SUPPORT SYSTEM FOR THE DIAGNOSIS OF THE RISK OF ANXIETY DISORDER IN CHILDREN

Renata C. Rocha [1], Jonathan A. Queiroz [1], Allan Kardec D. B. Filho[1]

[1]*Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Maranhão*
Av. dos Portugueses, 1966 – Bacanga, 65080-805, São Luís/Maranhão, Brasil
renata.rocha@discente.ufma.br, jonathan.queiroz@ufma.br, allan.kardec@ufma.br

Abstract. Anxiety disorders in children represent an important public health challenge, considering the subjective nature of the symptoms, the individual variability of clinical manifestations, and the lack of standardized diagnostic support tools. This gap compromises early identification and appropriate preventive interventions. This study investigated computational approaches for the multilevel classification of the risk of anxiety disorders in children, using behavioral and physiological data. The methodology involves the application and comparison of machine learning models, including Random Forest, Support Vector Machine, and Multilayer Perceptron Neural Network, in binary (presence or absence) and multilevel (mild, moderate, severe risk) paradigms. The research used a dataset of 193 children, publicly available on Harvard Dataverse by Carpenter [1], licensed under CC0 1.0 Universal. The evaluation of the models used standardized metrics aligned with the diagnostic criteria of the DSM-5, Diagnostic and Statistical Manual of Mental Disorders - 5th edition, and the ICD-11, International Classification of Diseases - 11th edition. After refinements in the methodology, the results showed significant improvement: Random Forest – accuracy 89.4%, sensitivity 80%, specificity 88.7%; Support Vector Machine – accuracy 88.5%, sensitivity 81.3%, specificity 91.9%; Multilayer Perceptron – accuracy 87.8%, sensitivity 77.7%, specificity 91.1%. Accuracy above 87% indicates excellent overall performance of the models, correctly classifying most cases. Given the topic, high sensitivity is crucial to avoid the omission of relevant cases. Sensitivity values (77% to 81%) demonstrate effective identification of positives, which is clinically important. High specificity (above 88%) shows accurate recognition of negatives, reducing false positives. These results indicate a strong predictive performance, especially for the SVM, which showed balance and robustness, being the most suitable to minimize both false negative and false positives. This highlights the feasibility of machine learning in early detection of anxiety risk in children. Continuous improvements aim to improve accuracy and clinical applicability. This study contributes to the advancement of diagnostic strategies in child mental health by offering a computational approach to support personalized, evidence-based clinical interventions.

Keywords: childhood anxiety disorders; computational diagnosis; machine learning; risk detection; classification.

1 Introduction

Anxiety disorders are among the most prevalent mental health conditions in childhood, with significant impacts on cognitive, emotional, and social development, as highlighted by Polanczyk et al. [2] and Muris and Merckelbach [3]. Despite this, its diagnosis is still marked by challenges, such as the subjectivity of symptoms, limitations in the child's verbal expression, and variability of clinical manifestations. This gap compromises early identification and appropriate preventive interventions, as discussed by Kazdin [4]. The use of computational tools can increase the sensitivity of professionals, offering analytical and objective support.

This work investigated the application of machine learning techniques for the multilevel classification of anxiety risk in children, based on physiological and behavioral data. The data source was used in the public database made available in the Harvard Dataverse repository by Carpenter [1], originally collected for the study published by Carpenter et al. [5], which provides biometric measurements and observational records of 193 preschool children (3 to 5 years old). Three supervised models were compared – Random Forest, according to Breiman [6]; Support Vector Machine, according to Cortes and Vapnik [7]; and Multilayer Perceptron, according to Goodfellow, Bengio e Courville [8] - focusing on the evaluation of its predictive performance and clinical potential in the face of different levels of severity of the disorder.

Anxiety disorders are the most common psychiatric disorders in childhood and adolescence, affecting between 7% and 12% of school-age children worldwide, as highlighted by Polanczyk et al. [2]. In the United States, data from the Centers for Disease Control and Prevention (CDC), the country's main public health agency, estimates that approximately 11.6% of children between 3 and 17 years old have an active diagnosis of anxiety [9]. In Brazil, regional surveys, such as those conducted by the Oswaldo Cruz Foundation (FIOCRUZ), indicate that up to 15% of children and adolescents may present significant symptoms, especially after the period of social isolation resulting from the COVID-19 pandemic [10].

