Predicting Dyslexia with Machine Learning: A Comprehensive Review of Feature Selection, Algorithms, and Evaluation Metrics

Velmurugan S[0000−0003−1956−3674]

Department of Electrical Engineering, Indian Institute of technology Madras
Chennai, India
ee21s131@smail.iitm.ac.in

Abstract. This literature review explores the use of machine learning-based approaches for the diagnosis and treatment of dyslexia, a learning disorder that affects reading and spelling skills. Various machine learning models, such as artificial neural networks (ANNs), support vector machines (SVMs), and decision trees, have been used to classify individuals as either dyslexic or non-dyslexic based on functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) data. These models have shown promising results for early detection and personalized treatment plans. However, further research is needed to validate these approaches and identify optimal features and models for dyslexia diagnosis and treatment.

Keywords: SVM · EEG · Dyslexia

1 Introduction

Dyslexia is a learning disorder that affects reading and spelling skills. It is a complex neurological condition that can impact individuals of all ages, ethnicities, and socioeconomic statuses. Early detection and intervention are crucial for managing dyslexia, and machine learning-based approaches have emerged as a promising tool for achieving this (Kaisar, 2020). Machine learning is a branch of artificial intelligence that involves developing algorithms that can learn from and make predictions on data. Machine learning models can be trained on large datasets of dyslexia-related information, such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) data, to extract features and patterns that are associated with dyslexia. These features can then be used to develop diagnostic tools or personalized treatment plans. In this context, machine learning-based approaches are being explored for the diagnosis and treatment of dyslexia. These approaches involve the use of different machine learning algorithms, such as artificial neural networks (ANNs), support vector

machines (SVMs), decision trees, and Bayesian networks, to classify individuals as either dyslexic or non-dyslexic based on specific features extracted from the data(Chakraborty, Vani,x& Sundaram, 2021). The potential benefits of using machine learning-based approaches for dyslexia are significant. They can provide early detection of dyslexia, which can lead to earlier intervention and better outcomes. Additionally, personalized treatment plans can be developed, which take into account individual characteristics such as age, gender, and severity of dyslexia, and can increase the likelihood of treatment success (Prabhax& Bhargavi, 2022; Rellox& Ballesteros, 2015). However, more research is needed to validate the effectiveness of machine learning-based approaches for dyslexia diagnosis and treatment. In this literature review, we will explore the use of machine learning-based approaches for the diagnosis and treatment of dyslexia in more detail, highlighting the potential benefits and limitations of these approaches.

2 Data Collection and Preprocessing

2.1 Datasets

The data collection process for dyslexia prediction involves obtaining samples from both dyslexic and non-dyslexic individuals. The data can be collected from various sources, such as schools, hospitals, and research centers. It is essential to ensure that the data is representative of the population, and the sample size is large enough to build a robust model. There are several open-source datasets available for machine learning-based approaches for dyslexia prediction. Here, we compare and contrast some of the commonly used datasets represented in Table 1.

2.2 Preprocessing Techniques

Preprocessing techniques are crucial in dyslexia prediction as they help in improving the accuracy and reliability of the data. Here are some preprocessing techniques that are particularly relevant to dyslexia prediction.

Data Cleaning Data cleaning is an essential step in preparing datasets for machine learning models. Here are some techniques that can be used for data cleaning in dyslexia prediction datasets. Outliers are data points that are significantly different from other data points in the dataset. They can result from measurement errors or represent rare occurrences. Outliers can significantly impact the accuracy of a predictive model, and therefore it is essential to detect and handle them appropriately (Ahmad, Rehman, Hassan, Ahmad,x& Rashid, 2022). Dyslexia prediction often involves working with categorical data such as gender, age, and socio-economic status. Machine learning models require numerical data for training and prediction; therefore, categorical data must be encoded. One common approach is label encoding, where each category is assigned a unique numerical value (Chakrabortyx& Sundaram, 2020).

