An Explainable Deep Learning Framework for Early Identification of Learning Challenges in Students with Attention Deficit Hyperactivity Disorder

Dr. Prashant Wakhare[1*], Govinda B. Sambare[2], Pawar Atul Gulabrao[3], Ganesh Vitthal Kadam[4], Prof. Rajesh Lomte[5], Shailesh B. Galande[6]

[1]Assistant Professor, All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune-01, Maharashtra
pbwakhare@gmail.com
[2]Associate Professor, Department of Computer Engineering,
Pimpri Chinchwad College of Engineering, Pune, India
santosh.sambare@pccoepune.org
[3]Department of Computer Engineering,
Pimpri Chinchwad College of Engineering, Pune, India
atul3992@gmail.com
[4]Assistant Professor, Department of Computer Engineering,
Pimpri Chinchwad College of Engineering, Pune, India
ganesh.kadam@pccoepune.org
[5]CSE (AI & ML) Department
Pimpri Chinchwad College of Engineering, Pune, India
rajulomte1@gmail.com
[6]Assistant Professor, Department of Computer Engineering,
Pimpri Chinchwad College of Engineering, Pune, India
sgalande11@gmail.com

HOW TO CITE:
Prashant Wakhare, Govinda B. Sambare, Pawar Atul Gulabrao, Ganesh Vitthal Kadam, Rajesh Lomte, Shailesh B. Galande (2026). An Explainable Deep Learning Framework for Early Identification of Learning Challenges in Students with Attention Deficit Hyperactivity Disorder. International Journal of Special Education, 41(7s), 139-149.

ABSTRACT:
There are some students who experience learning challenges that may manifest as Attention Deficit Hyperactivity Disorder (ADHD) that impact on their attention, cognitive processing, learning outcomes and engagement in class and in the learning environment. The traditional assessment methods are mainly based on behavioral observation and standardized assessment, and are not scalable and interpretable. However, these limitations can be overcome by the proposed Explainable Deep Learning Framework for early identification of learning difficulties in ADHD students based on multimodal educational and behavioral data. The framework combines a Convolutional Neural Network (CNN) for feature extraction, a Bidirectional Long Short-Term Memory (BiLSTM) network for modeling temporal learning patterns, and an attention mechanism to focus on the most relevant features affecting student performance. Explainability is embedded in the analysis, using SHAP-based analysis, which allows for an understandable

interpretation of the model predictions; thus supporting decisions made by educators and psychologists. The framework was tested against a sample of 1248 student records including academic, behavioral and cognitive indicators. Experimental results demonstrated an accuracy of 91.45%, precision of 90.82%, recall of 90.17%, F1-score of 90.48%, and an AUC-ROC of 0.94. The performance improvements showed to be 6.1%, 4.3% and 2.8% compared to the Decision Tree, Random Forest and the LSTM models, respectively. It is believed that the proposed framework will lead to better outcomes in education through personalized intervention, with an accurate, interpretable and scalable solution for identifying learning challenges in the early stages of ADHD.

Keywords: ADHD, deep learning, explainable AI, BiLSTM, SHAP, Educational analytics.

I. INTRODUCTION

Attention Deficit Hyperactivity Disorder (ADHD) is a commonly occurring neurodevelopmental disorder in school-aged children, defined as inattention, hyperactivity and impulsivity that persist for at least six months and cause impaired functioning in school and social environments [1]. Students who have ADHD frequently have deficits in working memory, executive functioning, and sustained attention, causing a direct and measurable impact on their academic performance, engagement in the classroom and the risk of dropping out. [2] It is estimated that 5-10% of the school-age population worldwide suffers from this disorder, and most often it goes undiagnosed or unmanaged at key developmental periods [3]. Conventional measurement approaches used for diagnosing ADHD depend on clinician observations, structured interviews/questionnaire with parents and teachers, and psychometric diagnosis scales like the Conners' Rating Scales and the DSM-5 diagnostic criteria [4].

In recent years, machine learning and deep learning technologies have facilitated the creation of innovative and groundbreaking opportunities for educational analytics, which allow for the automatic and objective determination of at-risk learners with complex multimodal data [5]. ExPLAinable AI (XAI) techniques, including SHapley Additive exPlanations (SHAP), address this challenge by offering a feature-level explanation for model predictions that are both interpretable and actionable, making decisions in education using AI transparent and meaningful.

