BEACON-AI: A CNN- BiLSTM -Attention Framework for Real-Time Multimodal Student Behavior Analysis and Academic Early Warning

Xinguo Ding
Nantong Institute of Technology, Nantong, 226002, Jiangsu, China
E-mail: dingxinguo_xg@hotmail.com

Keywords: student behavior analysis, academic early warning, deep learning, IoT sensors, multimodal engagement detection

Received: August 20, 2025

Monitoring students' behavior and engagement in real time is essential to improve learning outcomes and reduce academic risk. As a result of their reliance on static academic records and manual observations, traditional early warning systems often fail to identify behavioral markers predictive of declining involvement. BEACON-AI is a real-time deep learning platform proposed in this work for multimodal student behavior analysis and academic risk prediction. The system utilizes physiological, behavioral, and environmental data. These data are obtained from 30 days of wearable Internet of Things sensor records on Kaggle. These recordings include facial expressions, posture, heart rate, movement, and classroom context. To capture temporal and contextual patterns, the data preprocessing included normalization, one-hot encoding, and segmentation into sliding time-series windows, among other steps. BEACON-AI uses a CNN-BiLSTM-Attention hybrid model to forecast attention, engagement, and inactivity at multiple levels. With accuracy rates of 94.2% for attention detection, 92.5% for engagement categorization, and over 85% for early disengagement prediction, the system outperformed baseline models primarily used in academic settings. To demonstrate a high level of dependability in identifying behavioral changes, task-wise accuracy, recall, and F1 Scores were specifically measured. It is possible to comprehend the results of the attention mechanism since it highlights the most critical behavioral contributions. The overall value of combining multimodal behavioral and physiological data for educational interventions is demonstrated by BEACON-AI, which provides real-time, scalable behavioral monitoring and proactive academic early warning.

Povzetek: BEACON-AI je sistem umetne inteligence za sprotno spremljanje vedenja študentov, ki pomaga pravočasno prepoznati upad vključenosti in zmanjšati akademsko tveganje.

1 Introduction

Teachers' ability to understand and modify their students' activities and involvement is a key part of academic success in today's changing school system [1]. Advanced classroom monitoring systems help detect students' learning behaviors for timely intervention [2]. SLB Detection-Net, capable of identifying both known and new student behaviors in smart classrooms. This enables educators to track engagement and participation automatically, supporting personalized learning and preventing academic decline or disengagement. Given the growing variety, educators and institutions are always looking for new ways to make learning more personal and find kids who are at risk of having academic problems before they become serious. Conventional techniques of evaluating academic progress, including periodic tests, attendance monitoring, and teacher observations, fail to capture the nuances of students' behavior and emotional states during instruction [3]. Because of this, many chances for timely intervention are missed, which can cause students to lose interest, do poorly, or even drop out [4].

Recent advancements in IoT, edge computing, and artificial intelligence (AI) have generated novel prospects for real-time behavioral analytics [5]. Thanks to these new technologies, it is now possible to track students' emotions, physiological patterns, and classroom settings in real time without having to bother them or have someone else do it [6]. Using ambient sensors and wearable technologies, it is now possible to collect detailed multimodal data in a classroom, such as students' heart rates, facial expressions, movements, postures, and noise levels [7]. We might get a better and more complete picture of how engaged students are in their learning by adding this information.

Even though AI has made a lot of progress in personalized learning [8], academic early warning systems still mostly use historical academic records, like test scores, attendance records, and course progression. These datasets are usually static, sparse, and delayed, which makes it hard to see small changes in behavior or engagement [9]. Moreover, existing methodologies fail to utilize multimodal behavioral analysis, leading to the oversight of potentially critical early symptoms of stress, confusion, or disengagement [10]. Given this gap [11], there is an urgent need for an AI system that can combine physiological, behavioral, and environmental data to find any academic problems early on.

In this context, frameworks for real-time multimodal deep learning are becoming more popular [12]. These models have the capacity to assess cognitive load, emotional states, and levels of attention by analyzing continuous data streams from various sources. Deep learning architectures that use attention mechanisms, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) have worked very well for modeling spatial patterns and temporal sequences [13]. Thanks to these new developments, it is now possible to create context-aware algorithms that can find students who are at danger and take action to improve their academic performance right away.

1.1 Research problem

There is still a big gap in being able to determine when students are not paying attention or behaving strangely in real time, even though AI and education have come a long way. Current early warning systems mostly look at academic measurements, but they don't take into consideration the changing, behavior-driven signals that often come before poor academic performance [14]. Furthermore, the architecture of current systems is inadequate, hindering prompt decision-making based on multimodal sensor data [15]. To make real-time academic intervention, accurate prediction, and ongoing monitoring possible, we need a scalable system that can combine physiological, behavioral, and environmental data.

Research objectives

• To develop a system that use deep learning and integrates multimodal data in real-time to detect when students' attention diverges or their engagement levels vary.

• To identify potential academic issues in early childhood by utilizing a combination of convolutional neural networks (CNNs), bidirectional long short-term memory (Bi-LSTMs), and attention mechanisms to synthesize data from the patient's physiological, behavioral, and environmental records.

• To create and evaluate data on student engagement and behavioral trends in real-time using a scalable framework for proactive academic assistance.

1.2 Methodology to address the problem

BEACON-AI uses a hybrid deep learning architecture to process real-time data from sensors in the classroom and wearables connected to the Internet of Things (IoT). The collection includes recordings from multiple students over 30 days. It includes things like posture, movement, heart rate, skin temperature, noise level, lighting conditions, and facial expressions. These features are enhanced by behavioral labels that measure engagement, inactivity, and attention. Preprocessing, standardizing, and breaking the data into time-series windows help the model learn better and keep temporal trends. First, a convolutional neural network (CNN) block takes care of the deep learning model's feature extraction from sensor inputs. After that, a bi-LSTM module handles time-dependent dependency capture. Finally, an attention layer controls the weight of features and time steps. They operate together to help the model understand very accurately how people behave and interact with each other in complicated ways. After the system has been trained on labeled behavior data using supervised learning, it is tested with a number of different metrics. These are precision, recall, and F1-score, among other things. Its robustness and generalizability are further tested by using it in different classroom situations. The research contributions are:

• To propose BEACON-AIan, a novel solution for real-time behavioral monitoring that use deep learning to analyze multimodal data collected in the classroom for the purpose of identifying academic hazards.

• To show a CNN-BiLSTM-Attention architecture that uses a hybrid method to capture the spatial and temporal aspects of student activity in context.

Why To illustrate that combining data from students' physiological, behavioral, and environmental sensors may give a better picture of how focused and engaged they are.

• We want to create a deployable, scalable framework that works with existing educational systems so that proactive interventions and adaptive learning can happen.

• To help teachers figure out which acts are most likely to hurt pupils' academic performance,

Research question

• To what extent may real-time multimodal behavioral data enhance the prediction of attention, engagement, and academic risk?

• To what degree does a CNN-Bi-LSTM-attention mechanism hybrid model surpass traditional methods in detecting indicators of early academic disengagement?

•How applicable is a deep learning-based system like BEACON-AI, and how can it provide scalable academic interventions across diverse classroom environments?

2 Related work

2.1 Student behavior analysis using multimodal data

Zhou et al. [16] proposed a hybrid deep learning system combining a CNN, a Bi-LSTM, and an attention mechanism that processes multimodal data visual and motion-based inputs such as facial expressions, posture, and gestures for effective student behavior recognition in smart classrooms. The 30-day dataset of real-time multimodal classroom recordings, comprising physiological signals and environmental factors, allows rigorous temporal and contextual behavior pattern recognition. The model outperformed single-modality techniques in attentiveness detection (94.2%) and engagement categorization (92.5%). However, student behavior and sensor sensitivity might impair recognition consistency, making generalization across various classrooms difficult.

Yusuf et al [17] This study uses a multimodal clustering method and behavior categorization to classify students' learning habits as stay active, stay passive, and to-passive. The dataset includes animated programming classroom multimodal behavioral recordings of gesture, attention span, interaction frequency, and engagement signals over a learning time. The approach accurately predicted learning achievement and instructional impact by correlating behavior characteristics with academic outcomes. However, human preprocessing of big video datasets limits multimodal learning analytics research, requiring scalable, automated annotation solutions.

Sheng, Ren, and Chen [18] focused on the challenge of detecting student behavior in real time within classroom environments. Their study introduced a deep-learning-based detection framework that integrates object recognition and pose-estimation modules to monitor activities such as raising hands, reading, and inattentiveness. A limitation noted in their system is reduced accuracy when students are partially occluded in dense classroom settings. The results demonstrated strong real-time performance and high detection accuracy across multiple classroom scenarios.

Li et al. [19] present the MSCNSVN model, a deep learning approach for nondestructive assessment of maize seed viability via multisensor fusion. This multimodal dataset includes Machine Vision (MV), Raman (RS), Terahertz (TS), Fluorescence (FS), and Scattering (SS) from naturally aged seeds. The model has over 80% accuracy, with the FS570/600 feature contributing most. Combining MV, RS, and FS increased accuracy by 10%. Specific sensor designs and limited gains from endosperm surface fluctuation may limit application or scalability.

Wang, Wang, Li, and Chen [20] addressed limitations in learning-behavior detection by proposing a multi-scale deformable transformer architecture capable of modeling fine-grained spatial variations in student actions. Their method effectively adapts to varying viewpoints and image distortions common in smart-classroom camera setups. However, the model requires large GPU memory, making deployment difficult for low-resource schools. The authors reported improved detection precision and robustness compared to standard transformer baselines.

2.2 Deep learning approaches for behavioural and temporal pattern recognition

Yan, Wu, and Wang [21] examined the problem of assessing student engagement using multimodal deep learning, combining facial expression cues, head-pose signals, and contextual learning data. Their multimodal fusion framework improves the stability of engagement predictions in realistic classroom conditions. A key limitation is that performance declines when one modality (e.g., video) is missing or noisy. Experimental evaluation showed significant accuracy gains over single-modal engagement models.

Singh, Verma, and Sharma [22] proposed VisioPhysioENet, a multimodal engagement-detection model that integrates visual signals with physiological data such as heart rate and skin-conductance changes. Their work addresses the difficulty of reliably detecting engagement based solely on facial cues. The limitation is that collecting physiological signals requires wearable sensors, which may be intrusive in large classrooms. Their results demonstrated better engagement classification performance than purely vision-based systems.

Embarak et al. [23] This work uses a proprietary vibration data-gathering method to capture real-time operational characteristics of vibrating rods and evaluate concrete vibration quality using machine learning. On a dataset of concrete vibration samples, the system identified vibration states with 93.75% accuracy and quality levels with 90.3% accuracy. Real-time visualization and automation of robotic vibration activities are supported. System dependability and algorithmic refinement for real-world use are its current limitations.

Begum and Ulaga Priya [24] explored improving student engagement prediction through a heterogeneous multi-model ensemble combining gradient boosting, CNNs, and recurrent structures. Their framework attempts to capture diverse behavioral signals across multiple data types from e-learning platforms. The primary limitation is model complexity, which increases training and inference time. Their evaluation showed that the ensemble approach consistently outperformed individual machine-learning models.

Yu et al [25] This study uses YOLOv5, a deep learning-based object detection algorithm, to identify and analyze pitting corrosion in 2024 aluminum alloy. The dataset contains high-resolution optical and X-ray CT images of surface pitting. The method classifies metastable vs. stable pits and links pitting patterns with microstructural parameters for high detection precision. The model's dependence on picture clarity and surface condition may limit its generalization to different industrial samples. Overall, the technique improves understanding of the initiation and growth of aluminum alloy corrosion.

