Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

AN INVESTIGATION INTO PREDICTIVE MODELS FOR EARLY DETECTION MACHINE LEARNING BASED DYSLEXIA PREDICTION

Dr.M.Sumithra[1,a], B. Buvaneswari[2,b], Gayathri V R[3,c], Dhaarani S[3,d], Rajalakshmi R[3,e]
[1]Associate Professor, Department of Information Technology, Panimalar engineering college
[2]Professor, Department of Information Technology, Panimalar engineering college
[3]Students, Department of Information Technology, Panimalar engineering college
sumithram.id@gmail.com[a], buvanrajan16@gmail.com[b], gayathrivemula2003@gmail.com[c], dhaaranis19@gmail.com[d], rajirajasekaran2628@gmail.com[e]

ABSTRACT

A neurological condition called dyslexia is typified by a lack of precision in word comprehension and generally subpar reading abilities. It impacts a sizable portion of school-age children, more often affecting boys, and puts them at risk for lifelong low self-esteem and subpar academic performance. The ultimate goal is to develop a dyslexia diagnostic approach based on brain biomarkers. Numerous machine learning techniques and, more recently, deep learning techniques have been applied to diverse dataset types with above-chance classification performance in this regard. This tutorial provides a thorough resource for both inexperienced and seasoned developers, outlining a methodical approach to creating a Dyslexia Detection Web App using Flask and Boosting techniques. The abstract summarizes two main points of emphasis: using sophisticated Boosting techniques to identify dyslexia accurately and utilizing Flask's flexibility to create web apps quickly and easily. Through an exploration of the complexities involved in dyslexia identification and the provision of useful insights regarding algorithm implementation, the goal of this book is to enable readers to develop an advanced tool for supporting dyslexics. The guide's dedication to increasing accessibility, cultivating inclusion through technology, and expanding the field of assistive applications is emphasized in the abstract.

Keyword: Dyslexia detection, Biomarkers, Feature Extraction, Classification Algorithms, Machine learning and Deep Learning.

I. INTRODUCTION

Experts in contemporary neuroscience are paying close attention to dyslexia, a very complex neuro-developmental brain condition. The majority of the cultures and languages possess a neurological disease that affects 5–17% of people in general. It is characterized by slow and erroneous word understanding as well as phonological impairment. This syndrome can be detrimental to academic achievement and often develops from childhood into puberty. Dyslexia may also have a major detrimental impact on children's development of self-perception and self-esteem. In academic contexts, dyslexic students face high rates of bullying in addition to feelings of isolation and alienation. According to the definitions provided by the British Dyslexia

Association and the International Dyslexia Association, some dyslexic youngsters struggle with executive control or higher-order processing. They also suffer from a visual attention impairment, which has a major impact on their reading comprehension. They might also struggle with letter recognition and memorization. Dyslexics therefore have severe deficiencies in verbal memory, phonological processing, and communication speed in both adults and children. Depending on what the victim's brain experienced throughout growth or whether they had a serious damage like a stroke, dyslexia can be either acquired or developmental. This project aims to create a dyslexia detection model by utilizing a boosting technique. Boosting strategies in machine learning solutions might be beneficial for individuals with dyslexia, a common learning impairment. The project's aim is to develop a model that can correctly diagnose dyslexia based on pertinent characteristics. This model was then included into a Flask web application to improve accessibility, giving users access to real-time predictions and maybe aiding in the early detection of dyslexia.

II. LITERATURE SURVEY

Velmurugan S. "A thorough analysis of feature selection, algorithms, and evaluation metrics for machine learning in dyslexia prediction" (2022). This review of the literature looks at the diagnosis and treatment of dyslexia, a learning disability that affects spelling and reading comprehension, using machine learning techniques. Based on functional Magnetic Resonance Imaging (fMRI) and electroencephalography (EEG) data, a variety of machine learning models, including Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and decision trees were used to categorize people as dyslexic or non-dyslexic. [1]

Yih-Choung Yu, Lisa Gabel, Amanuel Zewge, and KhaknazarShyntassov "Classification Predictive Modeling of Dyslexia" (2022). Despite having a normal IQ and access to education, children with dyslexia struggle with reading comprehension in all language orthographies. Untreated learning difficulties can lead to long-term social and emotional issues in children, which can negatively impact their ability to succeed in many facets of their lives in the future. The learning gap between children who are ordinarily developing and those who have reading impairments will be reduced with early detection and remediation. We have shown that on a virtual Hebb-Williams maze test, children with specific reading impairment, genetic models based on putative genes associated with dyslexia susceptibility, and animal models of dyslexia all exhibit a shared disadvantage.[2]

