International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.13, No.1 (Apr-23)

http://dx.doi.org/10.12785/ijcds/130162

Eye Movement Interpretation for Detecting Dyslexia Using Machine Learning Techniques

S R Mani Sekhar[1], Swathi Chandrashekar[2] and Siddesh G M[3]

[1]*Department of Information Science and Engineering, M.S. Ramaiah Institute of Technology, Bangalore, India*
[2]*Department of Information Science and Engineering, M.S. Ramaiah Institute of Technology, Bangalore, India*
[3]*Department of Computer Science and Engineering(Cyber Security), M.S. Ramaiah Institute of Technology, Bangalore, India*

Received 13 Mar. 2022, Revised 26 Dec. 2022, Accepted 6 Feb. 2023, Published 16 Apr. 2023

Abstract: Dyslexia is a learning disorder. It hampers an individual's ability to comprehend the words, making it difficult for the person to read, spell, and write. The studies exemplify that timely detection and support proves as an advantage in mitigating the negative effects of dyslexia. the traditional manual method takes more time and effort to identify dyslexia, to overcome these issues the presented paper incorporated a machine learning based screening technique to detect dyslexia.

As eye movements possess the ability to provide insights into reading disorders. Understanding the patterns that eye movements make while reading a text paragraph can help distinguish between dyslexic and non-dyslexic readers. The raw data consist of right-eye and left-eye movement positions along the x-axis and y-axis of 185 students were captured while reading a text paragraph. These eye movements were captured using an eye tracker based on the principle of human-computer interaction. The features such as fixation, saccadic movements were extracted for better prediction, later the classification was performed using XGBoost, support vector machine (SVM) and random forest (RF).

The results show that XGBoost provides an accuracy of 95%, SVM 94% and RF 91%. To further validate the machine learning models author has used the performs measured called confusion matrix, precision, recall and F1 scores. The obtained results shows that the SVM achieved an F1 score of 94%, Recall of 94.5% and precision of 96%, whereas RF achieved an F1 score of 90%, Recall of 92% and precision of 89%. Finally, XGBoost achieved an F1 score of 95%, Recall of 95.5% and precision of 95%. The results imply that XGBoost achieves better result compare to other models.
Keywords: Dyslexia, Reading Disorder, Eye Movement Interpretations, Saccades and Fixations, Support Vector Machine Classification, Random Forest Classification, XGBoost

1. INTRODUCTION

Individuals suffering from dyslexia find it challenging to perform lingual processing that results in perplexity in reading [1]. Our brains are fractionated into two fractions. The left hemisphere is responsible for analytical thoughts, logical reasoning, facts and science, and language [2]. And right hemisphere deals with spatial activities. It is the creative section of the brain that indulges with art, music, intuition, and holistic thoughts. FRMI research indicates that brains of individuals with dyslexia depend majorly on the right hemisphere, frontal lobe than the brains of the individuals without dyslexia. This indicates that when people with dyslexia indulge in reading, it takes a longer path through their brain and is possibly deferred in the frontal lobe[3] . Such neurological flaw makes it very challenging for them to read fluently. Hence, their expertise in reading and comprehending text is considered to be very low [4].

Tracking eye movements possess an ability to provide in-depth knowledge about a cognitive task such as reading [5]. Although there are several different reading techniques, there is an underlying principle eye movement pattern and processing that all the eye movements follow while reading. And they include fixations and saccades majorly [6]. Saccades are the sporadic eye movements while reading one text to the other. It is a continuous rapid eye movement. Fixation, on the other hand, are the longer pause or longer gaze on a particular area of the text. Use of an eye tracker while reading has the ability to produce a humongous quantity of data in a very short span of time considering fixations and saccades as feature points.

