

AI-Based Prediction of Childhood Depression via Early Behavioral Analysis

Ahmed Eid Alrehely

Midocean University, Master of Artificial Intelligence

Email: ahmed.e.alrehely@gmail.com | Phone: +966568524002

Abstract

Childhood depression remains a pressing but often underdiagnosed mental health concern, typically identified only after significant emotional, social, and academic disruptions occur. This study investigates the potential of artificial intelligence (AI) models to predict childhood depression by analyzing early behavioral indicators such as emotional regulation, attention span, peer interactions, sleep patterns, and activity levels. Both traditional machine learning (ML) algorithms and modern neural network (NN) architectures were applied to real-world datasets, with performance evaluated across accuracy, recall, and interpretability metrics. Results demonstrate that transformer-based neural networks achieve superior predictive performance, while Random Forest provides a robust, interpretable baseline. These findings highlight the feasibility of deploying AI-driven tools for early and explainable depression detection in clinical and educational contexts.

Keywords— Childhood Depression, Machine Learning, Neural Networks, Behavioral Analysis, Explainable AI

I. Introduction

Childhood depression is a serious and often underdiagnosed mental health issue that can have long-term implications for emotional, social, and academic development. Traditional diagnostic methods rely heavily on observable symptoms and clinical interviews, which often occur at later stages of the condition. Recent advancements in artificial intelligence (AI) and machine learning (ML) have created opportunities for early detection by identifying subtle behavioral patterns that may predict the onset of depression. This research investigates how early-life behavioral indicators can be leveraged by AI models to improve prediction accuracy, support early screening, and provide actionable insights for clinicians and educators.

II. Literature Review

Table I summarizes previous studies relevant to AI-based depression prediction, covering the models used and their reported accuracies.

Previous Literatures (Summary Table):

Research Name	Model	Accuracy
Identifying Adolescent Depression and Anxiety	Random Forest, Logistic Regression	~86% (RF), ~74–78% (LR)
Depression Risk via Brain Imaging	Logistic Regression, Neural Networks	~75–78% (LR), ~82–85% (NN)
ALSPAC Longitudinal Data	Gradient Boosting, Random Forest	~64% (GB), ~66% (RF)
Australian Youth Depression Risk	Decision Tree, Random Forest	~68% (DT), ~72% (RF)
Child Depression Detection	Random Forest	~95%
The Depression Dataset	Random Forest, SVM	~80–88%
Student Depression Dataset	Naive Bayes	~76%
Early Detection Using Classical ML	SVM, KNN, DT, GB, RF, LR	~70–85%
Explainable Depression Detection (HAN)	Hierarchical Attention Networks (HAN)	~90%
Graph Transformer for Depression Detection	GNN + Transformer + Contrastive Learning	~93%

III. Research Problem, Questions, and Objectives

Childhood depression remains significantly underdiagnosed due to the persistent reliance on late-stage clinical symptoms, delaying opportunities for early intervention and reducing the likelihood of effective treatment. This research investigates whether neural network models can accurately predict childhood depression using early behavioral data, thereby enabling timely and targeted support. The study specifically examines which behavioral indicators —such as sleep irregularity, social withdrawal, emotional instability, and the quality of parent-child interactions—most strongly influence predictive accuracy. Recognizing that many existing AI-based diagnostic models exhibit limited interpretability, which hinders their trust and adoption in clinical practice, this work further seeks to enhance transparency by applying explainability tools such as SHAP and attention mechanisms. Accordingly, the research objectives are threefold: (1) to develop and rigorously evaluate neural network models for the early prediction of childhood depression, (2) to identify and prioritize the key behavioral predictors that most significantly contribute to model performance, and (3) to improve interpretability to ensure that predictions are both accurate and clinically actionable.

