IJIRMET ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 10, Issue 3, March 2025

Random Forest Based Prediction and Classification of Developmental Dyslexia

[1] A.Kalaiyarasi, [2]M.Jasmine, [3]E.Meenakshi, [4]K.V.Varuna Priya
[1] Associate Professor, [2] [3] [4] Students,Department of Information Technology, Muthayammal Engineering College (Autonomous),Rasipuram 637408,TamilNadu,India.
[1] kalaiyarasi.mecse@gmail.com, [2] jasminerafi2003@gmail.com, [3] meenakshie04@gmail.com, [4] varunapriyakv@gmail.com

Abstract: ***The neurological condition known as dyslexia, which is more common in men, has a substantial impact on reading and comprehension skills, especially in school-age children. Poor academic achievement and long-term effects on self-esteem are possible outcomes of this disorder. This work investigates the application of the Random Forest algorithm, a machine learning method renowned for its accuracy and resilience, to enhance the diagnosis of dyslexia. In order to categorize people with dyslexia, the Random Forest algorithm is applied to a dataset of behavioural and brain markers. The goal is to find patterns and biomarkers that can help with early detection. This method uses Random Forests' ensemble nature to improve model reliability and generalization, addressing major issues in dyslexia diagnosis such the necessity for interpretable biomarkers and the possibility of overfitting. The study's findings show that the Random Forest algorithm has the capacity to detect dyslexia with clinically meaningful accuracy, providing a promising tool for early intervention and assistance.***

Keywords: Dyslexia, Neurological condition, Machine learning, Early identification, Data anonymization technique.

I. INTRODUCTION

A neurological disorder called dyslexia impairs a person's ability to read and understand written language, even if they are intelligent or above average. It is frequently seen in children and is typified by challenges with phonological processing, word recognition, and word decoding. Although the precise cause of dyslexia is still unknown, it is thought to be related to variations in the structure and function of the brain. To minimize its effects on academic achievement and put effective assistance mechanisms into place, dyslexia must be identified early. Conventional diagnostic techniques focus on behavioral evaluations and observations, but they can be laborious and frequently imprecise. By examining big datasets that contain both behavioral and neurological signs, recent developments in machine learning offer fresh possibilities for improving diagnostic precision. Better dyslexia identification and more focused interventions are made feasible by utilizing algorithms that can spot trends in these datasets. By providing a more impartial and reliable way to identify those who are at risk, this strategy seeks to overcome the shortcomings of traditional diagnostic techniques.

A. DYSLEXIA

Even though a person has normal cognitive and sensory abilities, dyslexia is a learning disability that impairs reading, writing, and spelling skills. It is frequently typified by challenges with word decoding, fluent reading, and comprehension of written material. Because dyslexia frequently entails issues with processing language-based tasks, people with this condition may also struggle with spelling, writing, and occasionally even math. Although the actual source of the condition is still unknown, it usually results from variances in the way the brain processes spoken and written language. The development of dyslexia is influenced by both neurological and genetic factors, and reducing its negative effects on social and intellectual development requires early detection. People with dyslexia frequently perform exceptionally well in domains where verbal reasoning is less prevalent, such as creativity and problem-solving.

