Detecting Risk of Dyslexia from Eye Tracking Measures Data

Saurabh Bade
Applied Modelling & Quantitative Methods – Big Data Analytics
Trent University
Peterborough, Canada

Dr. Zenab Noorin
Applied Modelling & Quantitative Methods
Trent University
Peterborough, Canada

ABSTRACT – Dyslexia is one of the most common Learning disabilities/disorders. This disorder should be diagnosed in the early stage of schooling. Diagnosing dyslexia is harder and needs a specialist to detect it. It is proven by many researchers that there is a difference in eye movement of the dyslexic and non-dyslexic person while reading the text. [1] ***In this experiment, we have used the data of the eye-tracking measure and tried to can make the detection task easier for the specialist to detect dyslexia with our eye-tracking model. We had 185 subjects containing 97 as dyslexic and 88 as non-dyslexic. From this data, we have created a machine learning model that detects the dyslexic person while they are reading the passage. Our model gives 87% average accuracy with K-nearest neighbors.***

KEYWORDS – Dyslexia, Eye Movement, Machine Learning, Detection, k-nearest neighbors.

1. INTRODUCTION:

Dyslexia is a Greek origin word, which means dysfunctionality in Language or Poor Language. [2]
The name "Dyslexia" was termed by Rudolf Berlin in 1887, however, it was first identified by Oswald Berkhan, who was a German in the year 1881 [3].
At least 5% and at most 17% population are dyslexic, as per the *Encyclopedia of Language Development*. Dyslexia is not easy to diagnose in the later stage, however, can be identified in the early stage of school life. Most of the student who has dyslexia get dropped from school/college as they find reading and writing too difficult.

Countries that are developed and have all facility, to get diagnosed have taken care of this learning disorder. Yet, getting diagnosed with a neurologist is quite expensive too. A per "The Reading Well" organization, for three sessions and assignments it costs 1500 dollars [4]. In third world countries or developing countries like China, Brazil, and India it is very difficult to provide information and detect dyslexia at an early stage. And hence it is necessary to identify such cases of students, teens, and adults who are Dyslexic.

Dyslexia is not a disease and cause of the dyslexia is not yet known, hence it cannot be cured. However, it is been proven that it is genetically based and it runs in the family from generation to generation [5]. This specific Learning Disability(LD) is a neurological disorder and can be found in people who are in-between age 6 to 60 years old. [6]

We know that hurdle racer will take time to make it to the finish line, also, hurdle racer will stumble, fall and again get up and hence the time span increases. Whereas Normal racer will run without any obstacles and will finish the line early. Here, we can compare the Hurdle racer with Dyslexic and Normal racer with the non-dyslexic person. Moreover, there are various people who were dyslexic, yet they achieved the greatest tasks in their fields. Albert Einstein, Walt Disney, Thomas Edison, Isaac Newton, and Pablo Picasso are some of the examples who were dyslexic and did great in their own field. This proved that IQ and EQ are not less of Dyslexic people and hence, they are not dumb. [7]

However, a dyslexic person should know the correct path to achieve his task/goals, for that it is necessary to detect dyslexia in those people as early as possible. If early detection is not done, people who face this issue get frustrated may cause social anxiety and depression.

So the questions come, can we detect dyslexia with machine learning, which gives high accuracy and in a more efficient manner?

2. PREVIOUS STUDIES ON EYE MOVEMENTS AND DYSLEXIA:

Rayner studied and shown that if eye movement is faster while reading and less pause is taken by the reader when he is non-dyslexic [8]. Hyona and Olson added that dyslexic people take a longer pause while reading the and the eye fixation is more for unfamiliar/new words and frequent words [9]. When a familiar word paragraph was given in the readable format, dyslexic people were reading those paragraphs similar to non-dyslexic people which was proven by Pirozzolo and Rayner [10]. And similarly, Non-dyslexic people showed the eye movement same as dyslexic people when they were asked to read the unfamiliar/ non-frequent words [11]. From all these factors, an eye-tracking movement experiment was done on the dyslexic and non-dyslexic people.