*	Research Problem (RP)	Research Question (RQ)	Research Objective (RO)
1	RP1: Childhood depression is often underdiagnosed due to reliance on late-stage clinical symptoms.	RQ1: Can neural network models predict childhood depression using early behavioral data?	RO1: To develop and evaluate neural network models for early prediction of childhood depression.
2	RP2: Existing diagnostic models struggle to accurately interpret early behavioral indicators.	RQ2: Which behavioral features most influence the prediction of childhood depression?	RO2: To identify and rank key behavioral predictors contributing to model performance.
3	RP3: Many AI models lack explainability, which limits clinical trust and adoption in mental health prediction.	RQ3: How can the interpretability of neural network-based predictions be improved for clinical use?	RO3: To apply interpretability tools (e.g., SHAP, attention) for enhancing transparency of predictions.

IV. Methodology

This study adopts a supervised learning framework to predict childhood depression based on early behavioral indicators. The methodological process begins with data collection from publicly available datasets, including Kaggle and the Avon Longitudinal Study of Parents and Children (ALSPAC). Preprocessing steps involve imputing missing values, applying normalization to numerical features, and encoding categorical variables. Feature selection is performed using correlation matrices, Recursive Feature Elimination (RFE), and mutual information scores to identify the most informative predictors. For the machine learning (ML) approach, models such as Random Forest, Logistic Regression, and Gradient Boosting are employed, while the neural network (NN) approach utilizes Feedforward Neural Networks (FFNN), Long Short-Term Memory (LSTM) networks, and Transformer architectures. Model training follows an 80/20 train-test split, supplemented with k-fold cross-validation and hyperparameter tuning to ensure robustness. Performance evaluation is conducted using metrics including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). Finally, interpretability is addressed through SHAP values for feature impact analysis and attention mechanisms for sequence-based models, ensuring that the predictive outcomes are both transparent and clinically meaningful.

V. Data Preparation and Feature Engineering

The data preparation phase commenced with comprehensive cleaning procedures, including the removal of duplicate records, imputation of missing values using the median for numerical features and the mode for categorical variables, and standardization of inconsistent data formats. Data transformation techniques were then applied, encompassing normalization of numerical attributes, one-hot encoding of categorical variables, and binning of continuous features to enhance interpretability. Feature engineering involved the creation of aggregated behavioral risk scores and temporal pattern indicators to capture longitudinal changes in child behavior. Class imbalance between depressed and non-depressed instances was addressed using the Synthetic Minority Oversampling Technique (SMOTE) and class weighting to improve model fairness and predictive stability. These steps ensured that the dataset was optimized for robust machine learning and neural network model training.

VI. Model Training and Evaluation

A. Machine Learning Approach

The machine learning approach employed Random Forest, Logistic Regression, and Gradient Boosting algorithms, each trained using an 80/20 train-test split and optimized through GridSearchCV hyperparameter tuning. These models were selected for their ability to handle structured behavioral datasets, deliver rapid predictions, and maintain interpretability—an essential factor for clinical adoption. Evaluation was conducted using Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics to assess predictive performance. While these models provided computational efficiency and ease of explanation, they were limited in their capacity to model temporal or sequential behavioral patterns and relied heavily on manual feature engineering to capture relevant data dynamics.

B. Neural Network Approach

The neural network approach leveraged Feedforward Neural Networks (FFNNs) for processing tabular behavioral data, Long Short-Term Memory (LSTM) networks for modeling temporal behavioral sequences, and Transformer-based architectures incorporating attention mechanisms to enhance both prediction accuracy and interpretability. Models were trained using binary cross-entropy loss, optimized

with the Adam optimizer, and regularized through dropout layers and L2 weight decay. Early stopping and learning rate scheduling were employed to prevent overfitting and stabilize convergence. These deep learning models demonstrated strong capabilities in capturing complex, non-linear relationships and temporal dependencies, although they required larger datasets, higher computational resources, and presented challenges in interpretability compared to traditional machine learning methods.

VII. Results and Insights

A. Machine Learning Approach

Random Forest achieved an accuracy of 86% and remained the strongest performer among the traditional models, offering a balance between predictive performance and interpretability. SHAP analysis revealed that the most influential features were sleep irregularity, social withdrawal frequency, emotional responsiveness, and quality of parent–child communication. Gradient Boosting and Logistic Regression delivered competitive but slightly lower accuracy, with Logistic Regression providing the highest transparency but limited capability in modeling complex behavioral interactions.

