AE-LSTM-Based Multimodal Sensing System for Real-Time Monitoring of Children's Play Behavior on Edge Devices

Yuexing Li[1*], Xiaojia Shen[1], Ye Yang[2]
[1]School of Education, Shaanxi Fashion Engineering University, Xi'an, Shaanxi, 712046, China
[2]Wenlin kindergarten, Xianyang, Shaanxi,712099, China
E-mail: yuexing_li@outlook.com

[*]Corresponding author

Keywords: children, play behavior, sensors, wearable devices, ResNet, Artificial Intelligence (AI).

Received: August 7, 2025

Children's play is a fundamental activity that supports emotional, cognitive, and social development. However, capturing and analyzing play behavior in real time is challenging due to its spontaneous, multimodal, and dynamic nature. Traditional observation methods are time-consuming, subjective, and lack real-time responsiveness. This research aims to design and implement a multimodal sensing and feedback platform that leverages edge computing and real-time Artificial Intelligence (AI) to monitor, interpret, and support children's play behavior. The platform collects multimodal play behavior datasets from various sensors, including action and posture recognition, microphones for speech and voice tone analysis, motion sensors to track physical activity, and wearable devices. An Autoencoder-based Long Short-Term Memory (AE-LSTM) network is used to analyze behavior in real time. Feature extraction is performed using a lightweight ResNet model to extract features. Data is pre-processed using Kalman filtering and normalization techniques to reduce noise and improve consistency. The entire system is deployed on edge devices to ensure low-latency processing, local storage, and privacy preservation. The system also provides real-time feedback through visual and haptic cues to enhance engagement. Implemented in Python, experiments have demonstrated that the proposed AE-LSTM model outperforms baseline architectures like LSTM, GRU, and BiLSTM+Attention, and the proposed model achieves higher results according to the F1-score (0.959), accuracy (0.975), recall (0.964), and precision (0.968). These findings offer robust performance in naturalistic settings and provide valuable applications for educators, therapists, and researchers who intend to support and understand child development through intelligent, responsive play environments.

Povzetek: Razvita je bila večmodalna robna AI-platforma za sprotno spremljanje in analizo otroške igre z namenom podpore otrokovemu razvoju.

1 Introduction

Playing is one of the most important childhood activities and the fundamental setting for young children's learning. Parents are crucial in supporting, guiding, and scaffolding children's play, and investigations showed that parent-child play is associated with kids' social competence and pro-social skill development [1]. Play has been described as self-motivated, player-controlled, process-oriented, compared to product-focused, non-literal, lacking rules imposed from outside, and involving active player participation. Play is a crucial component of the childhood curriculum as an educational resource for young children, with consequences for both academic and social-emotional growth [2]. Play enables children to develop cognitive abilities, language abilities, executive functions, and socio-emotional competency. Children's play behaviors represent the social-emotional growth, persistence, imagination, and inventiveness [3]. Primary behavior in society is exhibited through play and games; children improve their social abilities with other children. Children with disabilities need play behavior to maintain or improve social abilities [4]. Children's physical play, like running, jumping along with time spent outside, is a form of physical activity that assists in preventing obesity while supporting the mental and physical wellness of the children [5]. A range of social, cognitive, and physical/locomotor skills that children exercise during play is used to classify play behavior. The emphasis on play behaviors contributes to cognitive skill development [6]. The development and application of societal abilities and interests suffer significantly in children with illness, which can impact the social interactions and potentially lead to anxiety. Children's flexibility and satisfaction are emphasized through playtime. Specifically, 16% of young children's device usage is spent playing digital games. Twenty-three minutes per day on average are used for playing the games on a computer, tablet, Smartphone, or compact video game console [7]. The unlimited possibilities that the real-world circumstances compared with constructed circumstances are designed with particular objective, which are the contributing factors of

children [8]. The following requirements require being exceeded for an activity to be perceived as playful include enjoyment, active participation, significance or pertinence, social interaction, and repetition and diversity. Incorporating play into a process provides the greatest developmental benefits for young kids, and not all qualities are necessary for activities considered playful [9]. The play environment's physical attributes and materials provide more impact on playing behaviors. A variety of playing alternatives that are difficult to replicate indoors are provided by the distinct qualities and pressures of outdoor play areas [10].

1.1 Problem statement

Current advancements in AI-driven child behavior monitoring have greatly enhanced multimodal play data in real time. Nevertheless, several vital limitations exist in the current literature, which frequently fail to capture the spontaneous and dynamic nature of children's play, and maintaining privacy when analyzing sensitive behavioral information. The SOOPEN model relies heavily on manual observation, making it vulnerable to observe bias despite with high reliability scores. The utilization of class groups during observation might limit unplanned natural play behavior, which might affect the outcomes. The DNN model was capable of classifying CT characteristics, its practical value was limited. The DNN model fails to support the educators incorporated in actual classroom procedures. The small-sized dataset further limits its efficiency by the model's scalability, robustness, and classification findings. To address these problems, the AE-LSTM approach was used to accurately capture, analyze, and interpret children's play behavior in real time. An AE-LSTM model is used to manage diverse sensor inputs and ensure high-fidelity semantic understanding. The proposed solution supports real-time feedback and informed decision-making across educational, therapeutic, and developmental settings.

1.2 Aim and contributions of this research

The aim of this research is to design and implement a real-time, intelligent multimodal sensing and analysis platform capable of accurately monitoring children's play behavior in naturalistic environments by Autoencoder-based Long Short-Term Memory (AE-LSTM) model. The AE-LSTM model helps to learn compact representations and capture temporal dependencies in play sequences. The suggested model is deployed on edge devices, thus supporting privacy-preserving, real-time decision-making. The AE-LSTM model helps to identify emotional states, and social interactions of children's.

- ❖ The platform gathers information from a variety of sensors, such as movement sensors to track activity levels, microphones to analyze speech and voice tones, RGB-D cameras to recognize posture and action, and wearable technology that tracks physiological indicators like skin temperature and heart rate to determine emotional states.
- ❖ The obtained data are preprocessed by the Kalman filter and z-score normalization for noise reduction and consistency enhancement. Whereas, essential features are extracted through the Lightweight ResNet model.
- ❖ Effective performances of the playing behavior of the children are assessed by the AE-LSTM. According to experimental results, play behavior classification, emotional state detection, and social interaction identification were all accomplished with high accuracy.

2 Relevant articles

Using a group dynamics approach, the System for Observing Outdoor Play Environments in Neighborhood (SOOPEN) tool to evaluate school-aged children's play behavior and calculate its inter-observer reliability was developed [11]. Based on two thorough observation devices, SOOPEN was evaluated at eleven elementary schools. All variables showed strong consistency between observers, according to Kendall's tau b ($\tau b > 0.7$, p values < 0.05). Children had limited access to play in specific areas while in class groups.

The impact of emotional coaching and distraction techniques used by teachers on continuous development of societal and non-societal play behaviors was investigated [12]. 275 instructors and 487 children from 123 classrooms across 56 facilities were obtained. According to the findings, emotion coaching contributed to a sharper reduction in nervous behavior and a sharper increase in social play. The analysis did not provide direct comparisons across teachers from various cultural backgrounds.

Considering a specific emphasis on the nature of play, game creation and participants acting as facilitators in the play, the research [13] investigated the socio-dramatic play occurring in an early childhood educational environment. Results indicated that a key component of classroom play culture was that children established games with standards. The major limitation was that it was conducted with only 10 children from a single classroom.

Based on a seven-month ethnographic investigation, the exploration [14] described the efficiency of a robot that was implemented with two primary education children within the ages $1-2$ and $3-5$. To investigate the efficiency of the children's play with the robot, it descriptively combined the structure with qualitative interviews ($n = 6$) for children's play evaluation. However, only two distinct case groups were included in the limitations observed in the research. A summary of related works on Children's Play Behavior is illustrated in Table 1.