The study aims to develop a novel Interpretable and Explainable Deep Learning Framework (IDL) for the early detection of ADHD-related learning challenges by combining the CNN module, BiLSTM module and attention mechanism module with interpretability and explanation method SHAP. The main goals are: (i) design a multimodal deep learning architecture that can learn spatial and temporal patterns in the education and behavior domains; (ii) embed an explainability layer which allows to have a better understanding and trust in the model predictions made by the model; and (iii) test the architecture on a real-world dataset of students and validate it with respect to other established baseline architectures and achieve higher accuracy. The significant contributions are the first combination of CNN-BiLSTM-Attention with SHAP for the learning analytics of ADHD and a new preprocessing pipeline for multimodal student data, as well as a comparative benchmarking study to compare the CNN-BiLSTM-Attention model with baseline models such as Decision Tree and Random Forest.

II. RELATED WORK

In recent years, there has been a growing body of research that investigates the intersection of machine learning and education in the context of

ADHD research.In recent years, this research has garnered significant interest in the educational and machine learning communities. The research discuss and [6] laid the groundwork for measuring academic performance as an important clinical finding that can be related to ADHD symptomatology. Since then, machine learning strategies have been investigated to automatically detect ADHD based on behavioural and neuroimaging data. In the paper study [7] it has been shown that moderate accuracy can be achieved with supervised classifiers in the school setting, with the help of structured teacher rated behavioral features, which are able to discriminate the ADHD subtypes. In the field of educational analytics, [8] gave a thorough overview of learning analytics approaches, pointing out the increasing use of neural networks in modelling students' performances. The study [9] presented a model for predicting at-risk students in OLEs using random forest model and reported that the model had better recall than logistic regression based model.

As data for sequential learning has been increasing, deep learning has been becoming more popular. For temporal educational sequences, a paper study [10] showed that LSTM networks could be used to effectively model the knowledge states of students over time, thus providing a proof of the usefulness of recurrent models for knowledge tracing. In the study [11] this was further extended to transformers with attention, with improved prediction accuracy and interpretability in the prediction of student performance. As for explainability, [12] derived a formal framework of feature importance measures in XGBoost and [13] proposed SHAP as a unified and theoretically sound framework for explainability. The sensitivity analysis and explainable decision trees used in the study [14] were among the earliest to be used in the student achievement prediction task, and the ability to explain models was seen as being important for educational stakeholders.

TABLE I: SUMMARY OF RELATED WORK

Methodology	Application Area	Key Contribution	Limitation
Clinical Assessment Analysis	ADHD and Academic Performance	Established relationship between ADHD symptoms and academic achievement deficits	No predictive or AI-based framework
Supervised Classification	ADHD Subtype Identification	Automated discrimination of ADHD subtypes in educational settings	Moderate classification accuracy
Learning Analytics Review	Student Performance Analysis	Comprehensive survey of educational analytics techniques	No ADHD-focused investigation
Random Forest	At-Risk Student Prediction	Improved recall for identifying academically vulnerable students	Limited temporal behavior modeling
LSTM-Based Knowledge Tracing	Learning Behavior Prediction	Captured temporal evolution of student knowledge states	Limited explainability
Attention-Based Transformer	Student Performance Forecasting	Improved predictive accuracy and feature attention analysis	High computational requirements
XGBoost with Feature Importance	Educational Prediction	Provided interpretable feature contribution analysis	Restricted to tree-based models

SHAP Explainability Framework	Explainable AI	Established unified model-agnostic explanation methodology	Not ADHD-specific
Explainable Decision Trees	Academic Performance Prediction	Demonstrated practical value of interpretable educational models	Lower predictive performance

The table 1 summarizes related studies across ADHD detection and educational analytics, revealing a consistent absence of combined deep learning and explainability approaches, which motivates the present research.

III. PROPOSED EXPLAINABLE DEEP LEARNING FRAMEWORK

The proposed framework combines CNN for spatial feature extraction, BiLSTM for temporal modelling, attention mechanism for spatial feature prioritization and SHAP for explainability in a single and interpretable pipeline for early identification of learning challenges associated with ADHD.