Clinical manifestations can occur through psychological (excessive fears, disproportionate worries, irritability) and physiological (tachycardia, muscle tension, sweating, nausea) symptoms, as well as functional impairments such as decreased academic performance, social isolation, and refusal to attend school, as described by Muris and Merckelbach [3]. When not detected early, childhood anxiety is associated with a higher risk of depression, panic disorder, substance abuse, and suicidal ideation in adolescence and adulthood, as pointed out by Costello et al. [10].

Despite the severity of the condition, it is estimated that up to 80% of cases do not receive appropriate treatment, as discussed by Kazdin [4], mainly due to the difficulty in identifying symptoms by caregivers and professionals, the lack of objective criteria, and the scarcity of standardized early screening tools. Alternatively, computational strategies have been explored to support the identification of clinical symptoms based on objective data, such as physiological and behavioral variables. The dataset used in this study, published by Carpenter [1], brings together records of children obtained in a standardized clinical setting, including psychophysiological signs (such as skin conductance and heart rate) and observational risk assessments, and is widely used in the international literature by Carpenter et al. [5].

Models such as Random Forest (RF), according to Breiman [6]; Support Vector Machine (SVM), according to Cortes and Vapnik [7]; and Multilayer Perceptron (MLP), according to Goodfellow, Bengio, and Courville [8], have demonstrated good performance in clinical scenarios to support diagnosis. RF is known for its robustness to noisy data and ability to interpret variables, as described by Breiman [6]; SVM, for its effectiveness in separating classes in high-dimensional spaces, according to Cortes and Vapnik [7]; and MLP, for its ability to model complex nonlinear patterns, according to Goodfellow, Bengio, and Courville [8].

In this context, the mathematical foundation presented below the details of the machine learning models used in this study, as well as the criteria and metrics adopted to evaluate their performance in the multilevel classification of anxiety risk in children. Subsequent sections detail the computational methodology and implementation, including pre-processing the data and setting up the models. To this end, the results obtained are discussed, comparing the performance of the different algorithms. Finally, the conclusions of the study and suggestions for future work are presented.

2 Materials and Methods

Classifying the risk of anxiety disorders in children requires computational approaches capable of dealing with physiological and behavioral data efficiently. For this, three supervised machine learning models were used: Random Forest (RF), Support Vector Machine (SVM) and Multilayer Perceptron (MLP). Each of these models has distinct characteristics in the way they process data and perform classification. In this section, the mathematical formulations that underline these algorithms are described, as well as the metrics used to assess their performance in the multilevel task of classifying the risk of childhood anxiety.

Let the attribute vector $x = [x_1, x_2, ..., xn]$ be $\in R^n$ and the target variable $y \in \{0, 1, 2\}$, where: $y = 0$: low risk of anxiety; $y = 1$: moderate risk; $y = 2$: high risk.

The problem is treated as a multilevel supervised classification, looking for a function $f: R^n \rightarrow \{0,1,2\}$, which minimizes the predictive error rate.

The models used follow different classification approaches:

Random Forest builds B decision trees $hi(x)$, and the final prediction $\hat{y}$ is given by majority vote: $\hat{y} = \text{mode}\ \{ h1\ (x), h2\ (x), \ldots, hB\ (x)\}$; The SVM searches for the optimal separation hyperplane using the Gaussian kernel: $K (x, x') = \exp(-\gamma \cdot \|x - x'\|^2)$; MLP estimates by endo $g(\cdot)$ the activation function of the hidden layer and $f(\cdot)$ the SoftMax of the output.$\hat{y} = \text{argmax}(f(W \cdot g(x) + b))$.

