ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Development of a Preschool AI Interactive Tool Based on Contextual Pedagogy: Classroom Scenario Adaptation and Empirical Evidence for Young Children's Development

Ge Cui[1,*]

[1] Teacher Education Center, Anshan Normal University, Anshan, Liaoning, 114007, China

SUMMARY: *The preschool situational teaching method often has difficulty in making scenarios match children's physical and cognitive development. The present research carries out optimization on situational design through concentrating on children's gesture movements and voice emotions. For the purpose of improving complex-scene identification, a hand-movement identification model which is based on YOLOv7 is developed, and it is integrated with the CBAM attention module. One two-stage speech enhancement model which combines self-attention and spatial attention has been constructed for the reduction of noise in the speech of children. Mel frequency cepstrum coefficients and one residual neural network are then employed by us to conduct speech emotion recognition. After we have put this model into use in situational teaching, the children of Class E1 have showed markedly better self-control ($P<0.05$), thus this has verified the effectiveness of this method.*

KEYWORDS: *Scenario-based teaching; Gesture recognition; Speech emotion recognition; Speech enhancement; Preschool education*

1 Introduction

Early childhood education is a crucial component of basic education and serves as the foundation of the modern national education system [1, 2]. The extent and quality of early childhood education play a highly significant role in human development. From the reform and opening-up starting point, China has obtained break-through achievements on reforming the preschool education system, promoting teachers' educational background level, and developing education patterns. But, many difficulties still exist, among which regional development unbalance, big city-country gap, and resource shortage in country regions are especially serious [3-6]. Reference [7] points out that in recent years, China's preschool education has obtained very big promotions, which include the progress in physical environments and the ideas of teachers. But, this result shows that these progresses still are not even in different regions and between all kindergartens.Reference [8] has conducted an examination on the observed correlation which exists between preschool education quality and the overall development of children, and has carried out an analysis of how residential areas exert influence on this relationship. Through the comparison of children from Shanghai and Guizhou Province, it has been demonstrated that the quality of Shanghai's preschool education is superior to that of Guizhou, hence rural education is left behind by urban areas. The document [9] makes clear that China quickly expands the preschool cause, however, it puts more emphasis on getting more people access rather than on the level of quality.

*cuige@asnc.edu.cn
https://doi.org/10.65102/is2026316 Published: 30 April, 2026

According to child samples got from eight provinces, this paper therefore argues that we should establish strict national baseline quality standards for elevating rural preschool education and hence narrowing urban-rural disparities. The study [10] points out that although the preschool education teaching modes of China have obtained relatively great attention, they are still not sufficient. Through the inspection of preschool education, this text analyzes the obstacles in development and therefore puts forward the paths for progress.Document [11] has depicted the difficulties that preschool education workers encounter, which include low salary payment, not enough teaching abilities, and insufficient resources. Thus it put forward possible solutions such as teacher cultivation training and higher recognition of this teaching occupation. The blasting expansion of artificial intelligence (AI) technique brings chances to change this situation.

From one aspect, AI applications inside preschool education can promote infrastructure construction, give teachers power to promote teaching quality, and let parents conduct interaction with children in a more scientific way—thus promoting the comprehensive abilities of learners [12-14]. On the other side, the integrating of AI-supported interactive tools into scenario-based teaching further promotes education quality [15, 16]. Teaching which bases on scenarios simulates real world situations to build concrete learning environments in which little children gain knowledge and get experience through being immersed [17, 18]. The artificial intelligence interaction tools may produce the immersive study scenes, make the virtual study partners, and make the teaching interactions have more types [19, 20]. In the immersion experience that is formed by their combination, little children change from passive receivers to active joiners and explorers. They carry out interactions with elements in virtual situation, share what they see and their inner feelings with schoolmates, and hence effectively cultivate thinking and innovation capacities while increasing hands-on abilities and understanding of learning contents[21-25].