Table 1 Summary of literature review on children's play behavior and AI-based interaction systems

Ref	Technology Used	Objective	Result	Challenges / Limitations
[15]	Deep Neural Network (DNN)	To examine children's multimodal video-based Computational Thinking (CT)	AI models could classify numerous CT features, acting as an alternative team member in assessment	Lacked the ability to create an ML model to assist humans; provided limited data
[16]	Comparative observation	Compare screen time and playtime of preschool-aged children before and during COVID-19	Significant variations in screen time and playtime between weekdays and weekends (playtime: 3.55 ± 2.49 vs. 4.11 ± 2.58 h)	Findings limited to parents and children; not generalizable to other populations
[17]	AI-based educational games	To personalize educational game boundaries with player assessment identity	Instantaneous updates of game components; children performed more effectively	Lack of modifications despite positive results; limited pattern identification
[18]	Micro-longitudinal observation	Observe media influence on 150 children's play in a museum	Emotional and social expression of children not significantly influenced by screens integrated into monitors	Fails to provide continuous implications on children's play areas
[19]	Machine Learning (ML)	To identify and protect against child predatory behavior in online games	Examined risks children faced while playing online video games	Lacked coordinated responses to protect children using various digital platforms
[20]	Observational study, Executive Function (EF)	To investigate the relationship between play behaviors and EF components	EF development and play behavior were related	Due to Small sample size (97 children) lacks findings
[21]	Motor skill protocol / observational	To investigate connections between preschoolers' break behaviors and Foundational Movement Skills (FMS)	The motion time negatively correlated with total/locomotor skills; play without tools positively associated with other play behaviors	Focused on limited FMS types; observational and correlational findings only
[22]	Bi-directional-LSTM-Attention	Play behavior modeling and interaction system optimization in games	Outperformed traditional models in accuracy, click precision, response delay, and user satisfaction; improved adaptability and smoothness	Requires large-scale datasets; high computational requirements; implementation limited to gaming context

3 Research methodology

The use of multimodal sensing and edge AI technology, the research aims to develop a smart, real-time platform that monitors and supports children's play behavior. The research obtains the multimodal play behavior dataset.

The obtained data are preprocessed through the Kalman filter to reduce noise in the obtained information, and the z-score normalization is used to enhance the consistency of the data through the normalization process. The Lightweight ResNet approach is employed to extract the significant information from the processed data. To assess the children's playing behavior in real-time, the AE-LSTM is proposed in the research. Figure 1 depicts the process of methodology.

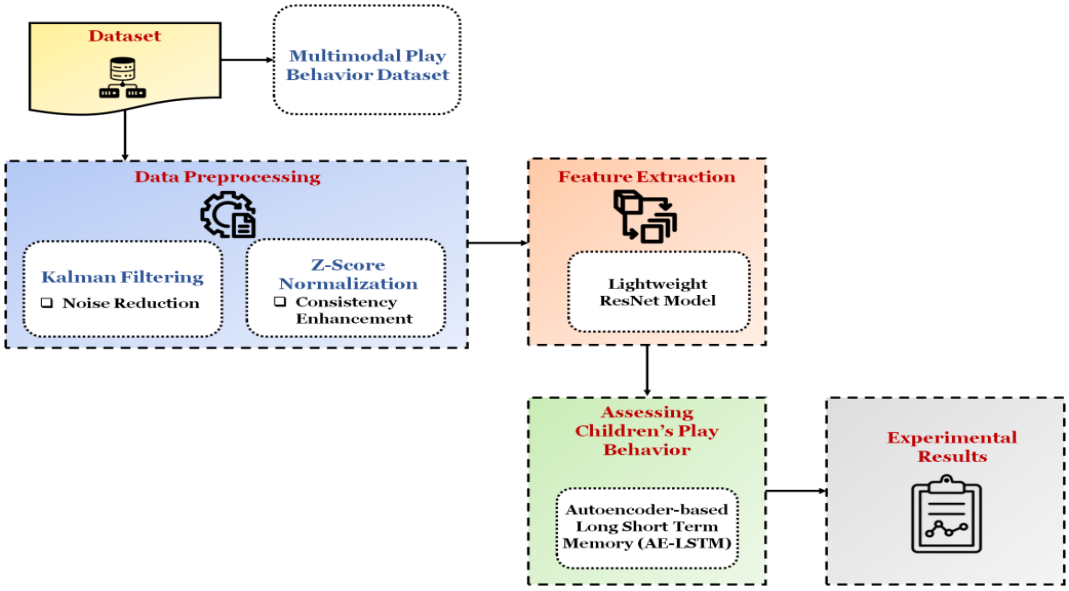

Figure 1: Proposed methodology processes

3.1 Dataset

The Multimodal Play Behavior Dataset is obtained [https://www.kaggle.com/datasets/ziya07/multimodal-play-behavior-dataset/data]. This dataset uses data from many real-world sensing sources to imitate children's play behavior. It consists of 12,480 synthetic multimodal time-series samples representing simulated children aged from 3–10 years. The dataset includes rich motion, acoustic, physiological, and contextual sensor streams across five annotated play behavior categories. The dataset simulates a naturalistic play environment with varying motion intensity, social proximity, and expressive characteristics. The data was split into 70% training and 30% testing. It consists of thorough, time-stamped recordings of children's vocalizations, body movements, and physiological conditions through different kinds of play. The information is arranged in synchronized records that show a child's vocalizations, body language, and emotional cues in a natural play environment. To facilitate research and learning in the fields of education, psychological development, and intelligent play circumstances, each entry is assigned a specific category based on play behavior. There are several key features, which are represented in Table 2.

Table 2: Significant features determined in the dataset

Features	Descriptions
Observations of Multimodal	It comprises motion, auditory, physiological, and physical data that illustrate different facets of play behavior.
Behavior Labels	Five categories of play behavior have been identified: Parallel play, cooperative play, playing alone, aggressive behavior, and inactive play.
Time-Series Information	Real-time observation is simulated by providing a distinct timestamp from 2024 to each record.
Signals of Emotion and Interaction	Body posture, verbal activity, social proximity, and emotional markers like heart rate are all represented by features.
Research-Focus	Established to support the comprehension of behavioral and interpersonal patterns in child development for educators, researchers, and developers.

3.2 Data preprocessing

The process of converting unprocessed data into a format that is more appropriate for modeling and evaluation is known as information processing. Obtained information requires being cleaned, transformed, and integrated to enhance its quality and facilitate the system's comprehension and processing. It fixes anomalies like missing data, inconsistencies, and noise to prepare the data for neural network algorithms. When examining children's play behavior, data preprocessing is essential. Academics and professionals acquire more information about the complicated dynamics of children's play behavior that is utilized for directing activities, learning techniques, and child development support. Two preprocessing techniques, such as Kalman filtering and z-score

normalization, are employed to evaluate the children's playing behavior.

3.2.1 Kalman Filtering to reduce noise

A Kalman filtering is an effective data preparation technique that smooth the noisy sensor streams, such as motion trajectories, object interactions, and ambient cues to ensure childrens play behavior. The Kalman filtering is appropriate for low-power edge devices used in classrooms or rehabilitation facilities due to its computational efficiency. The data instance's value is calculated by the Kalman filter using the observed value of the present instance and the known estimated value of the preceding moment. The Kalman filter is a probability distribution issue that determines the probability of the future by utilizing probability distribution and prior values. Using the state space technique, the Kalman filter characterizes the system's dynamic properties. The filter operates through two essential steps like prediction and correction. The state-space model in Equation (1).

$$A_{l+1} = XA_l + Y\mu_l + \omega_l \quad (1)$$

Where, ω_l is the process noise or disturbance, A is the actual system state, the state transition matrix is indicated as XA_l, and the control matrix is $Y\mu_l$, along with the control variable (μ).

Kalman filter allows the model to estimate child motion or posture even when sensors momentarily drop, fluctuate, or report inconsistent values. An error (f_{l+1}) calculation is indicated in Equation (2), and Equation (3) represents the uncertainty estimation ($Q_{l+1|l+1}$).

$$f_{l+1} = A_{l+1} - \hat{A}_{l+1|l+1} \quad (2)$$

$$Q_{l+1|l+1} = F(f_{l+1}f_{l+1}^S) = F\left(\left(A_{l+1} - \hat{A}_{l+1|l+1}\right)\left(A_{l+1} - \hat{A}_{l+1|l+1}\right)^S\right) \quad (3)$$

Exploring the position by employing the mathematical evaluation is presented in Equation (4).

$$\hat{A}_{l+1|l} = A_{l+1} - \widehat{XA}_{l|l} + Y\mu_l \quad (4)$$

The expected system state vector at time step l is represented by $\widehat{XA}_{l|l}$. The noise process uncertainty is denoted in Equation (5).

$$P_{l+1} = F(\omega_l + \omega_{l+1}^S) \quad (5)$$

The covariance matrix of the noise process is represented by P_{l+1}, whereas the noise process is represented by ω_{l+1}^S. Updated covariance is determined in Equation (6), and the uncertainty measure is calculate.

$$Q_{l+1|l+1} = (K - L_{l+1}G)Q_{l+1|l} \quad (6)$$

Where, covariance matrix is $Q_{l+1|l}$. It helps to refine the prediction by using real sensor measurements. Estimation of Kalman gain and the updated positions are indicated by Equations (7-8).

$$L_{l+1} = Q_{l+1|l}G^s\left(GQ_{l+1|l}G^s + Q_{l+1}\right)^{-1} \quad (7)$$

$$\hat{L}_{l+1|l+1} = \hat{L}_{l+1|l} + L_{l+1}\left(W_{l+1} - G\hat{A}_{l+1|l}\right) \quad (8)$$

Where, W_{l+1} is a measurement $\hat{A}_{l+1|l}$ is an anticipated system's state vector at time step l, and $\hat{L}_{l+1|l+1}$ is an assessed state vector of the system with time step $l+1$.

