IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# PREDICTION OF DYSLEXIA USING MACHINE LEARNING ALGORITHMS

**M.Mahalakshmi[1], Dr.K.Merriliance[2]**

**Department of Computer Applications, Sarah Tucker College, Thirunelveli-7.**

## ABSTRACT

Dyslexia is a learning disorder characterized by lack of reading and /or writing skills, difficulty in rapid word naming and also poor in spelling. Dyslexic individuals have great difficulty to read and interpret words or letters. The most common reading disability is dyslexia. This reading disability encompasses various symptoms such as poor spelling outcomes, reading fluency and difficulties in expressing oneself. Failure to diagnose children coping with dyslexia is a potential risk of discrimination and social exclusion. Dyslexia is the most widely recognized neurological learning disability. All these can influence scholastic achievement, confidence, and social-emotional development.

Studies have demonstrated that the earlier dyslexia is recognized and backing is given in education and training, the more its negative impacts can be alleviated. Subsequently, building up a solid and target screening technique to analyze dyslexia at an early age would be of most extreme significance. The work is carried out to classify dyslexic from non-dyslexics by various approaches such as machine learning, image processing, understanding the brain behavior through psychology, studying the differences in anatomy of brain. In this work, brain images are used for screening individuals who have high risk to dyslexia. This work also motivates the application of machine learning in distributed environment. The proposed predictive model uses the machine-learning algorithms like Decision Tree(DT) and Random Forest (RF). The model is classified using Weka tool and Python implementation.

## 1.INTRODUCTION

### 1.1Dyslexia

Dyslexia is a reading difficulty. Dyslexia is not a condition of vision impairment, mental acuity, or overall mental improvement. Dyslexics have a unique brain structure. The language processing region of the brain has also been affected. Dyslexia has a variety of side effects that differ by language. It's nothing more than a disease that has no cure. People who have dyslexia must learn to live with it. It affects people of all ages and is unrelated to IQ. It affects roughly 10% of the population. Predicting dyslexia is expensive and requires the assistance of a clinical or professional specialist. Machine learning, image processing in conjunction with machine learning, test-based evaluations, and other methods are used to detect dyslexia.

### 1.2. Machine learning

Machine Learning is a technique for ingesting and distinguishing new data patterns from large amounts of current data. It enables researchers and information professionals to recognize the strategies and plans that need to be developed in a viable manner. To create algorithms for grouping, regression, and classification using relevant data from massive data repositories.

Directed and unsupervised learning are the two types of machine learning algorithms. Supervised learning is a type of machine learning in which a function is inferred from data whose class labels are already known thanks to training instances. Unsupervised learning is a job in which data with unknown class labels is used to infer a function. Machine learning is a rapidly growing technology in healthcare that aids in the diagnosis and treatment of patients.

For improved research, prediction, and treatment of persons, medical specialists are needed. There are a variety of machine learning models available, each of which does prediction in a unique way. One of the most difficult and timeconsuming tasks is choosing a suitable machine learning algorithm .Navy Bayes is a pattern-based classifier that uses kernels to rank raw data, cluster it, and classify It. It builds a model that can classify a fresh dataset into a certain category based on a series of training examples. It's a non-parametric supervised learning method that's good for data with a lot of characteristics

The NB machine learning algorithm is used on brain pictures in this article. Dyslexia was diagnosed using MRI Functional Magnetic Resonance Imaging (FMRI) pictures. The activity of the cerebrum is estimated using these images. FMRI scans are used to track the brain's functioning by observing changes in blood oxygenation. It has been discovered that the control brain has a higher oxygen flow than a dyslexic brain. Positron Emission Tomography (PET),

Electroencephalography, and Computed Tomography are some of the other image acquisition techniques (CT) Storage and execution are major considerations when dealing with massive amounts of data. By adopting a distributed approach, these issues can be reduced to a minimum. Apache Spark is a lightweight open source big data analytics framework. It's an open source framework that's quick and easy to use.