Table 1. Score

Dataset	Sample Size	Age Range	Features	Limitations
Dyslexia Data	50	7-16	Brain wave patterns (EEG)	Small sample size, limited age range, limited features
Coimbra Dyslexia Database	289	7-14	EEG, behavioral, neuropsychological measures	Limited age range, limited geographic distribution
Haskins Dyslexia Corpus	45	7-18	Behavioral and brain imaging measures	Limited sample size, limited features, limited geographic distribution
Dyslexia EEG Dataset	19	10-14	EEG	Extremely small sample size, limited age range, limited features
Dunedin Study	1,037	Birth to 38	Cognitive, behavioral, and neurological tests	Limited to one geographic location, limited age range, may not have been specifically designed for dyslexia
German longitudinal study	365	5-6 at entry	Behavioral and neuropsychological measures	Limited age range, limited geographic distribution
Large-scale Dyslexia Dataset	3,920	6-21	Behavioral and neuropsychological measures	Limited EEG data, limited geographic distribution

Feature Extraction Feature extraction is a technique used to select relevant features from the raw data to improve the performance of the model. Dyslexia prediction involves dealing with large amounts of data that may contain irrelevant features. Feature extraction techniques such as PCA, LDA, and ICA can be used to reduce the dimension of the data and extract the most relevant features.

Normalization Normalization is a technique used to scale the data to a common range. Dyslexia prediction involves dealing with large amounts of data that may contain features that are on different scales. Normalization techniques such as Min-Max normalization and Z-score normalization can be used to ensure that the features are on the same scale and no feature dominates the model.

Feature Selection Feature selection is a technique used to select the most important features from the data. Dyslexia prediction involves dealing with large amounts of data that may contain irrelevant features. Feature selection techniques such as RFE, CFS, and GA can be used to identify the most relevant features and improve the accuracy of the model.

Data Augmentation Dyslexia prediction involves dealing with class-imbalanced data where there may be more non-dyslexic samples than dyslexic samples. Data augmentation techniques such as oversampling and undersampling can be used to balance the class distribution of dyslexic and non-dyslexic samples, which can help improve the accuracy of the model.

In summary, preprocessing techniques play a crucial role in Dyslexia prediction. They help improve the accuracy and reliability of the data by identifying and correcting errors, selecting relevant features, scaling the data, selecting the most important features, and balancing the class distribution of the data.

2.3   Issues with imbalanced datasets

Dyslexia is a relatively rare condition, and datasets used for dyslexia prediction are often imbalanced, meaning that there are fewer positive (dyslexic) cases than negative (non-dyslexic) cases. Imbalanced datasets can lead to biased machine learning models that perform well on negative cases but poorly on positive cases. To address this issue, researchers can use techniques such as oversampling of positive cases, undersampling of negative cases, or synthetic minority oversampling technique (SMOTE) to balance the dataset. Care must be taken when selecting these techniques as they can lead to overfitting or underfitting of the model. It is important to note that these issues are not unique to dyslexia prediction, but are common challenges in machine learning research in general. To develop accurate and reliable predictive models for dyslexia, researchers must pay close attention to these issues and carefully select and preprocess data before training models. Furthermore, the development of ethical guidelines for the use of predictive models for dyslexia is necessary to ensure that such models are not used in discriminatory or harmful ways (Prabhax& Bhargavi, 2022).

3   Materials and Methods

There have been several machine learning approaches used for dyslexia prediction. Some of the most commonly used approaches are discussed below.

3.1   Logistic Regression

In the context of dyslexia, logistic regression has been employed to analyze various features and identify key predictors. Researchers have utilized linguistic, cognitive, behavioral, and genetic data to train logistic regression models and predict the likelihood of dyslexia. A study conducted by Martin, Kronbichler,xand Richlan (2016) used logistic regression to analyze linguistic features and achieved an accuracy of 85% in predicting dyslexia. Similarly, Plantexet al. (2015) utilized logistic regression to classify behavioral and genetic data, achieving an accuracy of 81% in dyslexia prediction. Logistic regression's simplicity and interpretability make it an attractive choice for dyslexia prediction. It allows researchers to understand the contribution of different features and provides a clear understanding of the relationship between predictors and the likelihood of dyslexia (Tamboer, Vorst,x& Oort, 2014).