3.2.2 Z-score normalization to enhance consistency

The data preprocessing method, Z-score normalization, frequently referred to as standardization, converts data so that its mean is zero and its standard deviation is one. In this procedure, the
data is transformed into a unit variance and positioned at zero. Equation (9) denotes the z-score calculation.

$$Z - score = \frac{A-\mu}{\sigma} \quad (9)$$

Where, A represents the initial value, mean and standard deviation are represented by μ and σ.

3.3 Feature extraction

The process of turning incomplete information into a collection of new, important features that are more appropriate for predictive algorithms is known as feature extraction. It intends to enhance model performance, facilitate data representation, and lower dimensionality. Alternatively, selecting particular portions of the original features involves combining or altering existing characteristics to create new ones. The procedure of identifying and measuring specific features of a child's play that are subsequently utilized for evaluation, categorization, or other uses is known as feature extraction in terms of children's play behavior. Characteristics of the play activity, including play category, interactions with others, involvement level, and physical motions, are

represented by these features, which are easily obtained from a variety of data sources, including audio, video records, and sensor data. The Lightweight ResNet model is utilized in the research to extract the significant features from the children's playing behavior.

3.3.1 Lightweight ResNet model

Based on one of the existing training techniques, the ResNet50 model is employed in the playing behavior of children assessment. The collection of a pre-trained model is used to establish a model that fails to comprehend anything about images. By allowing training with fewer data sets, ResNet50 reduces the computing expenses. The ResNet50 model's input layer is configured to receive 224,224,1 values from the data set. After the input layer, the convolution layer's values are then updated. Figure 2 shows the entire architecture of ResNet50, collectively with the additional levels.

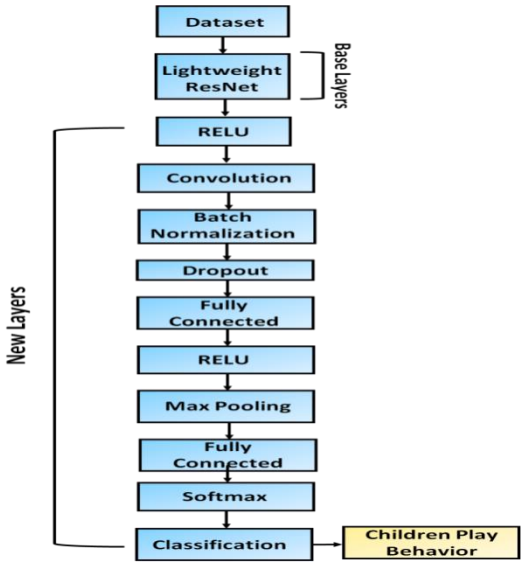

Figure 2: Entire Architecture of ResNet50

The new model is built from the ResNet50 model's input, convolution, activation, pool, fully-connected, softmax, and classification layers. Two new fully connected layers are generated; batch normalization is implemented for input values as well as stability and speed are improved. The output layer's fully connected structure uses Softmax activation for data classification. Dropout prevents the model from remembering training data.

Input Layer: This layer serves as the model's primary layer, and this layer's highest selection of input image sizes increased the amount of storage needed while extending the training and testing durations. As a result, all architectures of the input layer are determined to be $224 * 224 * 1$.

Activation Function: The activation layer is another designation for ReLU. Negative values in the input data are assigned to zero in the outcome. The network operates more quickly when its negative dimension value is zero. This investigation made use of the ReLU activation function. In Equation (10), the ReLU activation function is provided.

$$E(a) = \begin{Bmatrix} 0, a < 0 \\ a, a \geq 0 \end{Bmatrix} \tag{10}$$

By enabling its lower computational demands than other functions, the ReLU layer is more supported.

Layer of Convolution: The foundation of CNN networks is the convolution layer and also known as the transformation layer. Convolution is the procedure of applying filters to all layers. This layer's specified filters have $N \times N$ sizes. Equation (11) provides the convolution that consist of linear filters.

$$(g_l)_{ji} = (Z_l * a)_{ji} + y_{ji} \tag{11}$$

Where a represents the input data, (j, i) represents the pixel point index, l represents index of the feature map, Z and y represents weighing parameters, and $(g_l)_{ji}$ represents the feature map's output value.

Normalization: The network's efficiency is increased by the normalization procedure. The data on additional layers may have different dimensions. According to Equations (12) and (13), the normalization process is as follows:

$$a^l = \frac{a^l - F(a^l)}{\sqrt{Var(a^l) + \varepsilon}} \tag{12}$$

$$b^{(l)} = \gamma^l a^l + \beta^l \tag{13}$$

Where $b^{(l)}$ denotes the input's dimension and $F(a^l)$ denotes the dimension's average. The definition of the standard deviation is $\sqrt{Var(a^l)+\varepsilon}$. There are two learnable variables, γ and β.

Dropout Layer: A lot of data is used in deep learning to train networks. Therefore, the network has been trained when the memorization event is possible. It is necessary to remove certain nodes that stop the network from memorizing. Implement dropout to enhance network performance.

Fully-connected Layer: This layer is dependent on every field of the preceding layer. The Fully Connected Layer transforms the information from the previous layer into a one-dimensional matrix structure. There are possible variations in the variety of entirely interconnected layers that the architecture utilizes.

Pooling Layer: The input data size reduction and the computational complexity reduction are the primary objectives presented in this layer. The $N \times N$ size filters are selected in the pooling layer. The size of the completed image is determined by pooling, as demonstrated in Equations (14-16).

$$T = z2 * g2 * c2 \tag{14}$$

$$z2 = \frac{(z1-e)}{X+1} \tag{15}$$

$$g2 = \frac{g1-e}{X+1} \tag{16}$$

Where $z1$ represents the input width, $g2$ is the height, $c1$ indicates the image depth, e denotes the dimension of the filter, X determines the step counts, and the size of the data is indicated as T. In the suggested architecture, the pooling layer is maximum pooling.

Softmax Layer: In the classification process, it generates the probabilistic value using the previous layer's output. According to Equation (17), it calculates the values for every class. These possibilities estimate the classes using values ranging from 0 to 1.

$$Q(b = i|a; Z, y) = \frac{Exp^{A^{SZ_i}}}{\sum_{i=1}^{N} Exp^{A^{SZ_i}}} \tag{17}$$

Where the main class Z and y is a vector of weights. These processes make use of cross-entropy. Equation (18) provides the cross-entropy function that is most frequently used.

$$CrossEntropy = -\sum_a Q'(a) LogQ(A) \tag{18}$$

Whereas Q represents the actual production, Q' represents the expected output. Finally, images are categorized in the classification layer.

3.4 Assessing the children's play behavior through the autoencoder-based long short-term memory (AE-LSTM)

The integration of AE and LSTM model is used for analyzing and understanding children's play behavior. The AE-LSTM models help to adjust the children's specific variations by identifying distinctive patterns and behavioral abnormalities, which is crucial for customized monitoring and evaluation of children's. The AE-LSTM system on edge devices further enhances its practical utility of computational workload, performed close to the data source for reducing latency, and enabling real-time feedback. It makes the ability of the research to identify patterns in time-series data, such as the play behavior of children, effective and in real-time. In the AE-LSTM model, both encoder and decoder weights were jointly optimized during sequence learning. Two losses were trained together: (1) the autoencoder's reconstruction loss for learning compact temporal representations, and (2) the Softmax cross-entropy loss for play-behavior classification. Edge computing ensures to be local, for eliminating dependence on cloud connectivity and reducing response time for real-time feedback. Lightweight ResNet effectively extracts multimodal features while maintaining computational efficiency. The AE-LSTM model enhances sequential behavior analysis by combining dimensionality reduction with robust temporal modeling.