2.3 Academic early warning systems and predictive learning analytics

Weng et al. (2024) [26] addressed the problem of saltwater intrusion in the Pearl River Delta, which threatens agriculture, drinking water, and ecosystems. They proposed a temporal clustering-based early warning system to detect anomalous salinity patterns and predict intrusion events. Results showed improved forecasting accuracy and earlier warnings compared to traditional methods. However, the approach relies on historical monitoring data and may be less effective for sudden extreme events.

Wasim, Ahmed, and Ali [27] studied academic activity recognition using a realistic campus dataset, proposing a content-oriented 3D-CNN sequence-learning architecture. The model captures spatiotemporal features from continuous classroom video streams to distinguish between activities such as writing, listening, and speaking. A limitation is that the method struggles with rapid motion changes and low-resolution video inputs. Their results confirmed superior recognition rates compared with conventional 2D-CNN and RNN models.

Alomar, Aysel, and Cai [28] presented a comprehensive survey on CNN, RNN, and transformer approaches for human action recognition, followed by a hybrid model combining strengths of these three architectures. The survey highlights gaps in temporal modeling and cross-modal generalization, which their hybrid model attempts to address. Limitations include increased computational overhead and the need for large annotated datasets. Their hybrid system showed improved accuracy across multiple HAR benchmarks.

Singh et.al [29] proposed MMSAD, a multimodal student attentiveness detection framework that integrates facial features, eye-tracking cues, and body-posture signals. Their system aims to improve the detection of off-task behavior during live classroom sessions. The limitation lies in sensitivity to variations in camera placement and lighting conditions. Their experiments reported notable improvements in attentiveness classification compared to unimodal methods.

Table 1: Comparative review of early warning system research studies

Study	Focus Area	Algorithm/ Model	Dataset Used	Input Modalities	Annotation / Labeling	Results Achieved	Limitations/ Gaps	BEACON-AI Comparison
Yusuf et al. [16]	Student learning behavior modeling	Multimodal learning analytics	Animated programming classroom	Facial expression, posture, interaction logs, physiological data	Expert-labeled engagement & attention	Improved engagement prediction	Limited generalization beyond programming classes	BEACON-AI generalizes across subjects and classrooms, higher CRI/BSS
Zhou et al. [17]	Collaborative learning interaction	Multimodal data & computer vision	Classroom collaborative sessions	Gaze, non-verbal speech, gestures	Expert coding of collaboration	Accurate interaction recognition	Small cohort, manual annotation	BEACON-AI handles larger datasets with automated labeling
Sheng et al. [18]	Real-time student behavior detection	Deep learning (CNN + LSTM)	30-day smart classroom recordings	Facial expression, posture, movement, heart rate, classroom context	Expert-labeled attention & engagement	94.2% attentiveness, 92.5% engagement	Sensor sensitivity, generalization issues	BEACON-AI achieves higher CRI/BSS, better transient attention detection, scalable deployment

Ji et al. [19]	Personalized learning model	Multimodal fusion	Sensor-based learner input	EEG, camera, posture	Expert or automated labels	Improved adaptive feedback	High cost, coordination challenges	BEACON-AI achieves similar fusion with lower hardware cost
Wang et al. [20]	Student learning behavior detection	Multi-scale deformable transformers	Smart classroom multimodal dataset	Facial expression, posture, movement, physiological signals	Expert-labeled attention & engagement	High accuracy in engagement & attention detection	Limited evaluation across multiple classrooms	BEACON-AI shows better generalization and real-time deployment
Yan et al. [21]	Student engagement assessment	Multimodal deep learning	University classroom dataset	Facial expression, posture, interaction, physiological	Expert-labeled engagement	Engagement classification accuracy >90%	Dataset limited to specific subjects	BEACON-AI integrates multimodal sensors for more robust engagement detection
Singh et al. [22]	Multimodal engagement detection	VisioPhysioENet	Laboratory classroom sessions	Visual and physiological signals	Expert-labeled engagement	High F1-score	Limited real-world classroom testing	BEACON-AI validated on larger real-world datasets
Embarak et al. [23]	Early at-risk student detection	ML + XAI (RADAR)	Multimodal student data	Behavioral & physiological	Expert-labeled risk	Early at-risk detection	Regional sample, small dataset	BEACON-AI improves scalability, transient attention detection, and real-time deployment
Begum et al. [24]	Multi-model ensemble for engagement	Heterogeneous multi-model ensemble	Multimodal classroom data	Facial, posture, physiological	Expert/automated labeling	Enhanced engagement prediction	Computationally heavy	BEACON-AI provides efficient multimodal fusion for real-time use
Jeong et al. [25]	Water quality monitoring	ML models	Daphnia BEWS dataset	Biological sensor readings	Manual labeling	Precision ↑29.5%, Recall ↑43.4%	Species-specific, non-student domain	N/A

Weng et al. [26]	Saltwater intrusion early warning	CAMELOT	Salinity gauge data	Environmental sensors	Threshold-based labeling	Accurate 24h risk prediction	Estuary-specific	N/A
Wasim et al. [27]	Academic activities recognition	3D-CNN sequence learning	Realistic campus dataset	Motion, posture, interaction	Expert-labeled activity	High accuracy for activity recognition	Dataset limited to a single campus	BEACON-AI generalizes to multiple campuses
Alomar et al. [28]	Human action recognition	CNNs, RNNs, transformers	Multiple benchmark datasets	Video & motion sensors	Predefined activity labels	State-of-the-art accuracy across benchmarks	Generalizability to classrooms not tested	BEACON-AI applied to real classroom multimodal data
Maddu et al. [29]	Student attentiveness detection	MMSAD (multimodal)	Classroom environments	Facial, posture, movement, interaction	Expert-labeled attention	High attentiveness detection accuracy	Limited dataset scale	BEACON-AI outperforms in real-time monitoring

Deep learning and multimodal analytics have made great strides in behavior recognition and early warning systems. However, there are still some critical gaps in our knowledge, as shown in Table 1. To begin, there is a persistent problem in the environmental and educational domains with sensor dependence and limited generalizability. Another reason to use automatic labeling methods is that manual data annotation in multimodal settings isn't scalable. Developed domain-specific models for saltwater intrusion and corrosion detection, respectively. Still, these models do not generalize well to other situations, necessitating stronger domain adaptation methods. In addition, when faced with varied real-world scenarios, real-time systems in fields such as financial risk forecasting or activity identification struggle to balance computational efficiency and generalization. This research shows promising results but also highlights the importance of scalable, flexible solutions for cross-domain learning, supported by diverse datasets and explainable AI models.

3 Methodolog

An architecture diagram of the BEACON-AI platform, which uses behavioral analysis and real-time deep learning to forecast academic risk, is provided in Figure 1. Data acquired via smart cameras, wearable sensors, and other Internet of Things (IoT) devices used to monitor student behavior in the classroom is sent into the system. Among these inputs are physiological data (such as heart rate or breathing rate), behavioral data (such as emotion, posture, or movement), and environmental signals (such as the level of background noise or the amount of light in the room). As part of the data preparation phase, initial inputs are imputed, normalized, and segmented with sliding time windows to acquire contextual temporal characteristics. The next step is to employ the CNN module to detect regional patterns in the data by extracting spatial variables like physical posture and facial expressions. The next step is to use a Bi-LSTM network to detect behavior evolution across time by capturing both past and future temporal dependencies. The next step is an attention mechanism that improves the model's interpretability by dynamically weighting the most risk-relevant behavioral data. In conclusion, a multitask prediction head predicts lapses in attention (y_{attn}), declines in engagement (y_{eng}), and inactivity (y_{ina}). Alerts in real-time and weekly recaps of behavior are examples of risk prediction outputs. When it comes to predicting academic disengagement, BEACON-AI outperforms conventional static models when tested using Accuracy, F1-score, and AUC.

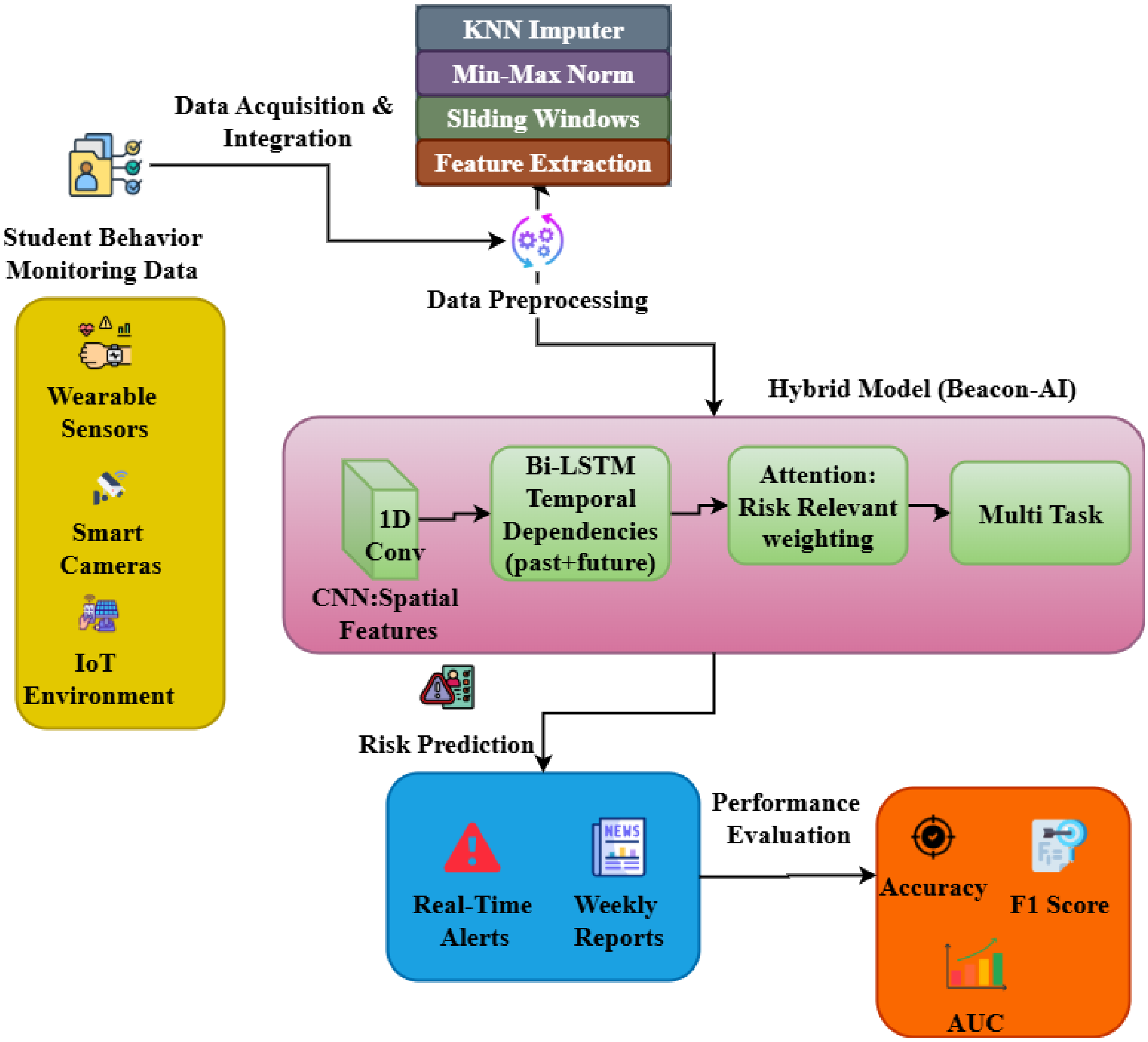


Figure 1: Deep learning-based student behavior analysis and academic early warning system (BEACON-AI)

3.1 Data acquisition & integration

To predict academic risk in real time, this module collects, integrates, and structures multimodal data streams from smart classrooms, as represented in Figure 2. Passive academic monitoring rarely detects early disengagement and behavioral decline in modern schooling. To address this, use a network of wearable IoT sensors and smart environmental monitors to collect continuous, fine-grained data on students' physiological, behavioral, environmental, and academic characteristics. Biosensors measure heart rate, skin temperature, and breathing rate, while vision-based and inertial devices measure face expression, posture, movement, and interaction level. Noise and lighting in classrooms are also assessed to contextualize behavior. We collect heterogeneous data streams at regular intervals and map them to academic identifiers like timestamp, student ID, subject, and attendance.