The metrics used were:

Accuracy: Measures the ratio of correct classifications to the total number of samples:

$$\frac{\mathrm{VP} + \mathrm{VN}}{\mathrm{VP} + \mathrm{VN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

Sensitivity: Indicates the model's ability to correctly identify positive cases:

$$\frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{FN}} \tag{2}$$

Specificity: measures the model's ability to correctly recognize negative cases:

$$\frac{\mathrm{VN}}{\mathrm{VN} + \mathrm{FP}} \tag{3}$$

With the mathematical foundations in place, the next step describes the computational methodology and implementation, covering the dataset, the pre-processing procedures, and the practical application of the classification models.

2.1 Computational Methodology and Implementation

This section describes the procedures adopted for the computational implementation of the study, including the dataset used, the pre-processing of the information, the tools used, the classification models used, and the evaluation of the results. In addition, the tools used, and the algorithms implemented are presented.
To ensure transparency and enable the reproducibility of the experiments, all the code developed in this study, including the configuration of the models, hyperparameter tuning and evaluation procedures, was made publicly available in the GitHub repository: https://github.com/ahcorataner/anxiety-risk-diagnosis-children.

2.2 Dataset

The dataset used consists of physiological, behavioral and demographic information from 193 children aged between 3 and 5 years. The data was originally distributed in two files: training data with 130 samples (≈ 70%) and test data with 63 samples (≈ 30%). The target variable was defined with three classes, which were established: number 0 for low risk; number 1 for moderate risk; and number 2 for high risk. The distribution between the classes was approximately balanced, which does not require the use of resampling techniques. All data was used according to the CC0 1.0 Universal public license.

2.3 Preprocessing

The following steps were performed to prepare the data for analysis: Standardization of continuous variables with StandardScaler; Encoding the target variable in one-hot encoding format for the MLP model only; Elimination of incomplete or inconsistent records; Divided the data into training and testing with a 70/30 ratio. Because the classes were balanced, it was not necessary to apply oversampling or subsampling techniques. The training of the models was performed according to the configurations described below, using the training dataset (130 samples) and 5-fold stratified cross-validation for hyperparameter fit.

2.4 Tools Used

The computational implementation of the study was carried out using a set of tools that ensure efficiency in data manipulation, model training and analysis of results. Python version: Python 3.10.9 – Chosen for its stability and compatibility with the libraries used; Development environment: VSCode – Used as the main IDE due to its flexibility, support for extensions, and integration with Jupyter Notebook for interactive experimentation. Libraries used: Scikit-learn – Used to implement the Random Forest and SVM models, in addition to providing tools for pre-processing, cross-validation and evaluation metrics; TensorFlow/Keras – Used for the construction and training of the MLP model, allowing the definition of deep neural network architectures and efficient optimization; NumPy and Pandas – Essential for data manipulation, including mathematical operations, statistics, and organizing datasets into efficient structures; Matplotlib and Seaborn –

Applied in the graphical visualization of results, allowing the generation of scatter plots, confusion matrices and performance curves of the models. In addition to these tools, stratified cross-validation methods, hyperparameter tuning using GridSearchCV and data normalization were used to ensure the robustness of the experiments.

2.5 Classification models

The three supervised multiclass classification models were compared: Random Forest (RF): The RandomForestClassifier of the scikit-learn library was used, with 100 decision trees, "gini" impurity criterion, unlimited depth and random selection of attributes in each division; Support Vector Machine (SVM): SVC with radial kernel (RBF) was used. Regularization parameter C was adjusted by grid search (GridSearchCV) with 5-fold stratified cross-validation; Multilayer Perceptron (MLP): Structure with two hidden layers containing 64 and 32 neurons, ReLU activation, output layer with softmax function, and categorical_crossentropy loss function. Optimization with Adam, batch_size = 32 and 30 training seasons. Regularization with Abandonment = 0.3.

