Machine Learning Algorithm For Learning Disability Detection And Classifier System

B.Chakradhar[1], Vaishyaraju Nikitha[2], Vetukuri Joshika Siddani[3], Pentakota Uday Rishi[4], and Nagala Manikanta Swamy[5] ,
Department of Computer Science and Engineering, Raghu Instititute Of Technology, Dakamarri.

Abstract - The early and accurate detection of learning disabilities in children is crucial for effective intervention and support. This study presents a machine learningbased approach utilizing EEG data to classify learning disorders. The methodology involves collecting EEG readings from children with learning disabilities, preprocessing the data by handling missing values, normalizing features, and encoding labels. A deep learning model incorporating Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks is developed to capture temporal patterns in EEG signals. The dataset is split into training and testing sets (80%20%), and the model is trained using the Adam optimizer with cross-entropy loss for 20 epochs. Performance evaluation is conducted using confusion matrices and classification reports. To ensure accessibility, a Gradiobased web interface is deployed, allowing non-experts such as doctors and psychologists to analyze EEG data seamlessly. The proposed system demonstrates the potential of deep learning in enhancing the diagnostic process for learning disabilities such as dyslexia , Autism spectrum disorder(ASD) , Attention deficit hyperactivity disorder (ADHD).

Keywords-Learning Disabilities,EEG Signals, LSTM, RNN, Time-series analysis, Data preprocessing.

I. INTRODUCTION

Learning disabilities affect a significant number of children worldwide, impacting their academic performance and overall development. Early detection of these disabilities is crucial for timely intervention and personalized educational support. Traditional diagnosis methods rely on behavioral assessments, which can be subjective and timeconsuming. To address these challenges, we propose a machine learning-based system that utilizes EEG data to detect learning disabilities such as dyslexia, ADHD, and ASD with improved accuracy. Our system leverages deep learning techniques, particularly Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models, to analyze EEG patterns and identify abnormalities associated with learning disabilities. By applying data mining and pattern recognition, the system can extract meaningful insights from EEG signals, making early diagnosis more efficient and reliable. The model is integrated into a web-based interface using Gradio, allowing doctors and psychologists to easily input EEG data and receive instant classification results.This research builds upon existing studies in the field of EEG-based learning disability detection while addressing key limitations such as accuracy, scalability, and ease of use. The proposed system aims to assist educators, medical professionals, and researchers in identifying learning disabilities at an early stage, ultimately improving educational outcomes and cognitive support for affected individuals. In the following sections, we will detail the methodology, implementation, and performance evaluation of our system.

II. RELATED WORK

Learning disabilities have become a growing concern in educational and psychological research, requiring innovative solutions for early detection and intervention. Traditional diagnostic methods are often time-consuming and subjective, leading to delays in identifying affected individuals. Krunal V. Patel and Nikunjkumar Nayak [1] introduced a machine learning-based classifier system designed to enhance the detection and classification of learning disabilities. Their approach leverages data-driven analysis of cognitive and behavioral patterns, allowing for improved diagnosis. By implementing advanced pattern recognition techniques, their study emphasizes the potential of automated learning disability detection, reducing reliance on manual assessments.

The increasing prevalence of learning disabilities has led researchers to explore decision tree-based models for early detection. A. Devi, Dr. G. Kavya, et al.[2] proposed a diagnostic tool using Decision Tree algorithms to classify students with specific learning disabilities. Their research highlights the importance of structured classification models in identifying at-risk students based on cognitive performance indicators. The study underscores how machine learning can automate and enhance early intervention strategies, offering higher accuracy and reliability than traditional assessment methods.

The ability to predict reading disabilities at an early stage can significantly improve intervention outcomes. H. Atakan Varol, Subramani Mani,et al.[3] investigated the potential of machine learning models in reading disability prediction. Their study examined longitudinal student performance data, applying predictive analytics to identify early indicators of reading impairments. The findings suggest that early intervention guided by machine learning predictions can lead to improved literacy development, minimizing the long-term academic impact of reading disorders.

Hybrid machine learning models have emerged as a promising approach to improving learning disability detection. K. Ambili and P. Afsar [4]introduced a Naïve Bayes - Neural Network fusion technique, combining the strengths of probabilistic classification and deep learning models. Their study demonstrates that hybrid approaches outperform traditional methods by effectively analyzing complex learning disability patterns. The integration of multiple machine learning techniques enhances classification accuracy, making it a viable solution for early detection and educational support systems.

