Application of Machine Learning in Learning Problems and Disorders: A Systematic Review

Mario Aquino Cruz[1], Oscar Alcides Choquehuallpa Hurtado[2], Esther Calatayud Madariaga[3]
Departamento Académico de Informática y Sistemas, Universidad Nacional Micaela Bastidas de Apurímac, Abancay, Perú[1, 2]
Departamento Académico de Ciencias Básicas, Universidad Nacional Micaela Bastidas de Apurímac, Abancay, Perú[3]

***Abstract*—Learning Disorders, which affect approximately 10% of the school population, represent a significant challenge in the educational field. The lack of proper diagnosis and treatment can have profound consequences, triggering psychological problems in those affected by disorders that impact reading, writing, numeracy and attention, among others. Notable among them are Attention Deficit Hyperactivity Disorder (ADHD) and dyslexia. In this context, a literature review focusing on Machine Learning applications to address these educational problems is addressed. The methodology proposed by Barbara Kitchenham guides this analysis, using the online tool Parsifal for the review, generation of search strings, formulation of research questions and management of information sources. The first findings of this research highlight a growing trend in the application of Machine Learning techniques in learning problems and disorders, especially in the last five years, as of 2019. Among the primary sources, the IEEE Digital Library emerges as a key source of information in this rapidly developing field. This innovative approach has the potential to significantly improve early detection, accurate diagnosis and implementation of personalized interventions, thus offering new perspectives in understanding and addressing the educational challenges associated with Learning Disorders.**

Keywords*—*Machine learning; learning disorder; deep learning; ADHD; dyslexia; learning impairment

I. Introduction

Learning Disorders (LD) are the most prevalent neurodevelopmental disorders in the population, affecting about 10% of the population at school age. AT cause that children with an adequate schooling and normal intelligence with difficulties of neurobiological origin, due to the lack of a proper diagnosis and treatment, suffer this problem of learning disorders causing psychological problems to those affected. These disorders affect reading, writing, calculation and/or attention, among others [1]. Some of the best known disorders are ADHD and dyslexia.

ADHD attention [2] deficit and hyperactivity disorder, considered as a specific learning disorder, as well as a behavioral disorder, has been one of the main reasons for consultation in pediatric neurology, although the clinical criteria are well established, probably the clinical history is not rigorously reviewed and the most determinant symptoms are not adequately noted.

Dyslexia is a brain condition that makes reading, spelling, writing and sometimes speaking difficult. The brains of people with dyslexia have difficulty recognizing or processing certain types of information. It occurs in people who do not have any physical, motor, visual or other disabilities. [3] Likewise, people with dyslexia have normal cognitive development.

Learning disorders are not considered problems, although they were for many years, but thanks to advances in neuroscience, it is now defined as a learning disorder, and the causes are neurobiological in origin. [3] The child is born and the cause is generally of genetic origin, although several institutions still maintain the name of learning disabilities, it must be clear that they are not problems but learning disorders.

It is therefore important to recognize the importance of approaching Learning Disorders from a holistic perspective. While science has advanced in understanding their neurobiological origin, it is essential to emphasize the need for early detection and a multidisciplinary approach to ensure accurate diagnosis and appropriate treatment. Lack of awareness and understanding of these disorders can lead to children facing unnecessary difficulties in their academic and emotional development. Therefore, advocating for greater awareness, education and support for those affected is key to building an inclusive society that recognizes and values diversity in learning abilities.

Machine Learning techniques can help in investigating and addressing learning disorders, such as Attention Deficit Hyperactivity Disorder (ADHD) and dyslexia, due to their unique ability to analyze vast amounts of heterogeneous data and extract complex patterns that may go unnoticed by conventional methods. These disorders, characterized by a diversity of clinical manifestations and associated factors, present significant challenges for accurate diagnosis and a thorough understanding of their underlying mechanisms. The application of these algorithms allows the integration of information from various sources, such as reading tests, magnetic resonance imaging and electroencephalography, thus facilitating a multidimensional approach. In addition, the classification and predictive capabilities of these techniques can be crucial to improve early detection, personalize intervention strategies and move towards the development of automated diagnostic systems.