## 2.LITERATURE SURVEY

This work critically analyzes recent machine learning methods for detecting dyslexia and its biomarkers and discusses challenges that require proper attentions from the users of deep learning methods in order to enable them to attain clinically relevance and acceptable level. The review is conducted within the premise of implementation and experimental outcomes for each of the 22 selected articles using the Preferred Reporting Items for Systematic review and MetaAnalyses (PRISMA) protocol, with a view to outlining some critical challenges for achieving high accuracy and reliability of the state-of-the-art machine learning methods. As an evidence-based protocol for reporting in systematic reviews and meta-analyses, PRISMA helps to ensure clarity and transparency of this paper by showing a four-phase flow diagram of the selection process for articles used in this review. It is therefore, envisaged that higher classification performance of clinical relevance can be achieved using deep learning models for dyslexia and its biomarkers by addressing identified potential challenge.

New diagnostic criteria for dyslexia and for specific learning disabilities on reading from DSM-5 no more considered discrepancy criteria (a discrepancy between IQ and scores on reading test), but the assessment of neuropsychological profile and cognitive profile by intelligence test and other neuropsychological tests of the child remains an important phase of clinical evaluation ((1), (2), (3)). Different studies analyzed performance on WISC test in children with reading disorders from different countries (for example in France and Portugal), using different versions of the instrument (Wisc-r, Wisc-III, Wisc-IV) ((4), (5), (6)...

Deep learning is being studied nowadays to overcome challenges in human health, and this paper proposes a pragmatic approach to leverage Deep learning Algorithms in teaching Numbers and Alphabets to Children. The Application, which was developed using the methodology described in this paper, introduces a smart way to have two-way interaction between the digital device and the student; the student gets Rewarded with points by drawing a similar Number or Alphabet as displayed on the screen using a blue, red-colored object; it also introduces different levels which seeds compulsion loop in the student. The Model harnesses the opensource OpenCV library to track blue, red coloured objects(Brobject) using Webcam and interprets the track produced using CNN to predict the A&N associated by implementing appropriate accuracy methodology. It utilizes Reward-based learning to trigger a compulsion loop in children, which could help make learning more interactive.

In this work, one of the first deep learning and machine learning based mobile applications, named “Pubudu” was developed for screening and intervention of dyslexia, dysgraphia and dyscalculia supporting local languages. In “Pubudu” we have followed up clinical screening and diagnostic procedures recommended by health professionals for screening and intervention. The screening of dyslexia, letter dysgraphia and numeric dysgraphia was carried out using deep neural network and the screening for dyscalculia was carried out using machine learning techniques. Intervention techniques are implemented using gamified environments. System testing was carried out using 50 differently abled children and 50 typical children. With the initial dataset 88%, 58%, 99% screening accuracies are achieved in neural networks for letter dysgraphia, dyslexia and numeric dysgraphia screening while dysgraphia, whereas 90% accuracy was achieved for dyscalculia. Handwritten letters and numbers were fed as inputs to CNN model in letter dysgraphia and numeric dysgraphia while embedded audio clips of letter pronunciation were fed in to voice recognition CNN model in dyslexia. “Pubudu” shows significant potential for screening and intervention of dyslexia, dysgraphia and dyscalculia.

## 3. METHODOLOGY

### 3.1. Architecture

The proposed architecture for dyslexia prediction is depicted in Fig. 1. The final section delves into the specifics of putting the proposed concept into action. Spark, ML, and NB all work in the same way.Apache Spark ML lib is a Spark-based distributed and scalable machine learning platform. It includes a number of machine learning algorithms and services that make large and difficult machine learning jobs easier to complete. The linear NB that was employed in this study is supported by ML lib. In general, pattern analysis algorithms are used to investigate many types of relationships in datasets, such as correlations and classifications.


Fig 1:Architecture diagram

The structures in photos must be converted into a two-dimensional vector. Each coordinate is given a name.

For image categorization in this study, linear NB is applied. On a training dataset with different classes, Linear NB looks for a hyper plane decision boundary to identify. The purpose of a hyper plane is to separate data points into different classes in a linear fashion.