Individuals with dyslexia produce a different eye

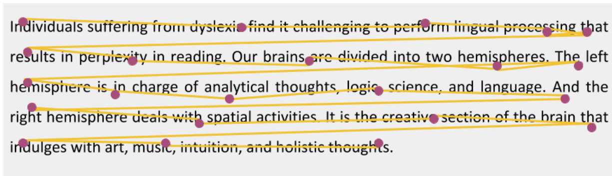

Figure 1. Eye movement pattern made by a non-dyslexic while reading

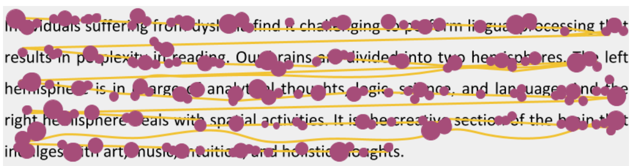

Figure 2. Eye movement pattern made by a dyslexic while reading

movement pattern during reading than the ones without dyslexia [6]. They are categorized by longer fixations occurring more frequently with minimum saccadic lengths with highly sporadic eye movements[7] [8]. Figure 1 and Figure 2 illustrate the eye movement patterns made by a non-dyslexic and dyslexic individual while reading a text. paragraph respectively. The yellow lines indicate the saccade movements and maroon circles indicate the fixations. The bigger the circle, the longer the fixation [9]. Figure 1 shows Eye movement pattern made by a non-dyslexic while reading. demonstrates that a non-dyslexic individual reads a text paragraph with a consistent path with minimal fixations. While Figure 2 shows Eye movement pattern made by a dyslexic while reading. It demonstrates that a dyslexic individual finds it highly difficult to follow a consistent path, and has a lot of fixations on multiple text area, eventually lowering the fluency and increasing the time taken to read the same text [10]. Due to the limited awareness of the learning disorder, ignorance and overlooking of the symptoms at school and home is very common. And the clinical evaluation for dyslexia is not only time consuming but also expensive which becomes a highlighted reason to ignore the disorder at an early stage [11]. And any delay in detection and confirmation of dyslexia in children becomes a major cause of depression and low self-esteem [12].

Hence, the main objective of this paper is to train a model that uses data obtained from a basic eye-tracking mechanism, to interpret eye movements, to understand the difference in patterns that the eye movements make while reading, to predict whether a child is suffering from dyslexia or not. This provides a real-time confirmation of the disorder using simple eye movement interpretation techniques, which is cost-effective and accurate.

2. LITERATURE REVIEW

In [13] an SVM classification model was developed by Rello and Ballesteros that used eye movements to detect dyslexia. They performed ten flod cross validation on 1135 person data with the help of eye tracker. It provided an accuracy of 81%.

In [14] taken 2165 primary school kids data for the detection of dyslexia. Their data consist of recording of less than 1 minute which was input to SVM- RFE classifier. In [15] author has used convolutional neural network for the detection of dyslexia on image dataset. One hundred fifty data where collected from the children between the age group of 3 to 8 years.

In [16] author has captured forty-four Childers data to analysis the dyslexia patients. They used SVM, K-NN models for the prediction of dyslexic. The results shows the accuracy of 95%.

These researches and studies conducted in literature review ensure that there exist several properties that can be extracted from eye movements and several computation methods can be applied to obtain better prediction values. The combination of features extracted and algorithms applied, produces a huge variation in performance result. Hence, machine learning models were trained to draw a conclusion regarding the better performance of the models.

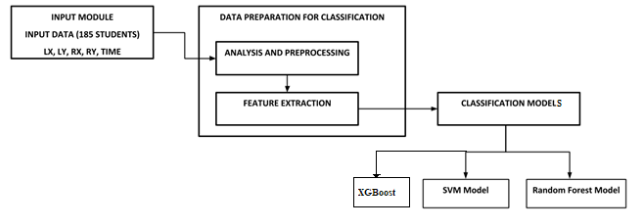

Figure 3. Proposed model for the identification of dyslxeia using learning methods.

3. METHODOLOGY

The objective of the paper is to build a predictive model for dyslexia which is a reading disorder using eye movements as the raw data. Figure 3 shows the proposed implementation model. Here the raw Eye-tracking data is obtained through figshare. It has a recording of eye movements of 185 students, consisting of both high-risk and low-risk group. The eye movements are recorded over a period of time both horizontally and vertically of both the eyes [17]. It is a supervised data set indicating whether the student belongs to a high-risk group or a low-risk group. The data is processed, and missing values are replaced with zero. Considering null values or missing values represent blinks, the raw is preprocessed and analyzed. Analyzing the raw data helps in identifying the essential features required to train the model for classification. Data is analyzed over time using statistical means. The average eye movements are graphically visualized with respect to time for both low-risk and high-risk students to understand the difference in eye patterns. Based on the analysis, the essential features are extracted. Features such as fixations and saccades are the primary distinguishing information that was drawn from the raw eye-tracking data. Features are extracted using a dynamic dispersion velocity-based algorithm. The extracted features are used to train the classification model [18]. We used three classification models for comparative performance analysis.