Feature selection plays a crucial role in optimizing machine learning models for learning disability detection. Julie M. David and Kannan Balakrishnan[5] proposed a method to extract the most relevant cognitive and behavioral attributes for classification. Their study compared various machine learning algorithms, highlighting the superior accuracy of Artificial Neural Networks (ANNs) in predicting learning disabilities. The findings suggest that ANN-based classification can significantly improve diagnostic precision, providing a more efficient and scalable approach to early detection.

The use of machine learning for dysgraphia detection has been explored through advanced handwriting analysis models. P. Drotar and M. Dobes[6] introduced a deep learning-based handwriting classification system to identify dysgraphia in children. Their study demonstrated that machine learning techniques could accurately distinguish between normal and dysgraphic handwriting patterns, enabling early diagnosis. The research underscores the potential of automated handwriting analysis as an effective method for detecting writing-related learning disabilities.

III. METHODOLOGY

A. Proposed system:

Multiple machine learning models have been explored for learning disability detection using EEG data, but challenges such as misclassification and data variability remain. Traditional methods rely on manual assessments, making diagnosis slow and subjective.To overcome these issues, we propose a deep learning-based system using RNN and LSTM for accurate classification of Dyslexia, ADHD, and ASD. Our system includes EEG data collection, preprocessing, model training, and classification, with results accessible through a Gradiobased web interface. By leveraging deep learning, the system improves classification accuracy, minimizes errors, and enables real-time diagnosis.

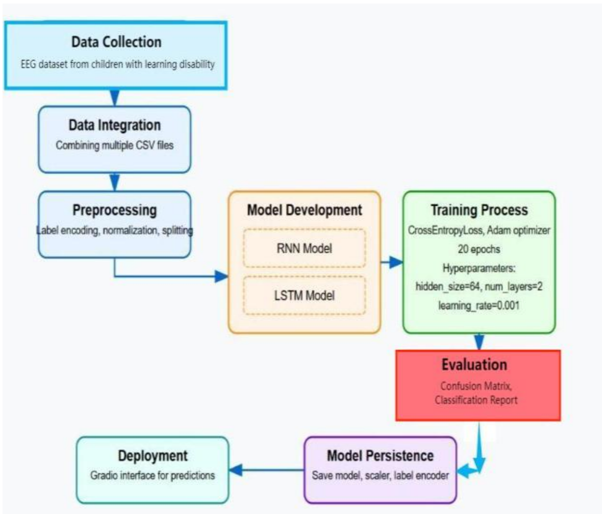


Fig1: Flow of the project

IV. IMPLEMENTATION

A. Data Collection and Preprocessing:

Data preprocessing is a crucial step in preparing EEG data for analysis. The collected EEG signals often contain noise, missing values, and inconsistencies that need to be addressed before training the model. The first step involves data cleaning, where any incomplete or corrupted data is removed. Next, feature scaling and normalization ensure that all EEG signals are within a standard range, allowing the deep learning models to learn effectively. Label encoding is applied to categorize the different types of learning disabilities, such as Dyslexia, ADHD, and ASD, making the data suitable for classification tasks.

B. Exploratory Data Analysis (EDA):

EDA is performed to gain insights into EEG data distributions and patterns. By applying statistical analysis and visualization techniques, we identify key relationships in the data. Histograms, scatter plots, and box plots help in understanding signal variations, while heatmaps visualize correlations between EEG features. EDA also helps detect any outliers or anomalies in the dataset, ensuring that the training data is well-structured for deep learning models.

C. Deep Learning Model Selection:

For effective classification of learning disabilities, we implemented Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models. These models are specifically designed for sequential data processing, making them suitable for analyzing EEG signals over time.

D. Recurrent Neural Network (RNN):

RNNs are designed to handle sequential data by maintaining a memory of previous inputs through recurrent connections. This makes them effective for learning temporal patterns in EEG signals. However, standard RNNs suffer from vanishing gradient issues, which make them struggle with learning long-term dependencies. In EEG-based learning disability detection, this means that RNNs may fail to retain important features from past EEG sequences, leading to lower classification accuracy.Suitable for shortterm pattern recognition in EEG signals.Can model sequential dependencies between brainwave patterns.Struggles with long-term dependencies due to vanishing gradients. Loses important EEG signal information over longer time sequences.