Data can be divided into two categories using Linear NB. The hyper plane serves as a decision boundary, dividing the dataset into two sections: one for class '0' and the other for class '1.' Only datasets with two data classes are affected by this. We utilize a '0' for non-dyslexic and a '1' for dyslexic in this example. This study's picture dataset consists of a collection of colorful images with uniform pixel sizes. Before submitting to NB, the photos are transformed to grey scale.

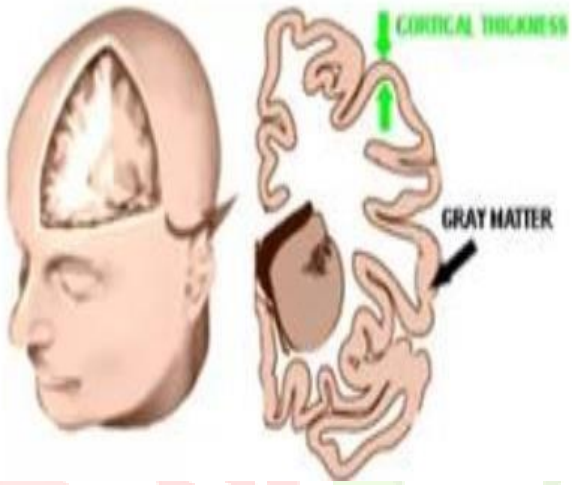



Fig 2:Sample Image

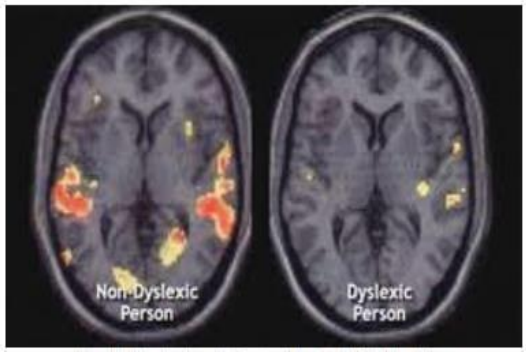



Fig3:Dyslexic and Control Brain While Reading

### 3.2 ALGORITHM

- **DECISION TREE**
- Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.
- The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data).
- In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

**Important Terminology related to Decision Trees**

1. Root Node: It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
2. Splitting: It is a process of dividing a node into two or more sub-nodes.
3. Decision Node: When a sub-node splits into further sub-nodes, then it is called the decision node.
4. Leaf / Terminal Node: Nodes do not split is called Leaf or Terminal node.
5. Pruning: When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
6. Branch / Sub-Tree: A subsection of the entire tree is called branch or sub-tree.
7. Parent and Child Node: A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.