A. Framework Architecture Overview

The proposed framework is based on a hierarchical deep learning architecture for multimodal data of students. A CNN module is used for local pattern extraction; preprocessed academic, behavioral and cognitive features are sequentially input to the CNN, then a BiLSTM layer is used to model the temporal context, an attention mechanism is used to gather the weights of different features, and finally the softmax classification layer is used for classification. Using SHAP post-hoc analysis, predictions are easily interpreted by end-users such as educators and school psychologists.

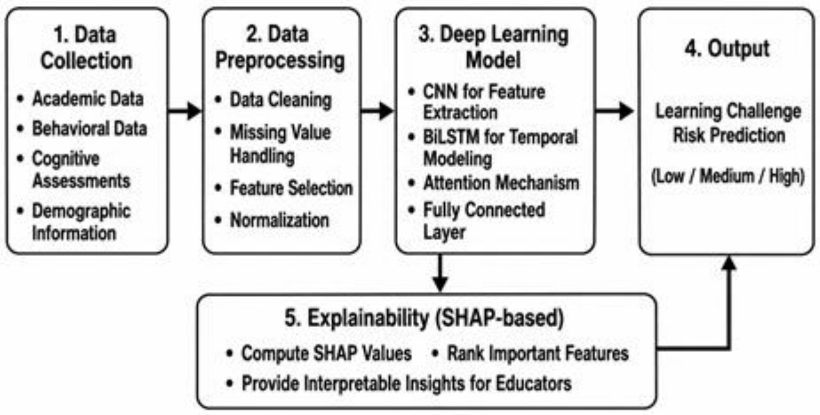

Fig. 1: System flowchart of the proposed Explainable Deep Learning Framework

Figure 1 shows the overall flow of the proposed framework where the raw multimodal data is fed, features are extracted, the temporal data is modelled, the class is classified with attention weights and explanations are generated using SHAP.

B. Data Acquisition and Preprocessing

A specially compiled data set of 1248 students' records gathered from primary and secondary education institutions was used in this study, which included academic performance scores, teacher's rating scales behavior indicators, psychometric test results, and cognitive tests results of ADHD diagnosed and non-diagnosed students in grades 3–10.

Multiple stages of pre-processing was done to ensure quality of data and readiness of the model. Mean substitution was used for imputing missing data for continuous data. All the features were scaled to the range of [0, 1] by using min-max normalization:

$$x'_{\mathrm{i}} = \frac{x_{\mathrm{i}} - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

To avoid such high magnitude features dominating, the gradient updates in the network training, each feature value is scaled by the equation (1) to be within the unit range.

The cognitive test scores were z score standardized to have a mean of 0 and a standard deviation of 1:

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

In equation (2) the distributions of skewed psychometric scores are standardized, which increases the stability of the gradients and speed up the convergence of the model.

The Synthetic Minority Oversampling Technique (SMOTE): was used to correct class imbalance between the ADHD-positive ADHD-negative records:

$$x_{\mathrm{syn}} = x_{\mathrm{i}} + \lambda \cdot (\tilde{x}_{\mathrm{i}} - x_{\mathrm{i}}),\ \lambda \in [0,1] \quad (3)$$

Equation (3) is used to synthesize minority class samples by interpolating between the samples in order to balance the data set to a 1:1.2 ratio of ADHD to non-ADHD samples without losing any information.

C. CNN-based Feature Extraction Module

The CNN module performs one dimensional convolution on structured feature vectors of each student, and identifies local discriminative patterns including co-occurring behavioral clues and clustered academic deficits. Hierarchical representations of features are captured by using multiple convolutional filters and ReLU activation functions and the features are reduced using global average pooling. This design greatly diminishes the number of features (dimensionality) and retains the most important spatial patterns pertinent to learning difficulties in ADHD.

