# Dyslexia Prediction Using Machine Learning Algorithms -A Review

**Research** · February 2023

CITATIONS
0

READS
1,315

**1 author:**

Kasthuri Stephen
Bishop Heber College
**4** PUBLICATIONS **1** CITATION

SEE PROFILE

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

# Dyslexia Prediction Using Machine Learning Algorithms – A Review

G. Vanitha[1], M. Kasthuri[2]

*[1]Research Scholar, Bishop Heber College (Autonomous), Tiruchirappalli- 620017*
*[2]Assistant professor, Department of Computer Science, Bishop Heber College (Autonomous), Tiruchirappalli-620017,Affiliated to Bharathidasan University.*

*Email: vanitha.it@bhc.edu.in[1], kasthuri.ca@bhc.edu.in[2]*

***Abstract: Dyslexia is a specific Learning disability that can cause difficulties in Reading, Writing and Spelling. It disturbs the parts of the brain that process linguistic This disease is passed in family lines through genes(hereditary) or through new genetic mutations. There are 6 different types of dyslexia's primary dyslexia, Secondary or Developmental dyslexia, Trauma dyslexia, Visual Dyslexia, Auditory dyslexia and Dysgraphia. Various machine learning algorithms to detect dyslexia they are Random Forest, Decision Tree, Support Vector Machine (SVM), Neural Networks and Bayesian classifiers. Many parameters are used to identify dyslexia eye tracking, fixation and saccadic eye movements and front face detected. This survey paper view at various dimensions of research toward dyslexia. This review finds the research holes, challenges and opportunities in this field. It also encourages to use Machine Learning (ML) algorithms in this research area.***

***Keywords: Dyslexia, Learning difficulty, Machine Learning (ML), Support Vector Machine (SVM).***

## 1. INTRODUCTION

According to the World Report on Disability published by the World Health Organization (WHO) the number of people with disabilities over the globe is almost 2 billion (37.5% of the world's population) [15]. Bestowing to the National Centre for Learning Disabilities (NCLD) dyslexia, dysgraphia and dyscalculia are among the five most common types of learning disabilities around the world [16]. Dyslexia involves the ways that the intelligence processes graphic symbols and the sounds of words. It commonly affects word recognition, spelling, and the skill to match letters to sounds. While it is a neurological condition, dyslexia have no relation to intelligence. Dyslexia are common. Researchers trust that 5%–10% of people have it, while some others estimate that the commonness is 17% [9]. Hereditary factors can play a major role in the cause of dyslexia. Dyslexia is usually diagnosed by conducting oral and written assessments.[16]

Teaching methods and tactics that can help people with dyslexia improve their reading skills and manage the challenges. People with dyslexia have normal cleverness and usually have normal vision. Most children with dyslexia can flourish in school with training or a specific education program. Emotional support also plays

an important role. Early diagnosis, guidance, and care can help and reduce the impact of the condition. Dyslexia is different for everyone. Roughly people have a minor form that they

ultimately learn how to manage. Though there's no medication for dyslexia, early assessment and intervention result in the best outcome.

Dyslexia can be identical from the difference in eye movements of the individual while reading [25]. Eye movements of French dyslexics were tracked while reading text and visual search. It has been observed that dyslexics had more fixation compared to normal readers in both visual tasks and reading [23].

Saccade pattern of dyslexia kids were analyzed while reading Chinese shown the fixations and gaze duration are more for dyslexics compared to normal readers. The fixation landing position was also different for both the group. A statistical model has been built to predict dyslexia from eye tracking movements [21]. An Accuracy of 80.18 percentile was achieved using Support Vector Machine (SVM) binary classifier. Neural Networks were also investigated to identify the gaze pattern in dyslexia and an accuracy of 78 percentile was achieved [18]. SVM gave a high accuracy compared to Neural Networks.This review paper critically replicates on current advancement in dyslexia detection using machine learning approaches and highlights the opportunity for future research.

## 2. MACHINE LEARNING ALGORITHMS

Four stages are used to detect dyslexia they are (i)Data Collection (ii) Pre-Processing, Feature Extraction and Selection (iii) System Testing and Classification and (iv)Performance Evaluation. Fig. 1. Shows the schematic Representation of the stages of dyslexia detection.