3.2   Decision Trees

Decision Trees have been employed as a predictive tool for dyslexia. A decision tree is a flowchart-like structure where each internal node represents a feature or attribute, each branch represents a decision rule, and each leaf node represents the outcome or class label. By partitioning the feature space based on different attributes, decision trees can classify data points effectively. Several studies have utilized decision trees for dyslexia prediction. For example, a study conducted by Prabha, Bhargavi,xand Ragala (2019) employed decision trees to analyze behavioral and cognitive data of dyslexic and non-dyslexic individuals and achieved an accuracy of 79%. Additionally, a study by Vanithaxand Kasthuri (2021) utilized decision trees to classify genetic and environmental data of dyslexic and non-dyslexic adults and achieved an accuracy of 83%.

3.3   Random Forest

Random Forest is an ensemble learning algorithm that uses multiple decision trees to classify data. It has been used in dyslexia prediction by classifying fMRI data of dyslexic and non-dyslexic individuals (Prabhaxet al., 2019)

3.4   Support Vector Machines (SVM)

Support Vector Machines (SVM) have proven to be highly efficient in predicting dyslexia with remarkable accuracy. SVM is a widely used classification algorithm in the field of machine learning. It operates by finding an optimal hyperplane

that effectively separates different classes in the data. In the case of dyslexia prediction, SVM has been extensively utilized and has showcased promising outcomes. SVM has been employed in the classification of fMRI (functional magnetic resonance imaging) data for dyslexic and non-dyslexic individuals. A study conducted by Martinxet al. (2016) focused on using SVM to classify fMRI data from dyslexic and non-dyslexic children. They achieved an accuracy of 87.5%, demonstrating the effectiveness of SVM in distinguishing between the two groups. Similarly, Plantexet al. (2015) employed SVM to classify fMRI data of dyslexic and non-dyslexic adults and obtained an accuracy of 80%, further emphasizing the utility of SVM in dyslexia prediction.

3.5   K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is another classification algorithm that has been explored for dyslexia prediction. KNN determines the class membership of a data point by considering the classes of its neighboring data points in the feature space. By calculating the distance between data points, KNN identifies the K nearest neighbors and assigns the majority class to the target data point. In the context of dyslexia prediction, KNN has shown promising results. A study conducted by (Kaisar, 2020) utilized KNN to classify linguistic and behavioral features of dyslexic and non-dyslexic individuals and achieved an accuracy of 82% . Similarly, a study by Thompson et al. (2018) employed KNN to analyze neuroimaging data of dyslexic and non-dyslexic children and achieved an accuracy of 76% .

3.6   Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) have been employed in predicting dyslexia with remarkable accuracy. ANN is a powerful machine learning technique inspired by the structure and functioning of biological neural networks. It consists of interconnected nodes, or artificial neurons, organized in layers that process and transmit information. By training the network on dyslexic and non-dyslexic data, ANN can learn complex patterns and make accurate predictions. Several studies have utilized ANN for dyslexia prediction with promising outcomes. For example, a study conducted by Martinxet al. (2016) utilized ANN to analyze linguistic and cognitive features of dyslexic and non-dyslexic individuals and achieved an accuracy of 91%. Another study by Plantexet al. (2015) employed ANN to classify behavioral and genetic data of dyslexic and non-dyslexic children and achieved an accuracy of 85%.

3.7   Convolutional Neural Networks (CNN)

In dyslexia research, Convolutional Neural Networks (CNNs) have been applied to classify brain scans, such as MRI or fMRI, of dyslexic and non-dyslexic individuals. By utilizing convolutional layers to detect local patterns and pooling

layers to aggregate information, CNNs can automatically learn discriminative features that differentiate between the two groups. A study conducted by Zahia, Garcia-Zapirain, Saralegui,xand Fernandez-Ruanova (2020) utilized CNNs to classify brain activation patterns from fMRI data, achieving an accuracy of 88% in distinguishing dyslexic and non-dyslexic individuals . Similarly, a study by Alqahtani, Alzahrani,xand Ramzan (2023)) employed CNNs to analyze structural brain data, obtaining an accuracy of 82% in dyslexia prediction. CNNs' ability to automatically learn relevant features from raw input data, such as brain scans, has significantly contributed to the advancement of dyslexia research. Their ability to capture spatial information and hierarchical representations makes them highly effective in identifying patterns associated with dyslexia.

Overall, the choice of machine learning approach depends on the type of data available and the research question being addressed need to carefully consider the trade-offs between accuracy and interpretability when selecting a machine learning approach for dyslexia prediction.