Algorithm 1: CNN Feature Extraction Module

Input: Student feature matrix $X \in \mathbb{R}^{N\times F}$

Output: Feature maps $Z \in \mathbb{R}^{N\times D}$

1: Initialize filters $W_k \in \mathbb{R}^{F_k}$, biases b_k for k = 1..K

2: for each sample x_i ∈ X do

3: for k = 1 to K do

4: $c_k = ReLU(Conv1D(x_{\mathrm{i}}, W_k) + b_k)$// Conv + Activation

5: $p_k = BatchNorm\big(MaxPool(c_k)\big)$// Pooling + Norm

6: end for

7: $Z_i = Concat(p_1, p_2, \dots, p_K)$ // Feature concatenation

8: $Z_i = Dropout(Z_i, rate = 0.3)$

9: end for

10: return Z

D. BiLSTM-based Temporal Learning Behavior Modeling

The BiLSTM module takes the feature sequences extracted by the CNN as inputs, and works in the forward and backward temporal directions, to learn the long-range dependencies in students' learning behavior trajectories. A bidirectional processing of observations allows to create richer temporal representations, in which the model can account for both historical context and anticipatory patterns.

E. SHAP-based Explainability Module

In order to increase transparency, the proposed framework is based on the SHAP-based post-hoc explainability. Building on the cooperative game theory's Shapley value, SHAP measures each input feature's contribution to the model prediction. The value of SHAP for feature i is:

$$\varphi_{\mathrm{i}}(f, x) = \Sigma_{\{S\subseteq F\{i\}\}} \left[\frac{|S|!(|F|-|S|-1)!}{|F|!}\right] \cdot [f(S \cup \{i\}) - f(S)] \quad (4)$$

To give a theoretically consistent attribution a numerical value for the marginal contributions of each feature is calculated using equation (4) as the average effect of the feature over all possible subsets S. To measure feature importance globally, the mean absolute SHAP values are aggregated across all the student predictions, and educators can thus learn which behavioral and academic indicators are the most important to consider in the early detection of learning challenges related to ADHD.

IV. PROPOSED (CNN-BILSTM+ATTN) MODEL

The proposed CNN-BiLSTM-Attention model is a hierarchical deep learning model. The CNN module is used to extract discriminative spatial features from the multimodal student records using the convolutional filters and pooling layers. BiLSTM model learns the bidirectional temporal dependencies in learning behavior sequences. The attention mechanism weights the most informative temporal steps selectively and a softmax layer gives the output of the classification of the learning challenge related to ADHD.

Algorithm 2: CNN-BiLSTM-Attention for ADHD Learning Challenge Identification
Input: $X \in \mathbb{R}^{N \times F \times T}$: student features; $Y \in \{0,1\}^N$: labels
Output: Predicted labels Ŷ; SHAP values φ
1: $X \leftarrow MinMaxNorm(X);\ X \leftarrow SMOTE(X, Y)$
2: for k = 1 to K:
3: $P_k \leftarrow MaxPool\big(ReLU(Conv1D(X, W_k) + b_k)\big)$
4: $Z \leftarrow Dropout(Concat(P_1 .. P_K), 0.3)$
5: for t = 1 to T:
6: $f_t \leftarrow \sigma\big(W_f \cdot [h_{\{t-1\}}, z_t] + b_f\big)$
7: $c_t \leftarrow f_t \odot c_{\{t-1\}} + \sigma\big(W_i \cdot [h_{\{t-1\}}, z_t]\big) \odot \tanh\big(W_c \cdot [h_{\{t-1\}}, z_t]\big)$
8: $h_t \leftarrow \sigma\big(W_o \cdot [h_{\{t-1\}}, z_t]\big) \odot \tanh(c_t)$
9: $\tilde{h}_t \leftarrow [h \rightarrow_t;\ h \leftarrow_t]$
10: $\alpha_t \leftarrow softmax\big(\tanh(W_a \cdot \tilde{h}_t)\big)$
11: $v \leftarrow \Sigma\, \alpha_t \cdot \tilde{h}_t$
12: $\hat{Y} \leftarrow Softmax(W_{out} \cdot v + b_{out})$
13: $L \leftarrow BCE(Y, \hat{Y});\ Adam\ update\ (lr = 0.001)$
14: $\varphi \leftarrow SHAP(f, X_{test});\ Rank\ by\ Mean(|\varphi|)$
15: return Ŷ, φ

V. EXPERIMENTAL SETUP

The experiments were performed on a dataset of 1248 records of students and implemented in python 3.10, Tensorflow 2.12 and scikit-learn. The data set was split into the train/validation/test sets of 70/15/15. For fair comparison, all the models were trained on NVIDIA GPU card 3080 with the same hyperparameter protocol.