In order to adapt to nowadays developments, early childhood education workers must again define their own roles, embracing new teaching duties and abilities. This requires changing from “leaders” to “helpers”, and carrying out transformation from “only teachers” to “co-learners” [26-29]. Reference [30] does exploration on AI usages in early childhood education, it puts focus on computer-based learning systems that combine intelligent methods and educational robots for little children.Reference points out China's conservative attitude on AI application in kindergarten environments, and finds out problems such as badly planned courses and teachers' distrust for AI techniques. A research work [32] employs a literature review to demonstrate that artificial intelligence (AI) can efficiently boost children's cognitive capabilities in fields like AI and robotics. Simultaneously, it can nurture skills such as creativity, emotional regulation, and reading and writing proficiency. Another piece of literature [33] offers a comprehensive in - depth account of the primary AI techniques utilized in preschool education. It presents perspectives from a historical context, summarizes notable accomplishments, and outlines unsolved issues, future development paths, and challenges.Reference [34] has put forward Kid Space, which is a technique that uses many kinds of sensor technologies to build immersive physical spaces. A research which was carried out for the purpose of verifying its effect showed that this technique decreased the usage of smart devices by children, at the same time it promoted their social communication and study results. The reference [35] has completed the development of one interactive learning application which is based on AI. Experiments that used children as participants have confirmed this program has obvious promotion effects on literacy and reading ability, hence hence proving AI can have effective utilization in the environment of preschool education. Literature [36] has carried out exploration on AI tools including augmented reality applications and holographic technology within preschool education, hence it indicates that

these tools can bring transformation to the essence of interactions between learners and learning materials, and thus provide obvious superiorities in early childhood education.

This present paper at first introduces the YOLOv7 network structure, puts forward an improved scheme for its deficiencies in target recognition, and builds an enhanced gesture recognition model that is based on YOLOv7. It then makes explanation on the structure and principle of single-channel time-domain speech enhancement, integrates self-attention and spatial attention to design a two-stage model of speech enhancement which is based on attention, and carries out analysis on its workflow. Based on this, it utilizes MFCC for the extraction of enhanced speech features, constructs a speech emotion recognition method which is based on residual network, and therefore forms a complete model of speech emotion recognition. At last, the performance of the two models is each undergone testing, and application experiments are conducted to prove that the situational teaching model within this framework can effectively push forward the growth of young children's self-control ability.

2 Gesture Recognition Model Based on an Improved YOLOv7 Network

When users interact with preschool education robots, the robot system learns specific interactive actions based on user needs, including grasping operations and dance movements. Therefore, action recognition serves as the primary algorithmic focus in constructing teaching scenarios. This paper refines action recognition specifically to gesture recognition.

The YOLO series of algorithms represents a class of object detection methods that are currently mainstream in the field. Among these algorithms, YOLOv7 is the most top deep learning-based object detection algorithm that people can get at present. Therefore, this method is utilized as the gesture identification approach inside the gesture-based identification model tool kit. Further optimization work for YOLOv7 has been carried out in order to promote the recognition performance.

YOLOv7 Network Architecture

The architecture of the YOLOv7 network is shown in Figure 1.

