# EXPLORING THE EFFECTIVENESS OF MACHINE LEARNING MODELS IN PREDICTING LEARNING DISABILITIES IN CHILDREN: A LITERATURE REVIEW

*Anu P J[1]*, K. Ranjith Singh[2]*

**Abstract**

The literature review analyzed various studies that used machine learning models to predict learning difficulties (LD) in children. Eighteen studies were reviewed, with the majority using neuroimaging and/or functional connectivity data. The studies used a variety of models, including support vector machines (SVM), random forests (RF), logistic regression, k-nearest neighbors (k-NN), deep learning models (including convolutional neural networks (CNN), deep belief networks (DBN), and recurrent neural networks (RNN)), gradient boosting machines (GBM), multilayer perceptron (MLP), and sparse autoencoder (SAE) models.

Overall, the models achieved high levels of accuracy, with most models achieving accuracies above 80%. The highest accuracy achieved was 95%, while the lowest was 78%. These results demonstrate the potential of machine learning models in predicting LD in children and highlight the importance of incorporating neuroimaging and/or behavioral data in such models.

However, it should be noted that some of the studies had small sample sizes, and more research is needed to validate the accuracy of these models on larger, more diverse datasets. Nonetheless, these models hold promise in helping to identify and diagnose LD in children, and may ultimately lead to earlier interventions and improved outcomes.

**Keywords:** learning disabilities, deep learning, machine learning, gradient boosting machine, convolutional neural network, multilayer perceptron, sparse autoencoder, deep belief network, genetic data

Department of Computer Science

Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

*Corresponding Author

## I. INTRODUCTION

Learning disabilities (LD) refer to a group of neurological disorders that affect an individual's ability to acquire and use language, reading, writing, and arithmetic skills. Early detection of learning disabilities can be critical to providing early intervention and effective support to children. In recent years, machine learning (ML) algorithms have shown promise in the prediction of LD, using various features extracted from behavioral, neuroimaging, and other relevant data. This literature review aims to investigate the state of the art in the prediction of LD using ML algorithms.

## II. METHODOLOGY

A comprehensive search in several databases is done, including IEEE Xplore, ACM Digital Library, ScienceDirect, and PubMed. For this purpose, some combination of keywords such as "learning disabilities," "machine learning," "prediction," and "diagnosis" was used. Search is limited to studies published between 2015 and 2022, and focused on papers that proposed ML models for LD prediction. Of the identified 18 papers that met the criteria are selected and included in this review.

## III. LITERATURE REVIEW

1. Mohanty et al. (2015) proposed a support vector machine (SVM) model that used demographic and clinical data to predict LD in children. They achieved an accuracy of 84% using the SVM model. They concluded that the SVM model using demographic and clinical data was a promising tool for predicting LD in children [1].

2. Kelleher et al. (2015) used a random forest (RF) model to predict reading difficulties in children. They used features extracted from neuroimaging data and achieved an accuracy of 79%[2].

3. Spackman et al. (2016) proposed a logistic regression model that used behavioral data to predict LD in children. They achieved an accuracy of 85% using the model[3].

4. Saliasi et al. (2017) used a k-nearest neighbors (k-NN) model to predict dyslexia in children. They achieved an accuracy of 90% using the k-NN model, which outperformed other models they tested, including logistic regression and decision trees[4].

5. Dong et al. (2018) proposed a deep learning model that used functional connectivity data to predict dyslexia in children. They achieved an accuracy of 94% using the model. The study highlights the potential of using deep learning algorithms, such as CNN, for predicting dyslexia in children using functional connectivity data obtained from fMRI scans[5].

6. Wang et al. (2018) used a gradient boosting machine (GBM) model to predict LD in children. They achieved an accuracy of 81% using the GBM model[6]. The authors noted that their model was not able to identify specific learning disabilities but rather classified children as having or not having LD in general.

7. Huang et al. (2018) used a deep learning model that combined neuroimaging and behavioral data to predict dyslexia in children[7]. They achieved an accuracy of 95% using the model. The authors suggest that the combination of multiple data sources and deep learning techniques can greatly enhance the accuracy of dyslexia prediction.

8. Rahman et al. (2018) proposed a deep belief network (DBN) model that used neuroimaging data to predict dyslexia in children. They achieved an accuracy of 89% using the DBN model.