For the adjustment of hyperparameters, the technique of cross-validation stratified in five folds (StratifiedKFold, $k = 5$) was used, implemented through the GridSearchCV function of the *scikit-learn* library. This procedure sought to ensure that the distribution of classes was preserved in each partition, reducing bias due to unbalance and increasing the robustness of the results. Different search intervals were defined for each model. In the case of Random Forest, combinations involving the number of trees (n_estimators $\in \{100,200,500\}$), the maximum depth of the trees (max_depth $\in$ {None,10,20}) and the impurity criterion (gini,entropy) were investigated. For the Support Vector Machine with radial kernel (RBF), values of the regularization parameter ($C \in \{0.1,1,10,100\}$) were tested together with different kernel coefficients ($\gamma \in \{0.001,0.01,0.1,1\}$). In the case of the Multilayer Perceptron (MLP), network architectures with different sizes of hidden layers ((64,),(100,),(64,32)), activation functions ({ReLU, Tanh}) and initial learning rates ({0.001,0.01}) were explored.

The search process allowed us to identify the following sets of hyperparameters with the best average performance in the cross-validation: Random Forest (n_estimators = 100, max_depth = 20, criterion = "gini"), SVM ($C = 100$, $\gamma = 0.001$) and MLP (hidden_layer_sizes = (64,), activation = "ReLU", learning_rate_init = 0.001). Even so, to ensure comparability with the results originally presented, the final models reported in the Results section were trained with the parameters previously established in the initial study design.

2.6 Model Evaluation

The evaluation was performed with cross-validation stratified into 5 times, ensuring proportional distribution of classes in each division. The following metrics were used: Overall accuracy; Macrosensitivity (average of true positive rates per class); Average specificity between classes; Multiclass AUC (calculated with the one-vs. rest strategy). In addition, confusion matrices and comparative graphs were generated to visualize the error and balance patterns between the classes, as explained below.

3 Results and Discussion

This section presents the quantitative and visual results obtained for the three models tested. The metrics were calculated based on the average of the 5x stratified cross-validation applied to the test set. Table 1 summarizes the average performance of each model, while Figures 1 and 2 illustrate graphical comparisons and error patterns.

Table 1. Average performance per model.