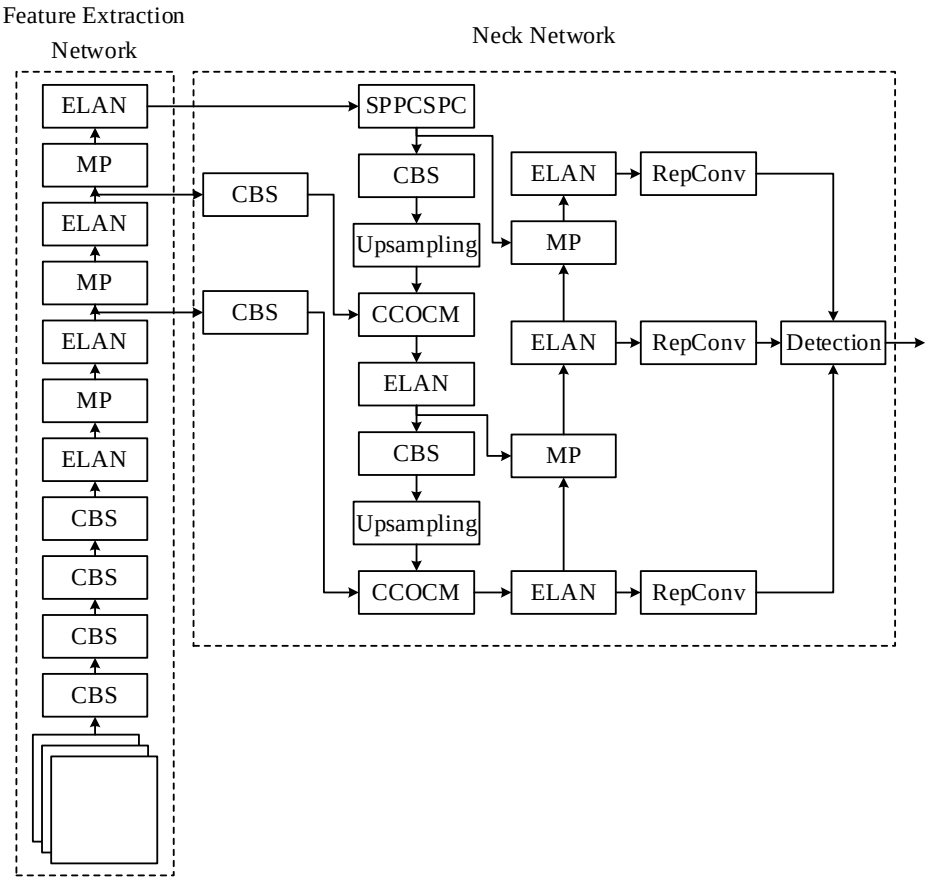


Figure 1: Schematic diagram of the YOLOv7 network structure

Inside the YOLOv7 network, images are firstly resized into a fixed size dimension when they are input. They afterwards pass through enhancement by Mosaic processing, prior to that they are input into the feature extraction network in batches of the appointed size, to carry out the feature extraction. The work of feature extraction has gotten three corresponding feature matrices. In the process of feature fusion, the smaller feature matrices are undergone 4-times and 2-times upsampling operations before they are concatenated together with the larger features. This procedure can guarantee that the output feature matrix holds more abundant semantic information.

The backbone module of the YOLOv7 network consists of multiple CBS layers, ELAN layers, and MP layers. The CBS layer comprises a convolutional layer, a Batch Normalization (BN) layer, and a Sigmoid - like Unilateral (SiLU) activation function. Its primary role is to extract image features at various scales. The ELAN layer is a type of convolutional layer that can improve the quality and performance of the extracted features while maintaining the original structure. The MP layer is formed by adding a max - pooling layer on top of the CBS layer.

Along with the continuous carrying out of convolutional operations, the depth of the network has the increase, this expands its receptive field, and therefore enables the extraction of more abundant semantic information. But, this therefore also brings the risk of losing shallow-layer semantic detailed information. For the solving of this problem, the YOLOv7 network, which gets inspiration from feature pyramid networks, has a top-down information flow channel built into its feature fusion layer. It also builds side connections among the two branch parts. This method can solve the problem that semantic information is lost, and at the same time, it can promote the accuracy of classification and regression work.

The loss function of the YOLOv7 network can be expressed as Equation (1):

$$L_{total} = \lambda_1 L_{cls} + \lambda_2 L_{box} + \lambda_3 L_{obj} \quad (1)$$

Inside Equation (1), gives the classification loss that belongs to the network, gives the localization loss, and gives the confidence loss.,, and are the individual weights that correspond to these component losses respectively. Specifically, and are computed using the BCEWithLogitsLoss method, while is computed using CIOU.

Gesture Recognition Using the Improved YOLOv7 Network

Although the YOLOv7 network has strong abilities for extracting object features and can reach effective object recognition and detection, it often has failure in more complex recognition situations, such as when detecting extremely small objects. Consequently, in order to further boost YOLOv7's efficacy in detecting tiny objects, a combined attention mechanism named CBAM (Channel and Spatial Attention Module) is incorporated to optimize the network. Inside CBAM, the channel attention mechanism enables the network to concentrate more on channels that hold crucial information and disregard those that are unimportant. The spatial attention mechanism, which builds on the channel attention, aids the network in pinpointing the areas where image features are densely packed, thus enhancing the detection precision even further.