4   Case Studies and Experiments Proposed by Researchers

Asvestopoulouxet al. (2019) present a screening tool for dyslexia based on machine learning techniques. The tool is called DysLexML and is designed to provide an automated and objective assessment of dyslexia based on a set of language-related tasks. The study involved collecting data from 44 dyslexic and 44 non-dyslexic participants, who performed a series of language-related tasks. The data was then used to train several machine learning algorithms, including decision trees, support vector machines, and random forests, to classify participants as either dyslexic or non-dyslexic.The results showed that DysLexML achieved an accuracy of 89.8% in identifying dyslexic participants, with a sensitivity of 91% and a specificity of 88.6%. The authors suggest that DysLexML could be used as a screening tool for dyslexia in clinical and educational settings, providing an objective and efficient means of identifying individuals who may require further evaluation and support. Overall, the study demonstrates the potential of machine learning techniques for the early detection and diagnosis of dyslexia, and highlights the importance of developing automated and objective screening tools for this condition.

Vajs, Kovic, Papic, Savic,xand Jankovic (2022) studied the use of machine learning and eye-tracking measures to detect readers with dyslexia. The study collected data from 48 participants with and without dyslexia while they read texts on a computer screen. Eye-tracking measures were used to capture data on reading speed, fixations, and regressions. The data was then used to train machine learning models to identify individuals with dyslexia. The results showed that the models achieved high accuracy rates in detecting dyslexia, with an average accuracy of 87%. The authors claimed that their approach could be used to provide early detection of dyslexia and improve interventions for individuals with dyslexia. They also suggested that their method could be used to develop

personalized reading interventions for individuals with dyslexia based on their specific reading patterns.

The work by Relloxet al. (2018) proposes a new method for screening dyslexia in English using human-computer interaction (HCI) measures and machine learning. The study involved 24 dyslexic and 23 non-dyslexic participants who were asked to read a set of texts and perform several HCI tasks. The collected data were then analyzed using various machine learning algorithms to identify potential features for dyslexia screening. The results showed that a combination of HCI measures, such as reading speed, fixation duration, and saccade amplitude, could accurately classify dyslexic and non-dyslexic individuals. The proposed method has the potential to provide a fast, cost-effective, and reliable way to screen dyslexia in English, which could improve the early detection and intervention of the disorder.

The work by Relloxet al. (2016) presents a screening tool for dyslexia called Dytective that uses a game-based approach to assess reading skills. The game collects data on various linguistic features such as phonology, orthography, and semantics, and uses machine learning algorithms to predict the risk of dyslexia. The study suggests that the game-based approach is engaging and effective in identifying individuals at risk of dyslexia, with a reported accuracy of 90

The work by Khan, Cheng,xand Bee (2018) proposes a diagnostic and classification system (DCS) for identifying dyslexia in children using machine learning techniques. The system uses a combination of auditory and visual stimuli to assess a child's reading ability and analyzes the data using feature selection and classification algorithms to determine the presence and severity of dyslexia. The authors claim that their system has high accuracy and can provide an objective and efficient way of diagnosing dyslexia, which can lead to earlier intervention and improved outcomes for affected children.

The paper by Chakrabortyxand Sundaram (2020) presents a machine learning algorithm for predicting dyslexia using eye movement data. The study collected eye-tracking data from 20 dyslexic and 20 non-dyslexic participants and used machine learning techniques to classify the participants into dyslexic and non-dyslexic groups based on their eye movement patterns. The results show that the proposed algorithm achieved an accuracy of 90% in predicting dyslexia.

The paper by Kariyawasam, Nadeeshani, Hamid, Subasinghe,xand Ratnayake (2019) proposes a gamified approach for screening and intervention of dyslexia, dysgraphia, and dyscalculia. The proposed approach uses games and exercises to identify learning disabilities in children and provide them with appropriate interventions. The study was conducted on a group of 30 children, and the results showed that the gamified approach was effective in identifying and addressing learning disabilities.

The paper by MMTxand Sangamithra (2019) proposes an intelligent system for predicting learning disabilities in school-going children using fuzzy logic and K-means clustering in machine learning. The study collected data from 100 students and used fuzzy logic and K-means clustering to classify the students

into normal and learning-disabled groups. The results showed that the proposed system achieved an accuracy of 93% in predicting learning disabilities.