For the first time, Luz Rello *et al* from Carnegie Mellon University, Barcelona, Spain came up with the idea to record the people's eye movement while reading and try to detect the risk of dyslexia. They took 1135 recording of people who has dyslexia and who don't have dyslexia. And with this big sample, they were able to achieve 80.18% of accuracy to detect the risk of dyslexia. They created their model with the Support Vector Machine (SVM) classification algorithm with 10 fold cross-validation. [12]

M. N. Benfatto *et al* from Karolinska Medical Institute, Sweden, also done similar kind of the experiment. Where they used SVM and created a dataset from 185 students consist of 97 dyslexic students and 88 non-dyslexic students. The accuracy they achieved is 96%, however, they got this accuracy on 48 features, which they selected from 168 features [13].

On the same dataset, of Benfatto, Isha Puri, a student from Harvard University, New York, USA created an application named "A Scalable and Freely Accessible Machine Learning Based Web Application for the Early Detection of Dyslexia". She achieved an accuracy of 90.18% which is greater by 10% from Luz work's accuracy. Also, her final application was giving an accuracy of 85.03% for the detection of the non-dyslexic person. Isha took the right eye reading and left eye reading and combined it in one feature so that more accuracy and less computational power should be used [14]. Moreover, Isha was awarded the Davidson fellowship of 25000 scholarship and was also awarded the Neuroscience Research Award. [15] [16]

3. DATASET:

Data Collection:

The dataset was created by M. N. Benfatto *et al* from Karolinska Institute, Sweden for the research purpose of detecting dyslexia. This dataset is open for everyone, with some conditions on the Figshare website [17]. In the experiment, they have taken a test which consists of a paragraph of 10 sentences which covered in 8 lines, and an average of 4.6 words per sentence. Then this paragraph was asked to read by the Swedish school children of 2nd grade (age 8-9 years old). In total 185 students were selected randomly (excluding some other students) for the test from which 88 were at low risk of dyslexia(LRD) and 97 were at high risk of dyslexia(HRD). The ratio of male to female in the dataset is 29:8, where 76 males and 21 females were in the HRD group and 69 males with 19 females in the LRG group.

In the experiment "Corneal Reflection System, Ober-2TM" was used for eye tracking with respect to time and record the eye movement and represent it in the cartesian coordinate system with an x-axis and y-axis measures. Every recording was taken after every 20 milliseconds time.

Once the was collected, then each student has his own file, and combining each file our dataset has columns which are given below:

- Time: Time recording for every 20 Milliseconds.
- LX – Coordinate points for Left eye and X-Axis
- LY – Coordinate points for Left eye and Y-Axis
- RX – Coordinate points for Right eye and X-Axis
- RY – Coordinate points for Right eye and Y-Axis
- Gender: Is a categorical variable, where 1 = Male, 2 = Female.
- Class – It is a Gold Standard(categorical var), where 0 = LRD and 1 = HRD.
- Subject – is categorical var, Unique ID of the students.

Data Analysis:

We did data analysis on the raw data set. We plotted the x-axis and y-axis reading on graphs with respect to time, from that we get the output shown in the figure below (Fig 1.a, Fig1.b). this figure shown is for one student with HRD for Fig 1.a and one student with LRD.

From the figure we can say that we get sawtooth waveform for X-axis recording, i.e. as we read from left to right, our eyes move accordingly. And hence when we plot the same with respect to time, the whole sawtooth wave is been created. As there were 8 lines, and hence we will get 8 sawteeth.

When we compare both plots, i.e. Fig 1.a and Fig1.b, then we get that there difference between both of the sawtooth waves. For the student who is non-dyslexic, is taking 14 seconds, to complete the paragraph while reading (Fig 1.b). Whereas, student having dyslexia is taking 40 seconds to complete the paragraph reading (Fig 1.a).