3.4.1 Long short-term memory (LSTM)

The Recurrent Neural Networks (RNNs) of the LSTM type are made to recognize and remember long-term dependencies in sequential input. It manages information flow through storage cells and gating mechanisms that make it useful for behavior evaluation tasks that involve the modeling of time-series data like voice, movement, or physiological signals from the children during playtime. Time series data is interpreted using a particular kind of computer-based RNN architecture called LSTM. RNNs have difficulty with gradient difficulties and long-term dependencies, which affects the capacity to accurately analyze complicated and sequential data on children's play activity. By utilizing gated memory cells, LSTM addressed the RNN's gradient difficulties and made it possible to accurately represent long-term dependencies. The infrastructure is more capable of detecting irregularities in assessing the children's behavior during play when the LSTM architecture is combined with an AE that assigns significance to important sequences.

As compared to conventional transfer networks, LSTM's feedback interactions between hidden components linked to certain time steps allow for the development of long-term sequence dependency and the

forecasting of interaction labels determined by the sequence of previous activities. To address the issue of diminishing and exploding gradients that occur during the training of conventional RNNs, LSTMs were developed. Updates are made to the data stored in the memory cell of the LSTM unit through the input, forget, and output gates. During random intervals, the factor maintains values, and the three gates control the flow of information from and to the factor. The LSTM's single unit is represented in Figure 3.

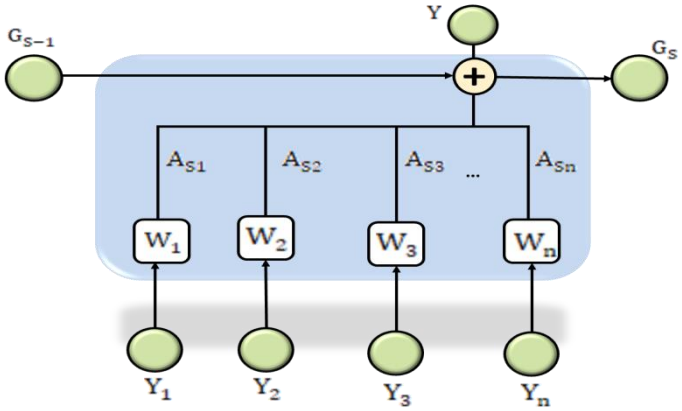

Figure 3: Design of LSTM Unit

Following Equations (19-24) is an approach to computing each cell in an LSTM.

$$a = \begin{bmatrix} G_{S-1} \\ A_S \end{bmatrix} \tag{19}$$

$$G_S = \sigma(W_G \cdot a + Y_G) \tag{20}$$

$$J_S = \sigma(W_J \cdot a + Y_i) \tag{21}$$

$$R_S = \sigma(W_R \cdot a + Y_R) \tag{22}$$

$$D_S = G_S \odot D_{S-1} + J_S \odot \tanH(W_D \cdot a + Y_D) \tag{23}$$

$$G_S = R_S \odot \tanH(D_S) \tag{24}$$

Where $W_J, W_G, W_R \in \mathbb{R}^{D \times 2D}$ are positioned in training, the weighted measures and Y_J, Y_E, $Y_R \in \mathbb{R}^D$ biases of the LSTM are learned, comprising three gates' transformations. Variable σ is the sigmoid function and element-wise multiplication is represented by $\odot$. The LSTM cell unit's inputs are contained in the vectors A_S. The vector of the hidden layer is Z_S. After linearizing the sentence into a vector with a size equal to the number of class labels, insert the final hidden vector n to indicate the phrase as a Softmax layer. Class labels that are neutral, negative, and positive are utilized.

3.4.2 Autoencoder (AE)

To improve proactive children's play behavior in real-time environments, the AE is used to obtain compact and resilient representations of time-series play behaviors. An AE is ideal for this task. It encodes the input sequences into a compact latent space while removing inconsequential information, and then recovers the original input data while reducing reconstruction loss via the decoding structure. The AE's efficient representation allows it to catch hidden patterns, which can be useful in spotting performance abnormalities. The technique is divided into three stages: encoding the input into a compressed latent space, decoding it to rebuild the input, and decreasing reconstruction loss.

Encoding: To encode high-dimensional input for assessing the children's play behavior in real-world environments, the encoder maps the input vector $x \in R^m$ into a compressed latent representation h. It is achieved using Equation (25).

$$h = f_1(w_i x + b_i) \tag{25}$$

The encoder weights and biases are denoted by wi and b_i, respectively, while the activation function is represented by f_1. This stage removes inconsequential differences while preserving fundamental structural elements.

Decoding: The decoder remaps the compressed representation into a reconstructed input $\hat{x}$ to recognize children's play behavior. This transformation is represented as follows in Equation (26).

$$\hat{x} = f_2(w_j h + b_j) \tag{26}$$

Where w_j and b_j are the weights of decoders and bias, and the activation function (f_2) is applied to reconstruct the input structure.

Reconstruction Loss: In quantifying reconstruction deviations for assessing children's play behavior in real-world conditions, the model computes loss (L) between inputs x and its reconstruction $\hat{x}$ as indicated by Equation (27).

$$L(x - \hat{x}) = \frac{1}{n}\sum_{n=1}^{n} |\hat{x}_t - x_t| \tag{27}$$

Where, x is the actual input data disregarding the observed system behavior $\hat{x}$ is the reconstructed output from the autoencoder, and n is the quantity of training illustrations. The loss function supports quantifying differences between actual and predicted behavior, enabling the detection of children's play behaviors relatively earlier than otherwise observable. This is refined by a position-aware formulation, Equation (28).

$$x_i = \frac{1}{n}\sum_{n=1}^{n} |\hat{x}_i - x_i|, \ \ where\ n = \begin{cases} N & if\ i \le \frac{N+1}{2} \\ n - i + 1 & if\ i > \frac{N+1}{2} \end{cases} \tag{28}$$

Variable x_i represents reconstruction error for the i model, and $\hat{x}_i$ is the predicted rate, n is the total sequence length, and N is adjusted per contextual importance over time. This weighted-based mechanism is validated by providing attention to children's play behavior in real-world circumstances. Overall reconstruction loss across time series is presented by Equation (29).

$$loss = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{29}$$

Where, x_i is the reconstruction loss and N is the full sequence length. Algorithm 1 shows the AE-LSTM algorithm.