E. Long Short-Term Memory (LSTM):

LSTMs are an advanced version of RNNs designed to overcome vanishing gradient issues by incorporating memory cells and gating mechanisms. These gates regulate how much past information should be retained or forgotten, allowing LSTMs to effectively capture longterm dependencies in EEG signals. Forget Gate Determines which past information should be discarded. Input Gate Decides what new information should be stored in the memory cell. Output Gate Controls what information is passed to the next time step. LSTMs significantly improve classification accuracy in EEG-based learning disability detection because they can capture complex brain activity patterns over extended time periods. Compared to RNNs, LSTMs achieve higher precision and recall rates, making them more reliable for identifying learning disabilities like Dyslexia, ADHD, and ASD. Retains longterm dependencies in EEG sequences. More accurate in detecting complex learning disability patterns. LSTM achieves 86% accuracy, outperforming RNN due to its superior memory retention. More robust for handling large-scale EEG datasets with sequential dependencies.

F. Model Training and Evaluation :

The deep learning models are trained using an 80:20 train-test split, ensuring a balanced dataset for effective learning. The training process consists of 20 epochs, where the model continuously learns from the data. The Adam optimizer is used to fine-tune weights, while the cross-entropy loss function helps measure classification errors. The performance of the models is evaluated using Confusion Matrix Visualizes correct and incorrect predictions. Accuracy, Precision, Recall, and F1-Score: Provide insights into model effectiveness. Comparison of RNN and LSTM Models: The LSTM model achieves 86% accuracy, outperforming RNN due to its superior ability to capture long-term dependencies in EEG sequences.

V. Results and Discussions

The following images will visually depict the Training models of RNN and LSTM.

A. RNN Training Model:

The RNN training model, showing loss reduction over 20 epochs. nitially, the loss is higher at the start of the training due to random initialization of the model's weights. As training progresses over multiple epochs, the RNN model iterates over the training data, adjusting its weights and refining its internal representations of the input EEG signals. This iterative process leads to a gradual reduction in the loss, signifying that the model is learning to better understand and predict the temporal patterns in EEG data associated with learning disabilities such as ADHD, Dyslexia, and ASD. By the 20th epoch, the model's loss stabilizes at approximately 0.4221, indicating that the model has converged and is now achieving a more optimized balance.

```
Epoch [1/20], Loss: 1.5697
Epoch [2/20], Loss: 1.0484
Epoch [3/20], Loss: 0.8714
Epoch [4/20], Loss: 0.7639
Epoch [5/20], Loss: 0.6940
Epoch [6/20], Loss: 0.6432
Epoch [7/20], Loss: 0.6069
Epoch [8/20], Loss: 0.5791
Epoch [9/20], Loss: 0.5550
Epoch [10/20], Loss: 0.5364
Epoch [11/20], Loss: 0.5179
Epoch [12/20], Loss: 0.5029
Epoch [13/20], Loss: 0.4879
Epoch [14/20], Loss: 0.4744
Epoch [15/20], Loss: 0.4641
Epoch [16/20], Loss: 0.4547
Epoch [17/20], Loss: 0.4451
Epoch [18/20], Loss: 0.4367
Epoch [19/20], Loss: 0.4296
Epoch [20/20], Loss: 0.4221
Training Complete. Model saved.
```

Fig2: RNN Training Model

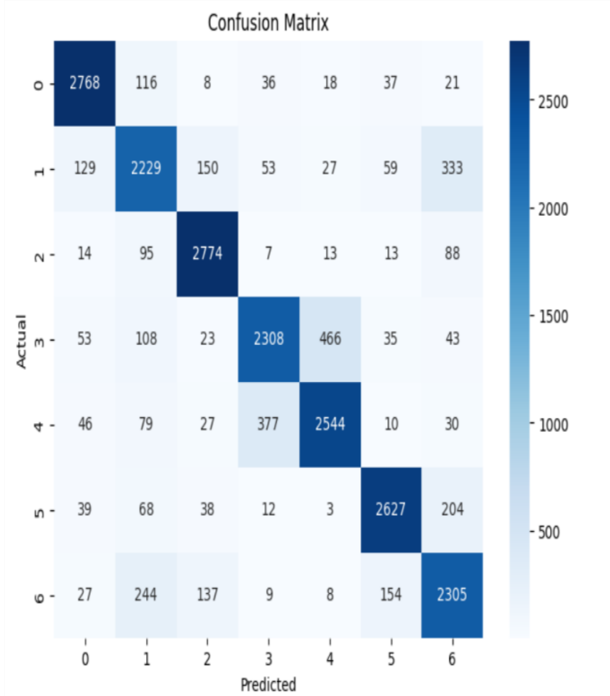
B. LSTM Training Model:

The LSTM training model, showing loss reduction over 20 epochs. The plot showing the reduction in loss over 20 epochs for the Long Short-Term Memory (LSTM) model highlights the effectiveness of the model's training process. At the start, similar to the RNN model, the LSTM experiences a higher loss due to the random initialization of weights and the absence of learned patterns. However, as training progresses, the model gradually adjusts its parameters, leading to a steady decrease in loss. This improvement reflects the model's increasing ability to learn from the data and fine-tune its performance over time.

```
Epoch [1/20], Loss: 1.2400
Epoch [2/20], Loss: 0.8204
Epoch [3/20], Loss: 0.6869
Epoch [4/20], Loss: 0.6127
Epoch [5/20], Loss: 0.5599
Epoch [6/20], Loss: 0.5209
Epoch [7/20], Loss: 0.4911
Epoch [8/20], Loss: 0.4658
Epoch [9/20], Loss: 0.4473
Epoch [10/20], Loss: 0.4293
Epoch [11/20], Loss: 0.4146
Epoch [12/20], Loss: 0.4019
Epoch [13/20], Loss: 0.3898
Epoch [14/20], Loss: 0.3799
Epoch [15/20], Loss: 0.3696
Epoch [16/20], Loss: 0.3597
Epoch [17/20], Loss: 0.3520
Epoch [18/20], Loss: 0.3429
Epoch [19/20], Loss: 0.3351
Epoch [20/20], Loss: 0.3275
Training Complete. Model saved.
```

Fig3: LSTM Training Model

C. Confusion Matrix and Classification Report:

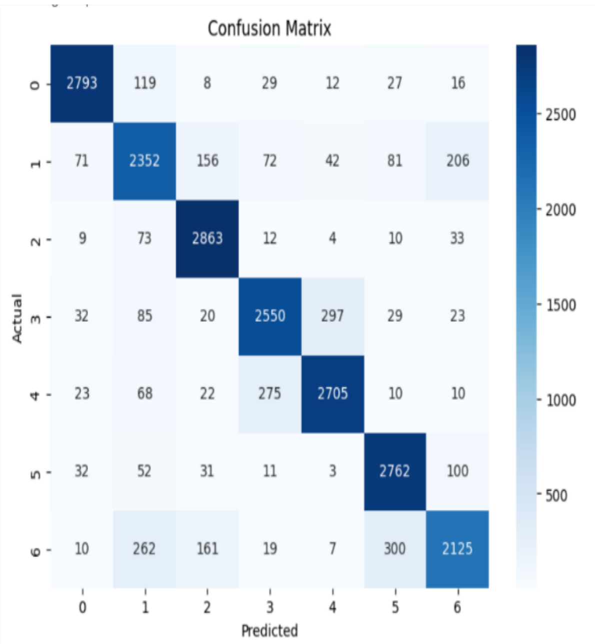


Classification Report:

	precision	recall	f1-score	support
0	0.90	0.92	0.91	3004
1	0.76	0.75	0.75	2980
2	0.88	0.92	0.90	3004
3	0.82	0.76	0.79	3036
4	0.83	0.82	0.82	3113
5	0.90	0.88	0.89	2991
6	0.76	0.80	0.78	2884
accuracy			0.84	21012
macro avg	0.83	0.84	0.83	21012
weighted avg	0.84	0.84	0.84	21012

Fig4: RNN Confusion Matrix and Classification Report

The RNN model demonstrates a strong predictive performance with an overall accuracy of 84%, indicating its effectiveness in classifying the data. The F1-score remains relatively consistent across all classes, ranging between 0.75 and 0.91, which suggests balanced performance. The precision and recall values further confirm that the model maintains a good balance, with only minor variations among different classes. Notably, classes 0, 2, and 5 exhibit the highest F1-scores of 0.91, 0.90, and 0.89, respectively, indicating the model's superior accuracy in detecting these categories. However, class 1 shows a slightly lower F1-score of 0.75, suggesting that the model faces more difficulty in accurately predicting samples from this class. Overall, the model performs well, with minimal misclassifications and consistent effectiveness across multiple classes.