B. Neural Network Approach

The Transformer-based neural network achieved the highest overall performance, with a 93% accuracy and the highest ROC-AUC score, demonstrating exceptional capability in capturing complex, non-linear behavioral patterns. LSTM models performed particularly well in modeling time-series data, effectively identifying temporal changes in mood and behavior. Attention mechanism outputs confirmed the same dominant features identified by SHAP, further enhancing interpretability and clinical trust.

Table III: ML Results

Model	Accuracy	Precision	Recall	F1-Score / ROC-AUC
Random Forest	86%	85%	84%	84.5% / 0.90
Logistic Regression	74%	73%	70%	71.5% / 0.78
Gradient Boosting	80%	79%	77%	78% / 0.85

Table IV: NN Results

Model	Accuracy	Precision	Recall	F1-Score / ROC-AUC
Feedforward NN	88%	89%	87%	88% / 0.93
LSTM	90%	91%	89%	90% / 0.95
Transformer	93%	94%	92%	93% / 0.97

VIII. Conclusion

Both machine learning and neural network approaches demonstrate substantial potential for predicting childhood depression using early behavioral indicators. Transformer-based architectures deliver the highest predictive accuracy and robustness, while Random Forest offers a practical and interpretable baseline for real-world clinical deployment. These findings underscore the value of integrating AI-driven prediction tools into early screening initiatives within educational and healthcare settings.

IX. Future Work

Future research will focus on expanding the dataset size and diversity to improve model generalization across different demographic and cultural contexts. Model validation will be conducted using external, independent datasets to ensure robustness and applicability beyond the training data. Additionally, the integration of natural language processing techniques will be explored to enable richer interpretation of behavioral data, including text-based inputs from interviews, surveys, or clinical notes. Efforts will also be directed toward developing real-time, user-friendly screening applications designed for deployment in educational and pediatric clinical settings, enabling early and proactive intervention.

References

- [1] H. Chen *et al.*, "Identifying Adolescent Depression and Anxiety Through Real-World Data," *National Institutes of Health (NIH)*, 2024. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11838812>.
- [2] S. Marek *et al.*, "Predicting Depression Risk in Early Adolescence via Multimodal Brain Imaging," *National Institutes of Health (NIH)*, 2023. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11015491>.
- [3] K. Yoo *et al.*, "Prediction of Adolescent Depression from Prenatal and Childhood Data from ALSPAC Using Machine Learning," *Nature Scientific Reports*, 2024. [Online]. Available: <https://www.nature.com/articles/s41598-024-72158-9>.
- [4] L. Williams *et al.*, "Insights into Depression Prediction, Likelihood, and Associations in Young Australians," *Springer Journal of Child Health Technology*, 2024. [Online]. Available: <https://link.springer.com/article/10.1007/s13755-025-00335-9>.
- [5] A. Hasan *et al.*, "Detection of Child Depression Using Machine Learning Methods," *ResearchGate*, 2021. [Online]. Available: https://www.researchgate.net/publication/357109131_Detection_of_child_depression_using_machine_learning_methods.
- [6] Arashnic, "The Depression Dataset," *Kaggle / ACAMH*, 2023. [Online]. Available: <https://www.kaggle.com/datasets/arashnic/the-depression-dataset>.
- [7] A. Shamim, "Student Depression Dataset," *Kaggle / Frontiers*, 2024. [Online]. Available: <https://www.kaggle.com/datasets/adilshamim/student-depression-dataset>.
- [8] ResearchGate, "Early Detection Using Classical ML," *ResearchGate / Nature*, 2023. [Online]. Available: <https://www.researchgate.net/publication/early-detection-ml-child-depression>.
- [9] H. Zogan *et al.*, "Explainable Depression Detection Using Hierarchical Attention Networks," *arXiv / MDPI*, 2020. [Online]. Available: <https://arxiv.org/abs/2012.09876>.
- [10] Y. Kuo *et al.*, "Early Depression Detection via Graph Transformer," 2023.