The aim of this article is to make a systematic review on the contributions of Machine Learning and Deep learning in learning disabilities and learning disorders detection. For which a methodology inspired by Barbara Kitchenham's proposal was used, then in the results section the information extracted from the selected articles is analyzed and finally the

conclusions obtained are specified, the information collected is intended to serve as a basis for other studies related to the application of Machine Learning in the problems of learning disorders. It is also intended to motivate readers to delve deeper into this area of research.

II. METHODOLOGY

For the development of this article, Barbara Kitchenham's methodology was applied as an inspiration, it is important to highlight that it helps significantly in works related to the systematic review of the literature, the phases that helped to develop this study are described below.

Plan the systematic literature review using the Parsifal tool.

- Determine research questions.
- Establish search process.
- Define inclusion and exclusion criteria for articles.
- Select query sources.
- Create search strings.

Develop the systematic literature review with the defined planning.

- Search for articles.
- Selection of definitive articles for information analysis.
- Analysis and classification of information.

Document and interpret the results of the review.

- Develop the report of the research questions of the review.

The tools used for the development of this research are as follows:

- Mendeley: It is a tool that allows the management and administration of bibliographic sources. It allowed the management of the references mentioned in this research work.
- Parsifal: It is a web tool that helped with the creation of search strings, keywords and research questions.

III. DEVELOPMENT

A. Plan the Systematic Literature Review

Each of the tasks performed during the planning stage of the systematic literature review are detailed below:

1) Determine the research questions: Based on the purpose of this article, the following research questions are raised.

- In which repositories have most of the articles on the topic of study been published?
- In what year were most articles published, and how are they interpreted?
- How does Machine Learning help in the detection of learning problems and disorders?

2) Establish the search process: Base terms were identified by applying the PICOC method [4] to define the scope of the systematic review. Its components are population, intervention, comparison, results and contexts. This method made it possible to define the expressions that make up the search strings. They are detailed below:

- Population (P): “Learning disorder” OR “Learning Impairment” OR “ADHD” OR “Learning Deficit”.
- Intervention (I): “Machine learning” OR “Deep Learning”.
- Comparison (C): Not applicable.
- Outcomes (O): “Algorithms” OR “Classification” OR “Detection” OR “Methods” OR “Techniques”
- Context (C): “Artificial Intelligence”.

3) Define the inclusion and exclusion criteria for the articles: According to the objectives and scope of this article, it is necessary to establish four inclusion criteria (IC) and four exclusion criteria (EC), which are described as follows:

Inclusion criteria (IC):

- IC1: Articles containing information on Machine Learning or Deep Learning techniques.
- IC2: Articles written in English or Spanish.
- IC3: Articles published from 2019 onwards.
- IC4: Articles that have been published in scientific journals, scientific articles, and scientific articles.

Inclusion criteria (EC):

- EC1: Duplicate articles.
- EC2: Articles whose title is not related to the object of study.
- EC3: Articles that do not belong to the area of Science and Computing
- EC4: Book chapters, manuals, gray literature.

4) Select sources of consultation: Four search sources indicated in Table I were selected due to their accessibility and advanced query support.

TABLE I. SCIENTIFIC DATABASES

Database	URL
IEEE	https://ieeexplore.ieee.org
ScienceDirect	https://www.sciencedirect.com
Scopus	https://www.scopus.com
Doaj	https://doaj.org/

5) Search Strings: The keywords were considered on the basis of what was applied in the PICOC method, and the logical operators "AND/OR" were used to generate the search strings.

The Parsifal tool generated the general search string, which was modified according to each database, and all the keywords were defined in English, as shown in Table II.