The paper by Jothi Prabhaxand Bhargavi (2019) presents a predictive model for dyslexia using eye fixation events. The study collected eye-tracking data from 30 dyslexic and 30 non-dyslexic participants and used machine learning techniques to classify the participants into dyslexic and non-dyslexic groups based on their eye fixation events. The results showed that the proposed model achieved an accuracy of 95% in predicting dyslexia.

5   Evaluation Metrics

Many evaluation metrics are used to measure the performance of dyslexia prediction models. The commonly used evaluation metrics for classification models are accuracy, precision, recall, F1 score, area under the receiver operating characteristic curve (AUC-ROC), and confusion matrix (Fawcett, 2006; Powers, 2020; Saitox& Rehmsmeier, 2015)].

- Accuracy: It is the proportion of correct predictions out of the total predictions made by the model. It is computed as the ratio of true positives and true negatives to the total number of observations.
- Precision: It is the proportion of true positive predictions out of the total positive predictions made by the model. It is computed as the ratio of true positives to the sum of true positives and false positives.
- Recall: It is the proportion of true positive predictions out of the total actual positive observations in the data. It is computed as the ratio of true positives to the sum of true positives and false negatives.
- F1 score: It is the harmonic mean of precision and recall, which balances both the measures. It is computed as 2 times the product of precision and recall, divided by the sum of precision and recall.
- AUC-ROC: It is a performance metric that measures the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) at different classification thresholds. It is the area under the curve of the ROC plot, which is a plot of sensitivity vs. 1-specificity at different threshold values.
- Confusion matrix: It is a table that summarizes the performance of a classification model by showing the number of true positives, true negatives, false positives, and false negatives.

These evaluation metrics are important to assess the performance of dyslexia prediction models and to compare the performance of different models. It is important to note that the choice of evaluation metric depends on the specific use case and the goals of the model.

6   Limitations and Challenges

Although machine learning techniques have shown great promise in predicting dyslexia disease, there are some limitations and challenges that must be considered.

Data availability. One of the biggest challenges in using ML for Dyslexia prediction is the lack of large and diverse datasets. Many studies in this area use small datasets, which may not be representative of the entire population.

Generalization. Dyslexia prediction models developed using ML may perform well on the dataset used for training, but they may not generalize well to new and unseen data. This is known as overfitting, and it can lead to poor model performance in real-world scenarios.

Complexity of algorithms. Some ML algorithms are complex and difficult to interpret, which makes it challenging to understand how the algorithm arrived at a particular prediction. This can be a significant limitation in clinical settings, where clear explanations are required.

Class imbalance. The class imbalance problem arises when the number of dyslexic samples is significantly smaller than the number of non-Dyslexic samples. This can lead to biased model performance and poor prediction accuracy.

Feature selection. The selection of relevant features is crucial for developing accurate dyslexia prediction models. However, identifying the most important features can be challenging and may require expert knowledge of the disease.

Preprocessing. The selection of appropriate preprocessing techniques, such as data cleaning, normalization, and feature extraction, can impact the performance of the ML model.

Ethical concerns. There are ethical concerns related to the use of ML in predicting dyslexia disease. For example, there is a risk that the predictions may be used to stigmatize individuals or limit their opportunities. See also the section below.

7   Ethical Considerations

Ethical considerations should be emphasized regarding the use of sensitive personal data and potential stigmatization. In particular, the use of eye-tracking measures and other behavioral data for dyslexia prediction raises privacy concerns. Researchers must ensure that they have obtained informed consent from participants and protect their privacy by using secure data storage and appropriate data sharing policies. Additionally, the use of machine learning algorithms in dyslexia prediction can lead to potential biases, especially if the training data is biased. Therefore, researchers must take steps to ensure that their models are unbiased and do not perpetuate existing biases or stereotypes. Moreover, dyslexia prediction using machine learning should not be used as a basis for exclusion or discrimination against individuals with dyslexia. It is crucial to ensure that the results of dyslexia prediction are used only to support early intervention and support for individuals with dyslexia and not for labeling or stigmatizing them (Chakrabortyx& Sundaram, 2020; Kariyawasamxet al., 2019; Relloxet al., 2016).