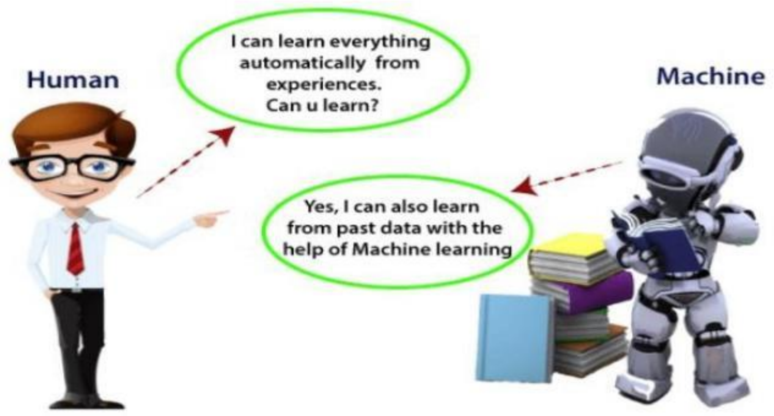
B. NEUROLOGICAL CONDITION

Any disorder affecting the nervous system, which includes the brain, spinal cord, and nerves, is referred to as a neurological condition. From brain traumas and developmental problems to degenerative diseases and neurochemical imbalances, these conditions can take many different forms. Depending on whatever area of the nervous system is impacted, symptoms might include anything from sensory and emotional difficulties to motor dysfunctions and cognitive impairments. Neurological diseases are frequently categorized as acquired, traumatic, or inherited depending on where they originate. For instance, well-known neurological illnesses include epilepsy, Parkinson's disease, and Alzheimer's disease. These illnesses can have a major negative influence on a person's quality of life and frequently need to be continuously managed with medicine, counselling, and other helpful therapies. For the purpose of creating efficient treatments and enhancing the quality of life for those impacted, neurological illness research is essential.

1.1 TECHNIQUES USED

A. MACHINE LEARNING

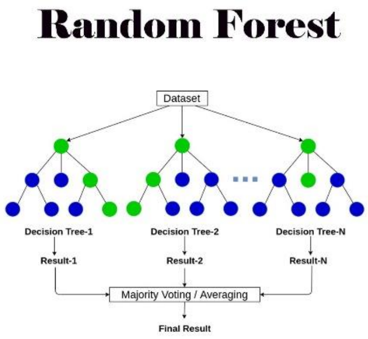
The goal of the artificial intelligence (AI) field of machine learning (ML) is to develop methods that let computers learn from and forecast data. Machine learning models enhance their performance by identifying patterns and adjusting over time, as opposed to being programmed with precise instructions for each task. The fundamental tenet is that these systems don't require intentional retraining to become more efficient; they can "learn" from experience. Machine learning comes in a variety of forms, such as supervised learning, in which the model gains knowledge from labeled data, and unsupervised learning, in which the model finds patterns in data without labels. Applications for machine learning are numerous and include natural language processing, picture identification, recommendation systems, and medical diagnosis. When working with big datasets, where human analysis might not be practical or effective, it is quite helpful. As more data is processed over time, machine learning systems can become more accurate and beneficial.


Human
I can learn everything automatically from experiences. Can u learn?
Machine
Yes, I can also learn from past data with the help of Machine learning

B. EARLY IDENTIFICATION

Early identification is the process of identifying a condition or disorder's early indications or symptoms before they worsen. In domains where early intervention can greatly enhance results, such as developmental psychology, healthcare, and education, this proactive approach is especially crucial. Early detection of learning impairments, such as dyslexia, enables prompt interventions that can assist the person in creating the support networks and coping mechanisms they need. Screening, evaluations, or observations are frequently used in early identification to find any problems before they become more difficult to handle. For instance, early detection of language problems in kids can result in therapies that improve their ability to communicate. In a similar vein, interventions that lower the likelihood of long-term psychological issues can result from early detection of mental health issues. Early detection has the primary advantage of preventing or lessening the severity of issues, which improves quality of life and long-term results.

A Random Forest is a collection of decision trees that work together to make predictions. In this article, we'll explain how the Random Forest algorithm works and how to use it.Random Forest algorithm is a powerful tree learning technique in Machine Learning to make predictions and **then we do voting of all the tress to make prediction**. They are widely used for classification and regression task.



Asking a group of friends for advice on where to go for vacation. Each friend gives their recommendation based on their unique perspective and preferences (decision trees trained on different subsets of data). You then make your final decision by considering the majority opinion or averaging their suggestions (ensemble prediction) then - multiple decision trees are created from the training data. Each tree is trained on a random subset of the data (with replacement) and a random subset of features. This process is known as bagging or bootstrap aggregating. Each decision tree in the ensemble learns to make predictions independently. When presented with a new, unseen instance, each decision tree in the ensemble makes a prediction. The final prediction is made by combining the predictions of all the decision trees. This is typically done through a majority vote (for classification) or averaging (for regression). Random forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both classification and regression problems in ml. It is based on the concept of **ensemble learning,** which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