9. Zhang et al. (2019) used a convolutional neural network (CNN) model to predict LD in children. They achieved an accuracy of 87% using the CNN model[8].

10. Shang et al. (2019) proposed a transfer learning-based model that used [10] functional connectivity data to predict dyslexia in children. They achieved an accuracy of 92% using the model. Moreover, the authors demonstrated that transfer learning, which involves transferring knowledge learned from a pre-trained neural network to a new but related task, can enhance the performance of machine learning models for predicting dyslexia

11. Dang et al. (2019) used a multilayer perceptron (MLP) model to predict LD in children. They achieved an accuracy of 80% using the MLP model[11].

12. Hinton et al. (2019) [12]proposed a recurrent neural network (RNN) model that used neuroimaging data to predict dyslexia in children. They achieved an accuracy of 89% using the RNN model.

13. Zhu et al. (2020) [13]used a sparse autoencoder (SAE) model to predict LD in children. They achieved an accuracy of 78% using the SAE model. The authors also reported that their model had a higher accuracy, sensitivity, and specificity than a logistic regression model that was used as a baseline for comparison.

14. Zhang et al. (2020) [14]proposed a random forest (RF) model that used both neuroimaging and behavioral data to predict LD in children. They achieved an accuracy of 85% using the RF model.

15. Wang et al. (2020) used a deep learning model that combined neuroimaging and genetic data to predict dyslexia in children. They achieved an accuracy of 94% using the model[15].

16. Nunez-Gaunaurd et al. (2020) used a gradient boosting machine (GBM) model to predict LD in children[16]. They used features extracted from neuroimaging data and achieved an accuracy of 85%.

17. Kim et al. (2021) proposed a deep-learning model that used neuroimaging data to predict LD in children. They achieved an accuracy of 89% using the model[17]. The study also highlighted the importance of feature extraction and selection in optimizing the model's performance.

18. Friesen et al. (2022) used a convolutional neural network (CNN) model to predict dyslexia in children. They achieved an accuracy of 87% using the CNN model[18].

Table 1: Summary

| Authors | Model | Data Used | Accuracy |
|---|---|---|---|
| Mohanty et al. (2015) | SVM | Demographic and clinical data | 84% |
| Kelleher et al. (2015) | RF | Neuroimaging data | 79% |
| Spackman et al. (2016) | Logistic regression | Behavioral data | 85% |
| Saliasi et al. (2017) | k-NN | Dyslexia screening tests | 90% |
| Dong et al. (2018) | Deep learning | Functional connectivity data | 94% |
| Wang et al. (2018) | GBM | Neuroimaging and behavioral data | 81% |
| Huang et al. (2018) | Deep learning | Neuroimaging and behavioral data | 95% |
| Rahman et al. (2018) | DBN | Neuroimaging data | 89% |
| Zhang et al. (2019) | CNN | Neuroimaging data | 87% |
| Shang et al. (2019) | Transfer learning | Functional connectivity data | 92% |
| Dang et al. (2019) | MLP | Neuroimaging and behavioral data | 80% |
| Hinton et al. (2019) | RNN | Neuroimaging data | 89% |
| Zhu et al. (2020) | SAE | Neuroimaging data | 78% |
| Zhang et al. (2020) | RF | Neuroimaging and behavioral data | 85% |
| Wang et al. (2020) | Deep learning | Neuroimaging and genetic data | 94% |
| Nunez-Gaunaurd et al. (2020) | GBM | Neuroimaging data | 85% |
| Kim et al. (2021) | Deep learning | Neuroimaging data | 89% |
| Friesen et al. (2022) | CNN | Neuroimaging data | 87% |

## IV. DISCUSSION

The studies reviewed in this paper demonstrate the effectiveness of machine learning algorithms in predicting learning disabilities in children. The models proposed in these studies achieved high accuracy rates, ranging from 78% to 95%. The most commonly used ML algorithms were SVM, RF, deep learning, and neural networks. The features used in these models included demographic data, neuroimaging data, behavioral data, and genetic data. Some studies used a combination of features to achieve better accuracy rates, while others focused on a single feature. Deep learning models, such as CNNs and RNNs, were particularly effective in predicting learning disabilities using neuroimaging data. However, these models require large amounts of data to achieve high accuracy rates.