TABLE II. SEARCH STRING

Database	Search Strings
Parsifal	("Learning disorder" OR "ADHD" OR "Learning Deficit" OR "Learning Impairment" OR "TDAH") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques")
IEEE	(("Learning disorder" OR "ADHD" OR "Learning Deficit" OR "Learning Impairment" OR "TDAH") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques"))
ScienceDirect	(("Learning disorder" OR "ADHD") AND ("Machine Learning" AND "Deep Learning") AND ("Algorithms" AND "Classification" OR "Detection" AND ("Methods" OR "Techniques"))
Scopus	TITLE-ABS-KEY(("Learning disorder" OR "Learning Deficit" OR "Learning Impairment" OR "ADHD" OR "TDAH") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" AND ("Classification" OR "Detection") AND ("Methods" OR "Techniques"))) AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"cp")) AND (LIMIT-TO (SUBJAREA,"COMP")) AND (LIMIT-TO (PUBYEAR,2023) OR LIMIT-TO (PUBYEAR,2022) LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019))
Doaj	"Learning disorder" AND "machine learning"

B. Develop the Systematic Literature Review

1) Search for articles: The search string was implemented in each of the selected databases (IEEE, ScienceDirect, Scopus and Doaj), where 413 articles were obtained. Of these, articles that did not meet the inclusion criteria were ignored and duplicate articles or those that had no relevance to the field of study were discarded. Of the 413 articles, 22 were selected for further review as they met the prerequisites.

2) Selection of definitive items for data analysis: The norm in this section, according to the methodology proposed by Barbara Kitchenham, is that the articles go through a series of quality questions in order to select the articles that contribute the most to the research objective. However, since there are only 22 articles and all meet the requirements for inclusion, the 22 definitive articles were considered for the analysis of information.

3) Analysis and classification of information: Sources of information: Thanks to the Parsifal tool, the following figure was created, showing the level of contribution of the sources of information in percentages in Fig. 1.

By observing and analyzing Fig. 1, the following can be determined:

- IEEE Digital Library information source with 45% obtains a higher percentage in the graph, in this group are [5] [6] [7] [8] [9] [10] [11] [12] [13] [14].
- Science@Direct obtains 36.4% of contribution level for the present article. In this group are [15] [16] [17] [18] [19] [20] [21] [22].
- Scopus, which is another of the repositories that made the greatest contribution to the research, contributed 13.6%. This group includes [23] [24] [25].
- Finally, 4.5% belongs to Doaj with the article [26].

Years of publication: Thanks to the Parsifal tool, a graph showing the number of articles per year was generated in Fig. 2.

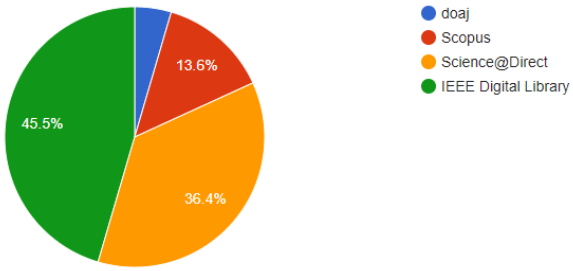

Fig. 1. Information sources.

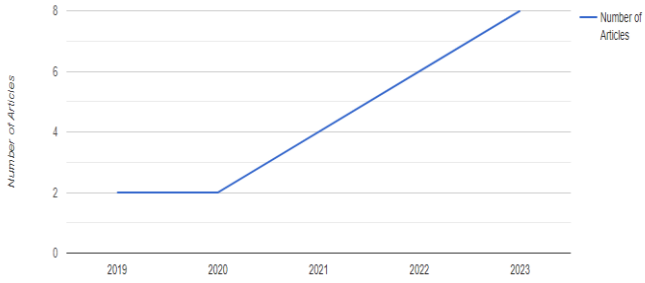

Fig. 2. Articles by year.

C. Document and Interpret the Results of the Review

The following section presents the answers to the research questions.

1) In which repositories have most of the articles on the topic of study been published?: Taking into account the information analysis and classification section, it was observed that the information source with the most articles published was the IEEE Digital Library with 10 articles, followed by Science@Direct with eight articles, these being the repositories with the greatest contribution in the field of Machine Learning in learning problems and disorders.

2) In what year were most of the articles published, and how are they interpreted?: Taking into account the section of analysis and classification of the information, it is possible to identify that the year 2023 was when more articles were published with a total of 8 articles and this was growing since 2019, this could indicate that interest in applying Machine

Learning techniques in areas such as learning disorders is growing.