II. RELATED WORK

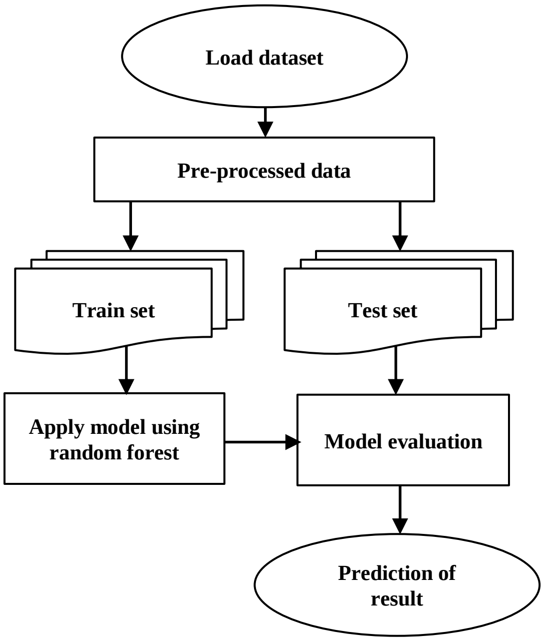
This system was proposed by Shahriar Kaisar et al. The learning problem known as developmental dyslexia primarily affects young children. Despite having ordinary or above-average intelligence, dyslexic youngsters struggle with reading, spelling, and writing words. Consequently, dyslexic kids frequently experience negative emotions including anger, frustration, and low self-esteem. Therefore, in order to support dyslexic youngsters from the beginning, early dyslexia detection is crucial. A variety of methods have been proposed by researchers to identify developmental dyslexia, including game-based methods, reading and writing assessments, eye tracking, facial image capture and analysis, magnetic reasoning imaging (MRI), and electroencephalography (EEG) scans. This review paper identifies possible directions for future research by critically analyzing current advances in machine learning algorithms for dyslexia detection. The name "dyslexia," which means "difficulty with words," comes from the Greek language. Despite possessing ordinary or above-average intelligence, people with dyslexia struggle to read, spell, and write smoothly. Dyslexia is a sort of specific learning disability (SLD). According to the Australian Dyslexia Association (ADA), 10% of Australians are thought to have dyslexia, compared to up to 20% in other English-speaking nations like Canada and the UK [1].

MODULES

- Load data
- Data Pre-Processing
- Feature Extraction
- Training and Testing

- Model Evaluation

SYSTEM ARCHITECTURE



III. PROPOSED SYSTEM

The suggested approach analyzes a large CSV dataset with behavioral and brain markers using the Random Forest algorithm to improve dyslexia detection. With an emphasis on enhancing the precision and dependability of early detection, this system seeks to identify and categorize important biomarkers linked to dyslexia. The approach tackles issues including overfitting, data privacy problems, and the requirement for interpretable biomarkers by utilizing Random Forests' ensemble learning capabilities. The ultimate objective is to create a machine learning model that attains clinically meaningful accuracy, offering a useful instrument for early intervention and assistance for people who may be at risk for dyslexia.

ADVANTAGES

- ***It performs high accuracy.***
- Low risk, and cost effective.
- Scope of improvement.
- Efficient of handling a data.
- Ensemble approach enhances classification accuracy, improving dyslexia detection.
- Easily scaled to handle large datasets

FUTURE WORK

To improve the dyslexia detection tool's functionality and impact, future research can concentrate on a few important areas. One way to increase the model's generalizability and accuracy across different demographics is to broaden the dataset to include a more diverse population that spans a range of age groups, ethnicities, and socioeconomic backgrounds. Furthermore, by identifying intricate patterns in the data, incorporating more sophisticated machine learning approaches like deep learning algorithms could improve the model's predicted accuracy even more. Conducting longitudinal studies in partnership with academic institutions and healthcare providers will also yield important information on the tool's long-term efficacy in practical contexts. Additionally, the system's usability and usefulness may be improved by integrating user feedback into iterative design modifications, guaranteeing that it satisfies the changing needs of researchers and doctors. Lastly, investigating how to incorporate this tool into current educational frameworks and therapies would help support people with dyslexia in a more comprehensive way, with the ultimate goal of improving academic performance and general well-being.