Algorithm 1: AE-LSTM

```
Input:
  D = {Training, Validation, Test} datasets
  Model Parameters:
    Encoder: wi, bi
    Decoder: wj, bj
    LSTM gates:
      W_G, W_J, W_R, W_D
      Y_G, Y_J, Y_R, Y_D
  Hyperparameters:
    lr, epochs, batch_size, clip_norm, patience
Initialize:
  optimizer ← Adam({all parameters}, lr)
  best_val_loss ← ∞
  no_improve ← 0
Function Encode(x):
  For t = 1 ... N:
    h_t ← f1(wi * x[t] + bi)
  return {h_1 ... h_N}
Function LSTM_Forward(H):
  Z_0 ← 0 ; D_0 ← 0
  For S = 1 ... N:
    a ← concat(Z_(S − 1), H[S])
    G_S ← sigmoid(W_G * a + Y_G)
    J_S ← sigmoid(W_J * a + Y_J)
    R_S ← sigmoid(W_R * a + Y_R)
    D_S ← (G_S ⊙ D_(S − 1)) + (J_S ⊙ tanh(W_D * a + Y_D))
    Z_S ← R_S ⊙ tanh(D_S)
  return Z_S, {Z_1 ... Z_N}, {D_1 ... D_N}
Function Decode(H):
  For t = 1 ... N:
    x_hat_t ← f2(wj * H[t] + bj)
  return {x_hat_1 ... x_hat_N}
Function Position_Aware_Loss(x, x_hat):
  N ← length(x)
  loss_sum ← 0
  For i = 1 ... N:
    if i ≤ (N + 1)/2:
      n_i ← N
    else:
```

```
      n_i ← N − i + 1
    e_i ← mean(|x_hat[i] − x[i]|)
    x_i ← (1 / n_i) ∗ e_i
    loss_sum ← loss_sum + x_i
  return loss_sum / N
Training Loop:
For epoch = 1 ... epochs:
  Shuffle Training data
  For each batch B:
    X_batch ← inputs in B
    For each sequence x in X_batch:
      H ← Encode(x)
      Z_final, Z_seq, D_seq ← LSTM_Forward(H)
      x_hat ← Decode(H)
      L_seq ← Position_Aware_Loss(x, x_hat)
      Accumulate L_seq
    batch_loss ← mean(L_seq)
    optimizer.zero_grad()
    Backprop(batch_loss)
    Clip_Gradients(all_parameters, clip_norm)
    optimizer.step()
  val_loss ← Evaluate(D.Validation)
  If val_loss < best_val_loss:
    best_val_loss ← val_loss
    Save_Checkpoint(model_parameters)
    no_improve ← 0
  Else:
    no_improve ← no_improve + 1
    If no_improve = patience:
      Break
Evaluation (Test Phase):
Load_Checkpoint(best_model)
For each sequence x in Test set:
  Compute Encode → LSTM_Forward → Decode
  Compute Position_Aware_Loss
Return final test_loss and metrics
Final Evaluation:
  Load best checkpoint
  test_loss, test_metrics = Evaluate(D_test, model_parameters)
  RETURN best_model, test_loss, test_metrics
Procedure Evaluate(D_split, params):
  Set model to eval mode (disable dropout, etc.)
  losses = []
  pred_labels = [], true_labels = []
  FOR each sequence x (and optionally y) in D_split:
    compute x_hat_seq and final hidden Z_final (no gradient)
    compute per − sequence position − aware loss x_i as in training
    seq_loss = mean_over_i(x_i)
    append seq_loss to losses
    IF labels present:
      logits = Linear(Z_final)
      y_pred_label = argmax(softmax(logits))
      append y_pred_label to pred_labels
      append y to true_labels
  avg_loss = mean(losses)
  IF labels present:
    metrics = compute_metrics(true_labels, pred_labels)
  ELSE:
    metrics = {}
  RETURN avg_loss, metrics
```

Learning compact, sequential representations of temporal data while capturing long-range dependencies is accomplished with AE-LSTM. This research allows for real-time monitoring of children's play behavior by encoding multimodal sensor inputs and decoding patterns to accurately identify social interactions, activity categories, and emotional states in dynamic and noisy conditions. Several hyperparameters utilized in the research are explored in Table 3.

Table 3: Hyperparameters and Values for AE-LSTM model configuration

Training Epochs	Learning Rate	Batch Size	Latent Dimension (AE)	LSTM Hidden Units	Dropout Rate	Sequence Length
10 Epochs	0.001	32	32	64	0.1	30
20 Epochs	0.001	32	48	96	0.15	40
30 Epochs	0.001	64	64	128	0.20	50
40 Epochs	0.0008	64	64	128	0.25	60
50 Epochs	0.0005	64	64	128	0.30	60

4 Experimental results

The research intended to provide a real-time assessment of children's play behavior. The following phases provide a detailed explanation of the research results. The proposed platform is designed to run on lightweight edge devices, which minimizes hardware cost and eliminates the need for high-performance servers. Sensors such as RGB-D cameras, wearable units, and microphones are selected based on low-power, commercially available modules to ensure affordability for classrooms and homes.

4.1 Evaluation criteria

This evaluation criteria section provides the evaluation outcomes of the proposed AE-LSTM method in various feature parameters like hesart rate distribution through play behavior, skin temperature during activity levels, comparison of heart rate and activity levels, evaluation of activity level and proximity to peers, pitch mean and standard deviation distributions, and physiological signals distribution. The discussion below discusses the evaluation features of the proposed AE-LSTM approach.

Figure 4 represents the engagement and feedback relevance of children play behavior. It indicates the median and variable heart rates, providing information about how the body responds to various forms of play. The AE-LSTM model uses temporal patterns in multimodal data including heart rate, posture, speech tone, and motion to classify and interpret play behavior dynamically. It helps to physiological variation across spontaneous play states. By using the information to track mental and physical states, the AI platform provides real-time feedback while playing and correlates internal signals with behavior.

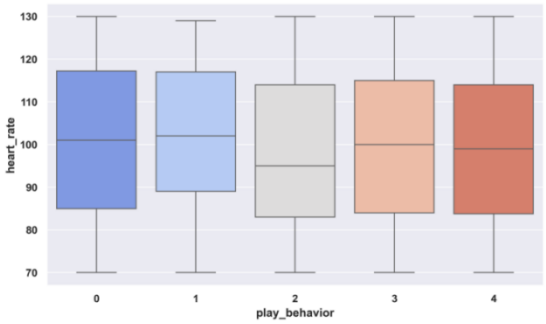

Figure 4: Engagement and feedback relevance of children play behavior

Depending on the type of play, heart rates might vary from 85 to 145 beats per minute. Running and other high-intensity activities exhibit median heart rates of about 130 bpm, whereas peaceful play activities have median heart rates of about 90 bpm. The inference of tension or excitement in real time is influenced by the values. The AI model's physiological-behavioral mapping is improved by significant interquartile ranges, which show behavioral variability.

Skin Temperature during Activity Levels: Figure 5 shows how children's skin temperatures change with different degrees of activity. Increased physical effort or emotional stimulation is frequently indicated by elevated temperatures. The AE-LSTM model model leverages temporal patterns in multimodal data including skin temperature to classify and interpret play behavior in real time. To determine emotional states, wearable sensors capture these physiological signals and incorporate them into the platform. This allows for real-time behavior classification and adaptive feedback.

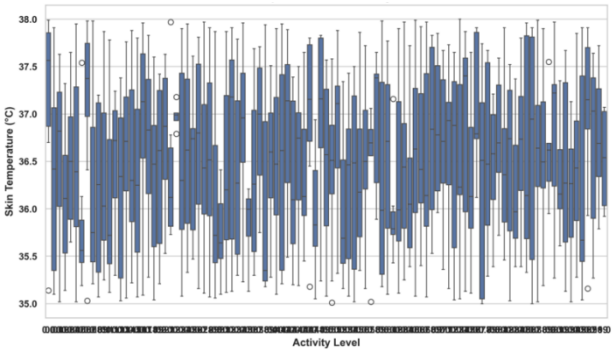

Figure 5: Evaluation of skin temperature through activity levels

Skin temperatures range from 32.0 to 36.5°C. The median temperature is near 36.0°C for high activity levels and closer to 33.5°C for passive behavior. Emotional changes or exertion are reflected in fluctuations. The adaptive feedback loop for children's comfort and stress detection is supported by these physiological readings, along with motion information for contextual behavior recognition.

Figure 6 depicts the bubble plot of physiological-behavioral mapping, which is categorized by color according to various play behaviors. The AE-LSTM model used to enhance the subtle temporal patterns that differentiate between levels of spontaneous play. It shows how multimodal physiological and behavioral information is used to enhance the cognitive performance.

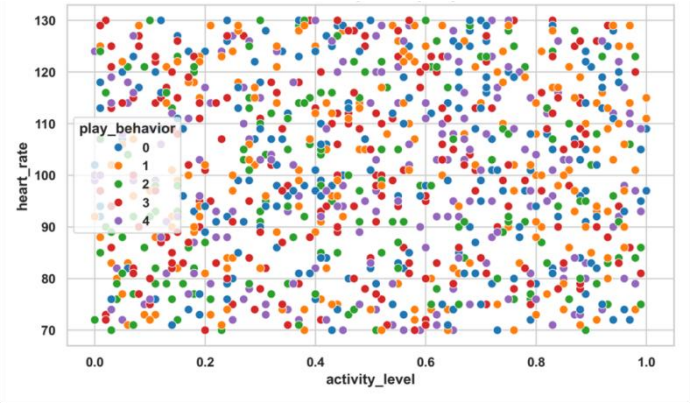

Figure 6: Bubble plot of physiological-behavioral mapping

There is a positive correlation between heart rate (85–145 bpm) and activity level (range: 0–10). Claiming and other play behaviors show top-right clustering (activity > 8, heart rate > 130 bpm), whereas quiet activities are located close to the origin. Physiological-kinematic synchronization is improved by the connection in the classification of multimodal behavior.

Activity Level and Proximity Peers: The connection between children's levels of activity and the way they are connected to their classmates is demonstrated in Figure 7. The AE-LSTM model learns temporal and spatial patterns to infer behavioral states in real time. By showing how the social and physical aspects of play co-vary, expose behavioral clusters that could represent cooperative, solitary, or transitional play styles. It allows the AI to identify social involvement levels in real time.