M. Mahalakshmi and Dr. K. Merriliance, "Machine Learning Algorithms for Predicting Dyslexia" (2022). Work is being done to separate dyslexics from non-dyslexics using a variety of methods, including machine learning, image processing, psychology, and an analysis of the anatomical variations in the brain. In this study, those at high risk of dyslexia are screened using brain scans. Additionally, this work encourages the use of machine learning in distributed environments. The suggested predictive model makes use of Machine learning techniques such as Random Forest

(RF) and Decision Tree (DT). Python implementation and the Weka tool are used to classify the model.[3]

Alqahtani, Norah Dhafer, Alzahrani, and Ramzan, Muhammad Sher, "Deep Learning Applications for Dyslexia Prediction" (2023). Several dyslexia-related datasets obtained from educational and medical institutions have been used to apply machine learning and deep learning techniques for dyslexia recognition. Researchers encounter many difficulties while attempting to diagnose and classify patients using deep learning models for dyslexia, which are summarized in this review study that examines the prediction performance of these models. Nineteen articles were reviewed and examined using the PRISMA technique. The studies concentrated on the performance of the prediction model, data extraction, preprocessing, and acquisition. With the help of existing dyslexia-related datasets, the aim of this review was to assist researchers in developing a dyslexia predictive model. [4]

A. Prabha and R. Bhargavi "Prediction of Dyslexia Using Machine Learning - Expedition Report" 2021. A common learning problem known as dyslexia is thought to be characterized by chronic spelling errors and rapid word recognition. It affects a person's ability to decode words and letters correctly and smoothly. The research community has attempted to distinguish individuals with dyslexia using a variety of machine learning strategies, image processing methods, design evaluation methods, and assistive technologies that support dyslexia. This study explores several aspects of dyslexia research. This study highlights the challenges, outstanding issues and research gaps in this field. In addition, it encourages the use of machine learning methods to predict early dyslexia.[5]

III. EXISTING SYSTEM

Traditionally, dyslexia is identified through educational and psychological evaluations. While some researchers are investigating machine learning for this purpose, there is no uniform system yet. To get the most up-to-date knowledge, look into recent research and open-source projects in the relevant domains. Ethical issues and stakeholder involvement are critical while creating treatments for learning difficulties.

Disadvantages

- Subjectivity in Evaluation
- Dependency on Human Expertise
- Limited Accessibility
- Costly Interventions
- Lack of Personalization

redicting Dyslexia with Machine Learning: A

Comprehensive Review of Feature Selection,

Algorithms, and Evaluation Me

IV. SYSTEM ARCHITECTURE

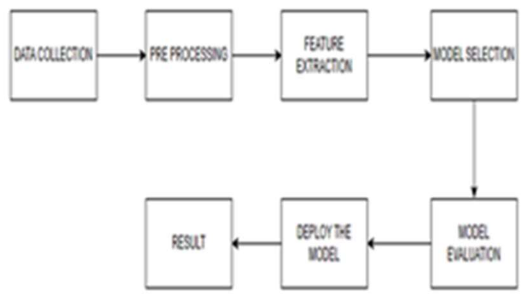

Figure 1: System Architecture

1.Data Collection

Compile a collection of medical records containing patient demographics, medical history, lifestyle choices, and stroke history. In order to train a reliable machine learning model, this dataset needs to be sizable and diverse. The above Figure 1 depicts the architecture of the system

2.Data Preprocessing

Prepare and organise the data. This include resolving missing values, normalising data, and putting numerical form on data that is categorical. While handling private medical data, make sure that data privacy and applicable requirements are followed

3.Feature Selection and Engineering

Identify relevant features (attributes) that can influence stroke prediction. These may include age, gender, hypertension, heart disease, smoking status, BMI, and more. Perform feature engineering to create new features or transform existing ones if needed.

4.Data Splitting

Split the dataset into training and testing sets. The training set is used to train the machine learning model, and the testing set is used to evaluate its performance. Figure 2 shows the labelled data set.

5. Machine Learning Model Selection

Choose the appropriate machine learning algorithm for classification. Common choices include logistic regression, decision trees, random forests, support vector machines and neural networks.

Evaluate and compare different algorithms to choose the one that works best. We use an AdaBoost classifier and a gradient boosting classifier.

6.Model Training

Train the selected machine learning model with the training data. The model learns to predict whether a patient is at risk of stroke based on predetermined characteristics.

7.Model Evaluation

Evaluate the model's performance on the testing set using appropriate metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC, to assess its predictive accuracy and reliability.