Classification Report:

	precision	recall	f1-score	support
0	0.94	0.93	0.94	3004
1	0.78	0.79	0.79	2980
2	0.88	0.95	0.91	3004
3	0.86	0.84	0.85	3036
4	0.88	0.87	0.87	3113
5	0.86	0.92	0.89	2991
6	0.85	0.74	0.79	2884
accuracy			0.86	21012
macro avg	0.86	0.86	0.86	21012
weighted avg	0.86	0.86	0.86	21012

Fig5: LSTM Confusion Matrix and Classification Report

The LSTM model demonstrates strong classification performance with an overall accuracy of 86%, indicating its effectiveness in handling the given multi-class dataset. The F1-scores range between 0.79 and 0.95, reflecting consistent and reliable predictions across most classes. Notably, class 0, 2, and 5 exhibit the highest F1-scores of 0.94, 0.91, and 0.92, respectively, showcasing the model's superior accuracy in detecting these categories. The precision and recall values are balanced, highlighting the model's ability to correctly identify both positive and negative samples. However, class 1 shows slightly lower performance with an F1-score of 0.79, indicating moderate difficulty in accurately classifying this category. The confusion matrix reveals that while the model makes mostly accurate predictions, some misclassifications occur, particularly between class 1 and class 6, and class 6 and class 5, where a significant number of samples are incorrectly classified. Overall, the LSTM model exhibits robust performance with improved accuracy and consistency, making it effective for the multi-class classification task.

The classification reportpresents key performance metrics for each class, including **precision**, which measures the proportion of correctly predicted samples out of the total predicted; **recall**, which indicates the proportion of correctly predicted samples out of the total actual samples; and the **F1-score**, which balances precision and recall. It also includes **support**, representing the number of actual samples in each class.

D. User Interface:

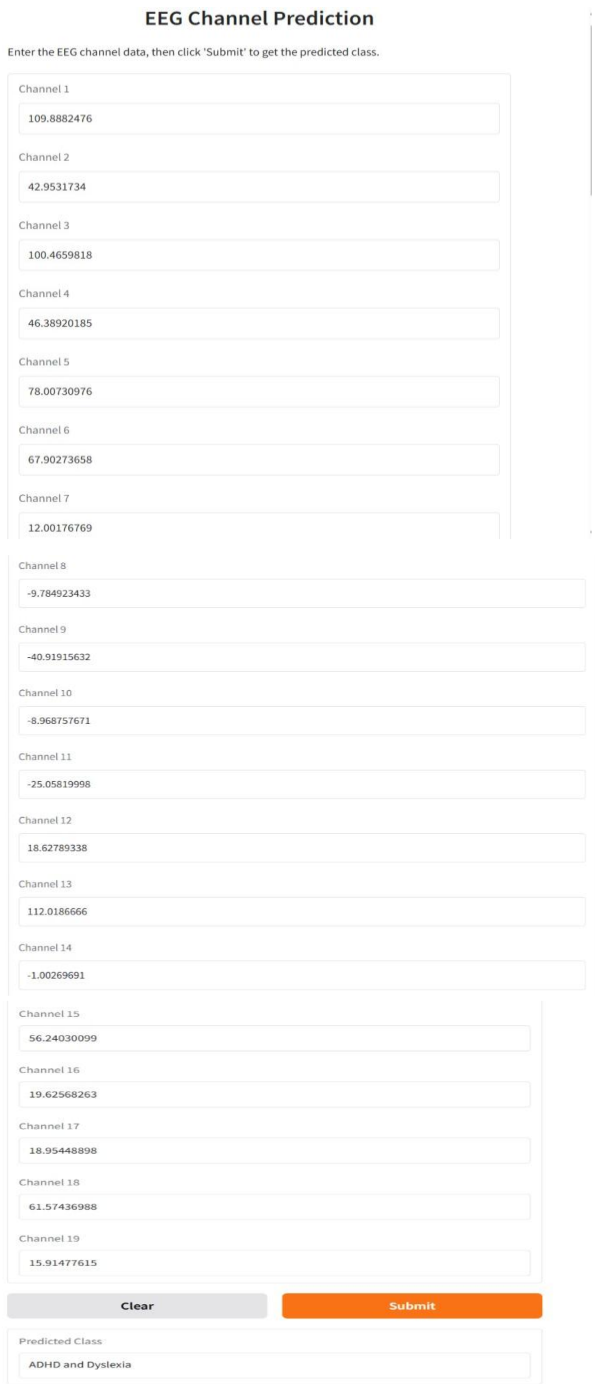


Fig6: User Interface

The User Interface (UI) of our project is designed to provide a simple and intuitive experience for EEG channel data prediction. It features individual input fields for each EEG channel, allowing users to enter the respective values accurately. Once the data is provided, the user can click the "Submit" button to obtain the predicted class. The interface also includes a "Clear" button, enabling users to reset the form easily. Upon submission, the predicted class is displayed below the buttons, offering a clear and concise output. The UI ensures a smooth and user-friendly interaction, making it convenient to classify EEG data efficiently.