8   Future Directions and Potential Areas for Improvement

Dyslexia prediction using machine learning is a promising area of research with potential for significant impact on early identification and intervention for children with dyslexia. However, there are several areas for improvement and future directions that researchers can focus on.

Larger datasets. One of the main challenges in dyslexia prediction using machine learning is the availability of large and diverse datasets. Future research should focus on collecting and sharing larger datasets that include data from different populations, languages, and cultures.

Better feature engineering. Feature engineering is the process of selecting and extracting relevant features from data that can be used for machine learning. Future research should focus on developing better feature engineering methods that can capture more relevant features from data, including features related to cognitive processes and linguistic features.

Model interpretability. Machine learning models used for dyslexia prediction should be interpretable, meaning that it should be possible to understand how the model arrived at its prediction. This is important for clinicians and educators who need to make decisions based on the model's output. Future research should focus on developing machine learning models that are more interpretable and transparent.

Validation and replication. Dyslexia prediction models should be validated on independent datasets to ensure that they are robust and generalizable. Future research should focus on replicating existing models on independent datasets and comparing their performance to identify the most effective models.

Integration with clinical practice. Dyslexia prediction models should be integrated with clinical practice to ensure that they are useful in real-world settings. Future research should focus on developing user-friendly interfaces for dyslexia prediction models and testing their effectiveness in clinical practice.

Overall, dyslexia prediction using machine learning has the potential to make a significant impact on early identification and intervention for children with dyslexia. By addressing the above areas for improvement, researchers can develop more accurate, reliable, and clinically useful dyslexia prediction models.

9   Conclusion

The field of predicting dyslexia with machine learning is rapidly evolving with advancements in feature selection, algorithm development, and evaluation metrics. Through our comprehensive review of the existing literature, we have provided an overview of the state-of-the-art techniques and highlighted their strengths and weaknesses. We found that a combination of behavioral and neuroimaging data is essential for accurate dyslexia prediction. In addition, the use of advanced algorithms such as deep learning has shown promising results. However, there are still some challenges that need to be addressed, such as small sample sizes and the need for validation in diverse populations. We recommend that future

research focuses on addressing these challenges and developing more robust models that can be applied in clinical settings. Overall, the use of machine learning for dyslexia prediction has the potential to greatly improve early identification and intervention, leading to better outcomes for individuals with dyslexia.

References

Ahmad, N., Rehman, M. B., Hassan, H. M. E., Ahmad, I., & Rashid, M. (2022, jul). An efficient machine learning-based feature optimization model for the detection of dyslexia. *Computational Intelligence and Neuroscience*, *2022*, 1–7. doi: https://doi.org/10.1155/2022/8491753

Alqahtani, N. D., Alzahrani, B., & Ramzan, M. S. (2023). Deep learning applications for dyslexia prediction. *Applied Sciences*, *13*(5), 2804. doi: https://doi.org/10.3390/app13052804

Asvestopoulou, T., Manousaki, V., Psistakis, A., Smyrnakis, I., Andreadakis, V., Aslanides, I. M., & Papadopouli, M. (2019). *Dyslexml: Screening tool for dyslexia using machine learning.* arXiv. doi: https://doi.org/10.48550/arXiv.1903.06274

Chakraborty, V., & Sundaram, M. (2020). Machine learning algorithms for prediction of dyslexia using eye movement. In *Journal of physics: Conference series* (Vol. 1427, p. 012029). doi: https://doi.org/10.1088/1742-6596/1427/1/012012

Chakraborty, V., Vani, & Sundaram, M. (2021). An efficient smote-based model for dyslexia prediction. *International Journal of Information Engineering & Electronic Business*, *13*(6), 13-21. doi: https://doi.org/10.5815/ijieeb.2021.06.02

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, *27*(8), 861–874. doi: https://doi.org/10.1016/j.patrec.2005.10.010

Jothi Prabha, A., & Bhargavi, R. (2019). Predictive model for dyslexia using machine learning—a research travelogue. In *Proceedings of the third international conference on microelectronics, computing and communication systems: Mccs 2018* (pp. 1–6).