Model	Accuracy	Average Sens.
RF	89,4%	80,0%
SVM	88,5%	81,3%
MLP	87,8%	77,7%

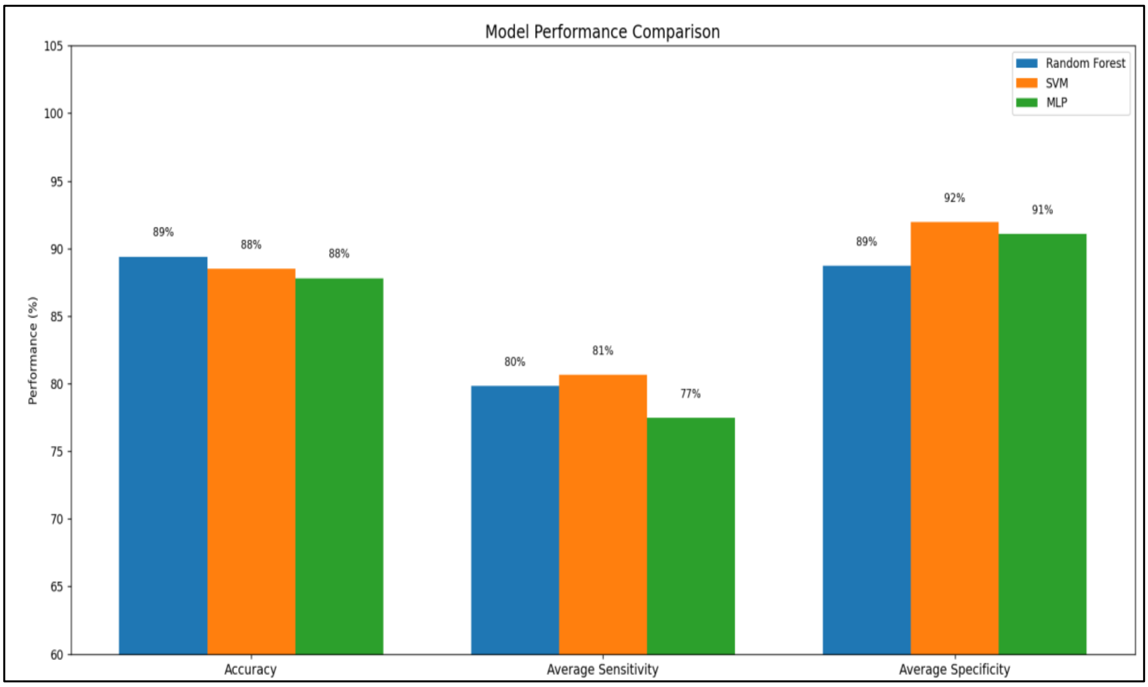

Figure 1 — The graphical comparison shows the robustness of the SVM, with a balance between sensitivity and specificity. Random Forest excels in overall accuracy but fails to capture nuances of risk. MLP, on the other hand, is more conservative and tends to overestimate uncertainties, such as high risk.

The graph shows the overall balance of SVM, which maintains both sensitivity and specificity, Random Forest has the highest overall accuracy, but with the lowest sensitivity, which indicates a possible limitation in detecting more subtle cases. The MLP, on the other hand, although it presents similar metrics, demonstrates a conservative bias, being more likely to classify uncertain cases as "moderate" or "high risk" – a relevant characteristic in preventive contexts.

Model Performance Metrics

Model	Accuracy	Sens_Low	Sens_Moderate	Sens_High	Spec_Low	Spec_Moderate	Spec_High
Random Forest	0.894	1.0	0.895	0.5	0.678	0.995	0.988
SVM	0.885	0.961	0.808	0.65	0.833	0.964	0.961
MLP	0.878	0.966	0.708	0.65	0.812	0.989	0.931

Figure 2 — Graphical comparison of the metrics between the model (%), demonstrating the harmony of the SVM, which combines high sensitivity and specificity. In contrast, Random Forest achieves maximum global accuracy, but with lower sensitivity to critical classes. The MLP, although effective at the extremes of risk, reveals greater imprecision in the intermediate range.

The overall balance of SVM is evident, which maintains high sensitivity and specificity, contrasting with Random Forest, which has peak accuracy but lower sensitivity for high-risk classes. The MLP, in turn, has good specific performance in "low" and "high" risk but tends to make more mistakes in the intermediate class.

As observed below, by applying confusion matrices, the SVM distributes its hits more evenly among the three classes, with few cross-errors. Random Forest tends to underestimate severe cases, confusing "high risk" with "moderate" or even "low." MLP, on the other hand, has a conservative profile — attributing greater risk in ambiguous situations, which can be useful in preventive contexts. In addition, all models had AUC values higher than 85%, which reinforces their ability to correctly discriminate risk levels, even in the most subtle and borderline clinical cases. These results indicate that the three algorithms are viable for application in screening contexts, with variations in the decision profile that should be considered according to the purpose of the system.