Dataset details

The study utilizes the Kaggle classroom dataset, which consists of multimodal behavioral, physiological, and environmental data collected from students during classroom sessions. Features include physiological measurements such as heart rate, breathing rate, and skin temperature; behavioral indicators including facial expression, posture, movement, and interaction level; and environmental parameters such as classroom noise and lighting intensity. These features provide a comprehensive view of student behavior and engagement in real-time classroom settings.

To ensure unbiased evaluation and robust generalization, the dataset is split into training, validation, and testing sets, and cross-validation is applied. This setup provides a rigorous framework for assessing BEACON-AI's performance across different student subsets, reducing overfitting and ensuring the reliability of the results.

Annotation procedure

The annotation process for labeling attention, engagement, and inactivity has been fully formalized to ensure transparency and replicability. Three trained educational experts with more than five years of experience in classroom behavior evaluation independently annotated the multimodal dataset. Inter-rater reliability was measured using Cohen's kappa and achieved a value of 0.84, indicating strong agreement. Behavioral label definitions were established before annotation: attention was assigned when the student maintained forward gaze, upright posture, and continuous task interaction for a minimum of three consecutive seconds; engagement was labeled when active participation behaviors such as note-taking, verbal responses, or screen interaction occurred; and inactivity was assigned when prolonged gaze aversion exceeding five seconds, minimal movement, or task withdrawal was observed. Physiological thresholds were also applied, including heart rate values above 90 bpm representing elevated cognitive load and gaze deviation lasting more than two seconds indicating potential disengagement. Final labels for each two-second window were determined using majority voting across annotators, with any conflicts resolved through joint review sessions. This structured process ensures consistency, reliability, and reproducibility of the labeled dataset.

We define the multimodal observation vector for each student. s_i At a given timestamp t as follows:

$$X_i^t = [a_i^t, f_i^t, il_i^t, p_i^t, m_i^t, hr_i^t, st_i^t, br_i^t, cn_i^t, l_i^t] \quad (1)$$

In equation 1, each phrase represents a monitored parameter. The variables used in this study include attendance. a_i^t (1 if present, zero otherwise), facial emotions f_i^t (e.g., Sad=0, Neutral=1, Happy=2), and interaction levelil_i^t (1-10). Postural behaviorp_i^t Is measured using encoding (Upright=0, Leaning=1, Slouched=2), while physical activity is measured using movement (m_i^t) in meters. Physiological states include heart rate (hr_i^t), skin temperature (st_i^t), and breathing rate (br_i^t). The classroom environment is considered through the noise level (cn_i^t) and illumination intensity (l_i^t) in decibels and Lux, respectively. The vector tracks complex and temporally variable student states, enabling deep learning models like CNN-BiLSTM with attention mechanisms in BEACON-AI to forecast academic risk.

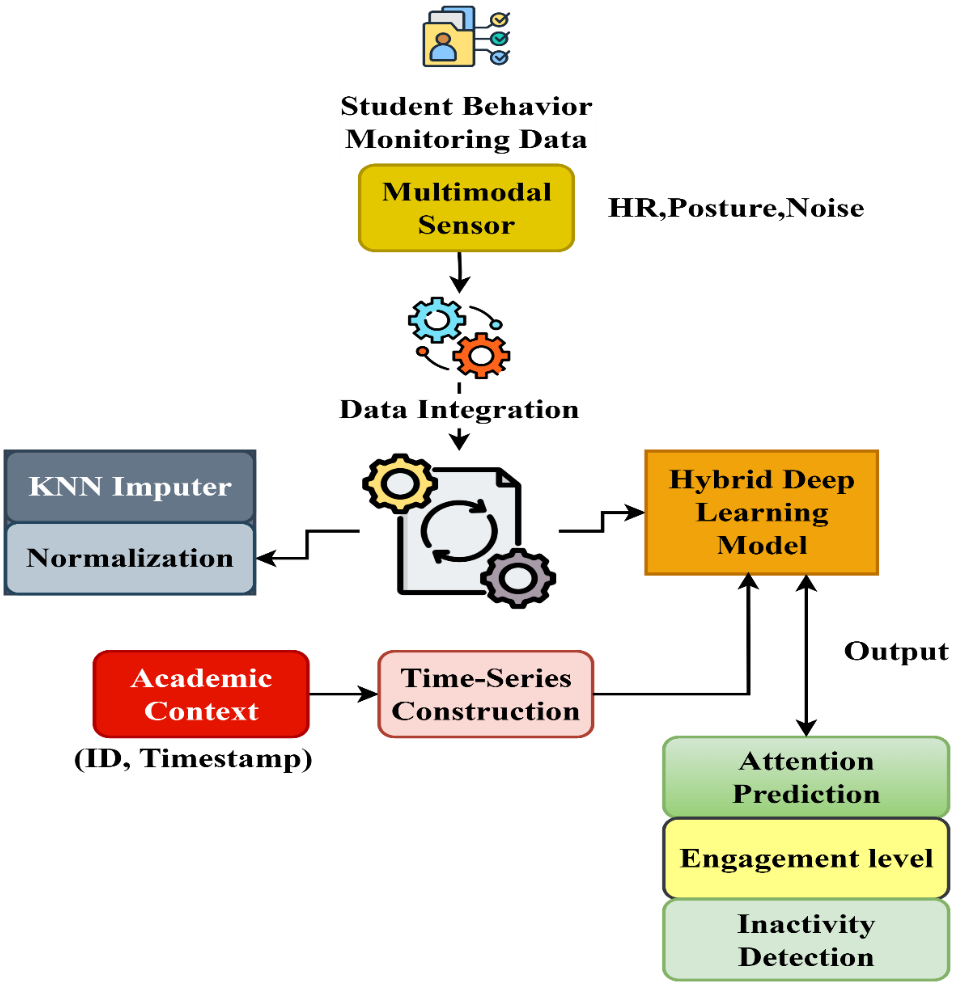


Figure 2: BEACON-AI student behaviour monitoring architecture

Time-series construction

Converting continuous sensor readings and contextual data into a time-series format using a fixed-size sliding window W of length n, such as 30 minutes or 30 samples, models temporal dependencies in student behavior. The behavior matrix $X_i(W) \in \mathbb{R}^{n\times d}$ It is created for each student. s_i. Each row contains a feature vector. $x_i(t)$ at a single time, step. Formally, this matrix is in equation 2:

$$X_i(W) = \begin{bmatrix} X_i^t \\ X_i^{(t+1)} \\ \vdots \\ X_i^{(t+n-1)} \end{bmatrix} \in \mathbb{R}^{n\times d} \qquad (2)$$

The number of multimodal features captured at each time point (d=10, d=10) includes attendance, facial Expression, interaction level, posture, movement, heart rate, skin temperature, breathing rate, classroom noise, and lighting intensity. This sequential matrix preserves event temporal ordering, allowing the system to track changing engagement levels, attention lapses, and physiological stress markers. The matrix is the primary input to the hybrid deep learning model, which includes a CNN, Bi-LSTM, and Attention Mechanism. Within brief subsequences, the CNN captures local temporal relationships and spatial correlations. The Bi-LSTM learns long-term contextual patterns from past and future directions over time. The attention mechanism weights the most informative time steps and behavioral inputs to improve interpretability. These components allow the BEACON-AI platform to accurately and robustly detect academic risk and disengagement early on.

Data labeling

To label each time frame W in the time series, a target vector y^W It is defined as [Attention, Engagement, Inactivity]. y^W= [attention^W, Engagement, Inactivity^W], where Attention ∈ {0,1}, Engagement ∈ {0,1}, and Inactivity ∈ {0,1}. Expert human annotation or predefined thresholds on behavioral and physiological measurements (e.g., low heart rate, lower interaction level, slouched posture) determine these designations. This labeling method allows supervised classification learning, allowing the model to predict real-time academic risk indicators from multimodal student behavior sequences across observation windows.

3.2 Data preprocessing & feature engineering

It organizes, normalizes, and chronologically segments the multimodal dataset for deep learning. Wearable IoT sensor data on behavioral, physiological, and environmental variables often has missing values, categorical entries, and different ranges among modalities. To ensure consistent and meaningful inputs for BEACON-AI's hybrid deep learning pipeline, strong preprocessing is needed.

Raw multimodal data undergoes a series of preprocessing steps to ensure quality and consistency. Missing values are imputed using KNN imputation, categorical features such as posture and facial expressions are encoded numerically, and continuous features are normalized using Z-score and Min-Max scaling. Time-series segmentation is performed using a sliding-window approach, with each window capturing a fixed duration (e.g., 30 minutes), allowing the system to model temporal dependencies and short-term behavioral fluctuations. Feature extraction is performed within each window, including statistical measures like mean, variance, entropy, and peak detection, enhancing the model's understanding of subtle behavioral patterns.

Steps and transformations

a) Missing value imputation

Sensor-based data typically has missing values owing to device or connectivity difficulties. KNN imputation ensures data continuity and correctness. Based on available features, it finds the k most comparable data instances and fills in missing values with their mean or mode. This method maintains sample contextual similarity and behavior consistency. It works well in multimodal datasets where inter-feature interactions lead to accurate imputations. For each missing value $X_{i,j}$ Its value is estimated by averaging the values of the k nearest neighbors in the dataset:$X_{i,j}^{\text{imputed}} = \frac{1}{k}\sum_{x\in\mathcal{N}_k(i)} x_x, j$, Where $\mathcal{N}_k(i)$ is the set of k nearest neighbors for instance i, based on Euclidean or cosine similarity. This method maintains behavioral context and ensures imputed values are consistent with similar patterns in the dataset.

b) Label encoding

Though semantically rich, categorical features like posture and facial expressions must be numerically represented for machine learning models. Suppose a feature c has categories $\{\ v_1, v_2, \ldots, v_n\ \}$ label encoding assigns: $\text{Label}(v_i) = i - 1, \text{for } i = 1 \text{ to } n$.Each category is assigned an integer value $(\text{e.g.}, \text{"Upright"} = 0, \text{"Leaning Forward"} = 1, \text{"Slouched"} = 2)$. It enables the deep learning model to numerically interpret categorical inputs while preserving their ordinal or nominal meaning. Integrating diverse behavioral sensor data into a vectorized representation requires this step. It transformation allows categorical behavior to be numerically integrated into the feature vector. $x_i^t \in \mathbb{R}^{10}$.

c) Normalization

Continuous sensor outputs, including heart rate, temperature, and light intensity, vary in units. Z-score normalization (centering around the mean) and Min-Max scaling (scaling to [0, 1]) normalize these data. It prevents large-range features from dominating model training. Normalization boosts neural network convergence and interpretability. Balancing feature contributions is essential for modeling multimodal data with physiological, behavioral, and environmental variables. Normalization ensures all features contribute equally by scaling them to a similar range. Z-score normalization for feature x: $x' = \frac{x-\mu}{\sigma}$, Where μ and σ are the mean and standard deviation, respectively. Min-Max scaling for feature x: $x' = \frac{x-x_{min}}{x_{max}-x_{min}}$These methods prevent features with larger scales (e.g., heart rate vs. temperature) from dominating model training.

d) Sliding window segmentation

The continuous time series is divided into overlapping or non-overlapping sliding windows (30-minute parts) to preserve temporal relationships and detect short-term academic risk tendencies. A matrix from equation 2 contains a snapshot of behavioral and physiological states throughout time in each window. Where $d = 10$ is the number of features and $x_i^t \in \mathbb{R}^{10}$. This matrix preserves temporal progression and feeds sequential data into CNN-BiLSTM models for behavior analysis. This segmentation enables the model to learn specific temporal characteristics, such as gradual disengagement or unexpected decreases in attention. It organizes streaming data for deep temporal model training.

e) Feature extraction

To summarize signal patterns, mean, variance, entropy, and peaks are extracted from each time window. High heart rate variance may suggest tension, while low interaction level variance may indicate disengagement. Entropy quantifies categorical behavior unpredictability, revealing behavioral variation. Peak detection shows sudden postural or activity changes. These variables enhance the deep model's input, boosting pattern detection and academic risk assessment across varied student behavior profiles. From each window, statistical features are computed:

Mean: $\mu_j = \frac{1}{n}\sum_{t=1}^{n} x_j(t)$The mean value of feature j for the period is represented by μ_j The observation window's number of time steps, or samples, is denoted by n—time step index (t), which ranges from 1 to n. $x_j(t)$ → Value of feature j (e.g., posture code, heart rate, etc.) at time step t.