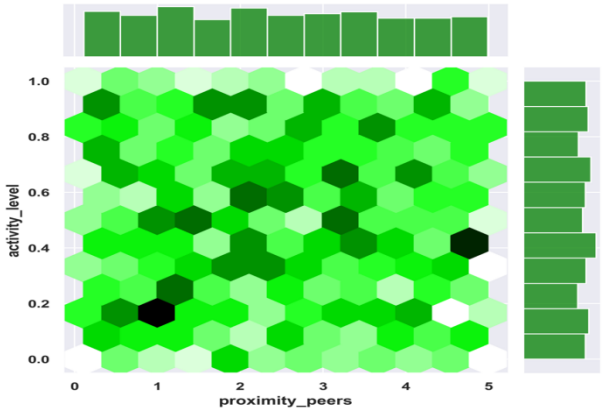

Figure 7: Results of activity levels and proximity peers

The range for proximity is 0.5–3.0 meters, while the range for activity level is 0–10. Activity > 7 and high densities at <1.5 m proximity indicate active, social play. Solitary or passive intervals are highlighted by sparse areas over long distances and low activity levels. Based on the outcomes, the platform categorizes different kinds of interactions, such as self-sustaining and collaborative.

Physiological Signals Distribution across play behavior types was illustrated in Figure 8.The AE-LSTM network leverages inputs to model temporal dependencies and reconstruct latent behavioral patterns. By capturing both individual and joint distributions of physiological metrics, the system enhances its ability to infer emotional and physical engagement levels. Significant differences in physiological indicators associated with various play behaviors are revealed by the diagonal plots, which show the kernel density calculation of all characteristics by behavior type.

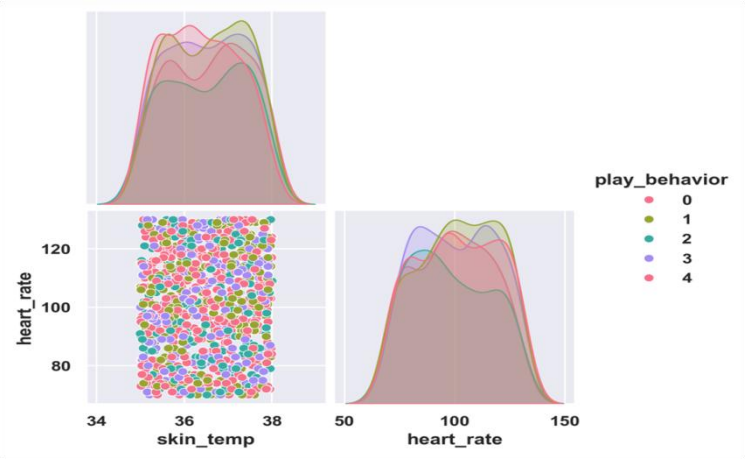

Figure 8: Physiological signals distribution across play behavior types

Five play behavior classes' relationships between skin temperature (34–38°C) and heart rate (60–150 bpm) are presented. There is a unique physiological characteristic for every behavior determined in the results. Behavior 0 is linked to higher heart rates, although behavior 2 leads to lower temperatures.

Pitch Mean and Standard Deviation Distributions: Pitch mean and pitch standard deviation distributions from children's speech data are displayed in Figure 9. By capturing both central tendency and variability in pitch, the system can detect shifts in emotional tone and engagement. This visualization shows how speech-based metrics enrich the system's ability to interpret and support child development through intelligent, responsive play environments. These audio features support real-time AI analysis by assisting in the inference of behavioral and emotional cues during play.

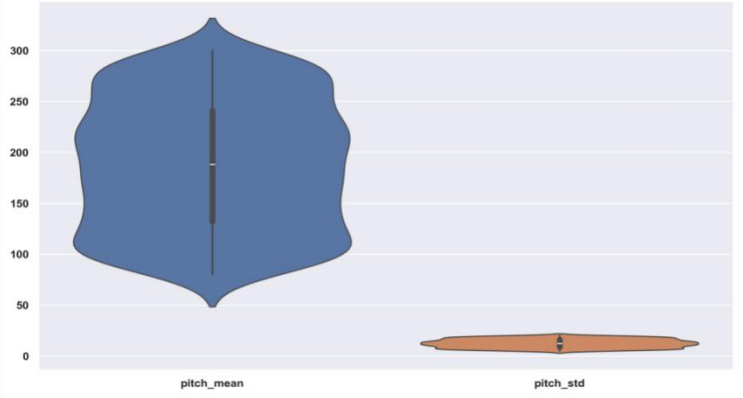

Figure 9: Determination of mean and standard deviation distributions

The voice pitch varies during play, as indicated by the pitch mean, which varies from roughly 75 Hz to 310 Hz. The pitch standard deviation (pitch_std) shows constant patterns of vocal variability, remaining densely emphasized between 5 Hz and 15 Hz.

Table 4 presents the performance of different models' confidence intervals. The confidence intervals further confirm the statistical reliability of these results, highlighting AE-LSTM's superior consistency and effectiveness. AE-LSTM outperforms all other models, achieving the highest scores in all metrics, indicating robust and reliable predictions.

Table 4: Evaluation metrics of AE-LSTM and baseline models confidence intervals

Model	Accuracy (95% CI)	Precision (95% CI)	Recall (95% CI)	F1-score (95% CI)
AE-LSTM	0.968 – 0.982	0.960 – 0.976	0.956 – 0.972	0.950 – 0.968
LSTM	0.833 – 0.865	0.819 – 0.851	0.805 – 0.839	0.811 – 0.845
GRU	0.816 – 0.848	0.800 – 0.834	0.788 – 0.822	0.794 – 0.828
BiLSTM+Attention	0.857 – 0.887	0.844 – 0.874	0.835 – 0.867	0.840 – 0.870

Table 5 presents the results of significance testing for the AE-LSTM model across using a significance level. These results confirm that the observed performance of the AE-LSTM model demonstrates robust and reliable predictive capability across all evaluated metrics.

Table 5: Statistical significance analysis of AE-LSTM performance metrics

Metric	AE-LSTM Value ($\hat{p}$)	z-Statistic	p-Value	Significance ($\alpha = 0.05$)
Accuracy	0.975	42.31	$p < 0.0001$	Significant
Precision	0.968	41.28	$p < 0.0001$	Significant
Recall	0.964	40.74	$p < 0.0001$	Significant
F1-Score	0.959	40.07	$p < 0.0001$	Significant

4.2 Comparison phases

The research compares the proposed AE-LSTM method with various existing techniques, such as Gated Recurrent Unit (GRU) [22], Bidirectional LSTM (BiLSTM) [22], and BiLSTM+Attention [22], to assess the playing behavior. Table 6 determines the comparison evaluation of proposed and existing methods with F1-score, precision, recall and accuracy. The performance matrix's formula and definitions are provided in Table 7.

Table 6: Formulas and definitions of performance matrices

Metrics	Definitions	Equations
Accuracy	The proportion of accurate true positive and true negative forecasts overall.	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	It is a proportion of real positive forecasting over all the positive predictions.	$\frac{TP}{TP+FP}$
Recall	The percentage of real positive cases that a prediction model accurately classifies as positive is known as recall.	$\frac{TP}{TP+FN}$
F1 Score	It is the harmonic mean of precision and recall. It measures the balance between both metrics.	$2 \times \frac{Precision \times Recall}{Precision+Recall}$

Table 7: Comparison of outcomes of AE-LSTM and exiting methods

Models	$Accuracy$	$Precision$	$Recall$	$F1-score$
LSTM [22]	0.849	0.835	0.822	0.828
GRU [22]	0.832	0.817	0.805	0.811
BiLSTM+Attention [22]	0.872	0.859	0.851	0.855
AE-LSTM [Proposed]	0.975	0.968	0.964	0.959

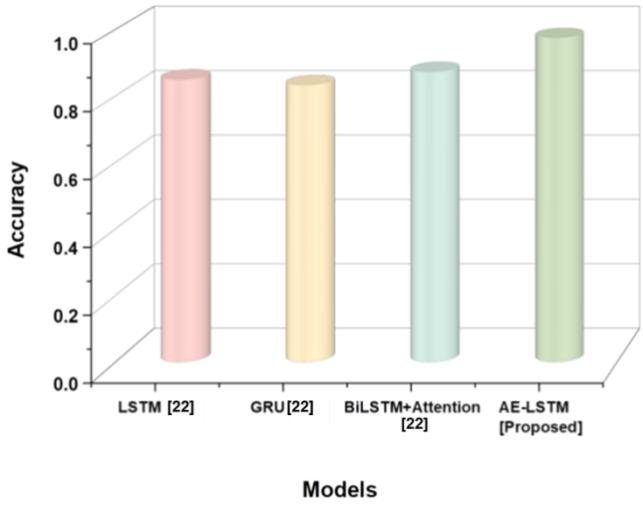

Figure 10: Outcomes of the accuracy metric in play behavior

Figure 10 depicts the outcomes determined by the accuracy. The proposed AE-LSTM approach provides an accuracy of 0.975, whereas GRU shows 0.832 accuracy, BiLSTM+Attention has 0.872, and LSTM provides 0.849 accuracy. Based on the results, the proposed AE-LSTM method has a high accuracy to compact latent representations from multimodal play behavior data and to effectively capture physiological signal distributions.