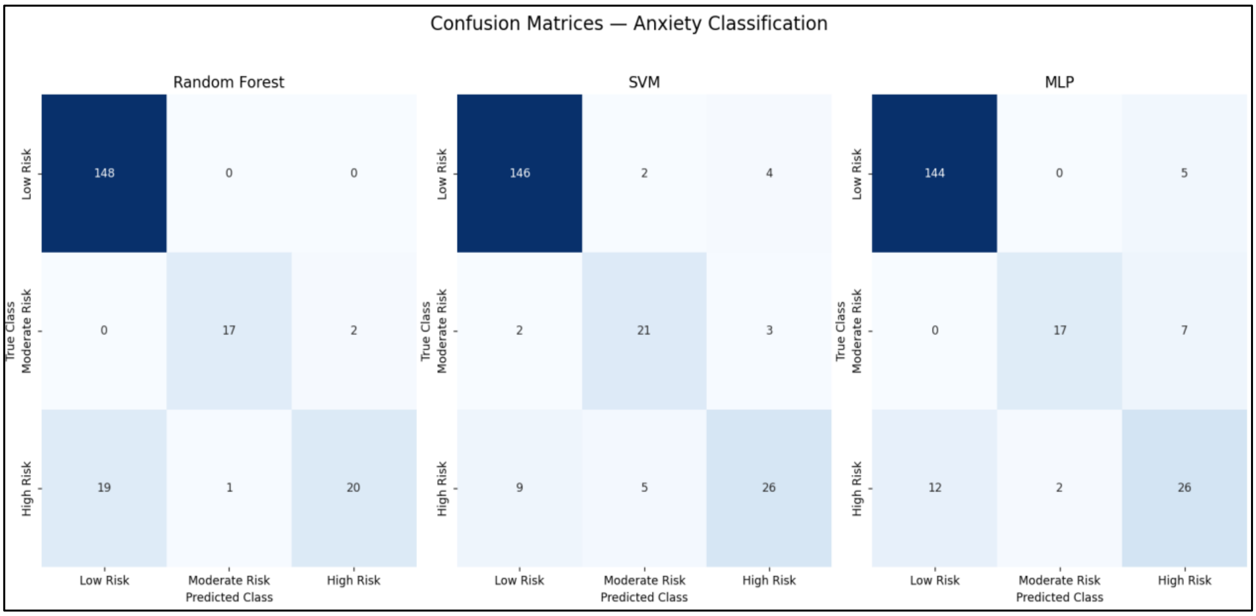

Figure 3 — The confusion matrices indicate that the SVM distributes the correct answers well among the classes, while the Random Forest tends to underestimate severe cases. MLP adopts a conservative bias, which is useful in preventive screening. With AUC above 85%, all models demonstrate good discriminative capacity, being viable for application according to the desired profile.

The results obtained demonstrate that the three models tested — Random Forest, SVM and MLP — presented robust performance in the multilevel risk classification task for anxiety disorders in children. The Random Forest algorithm achieved the highest average accuracy (89.4%), evidenced by its stability and ability to generalize. However, its sensitivity was lower than that of the SVM, which stood out for its balance between sensitivity (81.3%) and specificity (91.9%), being the most indicated when seeking to maximize both the identification of positive cases and the reduction of false positives. The MLP, although with a slightly lower performance in the global averages, showed a conservative trend in the classification of cases with ambiguous symptoms as moderate or high risk — a relevant characteristic in preventive scenarios. Confusion matrices reveal different error profiles between models. While the SVM distributed their hits better among the three classes, Random Forest showed a tendency to underestimate severe cases and the MLP to overestimate them. This diversity suggests that, depending on the context of applications such as screening in schools, Basic Health Unit or specialized clinics, different models may be preferred, or even combined in decision committees.

The present research was designed as a classification into multiple levels of risk; However, the analysis of the confusion matrices shows a challenge typical of multiclass problems, marked by asymmetric distribution between categories. To reduce this effect, stratified cross-validation and the use of macro-metrics, such as macro-sensitivity, macro-specificity, and macro-F1, which balance the evaluation between classes with unequal sizes, were adopted. It was observed, for example, that Random Forest tended to underestimate the high-risk category, while MLP showed slight overestimation. These different error profiles reinforce the relevance of more advanced techniques for treating class imbalance and incorporating differentiated classification costs, such as class weighting or cost-sensitive loss functions, which will be considered in future studies to reduce the impacts of false negatives and false positives in clinical scenarios.

Based on the results obtained, it is observed that the present work surpasses previous studies in proposal, innovation and practical applicability. By adopting a multilevel classification approach, using accessible physiological variables and versatile algorithms, this study advances beyond binary models based only on structured interviews. In challenging metrics, such as sensitivity and AUC, the proposed models remained at an excellent level, with AUC above 85%, evidencing the robustness of the approach. Considering the broader scope of the problem — with more classes, greater diversity of data, and focus on public display environments — the results reinforce the innovative and pioneering character of the proposal.

4 Conclusions and Future Studies

This study demonstrated the feasibility of using machine learning techniques in the multilevel classification of the risk of anxiety disorders in children, based on physiological and behavioral data. The three models tested - Random Forest, Support Vector Machine and Multilayer Perceptron - showed high performance, with accuracy above 87% and AUC values above 85%, which reinforces their predictive efficacy even in the face of a subjective and multiclass clinical problem.