Also, we can see the pattern, where LRD student's eye movement is like "reading a word and taking a pause, and there are not too many sharp spikes in the graph for the x-axis reading", however, when it comes to HRD student's graph, we can see there are more spikes in the sawtooth wave, are less inconsistent, and move too much even for recognizing the word/letter.

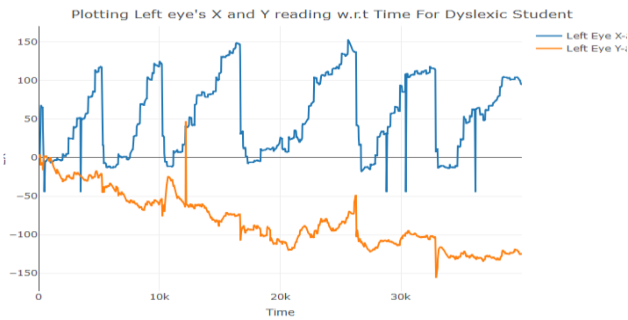

Fig 1.a: Sawtooth waveform for HRD(Dyslexic Student)

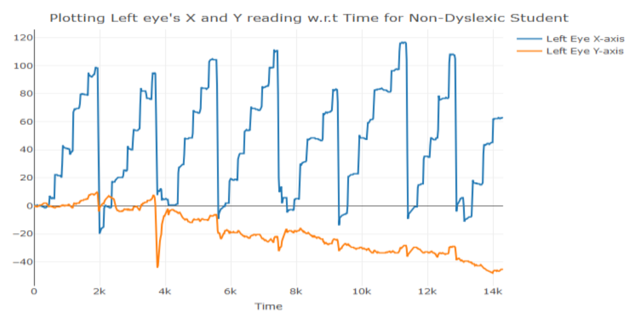

Fig 1.b: Sawtooth waveform for LRD(Non-Dyslexic Student)

Moreover, when we took the mean of the time taken by HRD and LRD to read the paragraph, which was more significant. To be exact, the average time taken to read the paragraph by LRD is 29582.27 ms (~30 sec), whereas the HRD took 39773.81 ms (~40 Sec). 107 students took 40 Sec to read the whole paragraphs, one student took 35 Sec, 56 students took 30 Sec, 10 students took 25 Sec and eleven students took 20 Secs.

Dataset Preprocessing:

In this setup, I have created extra attributes, removed noise, manipulated data and finally took the mean of the data with respect to each student recording. As the researchers have shown that there is high saccade in the dyslexic person when compared to the normal person while reading. And hence we will take them into the account, that how much time does a student took for the gaze of the word/letter [9]. So when there is no difference in the recording for last 20 ms, then we are taking count of that recording and considering as no difference (ND) and mentioned as 1 in the dataset, on the contrary, if there was the difference(DIFF) then we will consider as 0. Hence more four features are being placed for each cartesian coordinate and each eye. When data is plotted with respect to the time, it looks like as shown in Fig 2.

From these, we have counted the number of the continuous ND points so that we can get that how many times the student is gazing on the letter/word for how much time. *e.g.* from Fig 2, we can see that the first cycle shows that it too near about 80ms for the first gaze. And similarly, we will calculate it for the whole recording of each student. Once the eye fixation time was found for each student, we have to create a threshold, where it should be significantly different in LRD and HRD subject.after doing the statistical analysis, we were able to find the threshold from 60 ms, and hence we removed the count of 20 and 40 ms, as 60ms is the threshold, from which there was significant difference.

Moreover, when we checked if the Data is normal or not, then we found that reading of the X-axis and Y-axis is left-skewed. Hence, we tried to remove the outliers, however, all of the outliers cannot be removed, as some of the outliers are providing the whole data for specific students. Also, I have removed the noise, which is in most of the recording of the student's data. As we know that we will get the eight sawteeth waveform, and data after that is nothing but noise, which has to be removed. However, when we remove this noise manually, we automatically add bias in the dataset, and hence the result will be too optimistic. Furthermore, I have to take slice by reducing 2000 milliseconds to check if we can remove the biases. And finally, To avoid it, I reducted the data time frame to 30000 milliseconds. There might be some important data too which we might have removed yet, we are getting better results.