The computational process of the CBAM attention module is expressed as Equation (2):

$$\begin{cases} F' = M_c(F) \otimes F \\ F'' = M_s(F') \otimes F' \end{cases} \quad (2)$$

In equation (2), F denotes the input feature map to the attention module, F' denotes the feature map weighted by the channel attention module, F'' denotes the feature map processed by the spatial attention module, $M_C(F)$ denotes the extraction of F by the channel attention module, and $M_S(F')$ denotes the extraction of F' by the spatial attention module.

The YOLOv7 network architecture enhanced by CBAM is shown in Figure 2.

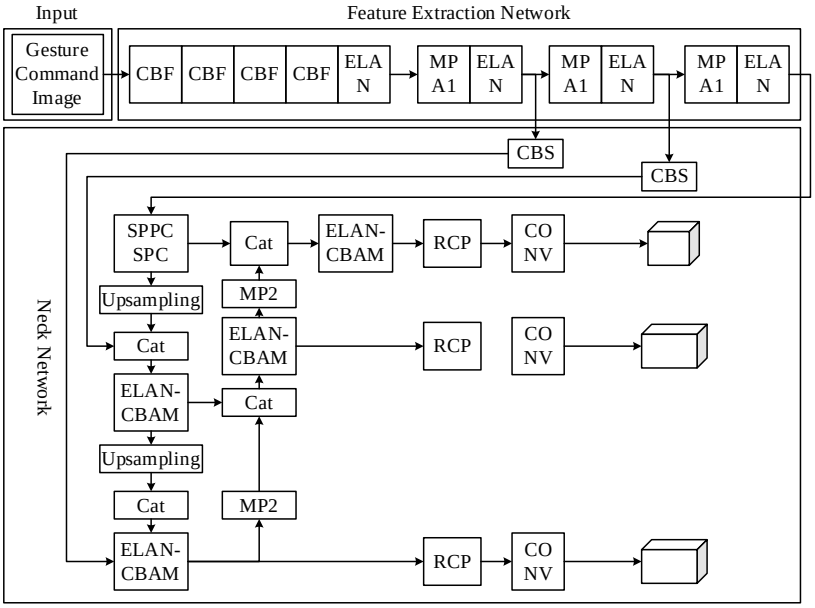


Figure 2: The YOLOv7 network structure improved based on CBAM

3 Speech Emotion Recognition Model Based on Speech Enhancement

Two-Stage Attention Mechanism Speech Enhancement Model

Single-Channel Time-Domain Speech Enhancement

The single - channel time - domain speech enhancement approach is characterized by an extremely straightforward architecture. It mainly employs a convolutional neural network (CNN) as a link. This framework, which consists of an encoder and a decoder, obviates the necessity for conversion between the time and frequency domains. As a result, it skips the procedure of reconstructing the original speech by using the phase information of noisy speech. This effectively deals with the human ear's lack of sensitivity to phase information.

Drawing upon the basic tenet of single - channel time - domain speech enhancement, which involves converting speech contaminated with noise into time - domain waveform data, sending it to the encoder, extracting and recovering data characteristics through a convolutional neural network, and ultimately producing enhanced speech for decoding and output, the time domain of noisy speech can be expressed by Equation (3):

$$y(m) = F_{\alpha}(y(m)) \quad (3)$$

In the given equation, stands for the initial noisy speech signal; denotes the denoised, clear speech signal; and represents pure noise. When it comes to time - domain speech enhancement, the process entails handling and making an estimation of the clean speech signal within the original data to acquire the estimated value . The calculation process can be presented as Equation (4):

$$\hat{s}(m) = F_{\alpha}(y(m)) \quad (4)$$

In the equation, stands for the feature mapping of the convolutional neural network training goal within the encoder - decoder. The value of this feature mapping needs the computation of the loss function. Its formula is presented in Equation (5).

$$E = \frac{1}{M}\sum_{m=1}^{M}\left\| F_{\alpha}(y(m)) - s(m)\right\|_2^2 \tag{5}$$

Two-Stage Attention Mechanism-Enhanced Network Model

Drawing on single-pass time-domain speech enhancement, this model integrates self-attention and spatial attention mechanisms during two separate stages to boost speech enhancement effectiveness. The structure of the model is depicted in Figure 3.