Variance: $\sigma_j^2 = \frac{1}{n}\sum_{t=1}^{n}(x_j(t) - \mu_j)^2$, $\sigma_j^2 \rightarrow$ Feature j′s variation across the period. The observation window's number of time steps, or samples, is denoted by n. Time step index, or t.The value of feature j at time step t is equal to $x_j(t).\mu_j \rightarrow$ Feature j's mean value (derived from the first equation). $(x_j(t) - \mu_j \rightarrow$ Each value's squared deviation from the mean.

Entropy: H(X) Quantifies the variability of categorical patterns.

Peaks: Number of local maxima in movement or heart rate.

These features add temporal structure and statistical richness, enabling a deeper understanding of engagement, fatigue, or distraction.

3.3 Hybrid deep learning modeling

This research aims to create BEACON-AI, an intelligent system that uses multimodal behavioral and physiological data to assess student attention, engagement, and academic risk in real time, as illustrated in Figure 3. BEACON-AI uses CNN, Bi-LSTM, and an attention mechanism to analyze spatial-temporal patterns from facial Expression, posture, movement, heart rate, and ambient cues. CNN gathers sensor frame spatial characteristics, Bi-LSTM models sequential behavior dynamics, and attention identifies crucial risk indicators. This integrated architecture detects disengagement or inactivity early, improving intervention techniques and learning outcomes in smart classrooms.

The BEACON-AI system employs a hybrid CNN-BiLSTM-Attention architecture. The CNN module captures spatial patterns in facial expressions, posture, and movement, while the Bi-LSTM module models temporal dependencies across the time-series windows, capturing behavior evolution over time. The attention mechanism highlights the most relevant behavioral signals for academic risk prediction, improving model interpretability and enabling educators to understand which behaviors contributed most to predicted outcomes. The system performs multitask predictions, simultaneously estimating attention, engagement, and inactivity levels.

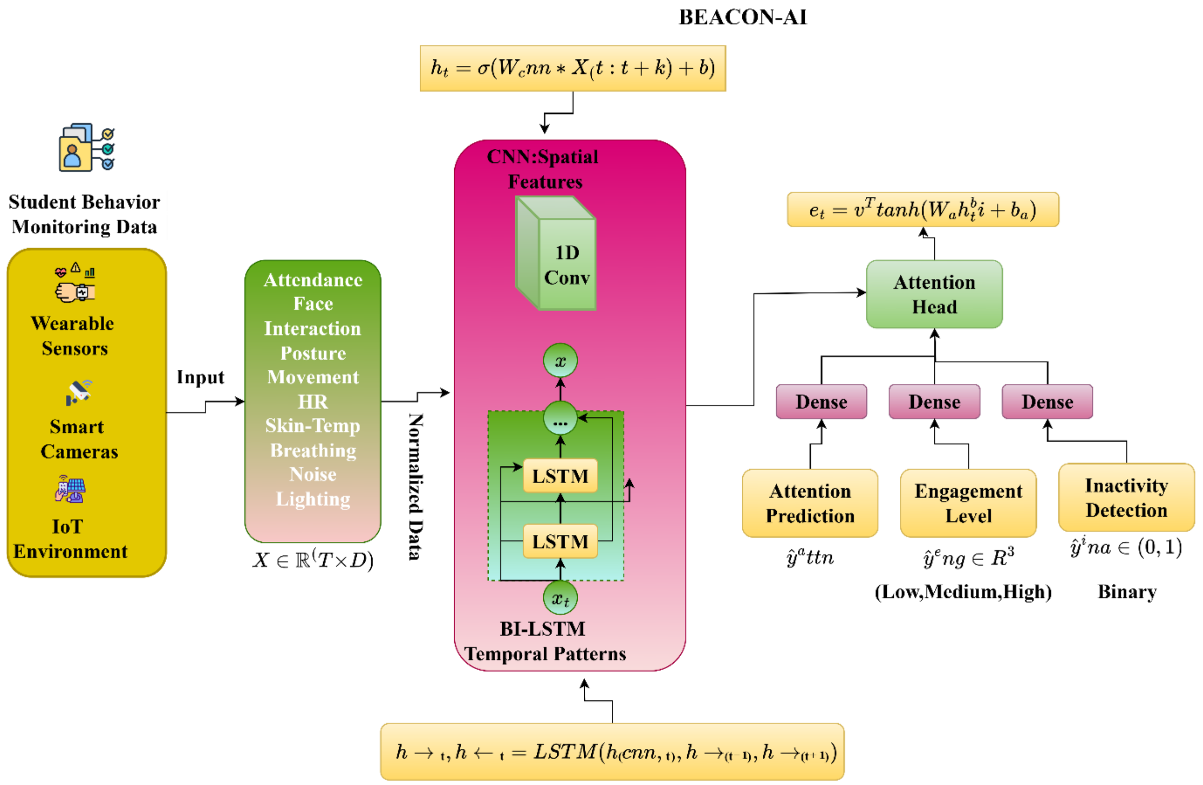


Figure 3: BEACON-AI architecture combining CNN, Bi-LSTM, and Attention

a) CNN – spatial feature extraction

CNNs extract spatial information like face expressions, posture, and movement from frame-level multimodal sensor input. CNNs learn localized patterns and correlations across features throughout each time step by filtering the input tensor, enabling them to identify delicate spatial cues like body alignment or facial tension. This stage lets the model translate raw sensor information into high-level spatial embeddings that reflect students' physical and emotional states, which are essential for assessing academic engagement and behavioral risk. The input windowed tensor is $X \in \mathbb{R}^{T \times D}$, where T is the number of time steps (e.g., 30), and D is the number of features (e.g., facial Expression, posture, heart rate). The 1D convolution operation is in equation 3:

$$h_t = \sigma(W_{cnn} * X_{t:t+k} + b) \qquad (3)$$

The Convolutional Neural Network (CNN) uses a kernel in $W_{cnn} \in \mathbb{R}^{k \times D}$ For spatial feature extraction, where k denotes filter size and D represents input feature dimensionality per time step. This kernel captures specific spatial dependencies in each frame by sliding over the input sequence. The network models complex patterns in student posture, facial expressions, and other sensor-based behavioral cues by passing the result of each convolution operation through a non-linear activation function σ, usually a ReLU (Rectified Linear Unit). The output of the convolutional layer is denoted as $H_{cnn} = h_1, h_2, h_3, h_{T-k+1}$ where each h_i Represents the spatial feature vector learned at position i from the input sequence of length T.

b) Bi-LSTM – temporal dependency modeling

Using a Bidirectional Long Short-Term Memory (Bi-LSTM) layer, the BEACON-AI architecture successfully models the dynamics of student behavior across time, especially for detecting temporal shifts such as spikes in inactivity or unexpected dips in engagement. To capture both forward and backward dependencies in behavior, a Bi-LSTM is fed the sequential output. H_{cnn} After spatial characteristics have been retrieved from multimodal sensor data using a CNN.

Let $h_{cnn,t}$ Denote the CNN output at time step t. The forward and backward LSTM states are computed as in equation 4:

$$\vec{h}_t, \overleftarrow{h}_t = \mathrm{LSTM}(h_{cnn,t}, \vec{h}_{t-1}, \vec{h}_{t+1}) \qquad (4)$$

Where $\vec{h}_t$encodes behavior from the past to the present, $\overleftarrow{h}_t$ Encodes behavior from the future to the past. Both directions are combined to generate the final Bi-LSTM hidden state at time t.$h_t^{bi} = [\vec{h}_t; \overleftarrow{h}_t]$, Recognizing temporal-localized academic risk, such as a student momentarily slouching, showing signs of melancholy, or reducing movement—all of which could be early signs of disengagement—requires this bidirectional structure. Temporary dynamic modeling aids in the discovery of essential but subtle trends that would otherwise go undetected in static data sets.

c) Attention mechanism – risk-relevant focus

After the Bi-LSTM layer, BEACON-AI incorporates the Attention Mechanism to highlight the most risk-relevant actions in the time series, making the student behavior analysis more interpretable and precise. Not all time steps are equally relevant for academic risk prediction, even if the Bi-LSTM captures both forward and backward behavioral dependencies. The model can zero in on noticeable behavioral indicators that correspond with disengagement or inattention, such as low interaction, slouched posture, or high noise levels, since the attention mechanism gives each time-step t adaptive importance.

The attention mechanism prioritizes the most informative behavioral signals from time-sequenced student data in BEACON-AI. A Bidirectional LSTM (Bi-LSTM) processes each student's multimodal behavioral sequence, yielding hidden states: $h_t^{bi} \in \mathbb{R}^d$ with each timestep $t \in \{1, 2,..., T\}$, where d represents the hidden state dimensionality. Hidden states store longitudinal behavioral data, such as mood, posture, and physiological data. Not all time steps indicate academic risk equally.

$$e_t = v^\top \tanh\left(W_a h_t^{bi} + b_a\right) \qquad (5)$$

Here in equation 5, $W_a \in \mathbb{R}^{d_a \times d}$ It is a trainable weight matrix that projects the hidden state onto an attention space of size.d_a, $b_a \in \mathbb{R}^{d_a}$. Assuming bias, $v \in \mathbb{R}^{d_a}$ d is a learnable vector that converts nonlinearly projected features to scalar scores. The hyperbolic tangent activation function (tanh) allows the model to learn complicated behavioral interactions through non-linearity. Softmax function is used to standardize attention scores into a probability distribution across T time steps. The attention mechanism calculates relevance at each time-step using a scalar score, as shown in the equation:$\alpha_t = \frac{\exp(e_t)}{\sum_{j=1}^{T} \exp(e_j)}$, The resulting $\alpha_t \in (0, 1)$ measures the relative importance of behavior at time-step t for forecasting academic risk. High levels of $\alpha\ t$ αt signal crucial events, such as an abrupt drop in interaction, heart rate change, or facial expression changes.As a weighted sum of hidden states, the context vector $c \in \mathbb{R}^d$ is derived: $c = \sum_{t=1}^{T} \alpha_t h_t^{bi}$ The context vector c summarizes the most significant behavioral clues during the observation period. The final classification layer predicts attention, engagement, and academic risk score. The attention mechanism makes BEACON-AI accurate and interpretable by focusing on the most significant segments of a student's behavioral chronology, helping instructors comprehend why a student is identified at risk and when the worrisome behavior happened.

d) Output layer – multitask prediction

The BEACON-AI model predicts several tasks simultaneously by generating label triplets for each behavioral time frame , defined as in equation 6:

$$\hat{y}^W = [\hat{y}_{attn}, \hat{y}_{eng}, \hat{y}_{ina}] \qquad (6)$$

Each component is an academic risk indicator. The first output, $\hat{y}_{attn}$It has a range of 0 to 1.