As we have, reading of each eye, and hence as done by Isha Puri [14] on her project, to merge Left eye and right eye, so that we will reduce the number of attributes and double the total number of records. So as we have 185 students, and taking a record of each eye will be a record of 370 eyes.
Finally, we have taken the Average of the reading of each student. By which we get 370 records and one record for each student's one eye. For the categorical

variable, we are using the mode as per the highest number of category occurred.

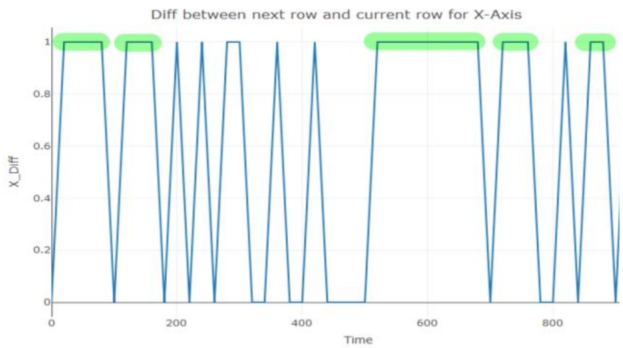

Fig 2: Gaze time plot of X-axis record of an eye while reading.

4. EXPERIMENTS:

In the experiment, I chose five classification algorithms, so that I can choose the best algorithms and do the validation on the best algorithm. The five classification algorithms for detection are:

- Logistic Regression(GLM)
- Linear Discriminant Analysis (LDA)
- K-Nearest Neighbors(KNN)
- Gaussian Naive Bayes(GNB)
- Decision Tree(DT).

We are not using SVM, as Luz and Benfetto both used SVM, and hence we will try here different algorithms.

Part a: I have split the data into Test and Training set in a 1:4 ratio randomly so that we can work on the algorithm, and check for the best algorithm. In training, we have 296 records and in the test set, we have 74 records out of 370 which has been shown in Table 1.a.

Part b: We are not supposed to touch Test set from part a, however, to test on the validation set, we have to split Training into training and validation set. The overall structure of the data distribution for this is shown in table 1.b.

Training Set	Test Set
80%	20%
296 records	74 Records

Table 1.a: Training set, test Set distribution.

Training Set	Validation Set	Test Set
70%	10%	20%
259 Records	37 Records	74 Records

Table 1.b: Training set, validation set, and test Set distribution.

All five Algorithms on Part a:

Here, we have trained the data on the five different classification algorithms, and we check the accuracy on the test data and training data, error on the test and train data, and Area under ROC(AUROC).

	LR	**LDA**	**KNN(k = 5)**	**DT**	**GNB**
Training Accu	85.5	84.8	89.89	100	88.4
Test Accu	84.94	83.87	91.39	90.32	87.0
AUROC	90.2	88.1	95.69	91.5	92.2
Training Error	0.14	0.18	0.101	0.0	0.11
Test Error	0.15	0.17	0.086	0.09	0.12

Table 2: Accuracy, Error, and AUROC of five classifiers.

Here, we will choose the highest accuracy on the test set, the highest value of AUROC and less difference between training and test error. As we are doing here, a binary classification, we are using AUROC for choosing the best Algorithm. As given in table 2, KNN is the best algorithm as it gave 91.398% accuracy and 95.70% is the area under the ROC. However, when we compare test and train error of KNN, with others it is not that good. Yet, I chose KNN, by this accuracy and AUROC, and check if we can get less error and higher accuracy with different values of K.

For the different values of K (*i.e.* 1 to 25), we checked the accuracy, error values for the test set, AUROC, and plot the graph for the same (Fig 3.a, 3.b, 3.c). From all the figures, we can see that for K = 1, we are getting max accuracy, least error, and best AUROC percentage. Followed by K = 2, and then k = 5, 10 which gives the same accuracy.