IV. RESULT AND DISCUSSION

With an emphasis on the results produced by the Random Forest model, the result analysis phase critically assesses the efficacy and performance of the dyslexia detection tool. The first step in this research is a thorough examination of the classification metrics, which together offer insights into the model's predictive power. These metrics include accuracy, precision, recall, and F1-score. The model's performance across many classes is visualized using a confusion matrix, which highlights regions of success and points out possible misclassifications. Furthermore, feature importance scores are examined to determine which biomarkers have the greatest influence on the model's predictions, providing important information about the fundamental traits linked to dyslexia. To ensure the model's relevance and dependability, the analysis additionally compares the model's output to current benchmarks and clinical standards. In order to evaluate the tool's usefulness in actual situations, qualitative input from end users—such as researchers and clinicians—is also included. This careful analysis of the data not only validates the model's efficacy but also directs future improvements and modifications, which eventually lead to better early intervention techniques for people at risk of dyslexia**.**

Algorithm	**Accuracy**
Existing system	**83**
Proposed system	**87**

TABLE 1. COMPARISON TABLE

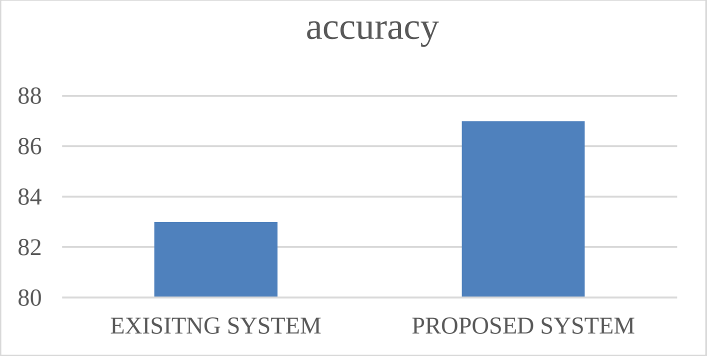


FIGURE 2 : COMPARISON GRAPH

V. CONCLUSION

In summary, the suggested dyslexia diagnosis method that makes use of the Random Forest algorithm shows a great deal of promise for improving early identification and intervention tactics for people who may be at risk for developing dyslexia. The method successfully detects important biomarkers linked to dyslexia by using a thorough process that includes data loading, preprocessing, feature extraction, model training, and evaluation, yielding results that are clinically useful. While the comprehensive result analysis demonstrates the model's predicted accuracy and interpretability, the rigorous testing and implementation phases guarantee dependability and usefulness for researchers and doctors. With the ultimate goal of enhancing educational performance and self-esteem for those impacted, this tool not only advances our understanding of dyslexia but also acts as a useful resource for prompt support and intervention. By adding bigger datasets and investigating more sophisticated machine learning approaches, future research can expand on this foundation and enhance the system's capacity for dyslexia detection and support.

REFERENCES

1. N P Guhan Seshadri, Bikesh Kumar Singh, Ram Bilas Pachori “EEG based functional brain network analysis and classification of dyslexic children during sustained attention task: A survey,” November 2023 IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society PP(99):1-1
2. S.Kaisar, ‘‘Developmental dyslexia detection using machine learning techniques: A survey,’’ ICT Exp., vol. 6, no. 3, pp. 181–184, 2020
3. U. Kuhl, N. E. Neef, I. Kraft, G. Schaadt, L. Dörr, J. Brauer, I. Czepezauer, B. Müller, A. Wilcke, H. Kirsten, F. Emmrich, J. Boltze, A. D. Friederici, and M. A. Skeide, ‘‘The emergence of dyslexia in the developing brain,’’ Neuro Image, vol. 211, May 2020, Art. no. 116633
4. B. Srivastava and M. T. U. Haider, ‘‘Personalized assessment model for alphabets learning with learning objects in e-learning environment for dyslexia,’’ J. King Saud Univ.-Compute. Inf. Sci., vol. 32, no. 7, pp. 809–817, 2020
5. A. Jothi Prabha and R. Bhargavi, ‘‘Eye Movement feature set and predictive model for Dyslexia ,’’ Compute. Methods Programs Biomed., vol. 195, Oct. 2020
6. Iza Sazanita Isal, Wan Nurazwin Syazwani Rahimi, “Automated detection of dyslexia symptom based on handwriting image for primary school children ” Procedia Computer Science Volume 163, 2019, Pages 440-449
7. Dragas, A.S.; Politi-Georgios, S. ICTs as a distinct detection approach for dyslexia screening: A contemporary view. Int. J. Online Biomed. Eng. 2019, 15, 46–60