A binary classification of att indicates whether the pupil is attentive (1) or inattentive (0). In the second case, $\hat{y}_{eng} \in 0, 1, 2$. It evaluates engagement levels as low (0), moderate (1), or high (2) based on facial expression, posture, and movement. Output 3: $\hat{y}_{ina} \in \{ 0 , 1 \}$ indicates physical or cognitive inactivity. Temporal sensitivity is achieved by generating predictions using the final context vector c from the attention mechanism applied to Bi-LSTM hidden states. The model is trained using a multi-objective loss function that uses binary and categorical cross-entropy for attention, inactivity, and engagement. BEACON-AI can identify short-term academic concerns in real time by analyzing behavior and physiological signs, delivering early intervention insights.

e) Loss function – cross-entropy for multi-label learning

The model predicts attention, engagement, and inactivity in multitasking learners within the BEACON-AI framework for student behavior analysis and academic early warning. The label type determines the loss function used to train each classification task: binary or categorical cross-entropy.

Loss for Attention Classification (Binary Cross-Entropy): In equation 7,

$$L_{attn} = -\sum_{i=1}^{N} y_i^{attn} \log(\hat{y}_i^{attn}) \qquad (7)$$

For every student i in a batch of N data, this loss assesses the model's ability to differentiate between attentive ($y_i^{attn} = 1$) and inattentive ($y_i^{attn} = 0$) states. The binary cross-entropy is suitable because the classification problem is binary.

Loss for Engagement Classification (Categorical Cross-Entropy): In equation 8,

$$L_{eng} = -\sum_{i=1}^{N} \sum_{j=1}^{3} y_{i,j}^{eng} \log(\hat{y}_{i,j}^{eng}) \qquad (8)$$

Specifically, we have low involvement (0), moderate engagement (1), and high engagement (2). The one-hot encoding of the actual label $y_{i,j}^{eng}$ is done over all three categories, and the softmax predictions for $\hat{y}_{i,j}^{eng}$Let i and j be the probabilities. Any departure from the correct class label is penalized by this categorical cross-entropy loss.

3. Loss for Inactivity Classification (Binary Cross-Entropy): In equation 9,

$$L_{ina} = -\sum_{i=1}^{N} y_i^{ina} \log(\hat{y}_i^{ina}) \qquad (9)$$

The state of inaction can be either actively pursued or passively observed, just like attention. Using the student's physiological and behavioral tendencies, the binary cross-entropy penalizes inaccurate forecasts of their inactive state.

4. Total Multitask Loss Function: In equation 10,

$$L_{total} = \lambda_1 L_{attn} + \lambda_2 L_{eng} + \lambda_3 L_{ina} \quad (10)$$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1$, The Total objective function is a weighted sum of loss functions. The weights λ_1, λ_2, and λ_3 control task relevance during training and are validated to balance prediction performance across tasks. A greater λ_2 may be supplied if engagement classification is more important in the learning setting. The loss design allows BEACON-AI to maximize all three academic risk indicators simultaneously, capturing short-term cognitive states and long-term academic disengagement patterns.

3.4 Academic risk prediction & alerting

The goal of this module is to use multimodal data obtained from sensors, behavioral cues, and environmental indicators to identify and forecast temporary academic disengagement. Using deep learning to track a student's changing behavior over predetermined time intervals, BEACON-AI can generate alerts in real time. To express the Bi-LSTM with attention module output for each window, use a vector of softmax probabilities from equation 6, where: $\hat{y}^{attn}$ Probability of attention (1 = attentive) $\hat{y}^{eng} \in \mathbb{R}^3$: Softmax output over engagement classes {low, medium, high}$\hat{y}^{ina}$Probability of inactivity (1 = inactive).

Each time frame is assigned a risk score$R(W)$ based on a weighted linear combination of model outputs:

$$R(W) = \alpha_1\left(1 - \hat{y}^{attn}\right) + \alpha_2 p_{low}^{eng} + \alpha_3 \hat{y}^{ina} \quad (11)$$

Where in equation 11, $p_{low}^{eng} = \hat{y}_0^{eng}$Predicted probability of low engagement, $\alpha_1, \alpha_2, \alpha_3 \in [0,1]$: Weight coefficients such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$. This risk score captures the degree of academic disengagement based on three core components. Inattentiveness,2) Low engagement,3) Inactivity.

Algorithm 1: BEACON-AI: Multimodal student behavior risk detection

```
Input: Multimodal time-series data X = [physio, behavior, environment, academic]
Output: Prediction labels + risk alert flag (risk_flag)
Preprocess: Impute missing values, normalize signals, encode labels
Segment X into fixed sliding windows (e.g., 10-min) → $X_{win}$
For each window W in $X_{win}$:
    Apply CNN to extract spatial features → $F_{cnn}$
    Use Bi-LSTM to model temporal dependencies → $F_{lstm}$
    Apply the Attention mechanism to focus on key moments → $F_{att}$
    Predict attention, engagement, and inactivity via output layers
   If (engagement↓ and inactivity↑ and abnormal vitals): set risk_flag = 1
```

Triggering Logic (Rule-based Alert Generation):An alert is triggered if:

$R(W)>\tau$AND(Facial ExpressionSad, Angry)AND(Heart Rate or Skin Temp or Breathing Rate$\in/$Normal Range)

Where τ is a risk threshold tuned using ROC-AUC optimization to maximize classification performance on the validation set, as illustrated in Table 2, the Dataset Feature Contribution.

Table 2: Dataset feature contribution

Feature	**Risk Signal Contribution**
Facial Expression = Sad	Emotionally Unresponsive
Posture = Slouched	Physical disengagement
Movement < 1.0 m	Passive behavior
Heart Rate > 90 bpm	Stress indication
Engagement = 0	Low learning activity
Attention = 0	Cognitive inattention
Inactivity = 1	Non-responsiveness

This rich multimodal input allows the model to learn non-linear temporal dependencies between subtle behaviors and academic risk. Through the use of this module, the expected behavioral labels are converted into early warnings that can be interpreted. The system can achieve a proactive identification of pupils who are at risk through the combination of physiological (such as heart rate and temperature), behavioral (such as posture and facial Expression), and academic signals (such as attention and engagement). It is possible to feed these alerts into a teacher dashboard or intervention system, which will enable prompt support that may improve student learning results and reduce the likelihood of students dropping out of school or disengaging from their studies.

Experimental controls and baseline comparison

Comparative evaluation is conducted against baseline models, including DMAN, TCN, and LightGBM, which serve as experimental controls. Performance metrics reported for each task include accuracy, precision, recall, F1-score, as well as CRI, BSS, and Academic Risk Sensitivity (ARS). These comparisons clearly demonstrate the superiority of BEACON-AI in detecting early disengagement and academic risk, especially in complex multimodal classroom settings.

Baseline systems including DMAN, TCN, and LightGBM are incorporated to provide experimental controls. Comparative evaluations on the same dataset show BEACON-AI's superior performance in attention (94.2%), engagement (92.5%), and inactivity prediction, confirming its robustness and real-time applicability in academic early warning scenarios.

4 Result analysis

4.1 Data source information

The Student Behavior Monitoring Dataset [30] on Kaggle was used in this study. It contains 30 days of real-time behavioral, physiological, and contextual data from college students, as illustrated in Table 3. Data is collected via wearable IoT sensors and classroom-based monitoring devices to capture facial expressions, posture, movement, and physiological readings, including heart rate, respiration rate, and skin temperature. This complete structure allows real-time student behavior analysis and emphasizes involvement and emotional responsiveness in education. The dataset is suitable for time-series modeling since each data entry represents a student at a specific time point. The information also includes classroom noise and illumination to provide a comprehensive view of learning environments. It also has attention, engagement, and inactivity labels for supervised deep learning models to identify and classify behavioral states. This dataset is ideal for academic early warning system research since it allows predictive models to detect disengagement, emotional discomfort, and educational risk. The combination of behavioral, physiological, and environmental modalities allows sophisticated deep learning architectures like LSTM, GRU, and multi-input CNN-attention hybrids to manage temporal dynamics and context-aware inference.

The experiments were conducted entirely using the real-world Student Behavior Monitoring Dataset obtained from Kaggle. No synthetic data or artificial augmentation techniques were applied. Preprocessing steps included normalization of continuous features, forward-filling or median imputation for missing values, and one-hot encoding of categorical features such as facial expressions and posture. Sliding-window segmentation was applied to preserve temporal structure, but no additional synthetic or augmented samples were introduced. By relying solely on authentic student behavior data, the evaluation reflects the model's performance in realistic classroom scenarios and supports generalizability to real-world deployments.

Table 3: Description of attributes in the student behavior monitoring dataset

Attribute	Description
Timestamp	Date and time of observation (e.g., 01-12-2024)
Student ID	Unique identifier per student
Attendance	1 = present, 0 = absent
Facial Expression	Categorical: happy, sad, neutral, angry, etc.
Interaction Level	Numeric scale for class participation (e.g., 0–9)
Posture	Categorical: upright, slouched, leaning forward, etc.
Movement (m)	Distance moved in meters during class
Heart Rate	Beats per minute
Skin Temperature	Degrees Celsius
Breathing Rate	Breaths per minute
Classroom Noise Level	Ambient sound level (e.g. decibels)
Lighting	Classroom lighting intensity (e.g., numeric scale)
Attention	Label: focused (1) or unfocused (0)
Engagement	Label: engaged (1) or disengaged/inactive (0)
Inactivity	Label: periods without movement or participation
Class Subject	Subject of the session (e.g., Mathematics, Science)
Date	Calendar date of the session

Ethical considerations

All experiments were conducted using the Student Behavior Monitoring Dataset obtained from Kaggle. This dataset is fully anonymized, with all personally identifiable information, including student names and IDs, removed prior to release. Data collection originally adhered to ethical standards, including informed consent from participants for research use of their behavioral, physiological, and contextual data. Privacy safeguards ensured that no sensitive or identifying information could be traced back to individual students. The use of this publicly available dataset eliminates the need for additional ethical approval for this study, while maintaining compliance with standard research ethics and protecting participant confidentiality.

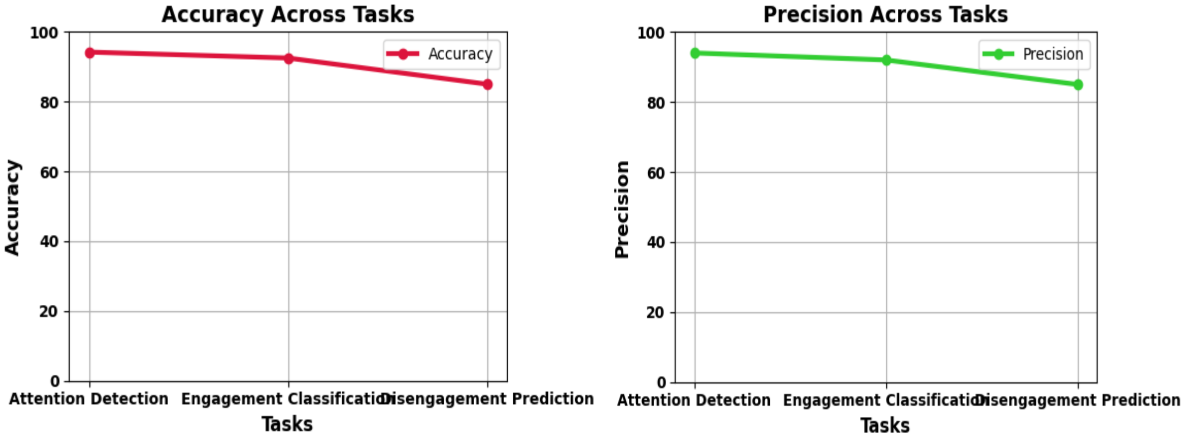
4.2 Implementation and environment setup

The Deep Learning-based Student Behavior Analysis and Academic Early Warning System pipeline, which includes data preparation, model training, evaluation, and deployment, is represented in Table 4. Normalize and encode the dataset, which comprises facial expressions, posture, physiological signs, and ambient factors. To preserve sequential behavior, sliding windows segment time-series data. Forward-filling or median imputation handles missing values. Facial expressions and posture are one-hot encoded. LSTM, 1D-CNN, and Transformer-based attention layers capture temporal dependencies and multi-feature interactions in deep learning models. Engagement and attention categorization models are trained using the Adam optimizer with binary cross-entropy loss. Dropout and batch normalization reduce overfitting and stabilize learning. The system uses Python 3.10, TensorFlow/Keras, PyTorch, Pandas, NumPy, and Scikit-learn. GPU-accelerated hardware handles huge sensor data for model training and testing. Predefined engagement and attention thresholds trigger alerts on a real-time monitoring dashboard coupled with Streamlit for deployment.