4. PROPOSED SOLUTION

Dyslexia hampers an individual's ability to comprehend words, making it difficult for the person to read, spell, and write. The proposed work uses SVM, RF and XGBoost methods for detecting of dyslexia. Author has used raw eye-tracking data obtained from figshare, which contained eye movements of 185 students.

A. Data Collection and Data Pre-Processing

An infrared eye-tracking goggle known as Ober-2 was used to track the eye movements of the participants. The eye positions were tracked over a duration in milliseconds. A total of 185 students' eye movements was recorded. Each participant was required to read a text paragraph and their eye motions were tracked and recorded while reading. Each eye movement of every participant was recorded along the x-axis and y-axis [19].

When the participants were asked to read a passage to capture the data during the event, horizontal eye movements of the left eye(LX) was followed by the horizontal eye movements of the right eye(RX). And similarly, the vertical eye movement of the left eye(LY) was followed by the vertical eye movements of the right eye(RY)[14] [20]. And to evaluate the comprehending and understanding skills of the participants, after the reading, a few questions were asked at the end. Figure 4 demonstrates that the left eye movements were followed by the right eye movements both horizontally and vertically.

It is also essential to consider null values in this experiment as when participants blink, the focus of the eye is out of the screen boundary and the horizontal values and vertical values of the eyes will be assigned as null. And these null or missing values are replaced by zero and must be considered during feature extraction [21]. Out of 185 participants, 97 participants were integrated into the high-risk group and 88 participants to the low-risk group. HR and LR are abbreviations given for high-risk and low-risk groups respectively.

The students whose performance was lower than average, faced difficulty in reading, decoding and comprehending letters and words were grouped under the HR group by their professors. And students whose performance was above average were placed in the LR group. However, the IQ of students or participants was not taken into consideration while grouping [22].

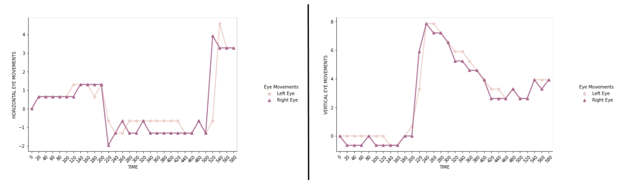

Figure 4. Horizontal and Vertical Eye Movements captured during reading

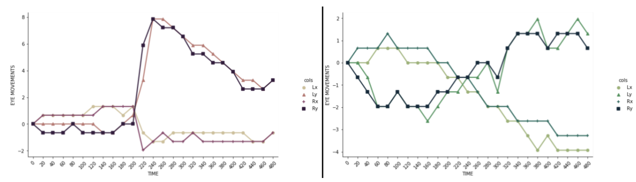

Figure 5. Raw data visualization of an LR Student(left) HR Student (right)

B. Data Analysis

As eye movements form an intrinsic part of any reading process, the link between eye movement patterns and reading is considered logical. Tracking eye movements while reading has the ability to provide sufficient knowledge about reading disorders such as dyslexia [23]. And the patterns formed by the eye movements during reading by a non-dyslexic individual is different from the patterns made while reading by a dyslexic individual.

It is essential to analyze the raw data obtained to identify and understand the patterns that the eye movements make while reading [24] [25]. The raw eye movements (Lx, Ly,Rx, Ry) captured with time is visualized using statistical means, to obtain a graphical representation of the raw data to differentiate between LR and HR students.

As seen in the Figure 5, the left eye movements are followed by the right eye movements. The diagrams also indicate horizontal and vertical eye movements. Figure 5 (left) displays the eye movements of an LR student. The reading path is consistent and steady. Figure 5 (right), the visual representation of an HR student indicates several deviations from the normal reading path.

To further simplify the data and extract features, advanced statistical measures were applied for analysis. The average of the left-eye and right-eye movements are calculated for visualization [26]. As seen in Figure 6, an LR student has a consistent steady eye movement pattern, following a particular path while reading from the left to the right. While figure 6(right) of an HR student indicate that the student follows an inconsistent, distracted path while reading for the same duration and same text passage.