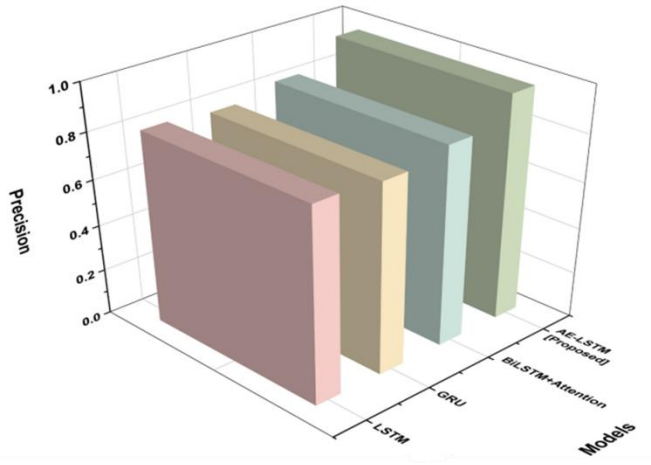

Figure 11: Result of play behavior with precision

Figure 11 shows precision findings. Existing methods provide precision outcomes (LSTM has 0.835, BiLSTM+Attention has 0.859, and GRU has 0.817). The precision of the AE-LSTM method is 0.968, and it indicates that the AE-LSTM technique has more efficiency, and reliability in children's play behavior than other existing models.

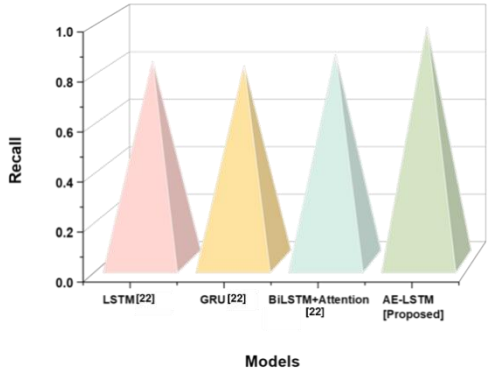

Figure 12: Visual depiction of recall results

Figure 12 demonstrates the results of recall. The proposed AE-LSTM approach indicates essential results in terms of recall (0.964), indicating its strong capability in accurately identifying children's play behavior pattern. Outcomes of the research demonstrate that the proposed AE-LSTM approach is more significant with a recall result the existing techniques like GRU (0.805), LSTM (0.822), and BiLSTM+Attention (0.851).

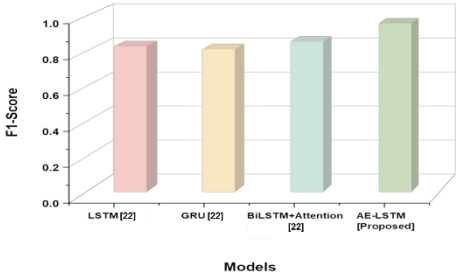

Figure 13: Evaluation outcomes of F1-score

Estimation of the F1-score is displayed in Figure 13. The AE-LSTM method has a 0.959 F1-score, BiLSTM+Attention has 0.855, GRU has 0.811, and LSTM has a 0.828 F1-score. The proposed AE-LSTM method explores superior results compared to the traditional approaches, as determined by the research findings. By effectively modeling temporal patterns and physiological signal distributions, the AE-LSTM captures subtle

behavioral variations, enabling more accurate classification.

Assessing the behavior of playing children was the focus of the research. The evaluation of the research compared the proposed method with existing techniques, and certain limitations were observed in the existing methods in assessing the play behaviors. The LSTM [22] model might have trouble in recognizing long-range temporal connections. The computing capacity of edge devices could limit the accuracy of real-time inference and model complexity. The computational depth of GRU [22] was insufficient to capture complex temporal connections. Although it has superiority in contextual training, BiLSTM [22] was less appropriate for real-time applications due to its increased computing complexity and latency. The BiLSTM+Attention [22] raises attention on vital features, whereas it was expensive for edge distribution and issues from over-fitting with small training data. These models have a general difficulty with dynamic and natural environments in children's play. Our system balances latency, memory usage, and real-time processing by deploying lightweight ResNet feature extraction and the AE-LSTM architecture directly on edge hardware. Unlike GRU, LSTM, or BiLSTM-Attention models, the AE-LSTM compresses multimodal inputs into compact latent representations, reducing memory load while maintaining strong temporal modeling. Because computation occurs locally, inference latency is significantly lower than cloud-based models, enabling immediate feedback. Existing observation methods for children's play are largely manual, subjective, and unsuitable for real-time interpretation. To address these shortcomings, the proposed AE-LSTM model enhances the real-time understanding of children's natural play. The AE model helps to compress and denoise multimodal features, whereas LSTM captures temporal patterns and emotional cues. Overall, the conceptualized model AE-LSTM will facilitate a real-time analysis of complex sensory play behavior, which provides robust temporal modeling.

5 Conclusions

1.The purpose of the research was to create and implement a multimodal sensor and feedback platform by utilizing edge computing and real-time AI to monitor, evaluate, and assist children's play behavior. The multimodal play behavior dataset with various sensors was obtained for the performance. Kalman filtering and normalizing techniques were used to pre-process data to increase consistency and minimize noise. Real-time behavior analysis was performed with an AE-LSTM network, while feature extraction was accomplished with a Lightweight ResNet model. To provide low-latency processing, local data storage, and privacy protection, the entire system has been deployed on edge devices. To increase engagement, the device provided real-time input via physical and visual signals. According to the experimental results, play behavior classification, emotional state detection, and peer interaction identification were all achieved with highly accurate results. Comparison of the proposed method demonstrates significant results in terms of accuracy (0.975), precision (0.968), recall (0.964), and F1-score (0.959). A real-time interpretation of children's play behavior, the proposed model supports educators in understanding engagement levels, social interaction patterns, and emotional cues during learning activities. Therapists gain continuous behavioral monitoring that supports early detection of developmental needs. Researchers benefit from reliable, unobtrusive multimodal analytics that capture natural play behavior accurately.

5.1 Limitations and future scopes

Children's behavior variability, sensor position limitations, and the requirement for frequent validation in dynamic playing environments limit the system's efficacy. The scalability of the system across different cultural or contextual settings remains challenge, which could affect its generalizability. The long-term deployment feasibility, include issues like battery life, device comfort, and overall cost, which are essential for practical implementation. Additionally, it collects sensitive data from children, ethical and privacy considerations lead to potential risks. Adaptability in a variety of play environments, emotional recognition skills, and the incorporation of adaptive learning models will represent the main areas of future research. To assess the platform's scalability and developmental impact in larger educational or clinical contexts, further long-term investigations are needed.

Data availability

The datasets generated and/or analysed during the current study are available in the Kaggle repository, [https://www.kaggle.com/datasets/ziya07/multimodal-play-behavior-dataset/data].

Ethics approval

The data analyzed in this study originate from the publicly available "Multimodal Play Behavior Dataset" on Kaggle. Therefore, no new ethics committee approval or informed consent was required for our analysis.

