLoS *ONE 11(12): e0165508.doi:10.1371/journal.pone.0165508
2. Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46: 389–422.
3. Jacobson C (1998) Reading Development and Reading Disability. Analyses of Eye Movements and Word Recognition. Dissertation, Lund University.
4. Fouganthine A (2012) Dyslexia from Childhood into Adulthood: A Developmental Perspective on Reading and Writing Disabilities. Dissertation, Stockholm University.
5. Samantha Thelijjagoda, Mithila Chandrasiri, Dilan Hewathudalla, Pasindu Ranasinghe, Isuru Wickramanayake The Hope: An Interactive Mobile Solution to Overcome the Writing, Reading and Speaking Weaknesses of Dyslexia The 14th International Conference on Computer Science & Education (ICCSE 2019) August 19-21, 2019. Toronto, Canada
6. Prof. Masooda Modak,Omkar Warade,G.Saiprasad,Shweta Shekhar Machine Learning based Learning Disability Detection using LMS 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA) Galgotias University, Greater Noida, UP, India. Oct 30-31, 2020
7. Ramis Ileri, Fatma Latifoglu, Esra Demirci New Method to Diagnosis of Dyslexia Using 1D-CNN New Method to Diagnosis of Dyslexia Using 1D-CNN
8. Zeeshan Bhatti Maymoona Bibi Naila Shabbir Augmented Reality based Multimedia Learning for Dyslexic Children 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)2020
9. Alae Eddine El Hmimdi Lindsey M Ward , Themis Palpanas and Zo¨ı Kapoula Citation El Hmimdi, A.E.; Ward, L.M.; Palpanas, T.; Kapoula, Z. predicting Dyslexia and Reading Speed in Adolescents from Eye Movements in Reading and Non-Reading Tasks: A Machine Learning Approach Brain Sci. 2021
10. Ioannis Smyrnakis , Vassilios Andreadakis , Vassilios Selimis , Michail Kalaitzakis , Theodora Bachourou , Georgios Kaloutsakis , George D. Kymionis, Stelios Smirnakis, Ioannis M. Aslanides ADAR: A novel fast-screening method for reading difficulties with special focus on dyslexia PLOS ONE 2017
11. YING WANG , XIAOMING WANG AND YAOWU WU A Simple Model of Reading Eye Movement Based on Deep Learning IEEE 2020
12. https://dataverse.harvard.edu/api/datasets/export?exporter=html&persistentId=doi%3A10.7910/DVN/3YCB56
13. https://github.com/wxmgo/eye_movement_in_reading/blob/master/Provo_Corpus-Eyetracking_Data(sub1-10).csv