V. PROPOSED SYSTEM

The suggested system takes an innovative approach to dyslexia diagnosis by using a boosting algorithm, which is specifically designed to improve accuracy and efficiency when compared to traditional methods. This method employs machine learning, taking advantage of the boosting algorithm's capabilities to automate the identification process. A Flask web application enables real-time predictions while overcoming subjectivity concerns and providing continuous monitoring. The system is intended to adapt to new technologies, providing a data-driven, accessible, and successful method of detecting early dyslexia.

Advantages

- Precision and Personalization.
- Efficiency and Time-Saving.
- Continuous Monitoring
- Real-Time Predictions
- Efficient Pattern Recognition

A. Analysis


Delimiter:

	Language_vocab	Memory	Speed	Visual_discrimination	Audio_Discrimination	Survey_Score	Label
1	0.5	0.6	0.5	0.8	0.6	0.7	1.0
2	0.6	0.7	0.8	0.9	0.5	0.8	2.0
3	0.6	0.4	0.3	0.3	0.4	0.6	1.0
4	0.3	0.5	0.2	0.1	0.3	0.5	0.0
5	0.7	0.6	0.7	0.8	0.9	0.5	2.0
6	0.4	0.1	0.0	0.1	0.4	0.2	0.0
7	0.8	1.0	0.8	0.9	0.6	0.6	2.0
8	0.5	0.3	0.5	0.4	0.7	0.4	1.0
9	0.6	0.5	0.5	0.4	0.6	0.5	1.0
10	0.6	0.7	0.7	0.8	0.7	0.6	1.0
11	0.7	0.7	0.5	0.7	0.8	0.8	1.0
12	0.5	0.6	0.6	0.6	0.7	0.6	1.0
13	0.9	0.6	0.5	0.9	0.7	0.7	2.0
14	0.6	0.4	0.4	0.6	0.5	0.6	1.0
15	0.7	0.6	0.4	0.3	0.4	0.3	1.0
16	0.6	0.5	0.5	0.5	0.4	0.4	1.0
17	0.0	0.2	0.3	0.0	0.3	0.2	0.0
18	0.4	0.3	0.3	0.1	0.4	0.4	0.0
19	0.7	0.7	0.8	0.7	0.7	0.9	2.0
20	0.5	0.6	0.6	0.4	0.5	0.3	1.0
21	0.5	0.6	0.3	0.4	0.3	0.6	1.0
22	0.7	0.6	0.6	0.9	0.7	0.7	2.0
23	1.0	0.8	0.9	0.5	0.7	0.8	2.0
24	0.6	0.3	0.4	0.5	0.5	0.5	1.0

Figure 2: Data Set

1.Language vocabulary

The language vocabulary feature can provide valuable insights into an individual's linguistic abilities and potential challenges with reading and language processing. Lower scores or performance deficits in language vocabulary measures may indicate difficulties with word recognition, comprehension, or language processing associated with dyslexia. These measures can be crucial for identifying individuals at risk for dyslexia and developing targeted interventions to support their language and literacy development.

2.Memory

In the context of predicting dyslexia, deficits in memory, particularly working memory and verbal memory, are commonly observed and may contribute to difficulties with reading fluency, comprehension, and language processing. Assessing memory function as part of a dyslexia prediction dataset can provide valuable insights into individuals' cognitive profiles and help identify specific areas of weakness that may require targeted interventions to support reading and language development.

3.Speed

In the context of predicting dyslexia, slower processing speed, particularly in tasks related to reading and language processing, is often observed in individuals with dyslexia. Deficits in speed can contribute to difficulties with reading fluency, comprehension, and academic achievement. Assessing speed as part of a dyslexia prediction dataset can provide valuable information about individuals' cognitive processing abilities and help identify potential risk factors associated with dyslexia. It can also inform the development of targeted interventions aimed at improving processing speed and reading fluency in individuals at risk for or diagnosed with dyslexia.

4.Visual discrimination

In the context of predicting dyslexia, deficits in visual discrimination, particularly in tasks related to letter and word recognition, visual pattern recognition, and visual similarity, are commonly observed in individuals with dyslexia. Difficulties with visual discrimination can contribute to challenges with reading fluency, accuracy, and comprehension. Assessing visual discrimination abilities as part of a dyslexia prediction dataset can provide valuable insights into individuals' visual processing skills and help identify potential risk factors associated with dyslexia. It can also inform the development of targeted interventions aimed at improving visual discrimination skills and supporting reading and academic achievement in individuals at risk for or diagnosed with dyslexia.