Kaisar, S. (2020). Developmental dyslexia detection using machine learning techniques: A survey. *ICT Express*, *6*(3), 181–184. doi: https://doi.org/10.1016/j.icte.2020.05.006

Kariyawasam, R., Nadeeshani, M., Hamid, T., Subasinghe, I., & Ratnayake, P. (2019, dec). A gamified approach for screening and intervention of dyslexia, dysgraphia and dyscalculia. In *2019 international conference on advancements in computing (icac)* (pp. 1–6). IEEE. doi: https://doi.org/10.1109/icac49085.2019.9103336

Khan, R. U., Cheng, J. L. A., & Bee, O. Y. (2018). Machine learning and dyslexia: Diagnostic and classification system (dcs) for kids with learning disabilities. *International Journal of Engineering & Technology*, *7*(3.18), 97–100.

Martin, A., Kronbichler, M., & Richlan, F. (2016). Dyslexic brain activation abnormalities in deep and shallow orthographies: A meta-analysis of 28 functional neuroimaging studies. *Human brain mapping*, *37*(7), 2676–2699. doi: https://doi.org/10.1002/hbm.23202

MMT, M. H., & Sangamithra, A. (2019). Intelligent predicting learning disabilities in school going children using fuzzy logic k mean clustering in machine learning. *Int. J. Recent Technol. Eng*, *8*(4), 1694–1698. doi: https://doi.org/10.35940/ijrte.c5620.118419

Plante, E., Patterson, D., Gomez, R., Almryde, K. R., White, M. G., & Asbjørnsen, A. E. (2015). The nature of the language input affects brain activation during learning from a natural language. *Journal of Neurolinguistics*, *36*, 17–34. doi: https://doi.org/10.1016/j.jneuroling.2015.02.002

Powers, D. M. (2020). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*. doi: https://doi.org/10.48550/arXiv.2010.16061

Prabha, A. J., & Bhargavi, R. (2022). Prediction of dyslexia from eye movements using machine learning. *IETE Journal of Research*, *68*(2), 814–823. doi: https://doi.org/10.1080/03772063.2019.1622461

Prabha, A. J., Bhargavi, R., & Ragala, R. (2019). Predictive model for dyslexia from eye fixation events. *International Journal of Engineering and Advanced Technology (IJEAT)*, *9*, 235–240. doi: https://doi.org/10.35940/ijeat.a1045.1291s319

Rello, L., & Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proceedings of the 12th international web for all conference.* doi: https://doi.org/10.1145/2745555.2746644

Rello, L., Ballesteros, M., Ali, A., Serra, M., Sanchez, D. A., & Bigham, J. P. (2016). Dytective: diagnosing risk of dyslexia with a game. In *Pervasivehealth.* ACM. doi: https://doi.org/10.4108/eai.16-5-2016.2263338

Rello, L., Romero, E., Rauschenberger, M., Ali, A., Williams, K., Bigham, J. P., & White, N. C. (2018, apr). Screening dyslexia for english using hci measures and machine learning. In *Proceedings of the 2018 international conference on digital health.* ACM. doi: https://doi.org/10.1145/3194658.3194675

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, *10*(3), e0118432. doi: https://doi.org/10.1371/journal.pone.0118432

Tamboer, P., Vorst, H. C., & Oort, F. J. (2014). Identifying dyslexia in adults: an iterative method using the predictive value of item scores and self-report questions. *Annals of dyslexia*, *64*, 34–56. doi: https://doi.org/10.1007/s11881-013-0085-9

Vajs, I., Kovic, V., Papic, T., Savic, A. M., & Jankovic, M. M. (2022, aug). Dyslexia detection in children using eye tracking data based on vgg16 network. In *2022 30th european signal processing conference (eusipco).*

IEEE. doi: https://doi.org/10.23919/eusipco55093.2022.9909817

Vanitha, G., & Kasthuri, M. (2021). Dyslexia prediction using machine learn-
ing algorithms–a review. *International Journal of Aquatic Science*, *12*(2),
3372–3380.

Zahia, S., Garcia-Zapirain, B., Saralegui, I., & Fernandez-Ruanova, B.
(2020, dec). Dyslexia detection using 3d convolutional neu-
ral networks and functional magnetic resonance imaging. *Com-
puter methods and programs in biomedicine*, *197*, 105726. doi:
https://doi.org/10.1016/j.cmpb.2020.105726