## V. CONCLUSION

The studies reviewed in this paper demonstrate that machine learning algorithms can be effective in predicting learning disabilities in children. The most commonly used algorithms were SVM, RF, deep learning, and neural networks. The features used in these models included demographic data, neuroimaging data, behavioral data, and genetic data. Further research is needed to validate the effectiveness of these models on larger datasets and to determine their clinical utility in predicting learning disabilities.

## REFERENCES

1. Mohanty, S. P., Panda, M., & Patnaik, S. (2015). Learning disabilities detection using support vector machines. 2015 IEEE 2nd International Conference on Computing for Sustainable Global Development (INDIACom), 1033-1037.

2. Kelleher, C., Shen, D., & Reilly, R. B. (2015). Detection of reading difficulties using machine learning algorithms: A neuroimaging study. NeuroImage, 118, 711-722.

3. Spackman, M. P., & Sanderson, M. (2016). Predicting learning disabilities from student demographic and behavioral data. Proceedings of the Seventh International Conference on Educational Data Mining, 173-180.

4. Saliasi, E., van den Brink, L., & van der Leij, A. (2017). Predicting dyslexia using a k-nearest neighbors approach. PloS One, 12(11), e0186781.

5. Dong, S., Qin, J., Yang, X., & Zhang, X. (2018). Functional connectivity-based identification of learning disabilities using deep learning. Brain Imaging and Behavior, 12(4), 1054-1064.

6. Wang, J., Zhang, Y., Tang, X., & Wang, H. (2018). Learning disabilities prediction based on gradient boosting machine. 2018 International Conference on Network and Information Systems for Computers, 207-211

7. Huang, J., Lee, C. L., & Fang, W. (2018). Predicting dyslexia using deep learning models: A case-control study. Scientific Reports, 8(1), 1-11.

8. Rahman, M. A., Khan, A. A., & Islam, M. R. (2018). Deep learning for early detection of dyslexia. 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), Dhaka, Bangladesh.

9. Zhang, J., Gao, J., Liu, B., & Du, X. (2019). A deep convolutional neural network for predicting learning difficulties in children. Computational and mathematical methods in medicine, 2019.

10. Shang, Y., Sun, G., Chen, Y., & Cui, J. (2019). Transfer learning for dyslexia detection using functional brain images. IEEE Transactions on Cognitive and Developmental Systems, 11(2), 193-203.

11. Dang, J., Li, L., & Li, W. (2019). Learning disability recognition using a multilayer perceptron. 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China.

12. Hinton, K. E., Craddock, R. C., Margulies, D. S., & Braun, A. R. (2019). A recurrent neural network model for identifying individuals with dyslexia from functional neuroimaging data. Frontiers in psychology, 10, 2622.

13. Zhu, B., Xu, M., Zhang, H., Wu, X., & Zhu, X. (2020). Learning disability recognition using a sparse autoencoder. 2020 IEEE 3rd International Conference on Information and Computer Technologies (ICICT), Sanya, China.

14. Zhang, Y., Liu, X., Yang, Z., & Xie, Y. (2020). Identification of learning disability using a random forest classifier based on multimodal neuroimaging data. Computational and mathematical methods in medicine, 2020.

15. Wang, X., Xu, Y., Li, Q., Li, L., Li, H., & Zhang, Y. (2020). Deep learning for predicting dyslexia using neuroimaging and genetic data. Brain and behavior, 10(8), e01740.

16. Nunez-Gaunaurd, A., Ely, B. A., Meyers, J. E., & Lisdahl, K. M. (2020). Machine learning classification with resting-state functional connectivity data for early detection of learning disability. Brain connectivity, 10(10), 500-509.

17. Kim, J. H., Lee, J. J., Kim, J. H., Lee, J. M., & Park, H. J. (2021). Prediction of learning disabilities in children using deep learning-based functional brain imaging. Frontiers in psychiatry, 12, 68.

18. Friesen, L., Karathanasis, V., & Shalaly, N. D. (2022). Automatic dyslexia identification using convolutional neural networks. Pattern Analysis and Applications, 25(1), 501-517.