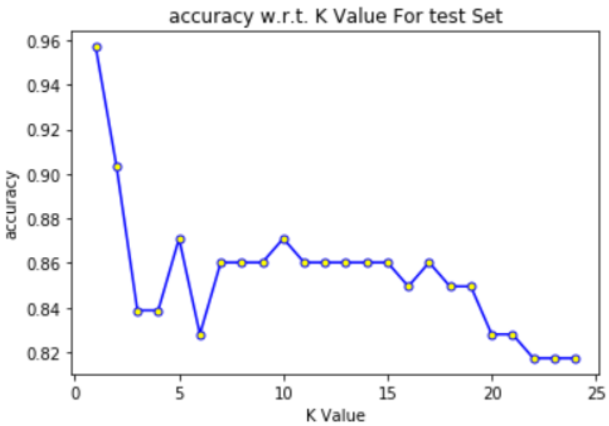

Fig 3.a: Accuracy of KNN for K = 1 to 25

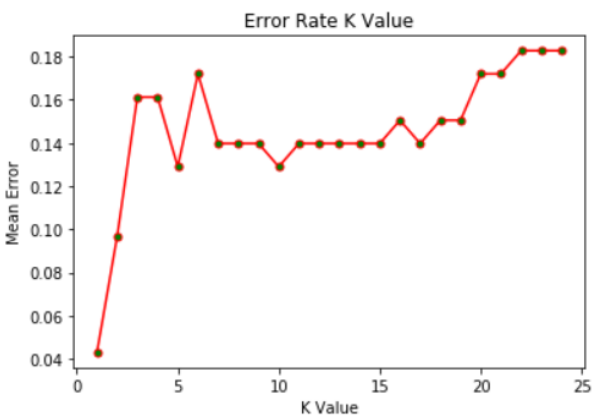

Fig 3.b: Error of KNN for K = 1 to 25

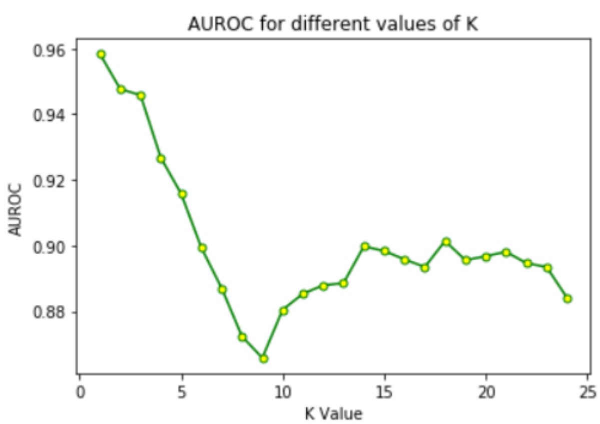

Fig 3.c: AUROC of KNN for K = 1 to 25

Part b:

Moreover, it is being said that K =1, is not a good option for when it comes to classification, because many times it gets complex, because of overfitting. And hence, we will do confirmation by validation set, if K = 1 is good or not. Hence for that, we are using the validation set if the model is correct or not. When we use the validation set on the trained data, we will analyze it with the below figures(Fig 4.a, 4.b, 4.c).

Fig 4.a and Fig4.b are exactly the mirror image of each other. From these figures, we can see that validation error is lowest at K = 3,4,5,14 and 18. Similarly, for the same value of K, validation accuracy is highest. Hence to select the best K-value, we have to look at Area under the ROC curve graph, i.e. Fig 4.c, where we can see that for the validation set AUROC was least for k = 1, and then it grew become highest at k = 18, and again came done. At K = 18, the AUROC was 90%. And hence our KNN model is the best fit model at K = 18.