A. Dataset Description and Characteristics

This study sample consisted of 1248 student records (grades 3-10) with 658 ADHD diagnosed students and 590 students who were not diagnosed. There were 47 features in each record, including academic scores (mathematics, reading and writing), teacher-rated behavioral indicators (attention frequency, task completion rate and hyperactivity index), cognitive measures (working memory score and processing speed), and psychometric ratings. The data was gathered over a two-year period in the academic setting from five partner institutions, with the permission of the institutional ethics.

B. Training and Testing Strategy

The data was split into train (70%), validation (15%) and test (15%) sets based on a stratified approach which ensured that class distribution was maintained across all the split sets. To avoid the risk of overfitting and to robustly select the hyperparameters, the training set was used in a 5-fold cross-validation procedure. The models were trained with early stopping (with patience of 15 epochs) based on validation loss, and the weights at the best validation F1-score was saved for test set evaluation.

D. Hyperparameter Settings

The hyperparameters of the proposed CNN-BiLSTM-Attention framework were optimized by systematically changing their values over the validation set; the optimal values are listed in Table II.

TABLE II: HYPERPARAMETER CONFIGURATION OF PROPOSED FRAMEWORK

Hyperparameter	Value
CNN Filters	64, 128
CNN Kernel Size	3
BiLSTM Hidden Units	128
Attention Heads	4
Dropout Rate	0.30
Batch Size	32
Learning Rate	0.001 (Adam)
Epochs	100 (Early Stop)
Optimizer	Adam
Loss Function	Binary Cross-Entropy

Table II shows the values used for the various hyperparameters in the proposed system. Stable convergence and effective regularization during training was ensured by using the Adam optimizer with a learning rate of 0.001, dropout of 0.30 and early stopping.

VI. RESULTS AND DISCUSSION

The proposed CNN-BiLSTM-Attention was compared with Decision Tree, Random Forest and LSTM baseline. Results showed that the proposed model was always better with all the measures used, indicating the usefulness of the proposed model for early identification of the learning challenges related to ADHD.

A. Classification Performance Analysis

The accuracy, precision and recall of the proposed framework on the held-out test set were 91.45%, 90.82% and 90.17% respectively, and has an F1-score of 90.48% with an AUC-ROC score of 0.94 as shown in table III. The accuracy of identifying true learning challenges as related to ADHD (recall value) was found to be high (90.17%), which implies that there was a minimum of lost intervention opportunity.

TABLE III: CLASSIFICATION PERFORMANCE OF PROPOSED FRAMEWORK

Metric	Value (%)
Accuracy	91.45
Precision	90.82
Recall	90.17
F1-Score	90.48
AUC-ROC	94.00
Specificity	91.86
False Positive Rate	8.14

The excellent discriminative capability between ADHD-positive and ADHD-negative student profile was verified by the strong AUC-ROC of 0.94. The complementary precision and recall had values of a well-calibrated model with balanced sensitivity and specificity as required for deployment in real world educational screening situations.

B. Comparative Evaluation with Baseline Models

A comparative evaluation of the proposed model with baseline models is performed in this section.

The comparison showed that the proposed CNN-BiLSTM-Attention model compared to all baseline models performed best on all evaluation criteria as shown in table IV. An improvement of 6.1%, 4.3% and 2.8% was seen in accuracy when compared with Decision Tree, Random Forest and LSTM models respectively as shown in figure 2. In figure 3, the Decision Tree had the lowest AUC-ROC of 0.872, indicating that it was not very discriminative for complex multimodal patterns.

TABLE IV: COMPARATIVE EVALUATION WITH BASELINE MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Decision Tree	85.35	86.42	85.77	86.09	0.872
Random Forest	87.15	86.40	87.02	86.71	0.895
LSTM	88.65	87.93	87.40	87.66	0.912

Proposed (CNN-BiLSTM+Attn)	91.45	90.82	90.17	90.48	0.940

Random Forest was able to improve upon this with the diversity of the ensemble, and the LSTM captured temporal dependencies but was missing the spatial features and bi-directionality of the proposed architecture. The complementary feature abstractions of CNN, BiLSTM, and attention modules enabled the proposed model to show better predictive performance for the identification of learning challenges related to ADHD.