When reading a paragraph, the horizontal and vertical positions of eyes follow a particular pattern. The eyes follow a consistent path from left to right while reading

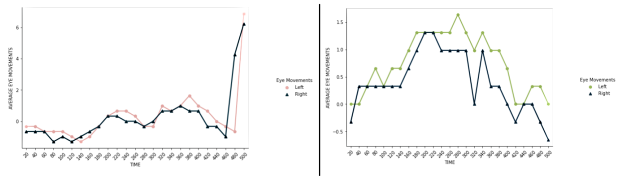

Figure 6. Average Eye movements of an LR Student(left) and HR Student (right) while reading

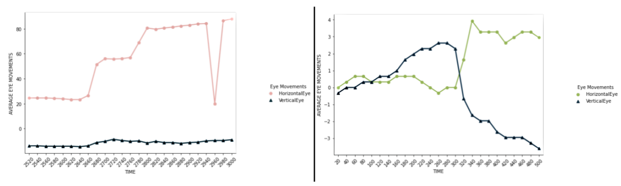

Figure 7. Average Horizontal and Vertical Eye movements while reading of LR Student(left) HR Student (right)

a sentence and then there is a depression or a drop when shifting to the sentence in the next line [27].

The Figure 7, illustrates that when a non-dyslexic individual is reading, the vertical eye movement follows a steady path indicating that the person is concentrating on a consistent path and the dip in the horizontal movement indicates that there is a shift to the next line. While figure 7(right), displays the inconsistent, distracted, horizontal and vertical movements of a dyslexic reader.

As mentioned earlier, students suffering from dyslexia find it difficult to read fluently, and hence consume more time to read the same text passage than a normal reader.

The Figure 8, the graphical visualization of an LR student indicates that there is a dip in the eye movement indicating the advancement to the next line around 2960ms. And in figure 8(right), representing an HR student indicates a dip around 3100ms. The time taken by an LR student to finish a sentence in a line and shift to another line, in this case, is 140ms faster. Based on the above in-depth analysis performed on the raw data, essential features were identified for feature extraction.

C. Feature Extraction

On the grounds of the previous analysis that was performed on raw data, essential eye movement features were extracted. These extracted features were then provided as input to train a classification model that classified whether a student belongs to the high-risk or low-risk group. Features were selected on the principle to preserve a maximum of the primal eye movement signals. Several types of events that could be extracted from eye movements such as saccades, distortions, fixations, and transient were derived using sta-

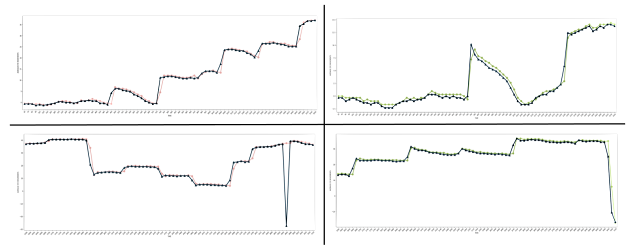

Figure 8. Average Eye Movements highlighting the time taken to go a new line of LR Student(left) HR Student (right)

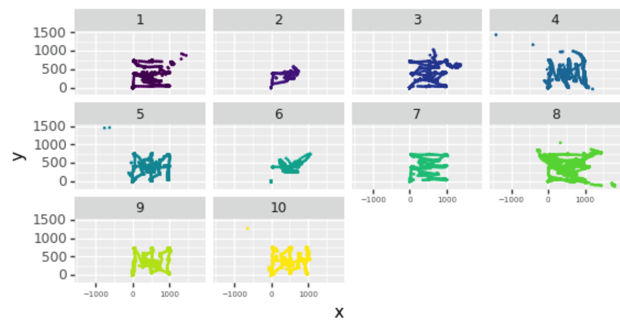

Figure 9. Visual representation of fixations per trail

tistical methods. To identify fixations in raw data, dynamic dispersion threshold algorithm which is a velocity-based algorithm is used. When the velocity of an eye movement goes beyond a certain threshold it is considered as saccades and the area between any two saccades is regarded as a blink or fixation [28]. Figure 9 illustrates the fixations detected for each trail over a period of time.

A fixation with an increased lower level of dispersion than a normal fixation is marked as blinks. So typically, a blink is a fixation on the axis with (0,0) coordinates, with zero dispersion, and followed by a saccade [29]. In figure 10, red coloured points represent eye movements along the a-axis and yellow coloured points represent the eye positions along the y-axis. Each of these points represents the raw data. The horizontal lines indicate fixations and the vertical lines indicate the start and an end of fixation.