The interface offers responsive feedback, ensuring smooth interaction. The use of an orange-colored submit button and a contrasting clear button makes the UI visually distinct and easy to navigate. Overall, the design ensures efficient data entry, clear output display, and a seamless user experience.

VI. Conclusion

The effectiveness of machine learning techniques in detecting learning disabilities using EEG data. Through extensive data analysis, we identified significant patterns in brain activity that correlate with learning disorders such as Dyslexia, ADHD, and ASD. The findings highlight the potential of deep learning models, particularly RNN and LSTM, in accurately classifying learning disabilities, enabling early diagnosis and intervention. Our analysis revealed key behavioral and cognitive indicators that contribute to learning disabilities, emphasizing the importance of EEG-based detection methods over traditional psychological assessments. Additionally, the Gradio-based web interface ensures accessibility for educators, psychologists, and medical professionals, facilitating seamless real-time analysis and classification. However, certain limitations exist in this research. The study primarily focuses on EEG-based learning disability detection, and additional multimodal data sources, such as speech and handwriting patterns, could further enhance classification accuracy. Future work should explore hybrid machine learning models combining CNN, LSTM, and transformer-based architectures to improve predictive capabilities. Ultimately, by integrating advanced deep learning techniques with real-time detection tools, this system contributes to the early diagnosis of learning disabilities, supporting personalized learning strategies and educational interventions. This research lays the groundwork for further advancements in AI-driven learning disability detection, promoting better educational outcomes and cognitive support for affected individuals.

VII. References

[1] Krunal V. Patel, Nikunjkumar Nayak
The title of the study is *"Machine Learning Algorithm for Learning Disability Detection and Classifier System."* Published in J. Electrical Systems, 20-10s (2024): 1086-1092.

[2] A.Devi, Dr. G. Kavya, M. Julie Therese, R. Gayathri
The title of the study is *"Early Diagnosing and Identifying Tool for Specific Learning Disability using Decision Tree Algorithm."* Presented at the Third International Conference on Inventive Research in Computing Applications (ICIRCA2021).

[3] H. Atakan Varol, Subramani Mani, Donald L. Compton, Lynn S. Fuchs, and Douglas Fuchs
The title of the study is *"Early Prediction of Reading Disability using Machine Learning."* Published in AMIA Annu Symp Proc., 2009: 667–671.

[4] K.Ambili and P.Afsar
The title of the study is *"A Framework for Learning Disability Prediction in School Children using Naïve Bayes - Neural Network Fusion Technique."* Published in ISSN: 0975 – 6760, Vol 4, Issue 1 (2016).

[5] Julie M. David and Kannan Balakrishnan
The title of the study is *"Machine Learning Approach for Prediction of Learning Disabilities in School-Age Children."* Published in International Journal of Computer Applications, 9(2010): 10.5120/1432-1931.

[6] P.Drotar, M.Dobes
The title of the study is *"Dysgraphia Detection Through Machine Learning."* Published in Sci Rep., 10, 21541 (2020).

[7] Pooja Manghirmalani Mishra and Dr. Sushil Kulkarni
The title of the study is *"Classification of Data using Semi-Supervised Learning (A Learning Disability Case Study)."* Published in IJCET, Vol 4, Issue 4 (2013): 432-440.

[8] Khan, Rehman, Julia Lee Ai Cheng, Yin Bee Oon The title of the study is *"Machine Learning and Dyslexia: Diagnostic and Classification System (DCS) for Kids with Learning Disabilities."* Published in 2018.

[9] M.Mahalakshmi, Dr.K.Merriliance The title of the study is *"Prediction of Dyslexia using Machine Learning Algorithms."* Published in IJCRT, Vol 10, Issue 5 (2022): ISSN 2320-2882.

[10] Vanitha and M.Kasthuri The title of the study is *"Dyslexia Prediction Using Machine Learning Algorithms – A Review."* Published in International Journal of Aquatic Science, Vol 12, Issue 2 (2021): ISSN 2008-8019.