5.Audio discrimination

In the context of predicting dyslexia, deficits in audio discrimination, particularly in tasks related to phoneme discrimination, word discrimination, and auditory pattern recognition, are commonly observed in individuals with dyslexia. Difficulties with audio discrimination can contribute to challenges with phonological processing, language development, and reading acquisition. Assessing audio discrimination abilities as part of a dyslexia prediction dataset can provide valuable insights into individuals' auditory processing skills and help identify potential risk factors associated with dyslexia.

6.Survey Score

Analyzing survey data alongside other features in a dyslexia prediction dataset can help identify patterns and associations between survey responses and dyslexia outcomes, facilitating the development of predictive models for identifying individuals at risk for dyslexia. Additionally, survey scores can inform the design and implementation of targeted interventions and support services to address individuals' specific needs related to reading difficulties and dyslexia.

7.Label

The label in a dyslexia prediction dataset serves as the target variable that predictive models seek to predict or classify, providing valuable information about individuals' dyslexia status, risk level, severity, progression, or related traits. Analyzing the label alongside other features in the dataset enables the development of predictive models for identifying individuals at risk for dyslexia, guiding early intervention efforts, and informing personalized interventions and support services tailored to individuals' specific needs.

B. Algorithm

Boosting Algorithms:

Boosting algorithms are a class of machine learning techniques that aim to enhance the predictive performance of models by combining the strengths of multiple weak learners. Weak learners, often simple decision trees, are sequentially trained on the data, with each subsequent learner focusing on the mistakes of its predecessors. The key idea is to assign higher weights to misclassified instances, thereby emphasizing their importance in subsequent iterations. Gradient Boosting, AdaBoost, and XG Boost are popular boosting algorithms. Gradient Boosting optimizes model errors by minimizing the gradient of the loss function, AdaBoost assigns varying weights to instances based on their classification success, and XG Boost extends these concepts with regularization and parallel processing capabilities. Boosting algorithms excel in improving predictive accuracy and are widely utilized in diverse machine learning applications.

C. Application

1.Exploratory Data Analysis:

We defined the target variable and carried out a descriptive analysis in this step. We also identified other variables that could be troublesome (high cardinality) and the number of categories that were targeted. In order to better comprehend the data distribution and help select the parameters, I also represented the target variable as a histogram.

2.Data Cleaning:

We dropped these large cardinal variables in this phase as a precursor to the preprocessing phase.

3.Pre-processing & Transformation

The categorical variable was transformed into a single hotcode template matrix, and the target variable was eliminated from the complete data set. Certain algorithms may occasionally be required to process data in the sparse matrix format. This stage of the model-building process is automated by other statistical tools, such R.

I used 0 for missing values. In order to prevent variables on various scales from significantly influencing the coefficients, I scaled the continuous variables using min-max normalisation, which converts the data to a 0–1 scale.

4.Data Partition:

The pre-processed data was divided into training and test sets. Modelling: Using Euclidean distance and ten neighbour classes, we constructed a k-NN classification model.

5.Evaluation:

We computed R-squared values for training and test data and assessed the classifier using unseen test data.

VI. RESULTS & DISCUSSION

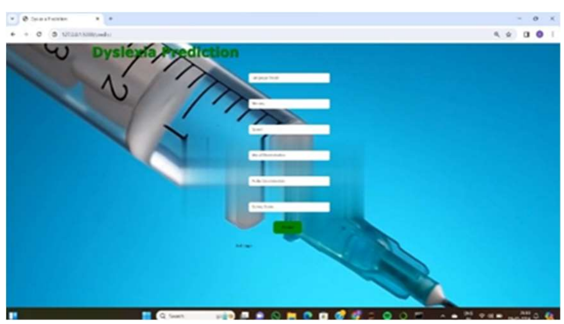

Figure 3: Web page for Dyslexia prediction

Possible future improvements to the dyslexia detection web application may include integrating new machine learning models to improve accuracy, include user feedback mechanisms to continuously improve and adapt the algorithm, improve the user interface for a more intuitive user experience, and explore the integration of new technologies. such as natural language processing or deep learning to improve the ability to detect blocking disorder. In addition, continued research and collaboration with experts in dyslexia and related fields can help improve the algorithm and expand the effectiveness of the program in real-world scenarios. Figure 3 Shows the web design for the dyslexia prediction.

In conclusion, the dyslexia detection system powered by the Flask web optimization algorithm provides efficient and accurate detection. With real-time predictions, continuous monitoring and user-friendly features, it is a promising solution for early detection of dyslexia. The system's scalability, adaptability and privacy protections contribute to its potential to positively impact education and support for people with dyslexia.

REFERENCE:

1.A. Elnakib, A. Soliman, M. Nitzken, M. F. Casanova, G. Gimel'farb and A. El-Baz, "Magnetic resonance imaging findings for dyslexia: A review", J. Biomed. Nanotechnol., vol. 10, no. 10, pp. 2778-2805, Oct. 2014.