3) How does Machine Learning help in the detection of learning problems and disorders?: According to the review of the articles, most of them are focused on the early detection of learning disorders, and another large percentage apply Machine Learning techniques for the classification of people with ADHD, followed by another group of articles related to the prediction of the presence of dyslexia in children. And a small percentage focus on optimizing the detection or classification of learning disorders using new and novel Machine Learning techniques, obtaining promising results.

IV. Results

The present review allowed answering the research questions, obtaining that the major source of information is the IEEE Digital Library and the year 2023 is the highest percentage of published articles indicating that more and more emphasis is being placed on the application of Machine Learning in learning problems and disorders. A brief summary of some articles will be given below:

They analyzed in [6] the classification performance of three machine learning algorithms (Naive Bayes, kNN, logistic regression) applied on the dataset of 157 children, of which 77 were ADHD patients and 80 healthy. The closest classifier k is able to predict with a high accuracy of 86% and is much better than Naive Bayes (52%) and logistic regression (66%).

This study in [15] focused on addressing data imbalance in prescreening tests for developmental dyslexia. It proposed an ensemble-based oversampling and machine learning technique to improve the detection of the minority class (dyslexic patients). The researchers used reading and writing tests, online games, magnetic resonance imaging (MRI), electroencephalography (EEG), photography, and video recording as input data. Simulation results showed that the proposed approach improved the detection accuracy of the minority class from 80.61% to 83.52%.

This research in [26] focused on identifying the neurocognitive characteristics of Attention Deficit Hyperactivity Disorder (ADHD), both in its pure presentation and in its comorbidity with other disorders. Using the Machine Learning Decision Tree (MLT) technique in the Rcran 4.2.1 software, probabilistic rules were constructed based on different neurocognitive variables. We found that children with pure ADHD showed poor performance on working memory and perceptual reasoning tasks, independent of IQ deficits. In addition, deficits in working memory were common across all ADHD presentations and comorbidities. The use of DTM allowed us to establish a clinical hierarchy and identify the most important variables in the different populations of children diagnosed with ADHD. This technique proved to be advantageous in differentiating the importance of dependent variables in the study of ADHD.

This project [5] aims to classify attention deficit hyperactivity disorder (ADHD) using machine learning techniques. The WEKA toolkit was used to perform a comparative analysis of the classification of four forms of ADHD using various machine learning algorithms. Different feature sets were experimented with, including those generated using the genetic algorithm from phenotypic data of the ADHD-200 dataset. The classification algorithms used were Logistic, Support Vector Machine (SVM), Decision Tree (DT) implemented using the J48 algorithm, Random Forest (RF), K-nearest neighbor (KNN) implemented using the instance-based learner (IBk) algorithm and multi-layer perceptron (MLP). Eight performance parameters were evaluated: accuracy, precision, recall, F-measure, Kappa statistic, root mean square error (RMSE), Mathew correlation coefficient (MCC) and area under the receiver operating characteristics curve (AUROC). The ultimate goal is to provide valuable information for the establishment of an automated diagnostic system for ADHD.

V. Discussion and Conclusion

The literature review reveals important points of convergence and divergence among the studies analyzed. First, there is a general consensus in the scientific community on the growing relevance and application of machine learning techniques, especially in the context of learning disorders and problems, such as Attention Deficit Hyperactivity Disorder (ADHD) and developmental dyslexia. All studies highlight the importance of using machine learning algorithms to improve the detection, classification and understanding of these disorders, leveraging diverse data sources, from reading and writing tests to magnetic resonance imaging and electroencephalography.

Different methodological approaches are also observed across studies. For example, while some focus on the performance of specific algorithms for ADHD classification, others address specific challenges such as data imbalance in prescreening tests for developmental dyslexia. In addition, the use of different feature sets and the application of specific techniques, such as the Machine Learning Decision Tree, highlight the diversity of strategies implemented in the research.