Fig 4.a: Error rate of test and Validation set for different values of K

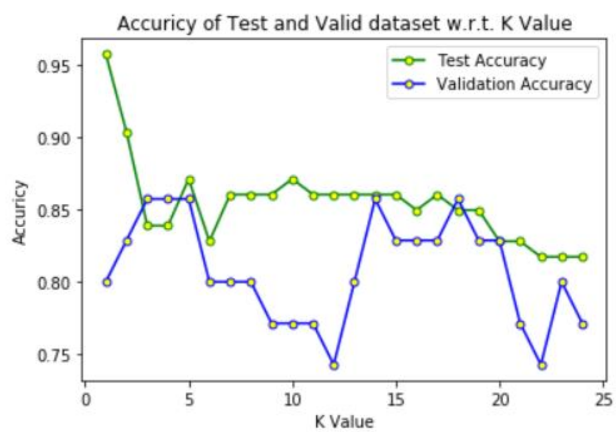

Fig 4.b: Accuracy rate of test and Validation set for different values of K

Once we found the values of K, for the best fit model we recalculated the test accuracy for the KNN model. Accuracy is 84.94%.

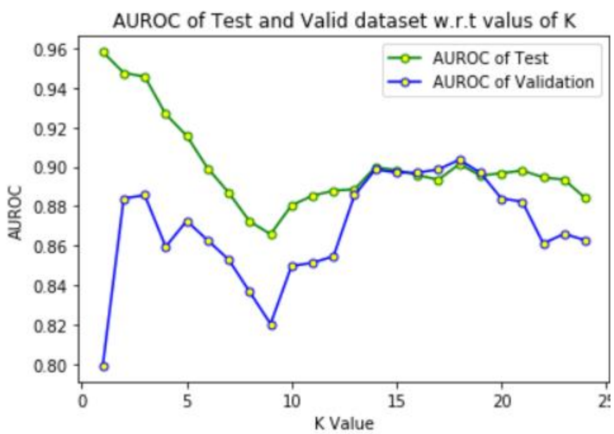

Fig 4.c: AUROC of test and Validation set for different values of K

We were getting different accuracy for each run, and hence we did the 10 fold cross-validation so that we can take the mean of the classifier accuracy. Table 3, shows the accuracy of the data for 10 fold cross-validation. The average accuracy we got like 87% for K = 18, in the KNN model to detecting dyslexia in young children *i.e.* from 74 children, we are detecting 64 children correctly. Moreover, the ratio of test to training error became 1.02, which was 1.17 for k = 5 at first.

Dataset	**Accuracy**
Fold 1	89.47
Fold 2	86.84
Fold 3	84.21
Fold 4	86.84
Fold 5	83.78
Fold 6	91.89
Fold 7	86.11
Fold 8	88.88
Fold 9	88.88
Fold 10	83.33
Average in %	**87**

Table 3: Accuracy on KNN for 10 fold cross-validation.

5. CONCLUSION AND FUTURE WORK:

We can detect the dyslexic person with the KNN model, which gives us an average of 87% accuracy after 10 fold cross-validation. This is not the end of the study here. We can try to add more derived attributes like the direction of the eye, removing the outliers or replacing it with other values and tuning more feature values with different classification algorithms. Moreover, as we can not detect 10 students correctly, this test should be done under guidance or specialist. Hence, we won't call it detecting dyslexia, but "Detecting Risk of Dyslexia in Children".

We can increase the dataset, add more attributes like age and include the dyslexic of different age groups. We can also train our model on different languages, as the different language has scripts. Some languages are supposed to be read from left to right, while some from right to left and few are to read as from top to bottom. For each language, we have to train a new model which might or might not gives us different results.

Reading Comprehension Impairment (RCI) is a disorder in which a person can read but cannot comprehend. [18] This experiment can be also performed on a person facing RCI. Similar to this study, first we have to collect the data from a few samples of people who face the RCI and train model. Eventually, we can also try to distinguish between from this combined model of detecting RCI and detecting Dyslexia, i.e. if a person is Dyslexic or having RCI or normal reader.

As we know that dyslexic students also face the issue of writing and to spell words [19]. Hence, to make our model more accurate we can create an advance model bu combining reading and spelling/writing tests. In this case, we can ask students to take another test which will be of written and Spelling test. The data from each student can be recorded and by combining the written data and reading data of the student and then create a model.