Concentrating on fixations and saccades features from the eye movements, a set of parameters were measured that is essential for classification models [30].

- The time duration of the event
- Average positions of the eyes during the event(reading), this consisted of average linear eye movements of both the eyes (Left-eye + Right-eye)/2 and average horizontal and vertical eye movements. The average horizontal eye movements were calculated as the sum of left-eye position along the x-axis and right-eye position along the x-axis [(Lx + Rx) / 2] and the average vertical eye movements were calculated as the sum of left-eye

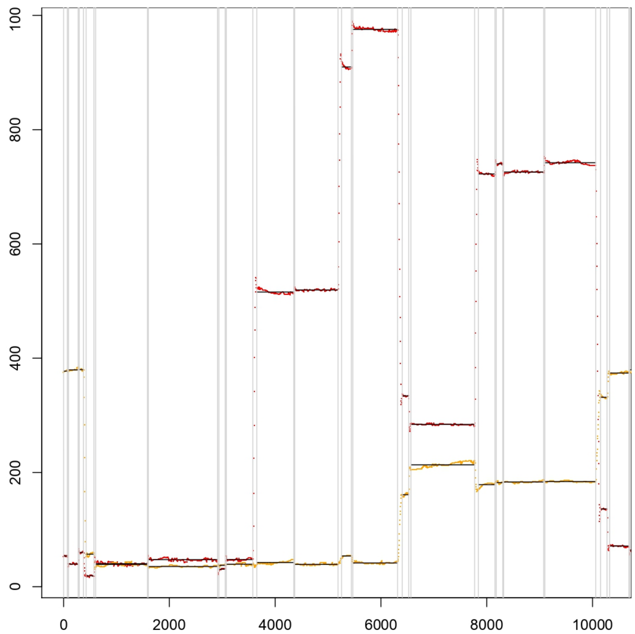

Figure 10. Graphical Visualization of Fixations

position along the y-axis and right-eye position along the y-axis [(Ly + Ry) /2]

- The standard deviation of the average eye positions
- The distance between the eye positions and the maximum value between any two positions [Left-eye - Right-eye].

All the parameters were calculated both horizontally and vertically. The standard deviation and mean were calculated for every parameter. All the parameters were extracted from each participating student.

D. Classification

Support Vector Machine Model: SVM Classification Model is chosen when the quantity of extracted features is high in number than the actual data in the data set [31]. It has the potential to provide high accuracy when the features and kernel is picked suitably. SVM along with the kernel functionality simplifies in analysing and drawing a relationship between the extracted features [32].

This classification model requires the input data to be within the standard range of 0 to 1 or -1 to 1. The eye-tracking data captured during reading was already in the range of -1 to 1 and hence, required no further normalization. This SVM classification model was trained to predict whether a student belongs to an HR Group or an LR Group [33], [34]. The extracted features and raw data were combined and divided randomly into two subsets mainly for training and testing. The classification procedure was repeated several times and the mean of the performance was calculated as the accuracy. 70% of features extracted with respect to fixations and saccades were used for training the SVM method and 30% of testing data was used to make predictions. Further, a confusion matrix was constructed to analysis the performance and accuracy of the model.

Random Forest Model: Random forest classification methodology is chosen as another classification model for this work. Random forest is used to model complex and intricate behaviors and, in this case, eye movements of students to identify and classify them to the risk group they belong to [35].

This classification algorithm develops multiple individual trees. And each of that individual tree is produced by selecting a subset of the features extracted from the raw eye movements data. It follows the principle of divide and conqueror. The Scikit-learn integrates the classifiers by averaging the probability of their prediction. This classification model consists of a combination of randomized decision trees, and the most popular class gets the highest vote. The accuracy and performance analysis are calculated by comparing the level of error obtained during prediction. The output of each individual tree with its testing data is used for measuring the accuracy of the classifier. If it matches, the error variable for that subtree is marked as 0, and if it fails to match then it is set to 1. The mean error value is each subtree is the overall error value of the classifier.

XGBoost Model: XGBoost [36] use the concepts of decision tree and gradient boost techniques in the construction of learner model. It have the capability to handle the missing data by just avoiding it during pre-processing.