However, aspects related to the integration and standardization of the proposed methods remain to be addressed, as well as the validation of the results in clinical settings and the consideration of ethical factors in the development of automated diagnostic systems. The variability in the feature sets and algorithms used suggests the need to establish common protocols for the comparison and replication of results. Furthermore, the practical implementation and acceptance of these technologies in clinical and educational settings requires careful consideration of the ethical and social implications. In this sense, future research could focus on addressing these aspects to move towards the effective application of machine learning techniques in the diagnosis and treatment of learning disorders.

References

[1] A. Sans, C. Boix, R. Colomé, A. López-Sala, y A. Sanguinetti, «Trastornos del aprendizaje | Pediatría integral», 2012. Accedido: 8 de julio de 2023. [En línea]. Disponible en:

https://www.pediatriaintegral.es/publicacion-2017-01/trastornos-del-aprendizaje-2017/.

[2] J. Vaquerizo-Madrid, «Clinical assessment of attention deficit hyperactivity disorder, interview model and controversial issues», Rev. Neurol., vol. 46, n.o SUPPL. 1, 2008, doi: 10.33588/rn.46s01.2008016.

[3] S. Mariza y E. O. Alvarado, «Trastornos de aprendizaje en la educación primaria», Educación, vol. 24, n.o 2, pp. 211-215, dic. 2018, doi: 10.33539/EDUCACION.2018.V24N2.1340.

[4] M. Petticrew y H. Roberts, Systematic Reviews in the Social Sciences: A Practical Guide. Blackwell Pub, 2008. doi: 10.1002/9780470754887.

[5] J. Singh, G. Kaur, y N. Kapoor, «Classification of Attention Deficit Hyperactivity Disorder Using Machine Learning», 2022 IEEE 3rd Glob. Conf. Adv. Technol. GCAT 2022, 2022, doi: 10.1109/GCAT55367.2022.9971947.

[6] S. Saini, R. Rani, y N. Kalra, «Prediction of Attention Deficit Hyperactivity Disorder (ADHD) using machine learning Techniques based on classification of EEG signal», en 8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 782-786. doi: 10.1109/ICACCS54159.2022.9785356.

[7] K. Banumathi, G. Sudhasadasivam, B. Banurekha, N. Shahana, y K. Vaishnavi, «AI Powered Screening Aid for Dyslexic Children in Tamil», 2023 Int. Conf. Adv. Intell. Comput. Appl. AICAPS 2023, 2023, doi: 10.1109/AICAPS57044.2023.10074214.

[8] C. Sharmila, N. Shanthi, S. Santhiya, E. Saran, K. Sri Rakesh, y R. Sruthi, «An Automated System for the Early Detection of Dysgraphia using Deep Learning Algorithms», 2nd Int. Conf. Sustain. Comput. Data Commun. Syst. ICSCDS 2023 - Proc., pp. 251-257, 2023, doi: 10.1109/ICSCDS56580.2023.10105022.

[9] G. P. Pralhad, A. Joshi, M. Chhipa, S. Kumar, G. Mishra, y M. Vishwakarma, «Dyslexia Prediction Using Machine Learning», Proc. - 2021 1st IEEE Int. Conf. Artif. Intell. Mach. Vision, AIMV 2021, 2021, doi: 10.1109/AIMV53313.2021.9671004.

[10] N. Gupte, M. Patel, T. Pen, y S. Kurhade, «Early detection of ADHD and Dyslexia from EEG Signals», 2023 IEEE 8th Int. Conf. Converg. Technol. I2CT 2023, 2023, doi: 10.1109/I2CT57861.2023.10126272.

[11] M. Maniruzzaman, M. A. M. Hasan, N. Asai, y J. Shin, «Optimal Channels and Features Selection based ADHD Detection from EEG Signal using Statistical and Machine Learning Techniques», IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3264266.

[12] S. Liu et al., «Deep Spatio-Temporal Representation and Ensemble Classification for Attention Deficit/Hyperactivity Disorder», IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 1-10, 2021, doi: 10.1109/TNSRE.2020.3019063.

[13] L. Shao, D. Zhang, H. Du, y D. Fu, «Deep forest in ADHD data classification», IEEE Access, vol. 7, pp. 137913-137919, 2019, doi: 10.1109/ACCESS.2019.2941515.