Machine Learning is a subset of Artificial Intelligence that learns and identifies new patterns from the past data. Learning algorithms can become more particular and truthful as they working with training data, allowing humans to gain extraordinary understanding into diagnostics, progressions, treatment variability and patient outcomes. There are many machine learning algorithms are available each and every one has its own pros and cons. Choosing a suitable algorithm is a vital role because there are many to select from.

Machine Learning algorithms are mainly classified into 3 categories, they are supervised, unsupervised and reinforcement learning. In supervised learning whose data is already trained and whose class category is already known. In unsupervised learning need not to train the model, it finds all kind of unknown pattern from the data. In reinforcement learning, it learns to performing set of actions and decisions by improving itself. Very frequently used machine leaning algorithms are Decision Tree, Random Forest, Neural Networks, Support Vector Machine (SVM) and Bayesian Classifier.

Decision Tree algorithm is a supervised learning algorithm. The plan of using a Decision Tree is to construct a training model that can be capable of predict the class or value of the target variable by learning simple decision rules inferred from training data.

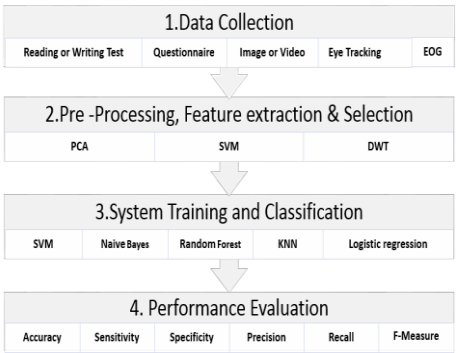

Fig.1. Stages of dyslexia detection using machine Learning Algorithms.

Random forest is one of the best classification algorithms for prediction. It is under supervised learning method, can be used in both classification and regression problems. Random forest consists of a huge number of individual decision trees that operate as a group. It takes the prediction from each tree and based on the majority votes of predictions and it predicts the final result.

Neural Networks imitates the functionality of brain which connecting with n number of neurons, which understands information and comes up with various solutions based on the information. Artificial Neural Network constructed form three layers they are input, hidden and output layer. The input layer has initial data, hidden layer has the highly messed neurons, and the output layer shows the solution to the initial data.

Support Vector Machine is a supervised learning method, which can handle both linear and non-linear classifications. It is a classifier using kernel for pattern analysis, classification, clustering, and ranking raw data. This model is suitable for data which has many parameters.

Naive Bayes algorithm is a classification technique based on Bayes theorem, it is supervised learning method, simple and most effective classification algorithms, this model helps to make quick predictions and also, it's a fast learning Algorithm.

**Related Work**

In this paper we conducted a literature survey to focus on Machine learning methods used in dyslexia prediction. The detailed survey information is shown in Table 1.

## 3. CHALLENGES AND OPPORTUNITIES

Dyslexia disease signs and features can vary based on the diverse languages, so language-based classification need to be improved. The severity of dyslexia can vary from minor to severe. The early detections methods are not analyzed all parts of the brain in image dataset. Dyslexia is not only affected elementary school children but also it will affect middle and high school students, so further research is need to be considered. Many Assistive tools can be designed or helping dyslexics to improve their reading and writing skills [17, 19, 20, and 22]. Many cases of dyslexia can't identify by parents, teachers and public, so spread awareness of this disease is needed. Early and better prediction of dyslexia is also challenging task.

## 4. FUTURE DIRECTION AND CONCLUSION

Machine Learning is a type of artificial intelligence that leaners and identifies new patterns from a huge amount of data. Generally, KNN, Random Forest and SVM are used for classification whose accuracy level is attained high. A combination of the above -mentioned methods is likely to provide better outcomes in detecting dyslexia. Identifying dyslexia children at an early age to provide them with appropriate learning facilities is highly important.

## 5. REFERENCES

[1] Gilles Richard and M. Serrurier, "Dyslexia and Dysgraphia prediction: A new Machine Learning Approach". In: ArXiv Journal, Volume: abs/2005.06401, 2020.