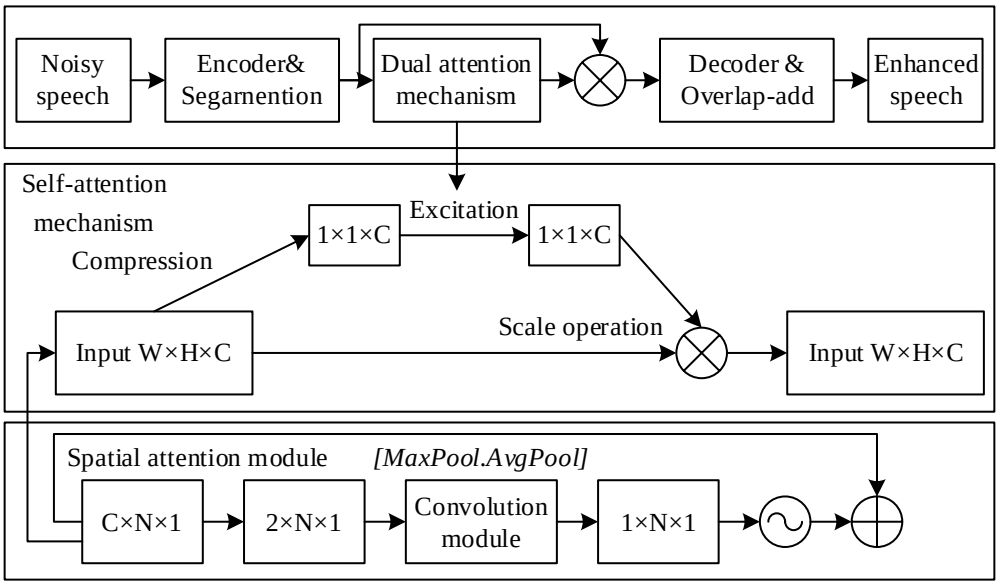


Figure 3: Two-stage attention mechanism speech enhancement model

Analysis of Figure 3 reveals that the speech enhancement process proceeds as follows:

(1) Prior to being fed into the encoder, the initial noisy speech is transformed into time - domain waveform data. Subsequently, this data is directly introduced into the self - attention mechanism module. This module then extracts global attention features from the initial noisy speech data.

(2) Within the spatial attention module, the positional data in the spatial domain of the features is retrieved. By performing global average pooling and maximum pooling operations, two feature maps are created and combined into a novel tensor. Subsequently, this tensor is fed into the convolutional layer module, where the spatial attention features are calculated using the sigmoid function.

(3) The characteristics acquired from the two attention modules are merged, and the combined characteristic data is sent to the decoder.

(4) In the decoder, the integrated attention features undergo a two-dimensional convolution operation, followed by a PReLU function to compute the attention feature masking, denoted as $M \in R^{C\times N\times 1}$. This masking is then applied to reconstruct the enhanced speech signal, with the reconstruction formula expressed as Equation (6):

$$M = f_{mask}(X_S) \quad (6)$$

In the equation, $f_{mask}(\cdot)$ represents the mapping function corresponding to the two-dimensional convolutional masking module.

Residual Network-Based Speech Emotion Recognition

Speech Signal Preprocessing

The speech signals obtained at the beginning cannot be used directly, either. To prevent problems like power reduction and poor resolution in the unprocessed speech signals, pre - processing procedures such as pre - emphasis, framing, and windowing need to be carried out prior to feature extraction.

Feature Extraction Based on Mel-Frequency Cepstral Coefficients

Mel - frequency cepstral coefficients represent a frequently employed approach for extracting features from speech signals. These coefficients characterize the sound perception abilities according to the operational mechanism of the human ear. The relationship between Mel frequencies and linear frequencies is presented by Equation (7):

$$Mel(f) = 2595 \times \lg_{10}\left(1 + \frac{f}{700}\right) \quad (7)$$

To enable the analysis of voice signals, the energy distribution of the time - domain voice signal is converted into the frequency domain. The processed symmetric frame signal is subjected to a Fast Fourier Transform (FFT), which is in accordance with Equation (8):

$$FFT(k) = \sum_{n=0}^{N-1} S_W(n) e^{-\frac{2\pi}{N}kn}, 0 \le n, k \le N-1 \quad (8)$$

The procedure for extracting speech features using Mel - frequency cepstral coefficients is detailed as follows:

(1) Convert each preprocessed frame of speech signal into an energy spectrum. Apply a set of Mel-scale filters to eliminate harmonic interference, achieving a smooth frequency representation that highlights the speech signal's formants. Typically, triangular bandpass filters are selected for this purpose, whose frequency response can be expressed by Equation (9):

$$H_m(k)=\begin{cases}0, k< f(m-1)\\ \dfrac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, \\ f(m-1)\le k\le f(m)\\ \dfrac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, \\ f(m)\le k\le f(m+1)\\ 0, f(m+1)\le k\end{cases} \tag{9}$$

In the equation, $f(m)$ represents the center frequency of the Mel filter; k represents the frequency.

(2) Compute the outcomes for each filter output independently to acquire the corresponding log - Mel energy spectrum. Concurrently, execute the discrete cosine transform to obtain the n - order Mel cepstral parameters, where the n - order represents the order of the Mel frequency cepstral coefficients. This stage conducts dimensionality reduction and abstraction on the speech signal data, extracting static characteristics.

(3) Insert dimensional data between two frames of the speech signal. To extract the dynamic features of the speech signal, utilize the first - order and second - order differences of the static features.

(4) Combine dynamic features with static features.

In summary, the Mel-frequency cepstral coefficient (MFCC) emotion features primarily consist of $1/3$ MFCC coefficients, $1/3$ first-order difference parameters, and $1/3$ second-order difference parameters.

Steps for Speech Emotion Recognition Based on Residual Networks

This speech emotion recognition model primarily consists of three components: The first component extracts emotional features from speech signals using the Mel-frequency cepstral coefficient method. The second component trains the model on the source domain dataset using a designed residual convolutional neural network. The third component transfers the knowledge model learned in the second component to the test dataset via transfer learning techniques to complete recognition. The specific network architecture is shown in Figure 4.

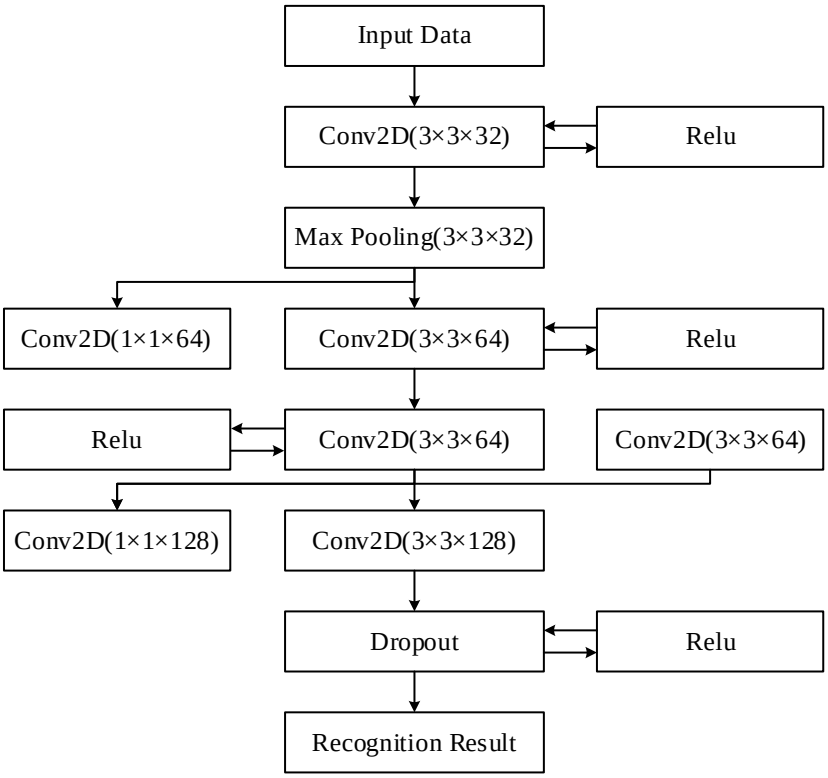


Figure 4: Transfer the neural network structure

In Figure 4, transfer learning techniques were selected due to the scarcity of continuous speech emotion datasets. Therefore, data was learned through transfer learning to enhance the accuracy of neural network recognition.