2.J. M. Fletcher et al., "Classification of learning disabilities: An evidence-based evaluation" in Identification of Learning Disabilities: Research to Practice, Washington, DC, USA:Erlbaum Associates Publishers, pp. 185-250, Jan. 2002.

3.P. Tamboer, H. C. M. Vorst, S. Ghebreab and H. S. Scholte, "Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia", NeuroImage: Clin., vol. 11, pp. 508-514, 2016.

4.S. O. Wajuihian, "Neurobiology of developmental dyslexia Part 1: A review of evidence from autopsy and structural neuro-imaging studies", Optometry Vis. Develop., vol. 43, no. 3, pp. 121-131, 2012.

5.S. O. Wajuihian and K. S. Naidoo, "Dyslexia: An overview", Afr. Vis. Eye Health, vol. 70, no. 2, pp. 89-98, Dec. 2011.

6.B. Sarah, C. Nicole, H. Ardag, M. Madelyn, S. K. Holland and H. Tzipi, "An fMRI Study of a Dyslexia biomarker", J. Young Investigators, vol. 26, no. 1, pp. 1-4, 2014.

7.Dyslexia in the Classroom—What Every Teacher Needs to Know, 2017, [online] Available: http://www.DyslexiaIDA.org.

8.N. A. M. Yuzaidey, N. C. Din, M. Ahmad, N. Ibrahim, R. A. Razak and D. Harun, "Interventions for children with dyslexia: A review on current intervention methods", Med. J. Malaysia, vol. 73, no. 5, pp. 311-320, Oct. 2018.

9.S. Kaisar, "Developmental dyslexia detection using machine learning techniques: A survey", ICT Exp., vol. 6, no. 3, pp. 181-184, 2020.

10.F. Ramus, I. Altarelli, K. Jednoróg, J. Zhao and L. Scotto di Covella, "Neuroanatomy of developmental dyslexia: Pitfalls and promise", Neurosci. Biobehav. Rev., vol. 84, pp. 434-452, Jan. 2018.

11. J. Light and D. McNaughton, "Sound blending," 2019

12. DSF, "Structured synthetic phonics: A guide for teachers and parents,"

Available:https://dsf.net.au/CMSPages/GetFile.aspx?

13. A. El Kah and A. Lakhouaja, "Developing effective educative games for

Arabic children primarily dyslexics," Educ. Inf. Technol., vol. 23, no. 6,pp. 2911–2930, 2018.

14.G. T. Gillon, Phonological Awareness: From Research to Practice. NewYork, NY, USA: Guilford Publ., 2017.

15.F. C. Blumberg and S. M. Fisch, "Introduction: Digital games as a context for cognitive development, learning, and developmental research,"

16. J. Andrew, S. Henry, A. Yudhisthira, Y. Arifin, and S. D. Permai,

"Analyzing the factors that influence learning experience through gamebased learning using visual novel game for learning Pancasila," Proc.Comput. Sci., vol. 157, pp. 353–359, 2019.

17. A. Pho and A. Dinscore, "Game-based learning," Tips Trends, 2015.

18. M. Dempsey, R. Riedel, and M. Kelly, "Serious play as a method for process design," in Proc. IFIP Int. Conf. Adv. Prod. Manage. Syst., 2014,pp. 395–402.

19. G. R. Gredler, "Snow, C.E., Burns, M.S., & Griffin, P.(Eds.)(1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press, 432 pp., $35.95," Psychol. Sch., vol. 39, no. 3, pp. 343–344, 2002.

20. M. Mengisidou and C. R. Marshall, "Deficient explicit access to phonological representations explains phonological fluency difficulties in

Greek children with dyslexia and/or developmental language disorder,"

Front. Psychol., vol. 10, 2019, Art. no. 638.

21.G. S. V. Padmakar, A. Khosla, and K. Chand, "LearnEasy-android application as a technological intervention for children with dyslexia," inEmerging Trends in the Diagnosis and Intervention of Neurodevelopmental Disorders. Hershey, PA, USA: IGI Global, 2019, pp. 236–248.

22. S. Franceschini et al., “Action video games improve reading abilitiesand visual-to-auditory attentional shifting in English-speaking childrenwith dyslexia,” Sci. Rep., vol. 7, no. 1, pp. 1–12, 2017.

23. R. T. Bigueras, M. C. A. Arispe, J. O. Torio, and D. E. Maligat, Jr.,“Mobile game-based learning to enhance the reading performance ofdyslexic children,” Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 1.3,pp. 332–337, 2020