[2] Geeta Atkar and Priyadarshini J, "Advanced Machine Learning Techniques to Assist Dyslexic Children for Easy Readability", In: International Journal of Scientific & Technology Research, (IJSTR), ISSN: 2277-8616, Volume-9, Issue- 03, pp.1655-1661, 2020.

[3] Jothi Prabha A and Bhargavi, "Predictive Model for Dyslexia from Fixations and Saccadic Eye Movement Events", In: IEEE Computer Methods and Programs in Biomedicine, ISSN: 0169-2607, pp.1-13, 2020.

[4] F. Latifoglu, R. İleri, E. Demirci and C. G. Altıntop, "Detection of Reading Movement from EOG Signals", In: IEEE International Workshop on Medical Measurement and Applications (MEMEA), pp.1-5, 2020.

[5] Michat Obidzinski, "Response Frequencies in the Conjoint Recognition Memory Task as Predictors of Developmental Dyslexia Diagnosis: A Decision-Trees Approach", In: WILEY, pp.1-12, 2020.

[6] N. Giri et al., "Detection of Dyscalculia Using Machine Learning", In: 5th International Conference on Communication and Electronics Systems (ICCES), pp.1-6, 2020.

[7] Viraj Trivedi et.al., "Detecting the Severity and the Type of Learning Disability with Pattern Extraction Using Machine Learning", In: International Journal of Computer Science Trends and Technology (IJCST), Volume 8, Issue 2, 2020.

[8] Ng Li Mun and Nur Anida Jumadi, "A Comparative Classification Models Study for Development of Early Dyslexia Screening System", In: Universal Journal of Educational Research, ISSN: 2332-3205, Volume-8, pp.1-15, 2020.

[9] https://www.medicalnewstoday.com/articles/186787, Medically reviewed by Timothy J. Legg, Ph.D., CRNP -Written by Yvette Brazier on February 25, 2020.

A. Jothi Prabha and R. Bhargavi, "Prediction of Dyslexia from Eye Movements Using Machine Learning", In: IETE Journal of Research, ISSN: 0377-2063, pp.1-10, 2019.

[10] Vani Chakraborty and Meentachi Sundaram, "Machine Learning Algorithm for Prediction of Dyslexia Using Eye Movement", In: 3rd National Conference on Computational Intelligence (NCCI) Journal of Physics, ISSN: 1742-6588, pp.1-9, 2020.

[11] R. Kariyawasam, M. Nadeeshani, T. Hamid, I. Subasinghe and P. Ratnayake, "A Gamified Approach for Screening and Intervention of Dyslexia, Dysgraphia and Dyscalculia", In: International Conference on Advancements in Computing (ICAC), ISBN: 978-1-7281-4171-8, pp.156-161, 2019.

[12] Margaret Mary T, Hanumanthappa M, and Sangamithra A, "Intelligent Predicting Learning Disabilities in School Going Children Using Fuzzy Logic K Mean Clustering in Machine Learning", In: International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-4, pp.1694-1698, 2019.

[13] Jothi Prabha A and Bhargavi, "Predictive Model for Dyslexia from Eye Fixation Events", In: International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249-8958, Volume-9, Issue-1S3,pp.235-240, 2019.

[14] Inclusive City Maker.2019. Disabled People in the World in 2019 facts and figures. Accessed from: https://www.inclusivecitymaker.com/disabled-people-in-the-world-in-2019-facts-and-figures.

[15] 5 Most Common Learning Disabilities-Masters in Special Education Degree Program Guide, Masters-in-special-education.com... [online]. Available: https://wwwmasters-in-special-education.com/lists/5-most-common-learning-disabilities/, 2019.

[16] M.Iwabuchi, R Hirabayashi, K.Nakamura and N.K.Dim, 2017. "Machine Learning based evaluation of reading and writing difficulties", In: Stud. Health Technol.Inform., Volume-242, pp.1001-4.