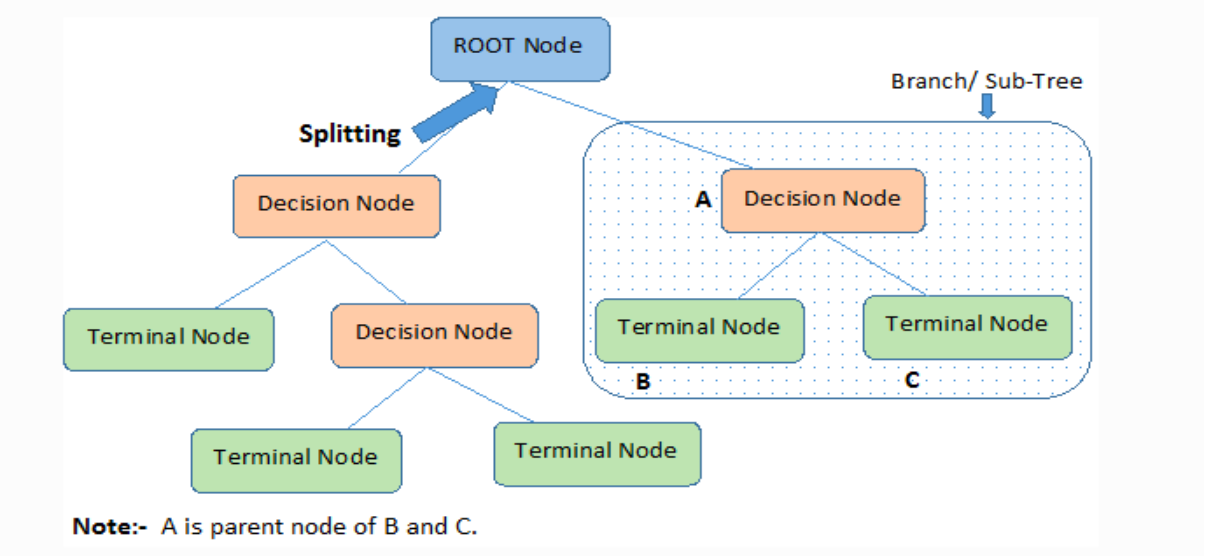



Fig 4:Decision Model

**Steps in ID3 algorithm:**

1. It begins with the original set S as the root node.
2. On each iteration of the algorithm, it iterates through the very unused attribute of the set S and calculates Entropy(H) and Information gain(IG) of this attribute.
3. It then selects the attribute which has the smallest Entropy or Largest Information gain.
4. The set S is then split by the selected attribute to produce a subset of the data.
5. The algorithm continues to recur on each subset, considering only attributes never selected before.

- **RANDOM FOREST**

- Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.
- One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems.

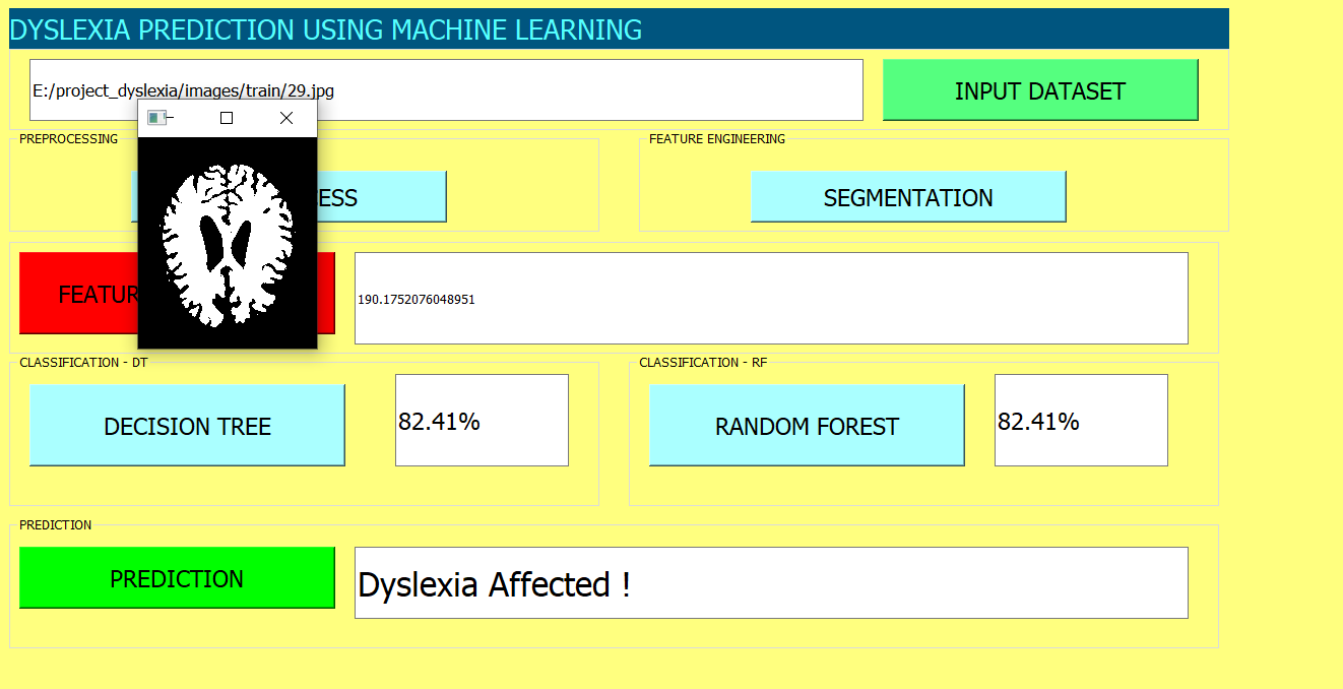
Steps involved in random forest algorithm:

- Step 1: In Random forest n number of random records are taken from the data set having k number of records.
- Step 2: Individual decision trees are constructed for each sample.
- Step 3: Each decision tree will generate an output.
- Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

## 4.EXPERIMENT OUTCOMES

NB is employed in this study to predict dyslexia using brain scans. White matter, grey matter, and cortical thickness are the features used in NB. This research project implements Apache SPARK, an in-memory framework that handles the storage, processing, and execution issues associated with large data sets. There were 150 brain MRI pictures in the sample, with ages ranging from 24 to 35. 50 of them have been diagnosed with dyslexia. Adult brain scans were chosen for the study because they would have progressed through the developmental reading stage, having been exposed to a variety of study materials and methodologies. K-fold cross validation is used as a validation model. Images are first transformed to grayscale for Dyslexia prediction. The scans are analyzed for three features: grey matter, white matter, and cortical thickness.

## 5.OUTPUT




## 6. CONCLUSION

The prediction of dyslexia in this study is based on brain scans. In a distributed or parallel approach, Linear NB from Apache Spark ML lib is utilized to build the prediction model. The accuracy reached was 92.5 percent with a high specificity, which is satisfactory. Medical professionals can use this categorization model to distinguish between dyslexics and non-dyslexics. Testing on a large dataset with a larger number of features can enhance the accuracy even more. The goal of this study's future work is to enhance accuracy by including various feature extraction methods and machine learning approaches. The prediction accuracy of 81.5% is achieved using Decision Tree and 97.6% using Random Forest Algorithm.This work concluded that the decision tree obtained 97.6%.

## REFERENCES

1.Advance Machine Learning Methods for Dyslexia Biomarker Detection: A Review of Implementation Details and ChallengeOpeyemi Lateef, UsmanRavie Chandren Muniyandi,,:IEEE Access( Volume: 926 February 2021 ,IEEE

2.Difficulties in reading and neuropsychological profile on WISC-IVin Italian childrenDonatella Rita Petretto,Paola Piras,Ilenia Pistis,Elena Tradori,Maria ValeriaCamboni,Federica Staico,Federica Palmas,Carla Lussu,Antonio Pre,,iCarmelo Masala., 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA),IEEE

3.An Interactive Alphabet and Number learning system using OpenCV and CNN ,2021 Asian Conference on Innovation in Technology (ASIANCON),27-29 Aug. 2021,IEEE

4.Deep Learning Based Screening And Intervention of Dyslexia, Dysgraphia And Dyscalculia,Ruchira Kariyawasam,Madhuka Nadeeshani,Tuan Hamid Inisha Subasinghe,2019 14th Conference on Industrial and Information Systems (ICIIS),18-20 Dec. 2019 ,IEEE

5.Learning Ecosystem to Overcome Sinhala Reading Weakness due to Dyslexia Lakchani Sandathara,Ransi Sathsarani, Shalini Tissera,Harini Hapuarachch,iSamantha Thelijjagoda,2020 2nd International Conference on Advancements in Computing (ICAC),10-11 Dec. IEEE , 26 February 2021

6.A Gamified Approach for Screening and Intervention of Dyslexia, Dysgraphia and DyscalculiaRuchira Kariyawasam;Madhuka Nadeeshani;Tuan Hamid;Inisha Subasinghe;Pasangi Ratnayake,2019 International Conference on Advancements in Computing (ICAC), 5-7 Dec.

7.An intelligent linguistic error detection approach to automated diagnosis of Dyslexia disorder in Persian speaking, IEEEFateme Asghari Tolami Mahsa KhorasaniMohsen Kahani Seyed Amir,2021 11th International Conference on Computer Engineering and Knowledge (ICCKE), 28-29 Oct. ,28 February 2022