ACKNOWLEDGMENT:

I would like to thank Dr. Noorian and Dr. Parker for supporting and providing guidance over the course of the project. Also, I would like to thank Mr. Gustaf OÈqvist Seimyr for clarifying the issue related to the dataset, which he has already worked up.

References

[1] F. J. Pirozzolo and K. Rayner, "Dyslexia: The Role of Eye Movements in

Developmental Reading Disabilities," in *Neuropsychology of Eye Movement*.

[2] L. Rello, M. Ballesteros, A. Ali, M. Serra, D. A. Sánchez and J. P. Bigham, "Dytective: Diagnosing Risk of Dyslexia with a Game," in *10th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2016.

[3] H. Selvi and M. S. Saravanan, "Comparative Study of Different Machine Learning Algorithm for Dyslexia Children Download," *Advanced Scientific Research,* 2018.

[4] "The Reading well," [Online]. Available: https://www.dyslexia-reading-well.com/dyslexia-test.html.

[5] A. Wilcke, M. B, G. Schaadt, H. Kirsten, and J. Boltze, "High acceptance of an early dyslexia screening test involving genetic analyses in Germany," *Eur J Hum Genet,* 24 Feb 2016.

[6] E. Saraee, C. MH and P. A, "Identifying trends concerning computer,," in *International Conference on Computers for People with Special Needs*, Las Vegas, 2005.

[7] ""Classification Techniques for Early Detection of Dyslexia Using Computer-Based Screening Test, " World Applied Sciences Journal, 2017".

[8] K. Rayner., "Eye movements in reading and information processing: 20 years of research.," *Psychological Bulletin,* 1998.

[9] J. H. R. and K. Olson., "Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency.," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 1995.

[10] F. J. Pirozzolo and K. Rayner, "The neural control of eye movements in acquired and developmental reading disorders.," in *Studies in Neurolinguistics*, 1978, pp. 1-27.

[11] K. Rayner, "Eye movements and the perceptual span in beginning and skilled readers," *Journal of experimental child psychology,* 1986.

[12] L. Rello and M. Ballesteros, "Detecting readers with dyslexia using machine learning with eye tracking measures," in *the 12th Web for All Conference*, 2015.

[13] M. N. Benfatto, G. Ö. Seimyr, J. Ygge, T. Pansell, A. Rydberg and C. Jacobson, "Screening for Dyslexia Using Eye Tracking during Reading," 9 Dec 2016.

[14] P. Isha, "A Scalable and Freely Accessible Machine Learning Based Web Application for the Early Detection of Dyslexia," Google Science fair, [Online]. Available: https://www.googlesciencefair.com/projects/2018/c8e0c790994dfe9aa3495467eacf084b5bbffdad2a46811f8c81567c7cc2bc25.

[15] "Chappaqua Central School District," [Online]. Available: https://www.ccsd.ws/district/departments/public-information/highlights/2018-2019/january-a/isha-puri.

[16] "2019 Davidson Fellows," Davidson Institute, [Online]. Available: http://www.davidsongifted.org/fellows-scholarship/2019-fellows/isha-puri.

[17] M. N. Benfatto, G. Ö. Seimyr, J. Ygge, T. Pansell and A. R. C. Jacobson, "Screening for Dyslexia Using Eye Tracking During Reading," Figshare, 3 Nov 2016. [Online]. Available: https://figshare.com/collections/Screening_for_Dyslexia_Using_Eye_Tracking_During_Reading/3521379/1.

[18] K. L. Werfel and H. Krimm, "A Preliminary Comparison of Reading Subtypes in a Clinical Sample of Children With Specific Language Impairment," *J Speech Lang Hear Res,* 17 Sept 2017.

[19] M. Hebert, D. M. Kearns, J. B. Hayes, P. Bazis and S. Cooper, "Why Children With Dyslexia Struggle With Writing and How to Help Them," 24 Oct 2018.