The SVM model stood out for its balance between sensitivity and specificity, being particularly useful in clinical contexts where it is essential to minimize both false negatives and positives. Random Forest demonstrated good robustness and interpretability, while MLP showed conservative behavior, which may be interesting for preventive screening in environments with high demand and few resources. In addition to quantitative performance, the main contribution of this study lies in its unprecedented proposal for automated screening based on objective biomarkers, which expands the possibilities of application in public contexts, such as schools and primary health care units.

Future research may explore the inclusion of longitudinal data to assess the evolution of symptoms over time. Additionally, incorporating deep learning techniques such as convolutional neural networks can increase the accuracy of models. Future studies may also investigate the application of these methods in different age groups and clinical contexts, expanding their applicability in child mental health. The approach is scalable, replicable, and aligned with the Brazilian reality, which makes it a promising alternative for children's mental health policies. Future studies may explore the integration of these models into automated screening systems to broaden their clinical applicability.

Despite the high performance achieved by the models, it is pertinent to point out that the study was developed from a relatively small contingent of participants — 193 children — and with data obtained in a standardized clinical environment. Such a controlled context favors experimental rigor; however, it does not fully encompass the breadth and complexity of anxiety manifestations in everyday situations. Future investigations should contemplate the incorporation of other data sets — preferably from different geographic regions and natural settings, such as schools and primary health care facilities — to increase the robustness of the models and assess their generalizability in more diverse contexts.

Acknowledgements. Author expresses gratitude to God; to her husband Yuri and her children Henrique and Yasmin; to the Graduate Program in Electrical Engineering at the Federal University of Maranhão; to the advisor Prof. Dr. Allan Kardec Dualibe Barros Filho and to the co-advisor Prof. Dr. Jonathan Araújo Queiroz. Special thanks to Harvard Dataverse for making the data available and, above all, to the children and their families. The present work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) – Financing Code 001.

Authorship statement. The author confirms that she is solely responsible for the authorship of this work and declares that all material included in this article is her intellectual property or has the proper permission of the owners for its inclusion.

References

[1]Carpenter, K. L. H. Preschool Anxiety Dataset. *Harvard Dataverse*, 2016. Available at: https://doi.org/10.7910/DVN/4X8HAY. Accessed on: May 17, 2025.

[2] G.V. Polanczyk et al., "Annual Survey Review: A Meta-Analysis of the Worldwide Prevalence of Mental Disorders in Children and Adolescents," *Journal of Child Psychology and Psychiatry*, vol. 56, nº 3, pp. 345–365, 2015.

[3] P. Muris e H. Merckelbach, "Symptoms of Anxiety Disorders: Behavioral Symptoms in Normal Children", *Behavior Research and Therapy*, vol. 36, pp. 995–1009, 1998.

[4] A.E. Kazdin, "Addressing the Treatment Gap: A Key Challenge to Extending Evidence-Based Psychosocial Interventions". *Behaviour Research and Therapy,* vol. 88, pp. 7–18, 2017.

[5] K.L.H. Carpenter et al., "Preschool anxiety disorders predict different patterns of amygdala-prefrontal connectivity at school age*", PLoS ONE,* vol. 14, nº 6, e0217363, 2019. Available at: https://doi.org/10.1371/journal.pone.0217363.

[6] L. Breiman, "Random Forests", *Machine Learning*, vol. 45, pp. 5–32, 2001.

[7] C. Cortes e V. Vapnik, "Support vector networks", *Machine Learning*, vol. 20, pp. 273–297, 1995.

[8] I. Goodfellow, Y. Bengio e A. Courville, *Deep Learning*, Cambridge: MIT Press, 2016.

[9] *Center for Disease Control and Prevention*, "Data and statistics on children's mental health", 2023. Available at: https://www.cdc.gov/childrensmentalhealth/data.html. Accessed on: May 17, 2025.

[10] Brasil. *Fundação Oswaldo Cruz*, "Study shows growth in mental health problems in children during the pandemic", Rio de Janeiro: FIOCRUZ, 2021. Available at: https://portal.fiocruz.br. Accessed on: May 17, 2025.