4 Classroom Scenarios and Early Childhood Development Based on Model Tools

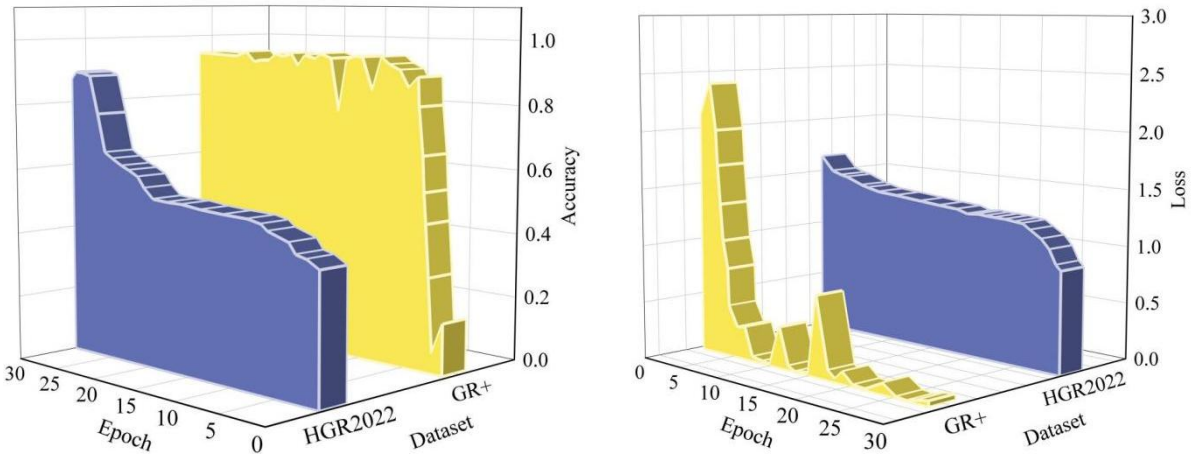
As previously mentioned, this research paper puts forward a gesture recognition model that relies on an enhanced YOLOv7 network, along with a speech emotion recognition model founded on speech enhancement. These models are designed to recognize gestures and speech emotions in young children. As AI tools, both models can assist in constructing, adjusting, and optimizing classroom scenarios and content within contextual teaching. This section evaluates the overall performance of the proposed speech recognition model using accuracy and loss metrics alongside confusion matrix analysis. To evaluate the overall proficiency of the suggested speech emotion recognition model, an assessment is carried out on its emotional recognition precision, recollection rate, and F1 score in real - world applications. Additionally, a practical experiment is carried out to evaluate the efficacy of the proposed model tools when creating preschool classroom scenarios within a context.

Performance Evaluation of Gesture Recognition Models

Loss Value and Accuracy Rate

Two large-scale, comprehensive gesture datasets—GR+ and HGR2022—were selected as experimental datasets. With 30 iterations set, the proposed gesture recognition model

performed gesture recognition tasks. The Figure 5(a)-(b) give the model's recognition accuracy and loss numerical values on the GR+ and HGR2022 data sets. Overall, the recognition accuracy of the proposed model increases with the number of iterations, while the loss value decreases. On the GR+ dataset, the model converged after 5 iterations, with recognition accuracy stabilizing at 80.00% or higher and loss values stabilizing between 0.2 and 0.3. On the HGR2022 dataset, the model exhibits slightly weaker performance without clear convergence characteristics, though recognition accuracy reaches a maximum of 89.53%. The loss value begins to converge after 5 iterations, stabilizing between 1.2 and 1.4.