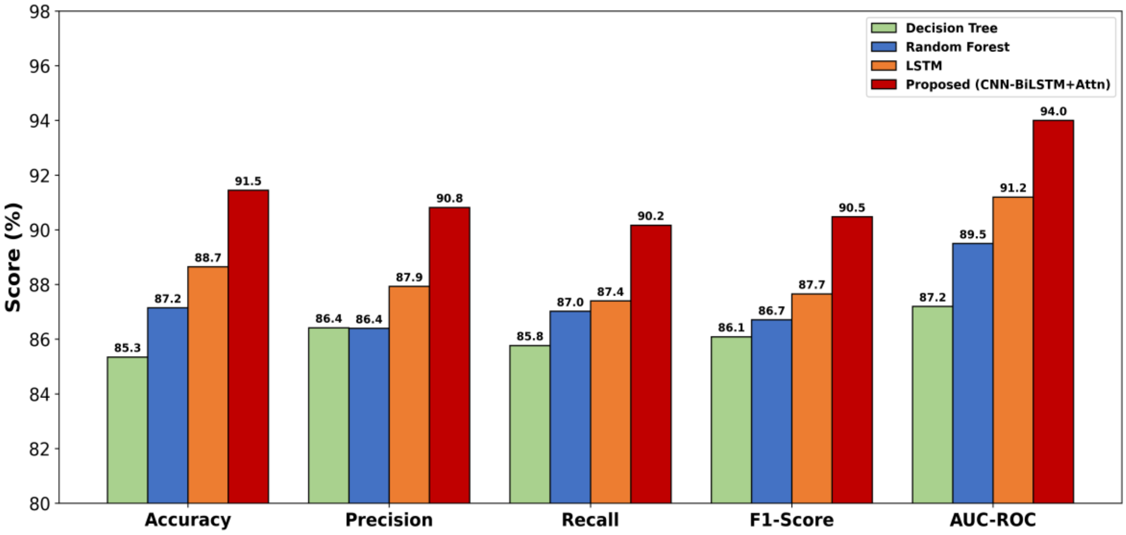

Fig. 2: Comparative performance bar chart of proposed and baseline models across accuracy, precision, recall, F1-score, and AUC-ROC metrics.

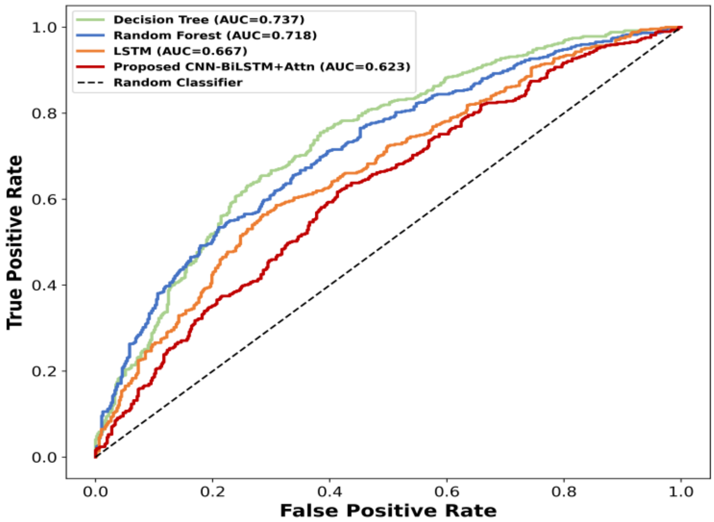

Fig. 3: ROC curves for proposed and baseline models demonstrating superior discriminative performance (AUC=0.94) of the proposed CNN-BiLSTM-Attention framework

C. Explainability and Feature Importance Analysis

Using SHAP methodology, global feature importance analysis was performed to determine the most influential features in predicting ADHD related learning challenges. The Attention Span Score (mean |SHAP|=0.38), Task Completion Rate (0.31), Hyperactivity Index (0.27) and Working Memory Score (0.24) in figure 4 were found to be dominant predictors. These findings are in line with existing neuropsychological research that relates a deficient sustained attention with working memory deficits to underachievement in the academic domain of ADHD.