[14] A. Mohd, A. M. Ali, y S. A. Halim, «Detecting ADHD Subjects Using Machine Learning Algorithm», en 2022 IEEE International Conference on Computing, ICOCO 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 299-304. doi: 10.1109/ICOCO56118.2022.10031796.

[15] S. Kaisar y A. Chowdhury, «Integrating oversampling and ensemble-based machine learning techniques for an imbalanced dataset in dyslexia screening tests», ICT Express, vol. 8, n.o 4, pp. 563-568, dic. 2022, doi: 10.1016/j.icte.2022.02.011.

[16] J. Kunhoth, S. Al Maadeed, M. Saleh, y Y. Akbari, «CNN feature and classifier fusion on novel transformed image dataset for dysgraphia diagnosis in children», Expert Syst. Appl., vol. 231, p. 120740, nov. 2023, doi: 10.1016/J.ESWA.2023.120740.

[17] N. P. Guhan Seshadri, S. Agrawal, B. Kumar Singh, B. Geethanjali, V. Mahesh, y R. B. Pachori, «EEG based classification of children with learning disabilities using shallow and deep neural network», Biomed. Signal Process. Control, vol. 82, abr. 2023, doi: 10.1016/J.BSPC.2022.104553.

[18] A. Devi y G. Kavya, «Dysgraphia disorder forecasting and classification technique using intelligent deep learning approaches», Prog. Neuro-Psychopharmacology Biol. Psychiatry, vol. 120, p. 110647, ene. 2023, doi: 10.1016/J.PNPBP.2022.110647.

[19] P. Raatikainen, J. Hautala, O. Loberg, T. Kärkkäinen, P. Leppänen, y P. Nieminen, «Detection of developmental dyslexia with machine learning using eye movement data», Array, vol. 12, p. 100087, dic. 2021, doi: 10.1016/J.ARRAY.2021.100087.

[20] Z. Mao et al., «Spatio-temporal deep learning method for ADHD fMRI classification», Inf. Sci. (Ny)., vol. 499, pp. 1-11, oct. 2019, doi: 10.1016/j.ins.2019.05.043.

[21] A. Riaz, M. Asad, E. Alonso, y G. Slabaugh, «DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI», J. Neurosci. Methods, vol. 335, p. 108506, abr. 2020, doi: 10.1016/J.JNEUMETH.2019.108506.

[22] B. TaghiBeyglou, A. Shahbazi, F. Bagheri, S. Akbarian, y M. Jahed, «Detection of ADHD cases using CNN and classical classifiers of raw EEG», Comput. Methods Programs Biomed. Updat., vol. 2, p. 100080, ene. 2022, doi: 10.1016/j.cmpbup.2022.100080.

[23] P. Drotár y M. Dobeš, «Dysgraphia detection through machine learning», Sci. Rep., vol. 10, n.o 1, pp. 1-11, dic. 2020, doi: 10.1038/S41598-020-78611-9.

[24] L. Devillaine et al., «Analysis of Graphomotor Tests with Machine Learning Algorithms for an Early and Universal Pre-Diagnosis of Dysgraphia», Sensors, vol. 21, n.o 7026, p. 7026, oct. 2021, doi: 10.3390/S21217026.

[25] M. Maniruzzaman, J. Shin, M. A. M. Hasan, y A. Yasumura, «Efficient Feature Selection and Machine Learning Based ADHD Detection Using EEG Signal», Comput. Mater. Contin., vol. 72, n.o 3, pp. 5179-5195, 2022, doi: 10.32604/cmc.2022.028339.

[26] C. Quintero-López, V. D. Gil-Vera, D. A. Landinez-Martínez, J. P. Vargas-Gaviria, y N. Gómez-Muñoz, «Predictive Neurocognitive Model of Attention Deficit Hyperactivity Disorder Diagnosis», Mediterr. J. Clin. Psychol., vol. 11, n.o 3, abr. 2023, doi: 10.13129/2282-1619/MJCP-3606.