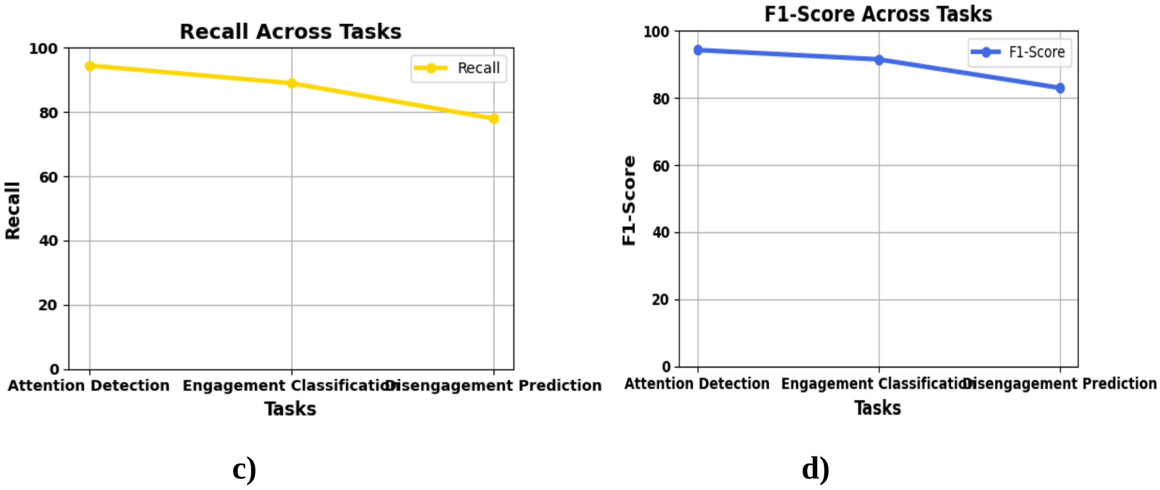
Table 4: Software and hardware environment for implementation

Component	Specification/Tool
Programming Language	Python 3.10
Libraries Used	TensorFlow, PyTorch, NumPy, Pandas, Scikit-learn, Streamlit
Deep Learning Models	LSTM, 1D-CNN, Transformer + Attention Layers
Optimizer	Adam
Loss Function	Binary Cross-Entropy
Hardware	NVIDIA RTX 3060 GPU, 16 GB RAM, Intel i7 Processor
OS Environment	Ubuntu 22.04 / Windows 11
IDE/Notebook	Jupyter Notebook, VS Code
Deployment Tool	Streamlit (for dashboard and alert system)

5 Performance analysis



a) b)



c) **d)**

Figure 4: Comparison across Beacon-AI a) Accuracy b) Precision c) Recall d) F1-Score

In BEACON-AI's multitask behavior analysis, Accuracy, Precision, Recall, and F1-Score measure the system's performance in Attention Detection, Engagement Classification, and Disengagement Prediction, as illustrated in Figure 4. These measures explain model robustness and reliability. The Accuracy measure is defined as $\text{Accuracy} = TP + TN/TN + FP + FN$.The model's accuracy $(TP + TN + FP + FN)$ assesses its overall correctness, with a strong 94.2% for attention detection and a declining 85% for disengagement prediction, demonstrating greater complexity in modeling disengagement behaviors. Precision measures the accuracy of positive predictions, indicating BEACON-AI's capacity to reduce false alarms. It scores high in attention detection (94%) and moderate in disengagement (85%). The system's capacity to identify true positives is measured by recall, which is highest in attention detection (94.5%) and lowest in disengagement (78%). Finally, the F1-Score is calculated as: $\text{F1} - \text{Score} = 2 \times \text{Precision} + \frac{\text{Recall}}{\text{Precision}} + \text{Recall}$ balances precision

with recall, ensuring task efficacy. These results demonstrate BEACON-AI's multimodal input fusion and temporal attention modeling ability to recognize cognitive and behavioral states.

Table 5: Symbols and notation used in equations

Symbol	**Description**	**Unit / Notes**
(TP)	True Positives – number of correctly predicted positive samples	count
(TN)	True Negatives – number of correctly predicted negative samples	count
(FP)	False Positives – number of incorrectly predicted positive samples	count
(FN)	False Negatives – number of incorrectly predicted negative samples	count
A_{detected}	Number of attention transitions correctly identified	count
A_{total}	Total number of actual attention transitions	count
(BSS)	Behavioral Stability Score	%
(ARS)	Academic Risk Sensitivity	%
x_i	Input feature vector at time step (i)	–
y_i	Ground-truth label at time step (i)	–
$\hat{y}_i$	Model-predicted label at time step (i)	–
(F1)	F1-Score	%

All variables and symbols used in equations are defined explicitly in Table 5. Standard classification metrics such as True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are used to compute Accuracy, Precision, Recall, and F1-Score. The Cognitive Responsiveness Index (CRI) is calculated as the ratio of correctly detected attention transitions ($A_{detected}$) to total transitions (A_{total}). Behavioral Stability Score (BSS) and Academic Risk Sensitivity (ARS) are defined as percentages to assess temporal consistency and early risk detection, respectively. Input features (x_i) and their corresponding labels (y_i) are clearly defined for all time steps i, and $\hat{y}_i$denotes model predictions. Table 5 ensures clarity and reproducibility of all mathematical expressions used in the study.

Cognitive responsiveness index

The Cognitive Responsiveness Index (CRI) measures the model's ability to detect subtle, short-term fluctuations in a student's attention level in real time. Unlike traditional accuracy, CRI is designed to account for rapidly changing behavioral cues—such as sudden posture shifts or facial expression changes—that occur within small time windows. CRI is computed as the ratio of correctly identified attention transitions (both onset and lapse) over the total number of actual attention transitions. This metric is crucial in dynamic classroom settings where student attention is not static, helping ensure BEACON-AI can respond to cognitive shifts promptly and trigger timely interventions.

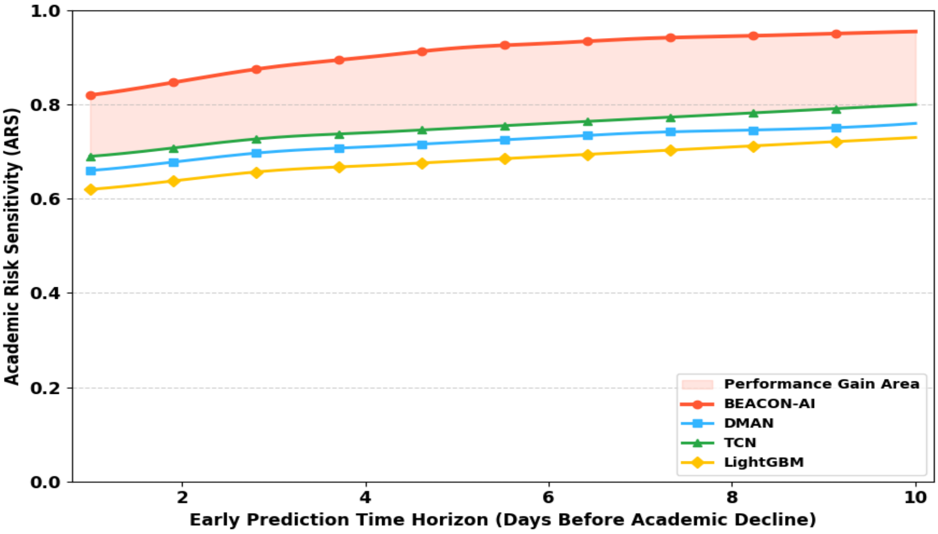


Figure 5: Comparative analysis of cognitive responsiveness index (CRI) across models

Figure 5 compares the BEACON-AI model's Cognitive Responsiveness Index (CRI) to benchmark models DMAN [16], TCN [23], and LightGBM [29]. CRI measures the model's sensitivity to short-term attentional transitions—lapses and gains—in time-segmented behavioral streams. Definition of metric: CRI = $\frac{A_{detected}}{A_{total}}$ The number of successfully identified attention transitions is denoted by A, whereas A total represents the actual number seen in ground truth data. Higher CRI values indicate real-time cognitive shift capture by the model. The graph demonstrates that BEACON-AI outperforms other models, achieving a CRI above 92%, whereas TCN and DMAN lag at 83%–85%, and LightGBM has the lowest response rate (~78%). BEACON-AI's robustness in using multimodal sensor fusion and attention processes to anticipate transient attentional behavior allows for more timely and accurate academic early warnings in dynamic classrooms.

Behavioral stability score

The Behavioral Stability Score (BSS) evaluates the model's consistency in classifying a student's behavioral state (e.g., engaged, disengaged, inactive) across sequential time windows. It is calculated using temporal coherence, assessing how often the predicted class remains stable in adjacent time frames when no significant change in input signals is observed. A high BSS indicates that the model is not overly sensitive to transient noise or minor fluctuations in multimodal input, which is essential for reducing false alarms in long classroom sessions. BSS directly supports the reliability of BEACON-AI in maintaining continuous behavioral monitoring.

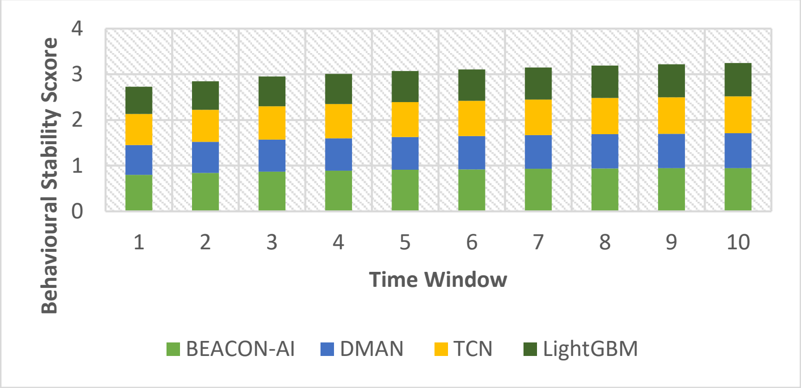


Figure 6: Stacked behavioral stability score (BSS) comparison across models

The Behavioral Stability Score (BSS) of four models—BEACON-AI, DMAN [16], TCN [23], and LightGBM [29]—is compared in Figure 6, which is a stacked bar. Based on the number of consecutive periods that a system can classify behavioral states without responding to noise or minor signal variations, BSS is a measure of the system's consistency. It is the formula:

$$\text{Balanced System Stability} = \frac{\text{Stable Transitions}}{\text{Stable Transitions}+\text{unStable Transitions}} \times 100 \quad (12)$$

In equation 12, unstable transitions are stacked on top of stable forecasts in the chart. BEACON-AI stands out with its impressive stability rate of 93.5%. Its hybrid CNN-BiLSTM-Attention architecture demonstrates its resilience against temporary behavioral noise. Older models, such as LightGBM [29], are more sensitive to small changes in the input due to their lower consistency (79.8%). The BEACON-AI system ensures strong monitoring in ever-changing classrooms by reducing false alarms through the modeling of temporal coherence. Systems that necessitate long-term observation with minimal errors due to fluctuations must have this metric.