[17] K.C.Fraser,K.LFors,D. Kokkinakis and A.Nordlund,"An analysis of eye movements during reading for the detection of mild cognitive impairment", In: Proceeding of the Conference on Empirical Methods in Natural Language Processing, pp.1016-26, 2017.
[18] Schmalz, X, Altoe, G. and Mulatti C, 2017, "Statistical Learning and Dyslexia: A Systematic Review", In: Ann. of Dyslexia, Volume-67, pp.147–162, 2017.
[19] J.Lusting, " Identifying Dyslectic Gaze Pattern: Comparison of Methods for Identifying Dyslectic Readers based on Eye Movement Patterns", 2016.
[20] S. I. Dimitriadis et al., "Classifying Children with Reading Difficulties from Non-impaired readers via Symbolic Dynamics and Complexity Analysis of MEG resting-state Data", In: IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp.333-336, 2016.
[21] Z. Mahmoodin et.al, "Electroencephalogram Electrode Localization in the Support Vector Machine Classification of Dyslexic Children", In: IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp. 296-300, 2016.
[22] L.Rello and M. Ballesteros, "Detecting Readers with Dyslexia using Machine Learning with Eye Tracking Measures", In: Proceedings of the 12th Web for All Conference ACM, 2015.
[23] Atakan VH et.al, "Early Prediction of Reading Disability using Machine Learning", In: Annual Symposium proceedings American Medical information association, 2009.
[24] C.Prado.M. Dubois and S. Valdois, "The Eye Movements of Dyslexic Children during Reading and Visual Attention Span", In: Elsevier, ISSN: 0042-6989, pp.2521-2530, 2007.

APPENDIX

TABLE 1: METHODS USED IN PREDICTION

| **Name of the Author** | **Year** | **Title** | **Proposed Method (Algorithm)** | **Dataset** | **Accuracy** | **Advantages** | **Disadvantages** |
|---|---|---|---|---|---|---|---|
| Gilles Richard et al. [1] | 2020 | Dyslexia and Dysgraphia Prediction A new machine learning approach. | Majority class, Naïve Bayes Logistic regression, Random Forest | 69 People including 41 People with Official dyslexia diagnosis | 90% | Accuracy can still be improved by gathering more data. | A better understanding of the correlation between the different disorders could also help in providing accurate predictions. |
| Geetha Atkar, et al. [2] | 2020 | Advanced Machine Learning Techniques To assist Dyslexic Children for Easy Readability | Mel Frequency Cepstral co-efficient (MFCC), Dynamic Time Warping (DTW) | 30 audios spoken by 20 different people dataset contains 200 words. Age between 5 | 90%-100% | Dynamic Time Wrapping techniques gives around 90% to 100% | System is tested with new user 30% accuracy |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | to 7 years | | | |
| Jothi Prabha Aet al.[3] | 2020 | Predictive model for dyslexia from fixations and saccadic Eye movement Events | Particle Swarm Optimization (PSO) based SVM Hybrid Kernel (Hybrid SVM -PSO), Support Vector Machine Random Forest (RF) Logistics Regression (LR), K-Nearest Neighbor (KNN) | Eye movements of 185 children whose age is around 9 to 10 years. 97 of them were dyslexia and 88 of them non-dyslexics. | 95.6 % | The proposed features such as avg.no of fixations, avg. fixation gaze duration, avg. saccadic movement duration and total no. of saccadic movements duration and the ML model can be used to design a screening tool for dyslexia. | The latest ML algorithms for improved predictive accuracy. |
| Fatma Latifogluet al. [4] | 2020 | Detection of Reading Movement from EOG Signals | J48, K-Nearest Neighbor (KNN), Random Forest (RF), Random Tree | The recorded EOG signals were obtained from 10 volunteers (5 female and 5 male), Age between 8 to 12 years old | 98% | High performance and highest accuracy of the Random Forest and KNN classifiers | EOG signals recorded from various disease groups can be analyzed. studies on modeling and detection of these two re-reading and skipping features automatically. |