E. Evaluation Methodology

The overall implementation and testing phase involved 5 major steps. Collecting the raw data and pre-processing it to extract features, followed by data analysis that helps highlight the essential features to train the classification model. Once the essential features were extracted, the classification model was trained. With the intention of providing comparative analysis of performance and accuracy, three models were trained namely Random Forest, SVM and XGBoost models The three classification models were applied to the extracted features to predict whether a student belongs to the high-risk group or low-risk group of dyslexia. The data set was divided into 70-30 proportions to train and test the model. The final step was to test the trained model.

Predicted Risk Group	Low	High
Actual Risk Group		
High	0	17
Low	15	3

Figure 11. Confusion Matrix to calculate the accuracy Random Forest Classification Model

Predicted Risk Group	Low	High
Actual Risk Group		
High	0.042	0.96
Low	0.94	0.062

Figure 12. Confusion Matrix to calculate the accuracy SVM Classification Model

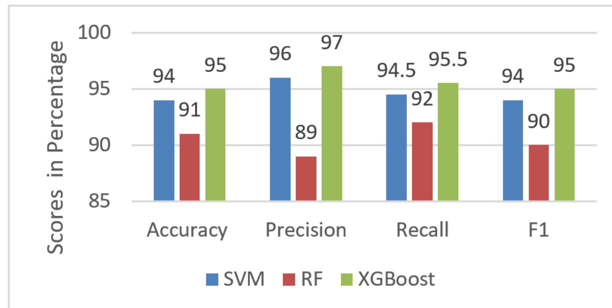

Figure 13. Accuracy, Precision, Recall F1 values in percentage for prediction of Dyslxeia using SVM, RF and XGBoost

To ensure that the implemented model is meeting the expectations, the evaluation is carried out. Evaluation mechanism also helps in improving the efficiency of the model. The evaluation technique used to calculate the efficiency and accuracy of the classification models used in this paper, is a confusion matrix [37]. Confusion matrix, also known as the error matrix, comprises four divisions: True positive, True Negative, False Positive, False Negative. As the name suggests, true-negative and true-positive values provide accurate prediction values. While the false-positive and false-negative provide the error values.

Accuracy defined as the measure of how close the actual output is to the expected output. Accuracy is in direct proportion to the performance of the model. It is calculated by using equation 1.

$$(TruePositives + TrueNegatives)/(TruePositive + FalsePositive + FalseNegative + TrueNegative) \quad ...1$$

5. RESULTS AND DISCUSSIONS

This section discusses the performance and prediction accuracy for dyslxeia using machine learning methods like SVM, RF and XGBoost

A. Detection of dyslxeia using Random Forest

With 97 HR students and 88 LR students, the model was trained with 70% of the total data set and testing was performed on the 30% of the dataset. Hence, the number of observations in the training data was 150 and the number of observations in the testing data was 35. Figure 11 illustrate the confusion matrix for random forest.

The expected and actual output, that is marked as true-positive, for high-risk group is 17, and for the low-risk group is 15, which is marked as true-negative. Error value on the false-positive was 3. Hence, the calculated accuracy is 91%. Later the RF model is validated using statistical measures and the obtained result shows that precision is 89%, recall is 92% and F1 score is 90%.

B. Detection dyslxeia Results using SVM and XGBoost Model

For SVM Classification model, the 70% of the data was used for training and 30% of the data was used for testing. The training size was 129 observations and testing size was 56 observations. The result of accuracy highlights that SVM classification model has a higher accuracy and prediction level than the random forestas shown in figure 12. Using the formula (1), the accuracy for SVM model is 94%.

Hence, Using the confusion matrix evaluation mechanism with accuracy equation 1. Table I and figure 13 summaries the accuracy, precision, recall, F1 scores obtained for XGBoost, SVM and random forest classification models. The result show that XGBoost performs better when compared with other models, it's because XGBoost uses distributed environment with more trees. Here XGBoost achieved an accuracy of 95% whereas SVM and Random forest achieved an accuracy of 94% and 91%.

Later the SVM and XGBoost model is also validated using statistical measures and the obtained result shows that precision is 96% for SVM and for XGBoost is 97%, recall is 94.5% for SVM and 95% for XGBoost. Finally F1 score is 94% for SVM and 95% for XGBoost.