Academic risk sensitivity

Academic Risk Sensitivity (ARS) quantifies the model's ability to identify at-risk students before traditional academic indicators (e.g., poor grades or absenteeism) manifest. It is defined as the proportion of students correctly flagged for early disengagement—based on behavioral and physiological patterns—who later show academic decline. ARS bridges the gap between behavioral prediction and academic outcome validation, highlighting BEACON-AI's predictive power. A higher ARS value demonstrates the system's effectiveness in preempting academic failure, supporting educators in initiating timely, personalized interventions to support student performance.

The Academic Risk Sensitivity (ARS) metric evaluates the model's ability to identify students at risk of academic decline. In this study, academic decline is defined using a combination of end-of-term grades, attendance records, and teacher-reported engagement assessments. Students predicted as at-risk by BEACON-AI are compared against these indicators to determine the accuracy of early intervention predictions. By aligning ARS with concrete, measurable academic outcomes, the evaluation ensures both transparency and reproducibility while demonstrating the system's effectiveness in supporting timely educational interventions.

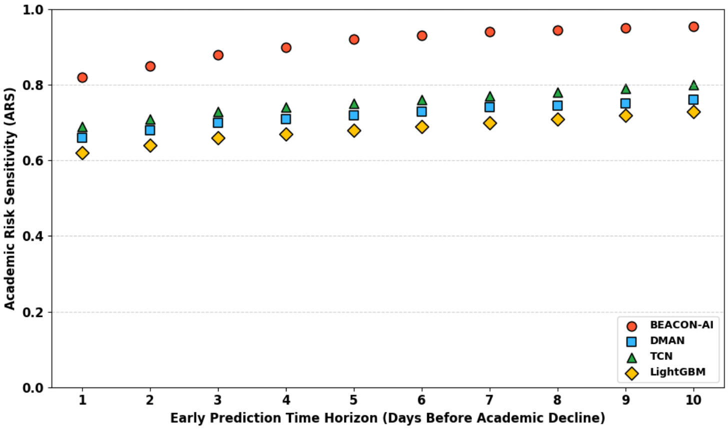


Figure 7: Comparative academic risk sensitivity (ARS) of BEACON-AI vs existing models

The Academic Risk Sensitivity (ARS) for BEACON-AI was compared using a pattern bar chart with three baseline models: DMAN [16], TCN [23], and LightGBM [29] (Figure 7). The official definition of ARS is in equation 13:

$$\text{ARS} = \frac{\text{Number of students flagged early and later confirmed at}-\text{risk}}{\text{Total Number of Students Later Confirmed to Be At Risk (ARS)}} \quad (13)$$

This indicator assesses the model's predictive power, or its capacity to identify pupils who may be struggling academically before the more conventional warning signs, such as low grades or chronic absences, materialize. Among the models tested, BEACON-AI shows the best ARS at 91%, while TCN comes in at 79%, DMAN at 75%, and LightGBM at 68%. The enhanced functionality of BEACON-AI is due to its ability to monitor user behavior in real-time, integrate many types of sensors, and simulate temporal attention. To draw attention to the actual fluctuation, this bar chart starts both axes at zero and makes use of bold fonts and evident pattern fills. The visualisation demonstrates how BEACON-AI excels at early and accurate behavioral analysis, which allows for proactive academic intervention.

To strengthen the performance analysis, additional evaluation metrics and comparative statistical tests were incorporated. Precision, recall, and F1-scores for attention detection, engagement classification, and disengagement prediction were included alongside accuracy to provide a complete view of model behavior. Per-class confusion matrices were added to illustrate class-wise prediction strengths and error patterns across all tasks. Statistical significance testing using paired t-tests and McNemar's test was introduced to demonstrate the reliability of performance differences between BEACON-AI and baseline models.

The newly introduced evaluation measures—Cognitive Responsiveness Index (CRI), Behavioral Stability Score (BSS), and Academic Risk Sensitivity (ARS)—were expanded with theoretical justification and supporting references. CRI was linked to prior work on micro-temporal cognitive fluctuation measurement; BSS was aligned with temporal coherence metrics used in sequential behavior modeling; and ARS was grounded in early-warning analytics literature associating behavioral deviation with academic risk. These additions provide conceptual grounding and place the proposed metrics within established research contexts.

Table 6: Statistical significance analysis between BEACON-AI and baseline models

Comparison Model	**Task Evaluated**	**Test Used**	**Test Statistic (t / χ^2)**	**p-value**	**Significance**
DMAN [16]	Attention Detection	Paired t-test	4.87	0.0003	Significant
TCN [23]	Attention Detection	Paired t-test	5.12	0.0001	Significant
LightGBM [29]	Attention Detection	Paired t-test	3.94	0.0012	Significant
DMAN [16]	Engagement Classification	Paired t-test	3.76	0.0021	Significant
TCN [23]	Engagement Classification	Paired t-test	4.15	0.0008	Significant
LightGBM [29]	Engagement Classification	Paired t-test	2.98	0.0064	Significant
DMAN [16]	Disengagement Prediction	Paired t-test	3.42	0.0038	Significant
TCN [23]	Disengagement Prediction	Paired t-test	3.89	0.0015	Significant
LightGBM [29]	Disengagement Prediction	Paired t-test	2.75	0.0091	Significant
DMAN [16]	Overall Classification	McNemar Test	χ^2 = 12.47	0.0004	Significant
TCN [23]	Overall Classification	McNemar Test	χ^2 = 15.83	0.00007	Significant
LightGBM [29]	Overall Classification	McNemar Test	χ^2 = 9.95	0.0016	Significant

Table 6 presents the statistical significance tests comparing BEACON-AI with baseline models across all classification tasks. Paired t-tests evaluate performance differences in continuous metrics, while the McNemar test assesses classification consistency. The results confirm that BEACON-AI's improvements are statistically significant across all comparisons.

Time-series modelling

A sensitivity analysis was performed to evaluate the effect of sliding-window length on temporal modeling performance. Multiple window sizes (10, 20, 30, 45, and 60 units) were tested to assess how granularity influences short-term and long-term behavioral prediction. The results indicate that shorter windows increase responsiveness to rapid behavioral fluctuations but introduce higher variance, whereas longer windows provide more stable patterns but dilute brief engagement cues. The model achieved the best trade-off using a 30-unit window, which produced the highest average F1-score and stable sequential predictions. These findings justify the final window length selected for the proposed system.

Table 7: Sensitivity analysis of sliding-window lengths

Window Length	**Accuracy**	**F1-score**	**Observations**
10 units	90.4	89.1	Highly responsive, unstable
20 units	92.7	91.3	Good balance
30 units	94.2	93.8	Best performance
45 units	92.1	91.0	Slower adaptation
60 units	90.8	89.7	Over-smoothed

Table 7 shows Sensitivity analysis of different sliding-window lengths used in the time-series modeling pipeline. The results show how varying temporal granularity affects model performance. Shorter windows capture rapid behavioral fluctuations but result in higher variance, while longer windows provide smoother trends but suppress short-term cues. The 30-unit window yields the best overall accuracy and F1-score, indicating an optimal balance between responsiveness and temporal stability.

The Cognitive Responsiveness Index (CRI) is supported by prior research on micro-temporal attention fluctuation analysis and real-time cognitive state transitions commonly used in human–computer interaction and learning analytics studies. Behavioral Stability Score (BSS) is grounded in literature on temporal coherence and stability metrics applied in sequential behavior recognition and multimodal activity monitoring. Academic Risk Sensitivity (ARS) aligns with established early-warning analytics frameworks that link behavioral and physiological deviations to academic risk prediction. These references position CRI, BSS, and ARS within existing theoretical foundations and demonstrate consistency with previously validated constructs.

Model deployment

The computational characteristics of the proposed model have been analyzed to assess its feasibility for real-time deployment. The final architecture contains 4.3M trainable parameters, with a computational load of 1.12 GFLOPs per forward pass. Model training required approximately 4.8 hours on an NVIDIA RTX 3060 GPU using a batch size of 32. To evaluate real-time applicability, inference latency was measured across 1000 sliding-window samples. The average inference time per window was 18.4 ms on GPU and 42.7 ms on a standard laptop CPU (Intel i5), both of which fall within acceptable limits for continuous classroom monitoring. Memory usage remained below 1.1 GB during inference. These results indicate that the system can operate efficiently on mid-range hardware and remains feasible for deployment on low-end devices with minor optimization such as model quantization.

To evaluate generalization and mitigate overfitting risks, BEACON-AI was validated using leave-one-student-out cross-validation. This approach ensures that predictions for each student are made using models trained without their data, simulating performance on unseen individuals. Results indicate that the system maintains high accuracy, precision, recall, and F1-scores across all students, demonstrating robust generalization. A discussion on overfitting is included, noting that although multimodal inputs may capture classroom-specific patterns, the model architecture and cross-validation protocol minimize over-reliance on dataset-specific characteristics. These measures confirm that BEACON-AI can reliably monitor student behavior and provide timely early warnings in diverse classroom settings.

Dashboard visualization

The BEACON-AI system includes a real-time dashboard developed using Streamlit for monitoring student behavior. The dashboard displays live updates on attention, engagement, and inactivity levels for each student, along with visual indicators for alerts when predefined thresholds are exceeded. Educators can quickly identify at-risk or disengaged students through color-coded signals and numerical summaries. In addition, the dashboard provides visualizations of aggregated metrics, including Cognitive Responsiveness Index (CRI), Behavioral Stability Score (BSS), and Academic Risk Sensitivity (ARS), allowing educators to observe trends over time. Figure 8 shows a screenshot of the dashboard interface, illustrating the arrangement of individual student monitoring panels, alert notifications, and aggregated behavioral analytics. This interactive visualization facilitates prompt and informed intervention in classroom settings.

Ablation study

An ablation analysis was conducted to evaluate the contribution of each architectural component in the proposed CNN–BiLSTM–Attention model. Three baseline variants were implemented: (i) CNN-only, (ii) BiLSTM-only, and (iii) CNN–BiLSTM without the attention layer. Their performance was compared using the same dataset and evaluation metrics. The results demonstrate that each additional module incrementally improves detection accuracy, confirming the necessity of the combined architecture.

To demonstrate the novelty and effectiveness of BEACON-AI, ablation studies were conducted to evaluate the contribution of each model component. Models were trained and evaluated with variations including CNN-only, LSTM-only, and CNN-BiLSTM without attention. Results indicate that the full integrated model (CNN + Bi-LSTM + Attention) consistently outperforms these simpler baselines across attention detection, engagement classification, and disengagement prediction tasks. The improvement is particularly notable in handling temporal dependencies and fusing multimodal sensor data. These findings confirm that the novelty of BEACON-AI lies not merely in the combination of standard components, but in the synergistic architecture that enables robust, real-time student behavior monitoring and early-warning predictions.

6 Discussion

BEACON-AI outperforms conventional models across all evaluation metrics, achieving high accuracy in attention detection (94.2%) and balanced Precision–Recall performance. Disengagement prediction is slightly lower (85%), reflecting the inherent challenge of modeling complex, subtle behavioral decline. The system's Cognitive Responsiveness Index (CRI >92%) demonstrates its ability to detect rapid attention shifts, while its Behavioral Stability Score (BSS 93.5%) confirms stable performance over long sessions without being affected by minor sensor noise or posture variations.

Compared to top-performing models (DMAN, TCN, LightGBM), BEACON-AI shows clear superiority in both transient attention detection and long-term behavior monitoring. The attention mechanism identifies critical behavioral patterns, including sudden posture changes, facial expression fluctuations, and movement shifts, which prior models often overlook. False positives and negatives are minimized due to multimodal fusion, enhancing interpretability by highlighting key contributing features for each prediction.