Consent to participate

Not applicable, as the study is entirely based on publicly available secondary data with no identifiable personal

Funding

This work was supported by 2024 construction project for the "1112" teaching project of Shaanxi College of Apparel Engineering, "Game Design for Preschool Children Offline Specialized Course" (No. 2024TSKC020)

References

[1] Hyun, S., McWayne, C.M. and Smith, J.M., 2021. “I see why they play”: Chinese immigrant parents and their beliefs about young children's play. Early Childhood Research Quarterly, 56, pp.272-280. https://doi.org/10.1016/j.ecresq.2021.03.014

[2] Passmore, A.H. and Hughes, M.T., 2021. Exploration of play behaviors in an inclusive preschool setting. Early Childhood Education Journal, 49(6), pp.1155-1164. https://doi.org/10.1007/s10643-020-01122-9

[3] Li, S., Sun, J. and Dong, J., 2022. Family socio-economic status and children’s play behaviors: The mediating role of home environment. Children, 9(9), p.1385. https://doi.org/10.3390/children9091385

[4] Wu, S., Pan, C., Yao, L. and Wu, X., 2022. The impact of the urban built environment on the play behavior of children with ASD. International journal of environmental research and public health, 19(22), p.14752. https://doi.org/10.3390/ijerph192214752

[5] Dodd, H.F., Nesbit, R.J. and Maratchi, L.R., 2021. Development and evaluation of a new measure of children’s play: The Children’s Play Scale (CPS). BMC Public Health, 21(1), p.878. https://doi.org/10.1186/s12889-021-10812-x

[6] van Dijk-Wesselius, J.E., Maas, J., van Vugt, M. and van den Berg, A.E., 2022. A comparison of children's play and non-play behavior before and after schoolyard greening monitored by video observations. Journal of Environmental Psychology, 80, p.101760. https://doi.org/10.1016/j.jenvp.2022.101760

[7] Lennon, M., Pila, S., Flynn, R. and Wartella, E.A., 2022. Young children's social and independent behavior during play with a coding app: Digital game features matter in a 1: 1 child to tablet setting. Computers & Education, 190, p.104608. https://doi.org/10.1016/j.compedu.2022.104608

[8] Dankiw, K.A., Kumar, S., Baldock, K.L. and Tsiros, M.D., 2024. Do children play differently in nature play compared to manufactured play spaces? A quantitative descriptive study. International Journal of Early Childhood, 56(3), pp.535-554. https://doi.org/10.1007/s13158-023-00384-9

[9] Yang, J.T., Chen, C.I. and Zheng, M.C., 2023. Elevating children’s play experience: a design intervention to enhance children’s social interaction in park playgrounds. Sustainability, 15(8), p.6971. https://doi.org/10.3390/su15086971

[10] Cakan, A. and Acer, D., 2024. Analysis of preschool children’s outdoor play behaviors. Journal of Outdoor and Environmental Education, pp.1-27. https://doi.org/10.1007/s42322-024-00174-4

[11] López-Toribio, M., Hidalgo, L., Litt, J.S., Daher, C., Nieuwenhuijsen, M., Márquez, S., Berrón, A., Franch, B., García, B. and Ubalde-López, M., 2025. SOOPEN: design and assessment of a tailored systematic observation tool to evaluate outdoor play behavior among schoolchildren groups. Cities & Health, pp.1-12. https://doi.org/10.1080/23748834.2024.2439645

[12] Wilhelmsen, T., Lekhal, R., Rydland, V. and Coplan, R.J., 2025. Exploring the role of early childhood educators’ emotion socialization strategies in the development of young children’s social and non-social play behaviors. Early Childhood Research Quarterly, 73, pp.92-100. https://doi.org/10.1016/j.ecresq.2025.06.005

[13] Yesil, R., Erdiller Yatmaz, Z. and Metindogan, A., 2025. Exploring Children’s Play Culture and Game Construction: Role of Sociodramatic Play in Supporting Agency. Early Childhood Education Journal, 53(3), pp.703-716. https://doi.org/10.1007/s10643-023-01621-5

[14] Samuelsson, R., 2023. A shape of play to come: Exploring children's play and imaginaries with robots and AI. Computers and Education: Artificial Intelligence, 5, p.100173. https://doi.org/10.1016/j.caeai.2023.100173

[15] Ocak, C., Kopcha, T.J. and Dey, R., 2023. An AI-enhanced pattern recognition approach to temporal and spatial analysis of children's embodied interactions. Computers and Education: Artificial Intelligence, 5, p.100146. https://doi.org/10.1016/j.caeai.2023.100146

[16] Nopembri, S., Mulyawan, R., Fauziah, P.Y., Kusumawardani, E., Susilowati, I.H., Fauzi, L., Cahyati, W.H., Rahayu, T., Chua, T.B.K. and Chia, M.Y.H., 2023. Time to play in Javanese preschool children—An examination of screen time and playtime before and during the COVID-19 pandemic. International Journal of Environmental Research and Public Health, 20(3), p.1659. https://doi.org/10.3390/ijerph20031659

[17] de Castro Rodrigues, D., de Siqueira, V.S., da Costa, R.M. and Barbosa, R.M., 2022. Artificial Intelligence applied to smart interfaces for children’s educational games. Displays, 74, p.102217. https://doi.org/10.1016/j.displa.2022.102217

[18] Shawcroft, J.E., Gale, M., Workman, K., Leiter, V., Jorgensen-Wells, M. and Jensen, A.C., 2022. Screen-play: An observational study of the effect of screen media on Children's play in a museum setting. Computers in Human Behavior, 132, p.107254. https://doi.org/10.1016/j.chb.2022.107254

[19] Faraz, A., Mounsef, J., Raza, A. and Willis, S., 2022. Child safety and protection in the online gaming ecosystem. Ieee Access, 10, pp.115895-115913. https://doi.org/10.1109/ACCESS.2022.3218415

[20] Gashaj, V., Dapp, L.C., Trninic, D. and Roebers, C.M., 2021. The effect of video games, exergames and board games on executive functions in kindergarten and 2nd grade: An explorative longitudinal study. Trends in Neuroscience and Education, 25, p.100162. https://doi.org/10.1016/j.tine.2021.100162

[21] Foweather, L., Crotti, M., Foulkes, J.D., O'Dwyer, M.V., Utesch, T., Knowles, Z.R., Fairclough, S.J., Ridgers, N.D. and Stratton, G., 2021. Foundational movement skills and play behaviors during recess among preschool children: A compositional analysis. Children, 8(7), p.543. https://doi.org/10.3390/children8070543

[22] Liang, R., Ye, Z., Liang, Y. and Li, S., 2025. Deep Learning-Based Play Behavior Modeling and Game Interaction System Optimization Research. https://doi.org/10.20944/preprints202505.2198.v1

Appendix

Symbol	Explanation / Meaning
A_l, A_{l+1}	Actual system state at time step l of $l+1$.
X	State transition matrix governing state dynamics.
Y	Control matrix.
μ_l	Control input at time step l .
ω_l	Process noise or disturbance at time step l.
f_{l+1}	Error between actual and estimated state at time $l+1$.
$(Q_{l+1\|l+1})$	Uncertainty estimation
$F(\cdot)$	Expectation operator or covariance function.
P_{l+1}	Covariance matrix of process noise.
K	Identity or transition matrix (context: covariance update).
G	Measurement matrix mapping states to observation space.
L_{l+1}	Kalman gain at time step $l+1$.
W_{l+1}	Actual measurement value observed at time step $l+1$.
A	Raw data value before preprocessing (Z-score).
μ	Mean value of data.
σ	Standard deviation of data.
$E(a)$	ReLU activation function output.
a	Input to activation function or normalization.
Z_l	Convolution filter kernel weights.
y_{ji}	Bias term of the convolution filter.
$(g_l)_{ji}$	Output feature map value at pixel
j, i	Pixel indices in image/feature map.
l	Feature map index or layer index.
a^l	Normalized activation values in layer l.
$F(a^l)$	Mean of activations in layer lll.
$Var(a^l)$	Variance of activations in layer l.
ε	Small constant for numerical stability during normalization.
γ^l, β^l	Learnable scale and shift parameters in normalization.
$z1, g1, c1$	Input width, input height, and input depth respectively.
$z2, g2, c2$	Output width, output height, and output depth respectively.
e	Size of the pooling filter (e.g., 2×2).
X	Stride value for pooling or convolution.
TTT	Output size after pooling.
Z_i	Weights corresponding to class iii.
y	Weight vector associated with softmax.
$exp(\cdot)$	Exponential function used in softmax.
N	Number of output classes.
$Q'(a)$	Ground-truth distribution (one-hot encoded label).
$CrossEntropy$	Loss function for classification.