| Name of the Author | Year | Title | Proposed Method (Algorithm) | Dataset | Accuracy | Advantages | Disadvantages |
|---|---|---|---|---|---|---|---|
| Obidzinski Michat[5] | 2020 | Response frequencies in the conjoint recognition memory task as predictors of developmental dyslexia diagnosis: A decision approach | Classification and Regression Tree (CART), Decision Trees | 71 High School students, 33 with developmental dyslexia took part in a memory experiment | 81.38 % | The advantages of this method are that is particularly eliminates possibility of disorder simulation. | Further studies implementing and testing the presented model are necessary to confirm its overall usefulness. |
| N Giriet al. [6] | 2020 | Detection of Dyscalculia Using Machine Learning | Random Forest, Decision Tree | 650 samples obtained from B.Y.L Nair Ch. Hospital | 99.96 % | This model will help doctors can diagnosis a speedy process. | Manual detection is a complex task |
| Viraj Trivedi etal. [7] | 2020 | Detecting the Severity and the Type of Learning Disability with Pattern Extraction Using Machine Learning | Decision Tree, Logistic Regression, Random Forest and XGBoost Classifier | Students studying standard 1 to 4 for MHSB. (Maharashtra State Board) | 0.70, 0.60, 0.50, 0.50 | proper support, guidance and right technology children can achieve greater success in school/college and be successful in life | Not focus Higher standard Students |
| Ng Li Mun et al. [8] | 2020 | A Comparative | Fuzzy Inference System | The data consists of 30 | the accuracy of | A rapid dyslexia screening | Further modification on rule statements by |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Classification Models Study for Development of Early Dyslexia Screening System | (FIS), Random Forest, Decision Table and Naïve Bayes | subjects: 17 dyslexiaand 13 slow learners with the age range of 6 to 10 years old. | 56.7 % and 100 % | system based on fuzzy logic | adding another IQ test. And the accuracy and robustness of the fuzzy inference system will be improved. |
| A.Jothi Prabha et al.[10] | 2019 | Prediction of dyslexia from Eye movement using Machine Learning | Principal Component Analysis (PCA), Particle Swarm Optimization (PSO) based Hybrid Kernel SVM-PSO | Raw eye tracking data recorded 185 subjects age of 9-10 Years. | 95% | Improved accuracy, sensitivity & specificity over Linear SVM | To improve (hybrid kernel SVM -PSO Model is validated) the predictive accuracy. |
| Vani Chakraborty et al. [11] | 2019 | Machine Learning Algorithms for prediction of dyslexia using eye movements. | K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest | Dept. of Psychology University of Jyvaskyla (Finland) | 89.8 % | The model is a prescreening instrument for dyslexia location | Utilizing the structure science standards in an iterative design, accomplish objective & furthermore acquire information regards to the issue |
| **Name of the Author** | **Year** | **Title** | **Proposed Method (Algorithm)** | **Dataset** | **Accuracy** | **Advantages** | **Disadvantages** |
| RuchiraKariyawasamet al. [12] | 2019 | A Gamified Approach for Screening and Interventi on of | Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor | Modified National Institute of Standards and Technology | Accuracy of 89%, 90%, 92%, 92% for Dysle | First game-based screening and interventio n tools for dyslexia, letter | Majority of applications test English Language only, Not focus children age between 6 to 7. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Dyslexia, Dysgraphia and Dyscalculia | (KNN) | (MNIST) numbers data set. | xia, letter dysgraphia, dyscalculia and numeric dysgraphia | dysgraphia, dyscalculia and numeric dysgraphia | |
| Margaret Mary T et al.[13] | 2019 | Intelligence Predicting Learning Disabilities in school going children using Fuzzy Logic K-Mean Clustering in Machine Learning | Fuzzy logic K- Mean Clustering | 620 datasets | 93.54 % | Effectively and accurately predict the educational inability in children. | Additional analysis needed to use similar approach for big dataset. |
| Jothi Prabha A et al. [14] | 2019 | Predictive model for dyslexia from Eye Fixation Events | Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Random Forest (RF) | The dataset contains raw recording data of 185 subjects studying grade 2. 97 were dyslexics and 88 were non-dyslexics | KNN achieved highest accuracy of 95% | Focuses on identifying features that contribute better prediction & build an appropriate prediction model. | Not to focus on multiple age groups. |