6. CONCLUSION AND FUTURE WORK

The growth of technology provides a path to simply and truncate the time required for the diagnosis of several disorders. In this paper, the author used the simplest and easily available raw data, that is, the eye movements based

TABLE I. Detection of Dyslxeia Accuracy Precision, Recall and F1 Using SVM, RF and XGBoost

Classification Model	Accuracy	Precision	Recall	F1
SVM	94%	96%	94.5%	94 %
RF	91%	89%	92 %	90 %
XGBoost	95%	97 %	95.5 %	95%

on human-computer interaction to propose a methodology for the early detection of dyslexia. The proposed eye movement interpretation system is used to detect dyslexia is a non-invasive and inexpensive system. The model is focus on the necessary features such as fixations and saccades, three models were trained to achieve the highest prediction level. From the above comparative analysis performed, the XGBoost model provides a better accuracy with 95% than the SVM and Random forest model that gives an accuracy of 94% and 91%. Similarly, it performs well with other validation measures like precision, recall and F1 scores. As a future work, multiple features can be extracted and can be applied to identify other disorders such as depression, autism, schizophrenia, ADHD.

7. ACKNOWLEDGMENT

Not applicable

References

[1] P. Ott, *Teaching children with dyslexia: A practical guide*. Routledge, 2007.

[2] E. Arzubi and A. Martin, "Overcoming dyslexia: A new and complete science-based program for reading problems at any level," *American Journal of Psychiatry*, vol. 162, no. 1, pp. 197–197, 2005.

[3] Boets, H. P. Op de Beeck, M. Vandermosten, S. K. Scott, C. R. Gillebert, D. Mantini, J. Bulthé, S. Sunaert, J. Wouters, and P. Ghesquière, "Intact but less accessible phonetic representations in adults with dyslexia," *Science*, vol. 342, no. 6163, pp. 1251–1254, 2013.

[4] J. Stein, "The magnocellular theory of developmental dyslexia," *Dyslexia*, vol. 7, no. 1, pp. 12–36, 2001.

[5] K. Fraser, K. L. Fors, D. Kokkinakis, and A. Nordlud, "An analysis of eye-movements during reading for the detection of mild cognitive impairment," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 1016–1026.

[6] R. Sandak, W. E. Mencl, S. J. Frost, and K. R. Pugh, "The neurobiological basis of skilled and impaired reading: Recent findings and new directions," *Scientific Studies of reading*, vol. 8, no. 3, pp. 273–292, 2004.

[7] M. A. Just and Carpenter, "A theory of reading: from eye fixations to comprehension." *Psychological review*, vol. 87, no. 4, p. 329, 1980.

[8] C. Prado, M. Dubois, and S. Valdois, "The eye movements of dyslexic children during reading and visual search: impact of the visual attention span," *Vision research*, vol. 47, no. 19, pp. 2521–2530, 2007.

[9] F. Hutzler and H. Wimmer, "Eye movements of dyslexic children when reading in a regular orthography," *Brain and language*, vol. 89, no. 1, pp. 235–242, 2004.

[10] A. Asvestopoulou, Psistakis, I. Smyrnakis, V. Andreadakis, and M. Aslanides, Ioannis, "Dyslexml: Screening tool for dyslexia using machine learning," *arXiv preprint arXiv:1903.06274*, 2019.

[11] S. E. Shaywitz, M. D. Escobar, B. A. Shaywitz, J. M. Fletcher, and R. Makuch, "Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability," *New England Journal of Medicine*, vol. 326, no. 3, pp. 145–150, 1992.

[12] R. L. Peterson and B. F. Pennington, "Developmental dyslexia," *The lancet*, vol. 379, no. 9830, pp. 1997–2007, 2012.

[13] L. Rello and M. Ballesteros, "Detecting readers with dyslexia using machine learning with eye tracking measures," in *Proceedings of the 12th International Web for All Conference*, 2015, pp. 1–8.

[14] M. Nilsson Benfatto, G. Öqvist Seimyr, J. Ygge, T. Pansell, A. Rydberg, and C. Jacobson, "Screening for dyslexia using eye tracking during reading," *PloS one*, vol. 11, no. 12, p. e0165508, 2016.

[15] K. Spoon, D. Crandall, and K. Siek, "Towards detecting dyslexia in children's handwriting using neural networks," in *Proceedings of the International Conference on Machine Learning AI for Social Good Workshop, Long Beach, CA, USA*, 2019, pp. 1–5.