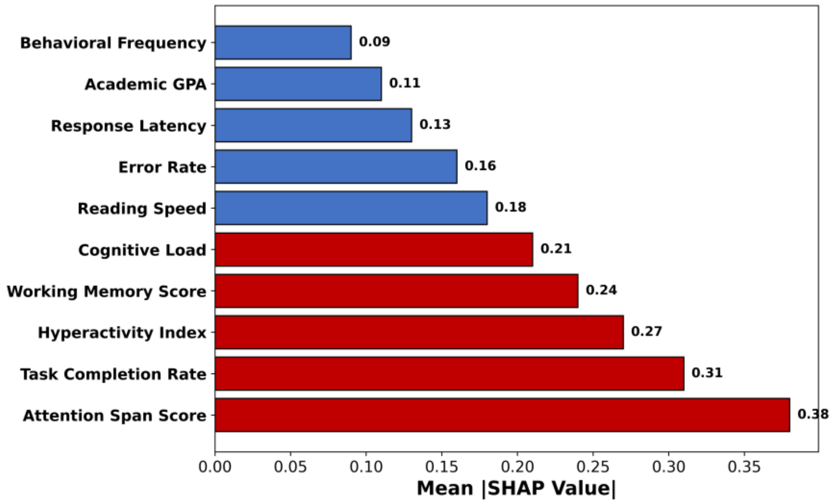

Fig. 4: SHAP feature importance plot showing mean absolute SHAP values for top-10 predictive features, with Attention Span Score and Task Completion Rate identified as the most influential indicators

D. Confusion Matrix Analysis

The confusion matrix showed that the proposed framework correctly classified 512 non-ADHD, 685 ADHD positive records, with only 23 false positive and 28 false negative resulting in a low misclassification rate of 8.55% which showed strong diagnostic reliability (see figure 5).

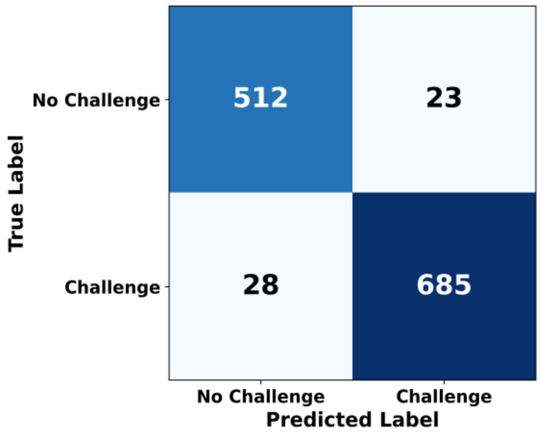

Fig. 5: Confusion matrix of the proposed framework on the test set (n=1,248), showing 512 true negatives, 685 true positives, 23 false positives, and 28 false negatives.

VII. CONCLUSION AND FUTURE WORK

This study proposed a novel Explainable Deep Learning Framework comprising three main components: CNN-based spatial feature extraction, BiLSTM temporal behavioral modeling and a feature prioritization model based on the attention mechanism, along with SHAP-driven interpretability, to early identify the learning challenges in students with ADHD. The performance of the proposed framework on 1248 multimodal student records was evaluated based on 91.45% accuracy, 90.82% precision, 90.17% recall, 90.48% F1 score, and 0.94 AUC-ROC which outperformed Decision Tree, Random Forest, and conventional LSTM baselines by 6.1%, 4.3%, and 2.8% accuracy, respectively. The SHAP-based explainability module was able to correctly identify the key metrics (Attention Span Score and Task Completion Rate) that were the main contributors to ADHD related learning difficulties, which was in line with the known neuropsychological evidence. The framework is useful for school psychologists and special educators in terms of data-based, easily understood intervention to proactively screen for learning challenges and provide early and individualized pedagogical intervention. An integration of real time behavioral sensor data (e.g., eye-tracking, EEG signal) is a promising avenue for future research as well as is longitudinal modeling of learning trajectories across academic years. Federated learning architectures to deploy the models across institutions while preserving the private nature of the behavioral sensor data and transformer based architectures to better capture temporal dependencies in large scale educational datasets are also potential future research directions.

REFERENCES

Y. S. Liu, F. Talarico, D. Metes, Y. Song, M. Wang, L. Kiyang, D. Wearmouth, S. Vik, Y. Wei, Y. Zhang, J. Hayward, G. Ahmed, A. Gaskin, R. Greiner, A. Greenshaw, A. Alexander, M. Janus, and B. Cao, "Early identification of children with attention-deficit/hyperactivity disorder (ADHD)," PLOS Digital Health, vol. 3, no. 11, p. e0000620, 2024, doi: 10.1371/journal.pdig.0000620.