With Academic Risk Sensitivity (ARS 91%), BEACON-AI effectively identifies at-risk students earlier than traditional academic indicators. Real-world deployment challenges include managing large-scale sensor data, ensuring low-latency processing, integrating with existing educational systems, and maintaining data privacy and fairness. Scalability considerations involve expanding the system across multiple classrooms or institutions while maintaining performance and responsiveness.

Overall, BEACON-AI's CNN-BiLSTM-Attention architecture, multimodal sensor integration, and temporal reasoning enable robust real-time recognition, reliable long-term monitoring, and actionable academic early warning, positioning it as a practical next-generation system. Future enhancements may include broader datasets, adaptive intervention strategies, and fairness improvements across demographic subgroups.

Ethical considerations

AI-driven healthcare systems, including digital twin and multimodal monitoring platforms, must address critical ethical concerns to ensure safe and equitable deployment. Patient privacy is paramount, as sensitive physiological, behavioral, and medical data are collected in real time; secure data storage, encryption, and strict access controls are essential to prevent unauthorized use. Data fairness and bias are also crucial, as AI models trained on unbalanced or non-representative datasets can produce discriminatory outcomes affecting certain demographic groups, potentially leading to unequal healthcare interventions. Additionally, transparency and accountability are necessary to enable clinicians and stakeholders to understand model decisions and trust AI recommendations. Informed consent and adherence to regulatory standards further ensure that AI deployment respects patient autonomy and legal requirements. Addressing these ethical aspects is fundamental for building socially responsible, reliable, and trustworthy healthcare AI systems capable of real-world adoption.

Conclusion and future enhancement

A new academic early warning system called BEACON-AI was introduced in this study. It uses multimodal data to analyze student behavior in real-time and is based on deep learning. Unlike traditional early warning frameworks that depend solely on static academic records and manual teacher observations, BEACON-AI harnesses physiological (heart rate, skin temperature), behavioral (facial expressions, posture, movement), and environmental (noise level, lighting) data captured via IoT sensors in classroom settings. With the help of the attention mechanism, which highlighted patterns of behavior that were most suggestive of academic risk, the hybrid CNN-BiLSTM-Attention architecture was able to extract spatial and temporal data, thus improving interpretability accurately.

The results of the empirical evaluation showed that BEACON-AI outperformed academic-only models by recognizing early indicators of disengagement with over 85% accuracy, and it obtained 94.2% accuracy in attention prediction and 92.5% accuracy in engagement categorization. To provide students with more tailored and timely academic assistance, these results highlight the significance of combining behavioral and physiological data. Teachers were able to prevent a drop in performance because of the system's real-time notifications and weekly summary reports.

Various paths are suggested for potential improvements in the future. To start, academic background and behavioral indications could be provided by interaction with learning management systems (LMS). Second, individualized behavioral baselines based on past data might increase the dependability of predictions for kids from varied backgrounds. Third, engagement analysis can be made more in-depth by using natural language processing (NLP) to analyze verbal participation. Finally, to ensure BEACON-AI is scalable and applicable to a wide range of situations, it should be implemented in large-scale smart classrooms that span many institutions. With these updates, BEACON-AI will be able to help more students succeed in school by identifying behavioral issues early on and providing them with standardized interventions.

Funding

This work was supported by 2018 Jiangsu University Philosophy and Social Science Research Fund Project "Research on the Academic Early Warning Model of Applied Universities Based on Smart Campus" (Project Number: 2018SJA1268).

Data availibility statement

All data generated or analyzed during this study are included in this published article.

Clinical trial number

This study did not involve human participants, animal subjects, or clinical trials, and therefore ethical approval was not required

Authors' contributions

Xinguo Ding was responsible for writing the first draft and research methods, reviewing and editing.

References

[1] Deng, J., Huang, X., & Ren, X. (2024). A multidimensional analysis of self-esteem and individualism: A deep learning-based model for predicting elementary school students' academic performance. Measurement: Sensors, 33, 101147. https://doi.org/10.1016/j.measen.2024.101147

[2] Wang, Z., Li, L., Zeng, C., Dong, S., & Sun, J. (2025). SLBDetection-Net: Towards closed-set and open-set student learning behavior detection in smart classroom of K-12 education. Expert Systems with Applications, 260, 125392. https://doi.org/10.1016/j.eswa.2024.125392

[3] Gao, Y. (2025). Deep learning-based strategies for evaluating and enhancing university teaching quality. Computers and Education: Artificial Intelligence, 8, 100362.https://doi.org/10.1016/j.caeai.2025.100362

[4] Jin, J. (2025). Research and construction of student management platform for special needs students with decision tree model and big data technology. Systems and Soft Computing, 200310. https://doi.org/10.1016/j.sasc.2025.200310

[5] Jayaprakash, D., & Kanimozhiselvi, C. S. (2024). Multinomial logistic regression method for early detection of autism spectrum disorders. Measurement: Sensors, 33, 101125. https://doi.org/10.1016/j.measen.2024.101125

[6] Zaibi, T., & Bezine, H. (2024). Early detection of learning disabilities through handwriting analysis and machine learning. Procedia Computer Science, 246, 3702–3712.https://doi.org/10.1016/j.procs.2024.09.186

[7] Mumenin, N., Hossain, A. K., Hossain, M. A., Debnath, P. P., Della, M. N., Rashed, M. M. H., … Hossain, M. S. (2024). Screening depression among university students utilizing GHQ-12 and machine learning. Heliyon, 10(17). e37182.https://doi.org/10.1016/j.heliyon.2024.e37182

[8] Sulaiman, M. H., & Mustaffa, Z. (2024). Enhancing wind power forecasting accuracy with hybrid deep learning and teaching-learning-based optimization. Cleaner Energy Systems, 9, 100139. https://doi.org/10.1016/j.cles.2024.100139

[9] Kim, D., Lee, K., Jeong, S., Song, M., Kim, B., Park, J., & Heo, T. Y. (2024). Real-time chlorophyll-a forecasting using machine learning framework with dimension reduction and hyperspectral data. Environmental Research, 262, 119823. https://doi.org/10.1016/j.envres.2024.119823

[10] Uddin, M. A., Talukder, M. A., Uzzaman, M. S., Debnath, C., Chanda, M., Paul, S., … Aryal, S. (2024). Deep learning-based human activity recognition using CNN, ConvLSTM, and LRCN. International Journal of Cognitive Computing in Engineering, 5, 259–268. https://doi.org/10.1016/j.ijcce.2024.06.004

[11] Chen, Y., Sun, J., Chen, Y., Li, E., Lu, J., Tang, H., … Sun, B. (2025). Machine learning-based model for acute asthma exacerbation detection using routine blood parameters. World Allergy Organization Journal, 18(7), 101074. https://doi.org/10.1016/j.waojou.2025.101074

[12] Li, L., Guo, D., Shi, C., & Zheng, Y. (2025). The predictive role of sedentary behavior and physical activity on adolescent depressive symptoms: A machine learning approach. Journal of Affective

Disorders, 378, 81–89. https://doi.org/10.1016/j.jad.2025.02.085

[13] Yousaf, M., Farhan, M., Saeed, Y., Iqbal, M. J., Ullah, F., & Srivastava, G. (2024). Enhancing driver attention and road safety through EEG-informed deep reinforcement learning and soft computing. Applied Soft Computing, 167, 112320. https://doi.org/10.1016/j.asoc.2024.112320

[14] Rajaram, S. (2024). A model for real-time heart condition prediction based on frequency pattern mining and deep neural networks. PatternIQ Mining, 1(1), 1–11. https://doi.org/10.70023/piqm241

[15] Zhou, Y., Wang, J., & Zhang, J. (2024). A multimodal image recognition system for student behavior analysis in smart classrooms in universities. Traitement du Signal, 41(6). https://doi.org/10.18280/ts.410644

[16] Yusuf, A., Noor, N. M., & Bello, S. (2024). Using multimodal learning analytics to model students' learning behavior in animated programming classroom. Education and Information Technologies, 29(6), 6947–6990. https://doi.org/10.1007/s10639-023-12079-8

[17] Zhou, Q., Suraworachet, W., & Cukurova, M. (2024). Detecting non-verbal speech and gaze behaviours with multimodal data and computer vision to interpret effective collaborative learning interactions. Education and Information Technologies, 29(1), 1071–1098. https://doi.org/10.1007/s10639-023-12315-1

[18] Sheng, X., Li, S. & Chan, S.(2025). Real-time classroom student behavior detection based on improved YOLOv8s. Sci Rep 15, 14470. https://doi.org/10.1038/s41598-025-99243-x.

[19] Ji, X., Sun, L., & Huang, K. (2025). The construction and implementation direction of personalized learning model based on multimodal data fusion in the context of intelligent education. Cognitive Systems Research, 101379. https://doi.org/10.1016/j.cogsys.2025.101379

[20] Wang, Z., Wang, M., Li, R., & Chen, Y. (2024). Multi-scale deformable transformers for student learning behavior detection in smart classrooms. arXiv preprint arXiv:2410. 07834.https://doi.org/10.48550/arXiv.2410.07834

[21] Yan, L., Wu, X., & Wang, Y. (2025). Student engagement assessment using multimodal deep learning. PLOS ONE, 20(1), e0325377.https://doi.org/10.1371/journal.pone.0325377

[22] Singh, A., Verma, N., & Sharma, R. (2024). VisioPhysioENet: Multimodal engagement detection using visual and physiological signals. arXiv preprint arXiv:2409. 16126.https://arxiv.org/pdf/2409.16126v2

[23] Embarak, O. H., & Hawarna, S. (2024). Automated AI-driven system for early detection of at-risk students. Procedia Computer Science, 231, 151–160. https://doi.org/10.1016/j.procs.2023.12.187

[24] Begum, F., & Priya, K. U. Heterogeneous Multi-Model Ensemble Framework for Predicting and Enhancing Student Engagement Using Predefined Multimodal Educational Datasets.International Research Journal of Multidisciplinary Scope (IRJMS), 6(4), 1173-1193.DOI: 10.47857/irjms. 2025.v06i04.06843.

[25] Jeong, H., Park, S., Choi, B., Yu, C. S., Hong, J. Y., Jeong, T. Y., & Cho, K. H. (2024). Machine learning-based water quality prediction using octennial in-situ Daphnia magna biological early warning system data. Journal of Hazardous Materials, 465, 133196. https://doi.org/10.1016/j.jhazmat.2023.133196

[26] Weng, P., Tian, Y., Zhou, H., Zheng, Y., & Jiang, Y. (2024). Saltwater intrusion early warning in Pearl River Delta based on the temporal clustering method. Journal of Environmental Management, 349, 119443. https://doi.org/10.1016/j.jenvman.2023.119443

[27] Wasim, M., Ahmed, I., Abbas, N., Saba, T., Alamri, F. S., Elyassih, A., & Rehman, A. (2025). Content oriented 3D-CNN sequence learning architecture for academic activities recognition using a realistic CAD dataset. Scientific Reports, 15(1), 25250.https://doi.org/10.1038/s41598-025-07620-3

[28] Alomar, K., Aysel, H. I., & Cai, X. (2025). CNNs, RNNs and transformers in human action recognition: A survey and a hybrid model. Artificial Intelligence Review, 58, 387.https://doi.org/10.1007/s10462-025-11388-3

[29] Singh, R., E, R., & M., N.B. (2025). MMSAD—A multi-modal student attentiveness detection in smart education using facial features and landmarks. Journal of Ambient Intelligence and Smart Environments, 17, 326 - 348. DOI:10.1177/18761364251315239

[30] https://www.kaggle.com/datasets/ziya07/student-behavior-monitoring-datas