[16] S. Kaisar, "Developmental dyslexia detection using machine learning techniques: A survey," *ICT Express*, vol. 6, no. 3, pp. 181–184, 2020.

[17] R. Elterman, L. A. Abel, R. Darof, L. Dell Oso, and J. Bornstein, "Eye movement patterns in dyslexic children," *Journal of learning disabilities*, vol. 13, no. 1, pp. 16–21, 1980.

[18] Q. Valizadegan and M. Hauskrecht, "Learning classification models from multiple experts," *Journal of biomedical informatics*, vol. 46, no. 6, pp. 1125–1135, 2013.

[19] M. Vidal, J. Turner, A. Bulling, and H. Gellersen, "Wearable eye tracking for mental health monitoring," *Computer Communications*, vol. 35, no. 11, pp. 1306–1311, 2012.

[20] H. Collewijn, C. J. Erkelens, and R. M. Steinman, "Binocular co-ordination of human horizontal saccadic eye movements." *The Journal of physiology*, vol. 404, no. 1, pp. 157–182, 1988.

[21] A. Jothi Prabha and R. Bhargavi, "Prediction of dyslexia from eye movements using machine learning," *IETE Journal of Research*, vol. 68, no. 2, pp. 814–823, 2022.

[22] M. Studdert-Kennedy and M. Mody, "Auditory temporal perception deficits in the reading-impaired: A critical review of the evidence," *Psychonomic Bulletin & Review*, vol. 2, no. 4, pp. 508–514, 1995.

[23] K. Rayner, "Eye movements in reading and information processing." *Psychological bulletin*, vol. 85, no. 3, p. 618, 1978.

[24] R. Leigh and D. Zee, "The neurology of eye movements, new york: Oxford univ," 2006.

[25] R. J. Little and D. B. Rubin, "Statistical analysis with missing data. john wiley & sons," *New York*, 2002.

[26] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: data mining, inference, and prediction. springer series in statistics," *Springer New York*, 2009.

[27] O. Zangwill and C. Blakemore, "Dyslexia: Reversal of eye-movements during reading," *Neuropsychologia*, vol. 10, no. 3, pp. 371–373, 1972.

[28] H. Aarts, E. Bijleveld, R. Custers, M. Dogge, M. Deelder, D. Schutter, and N. E. van Haren, "Positive priming and intentional binding: Eye-blink rate predicts reward information effects on the sense of agency," *Social neuroscience*, vol. 7, no. 1, pp. 105–112, 2012.

[29] M. Frens and J. Van der Geest, "Scleral search coils influence saccade dynamics," *Journal of Neurophysiology*, vol. 88, no. 2, pp. 692–698, 2002.

[30] P. Inchingolo and M. Spanio, "On the identification and analysis of saccadic eye movements-a quantitative study of the processing procedures," *IEEE Transt on Biomedical Engg.*, no. 9, pp. 683–695, 1985.

[31] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[32] A. Altaher, "Phishing websites classification using hybrid svm and knn approach," *International Journal of Adv. Computer Science and App*, vol. 8, no. 6, 2017.

[33] G. J. Zelinsky, H. Adeli, Y. Peng, and D. Samaras, "Modelling eye movements in a categorical search task," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 368, no. 1628, p. 20130058, 2013.

[34] Y.-W. Chang and C.-J. Lin, "Feature ranking using linear svm," in *Causation and prediction challenge*. PMLR, 2008, pp. 53–64.

[35] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern recognition*, vol. 44, no. 2, pp. 330–349, 2011.

[36] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd inter-national conf on knowledge discovery and data-mining*, 2016, pp. 785–794.

[37] S. Visa, B. Ramsay, A. L. Ralescu, and E. Van Der Knaap, "Confusion matrix-based feature selection." *MAICS*, vol. 710, no. 1, pp. 120–127, 2011.

S R Mani Sekhar currently working as an Assistant Professor in the Department of ISE, MSRIT, Bangalore. He has published a good number of research article and book. His area of interest is machine learning and bioinformatics.

Swathi Chandrashekar Department of ISE, MSRIT, Bangalore. Her area of interest is Machine learning.

Siddesh G M currently working as an Professor, in the Department of CSE(Cyber Security), MSRIT, Bangalore. He has published more than 50 research article in the peer reviewed journals. and also authored good numbers of book. His area of interest is cloud computing and data analytic