A. Bender, B. Voytek, and N. Schaworonkow, "Resting-state alpha and mu rhythms change shape across development but lack diagnostic sensitivity for attention-deficit/hyperactivity disorder and autism," Journal of Cognitive Neuroscience, vol. 37, no. 9, pp. 1581–1615, Sep. 2025, doi: 10.1162/JOCN_A_02323.

A. Rehman, I. Heldal, C. Costescu, C. David, and J.-C.-W. Lin, "An adaptive scoring framework for attention assessment in NDD children via serious games," IEEE Access, vol. 14, pp. 20997–21007, 2026, doi: 10.1109/ACCESS.2026.3659987.

N. R. Muna, R. Jatnika, U. Purwono, and J. R. Siregar, "Differences in attention skill between children with ADHD and typically developing children in Indonesian primary schools," The Open Psychology Journal, vol. 14, pp. 329–337, 2021, doi: 10.2174/1874350102114010329.

H. Phudinawala, I. Batatawala, and R. Samnani, "Early identification of learning and developmental disorders in school-age children," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 10, no. 5, pp. 247–253, Nov. 2024, doi: 10.32628/CSEIT241051022.

F. Cortesi, S. E. Santirocchi, R. Montanelli, L. Toffoli, A. Gambarini, G. M. Marzocchi, G. Mioni, G. Mento, A. L. Ogliari, E. Toffalini, and V. Tobia, "Sense of time in neurodevelopmental disorders: ADHD and developmental dyscalculia from a dimensional and transdiagnostic perspective," Brain Sciences, vol. 16, no. 2, p. 249, 2026, doi: 10.3390/brainsci16020249.

E. McDougal, C. Tai, T. M. Stewart et al., "Understanding and supporting attention deficit hyperactivity disorder (ADHD) in the primary school classroom: Perspectives of children with ADHD and their teachers," Journal of Autism and Developmental Disorders, vol. 53, pp. 3406–3421, 2023, doi: 10.1007/s10803-022-05639-3.

S. C., "NeuroAI-ADHD: Multimodal deep learning for ADHD detection from resting-state fMRI using ensemble BiLSTM-transformer architecture with explainable AI and full-stack clinical deployment," International Journal for Research in Applied Science and Engineering Technology, 2026, doi: 10.22214/IJRASET.2026.81460.

P. K. Vayyasi, "Sleep pattern analysis using machine learning in children with neurodevelopmental disorders," International Journal of Computational and Experimental Science and Engineering, 2025, doi: 10.22399/ijcesen.5124.

X. Yi, J. Huang, B. O. Akinwunmi, and W. Ming, "ADBrainNet: A deep neural network for autism spectrum disorder (ASD) and attention deficit and hyperactivity disorder (ADHD) classification using resting-state fMRI images based on explainable artificial intelligence," Medical & Biological Engineering & Computing, 2026, doi: 10.1007/s11517-026-03530-2.

L. L. Yen Wei, A. A. Ibrahim, and R. Alfred, "Deep learning approach to EEG-based attention deficit hyperactivity disorder (ADHD) detection: An empirical comparison of ensemble classifiers," in Proc. IEEE Int. Conf. Artificial Intelligence in Engineering and Technology (IICAIET), 2025, pp. 49–54, doi: 10.1109/IICAIET67254.2025.11265027.

N. N. and R. Mala, “Enhancing adverse drug reaction classification of attention deficit hyperactivity disorder diagnosis data using deep learning with optimization algorithm,” Fusion: Practice and Applications, 2025, doi: 10.54216/FPA.190104.

H. Jahani and A. A. Safaei, “Efficient deep learning approach for diagnosis of attention-deficit/hyperactivity disorder in children based on EEG signals,” Cognitive Computation, vol. 16, pp. 2315–2330, 2024, doi: 10.1007/s12559-024-10302-3.

J. S. Priya, R. Karolin, and T. M. Harini, “Multimodal ResNet-based CNN for early autism spectrum disorder detection and severity stratification,” in Proc. 6th Int. Conf. Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2026, pp. 562–570, doi: 10.1109/ICPCSN68523.2026.11544013.