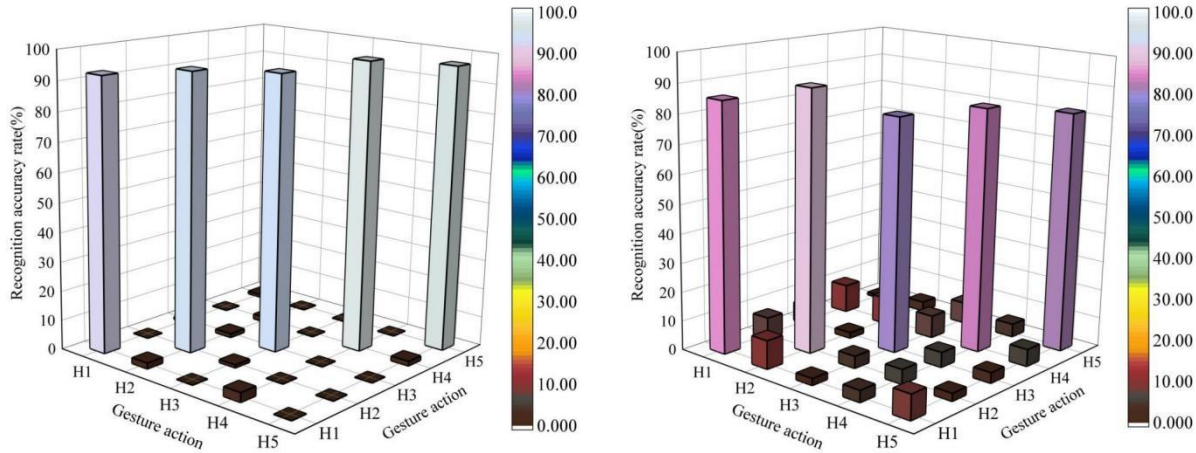


(a) Recognition accuracy rate (b) Loss value

Figure 5: The accuracy rate and loss value of gesture action recognition

Confusion Matrix Analysis

Five classic gesture images were extracted from two experimental datasets: (H1) fist, (H2) open palm, (H3) scissors, (H4) thumbs-up, and (H5) three-point. The confusion matrices for the gesture recognition model developed in this paper are shown in Figures 6(a)-(b). The gesture recognition model achieves recognition accuracy of 90.00% or higher for all five gestures on the GR+ dataset, with low misclassification rates between different gestures (0.00%–5.00%). However, on the HGR2022 dataset, recognition accuracy ranges only between 80.00% and 90.00%, with relatively high misclassification rates across different actions (0.00%–10.00%).



(a) GR+ Dataset (b) HGR2022 Dataset

Figure 6: A hybrid matrix for gesture action recognition

Based on the recognition accuracy, loss values, and confusion matrix performance of the proposed gesture recognition model across two experimental datasets, the model demonstrates high overall recognition accuracy and strong convergence. However, its performance on the GR+ dataset significantly outperforms that on the HGR2022 dataset. This disparity may stem from the HGR2022 dataset containing a larger number of images with higher similarity between them, making them more challenging to distinguish compared to the GR+ dataset.

Performance Evaluation of Speech Emotion Recognition Models

Practical Application Performance

E Kindergarten was selected as the experimental site, where the proposed speech recognition model was trained within its classroom environment. The trained model was then used to conduct speech recognition tests on 20 randomly selected preschoolers (divided into a girls' group and a boys' group), with a total of 20 tests performed. The female group utilized experiment serial numbers 1–10, therefore the male group utilized serial numbers 11–20. Table 1 presents the performance of the trained speech recognition model in terms of recognition accuracy.

Table 1: The recognition effect of the speech recognition model

Serial Number (Girls' Group)	Recognition accuracy rate (%)	Serial Number (Boys' Group)	Recognition accuracy rate (%)
1	97.9	11	89.28
2	91.49	12	92.47
3	96.22	13	94.04
4	93.11	14	94.75
5	94.4	15	92.64
6	88.69	16	86.96
7	97.83	17	93.24
8	85.12	18	94.8
9	96.58	19	94.82
10	95.85	20	89.06
Average	93.72		92.21

Following the training phase, the speech recognition model attained an average accuracy rate of 93.72% among female participants and 92.21% among male participants. Significantly, both of these rates surpassed the 90.00% threshold. These outcomes strongly indicate that this model is capable of accurately performing recognition and evaluation of the pronunciation of preschool - aged children.

Speech Emotion Recognition Performance

We have adopted the classic CSE2022 Children's Speech Emotion dataset to be our experimental dataset. We model for recognizing emotion in speech was utilized to distinguish six emotion kinds: (M1) Surprise, (M2) Fear, (M3) Anger, (M4) Sadness, (M5) Disgust, and (M6) Happiness. The whole identification correct rate confusion matrix is displayed in Figure

7, thus the detailed correct rate, recall, and F1 score performance is concluded in Table 2.

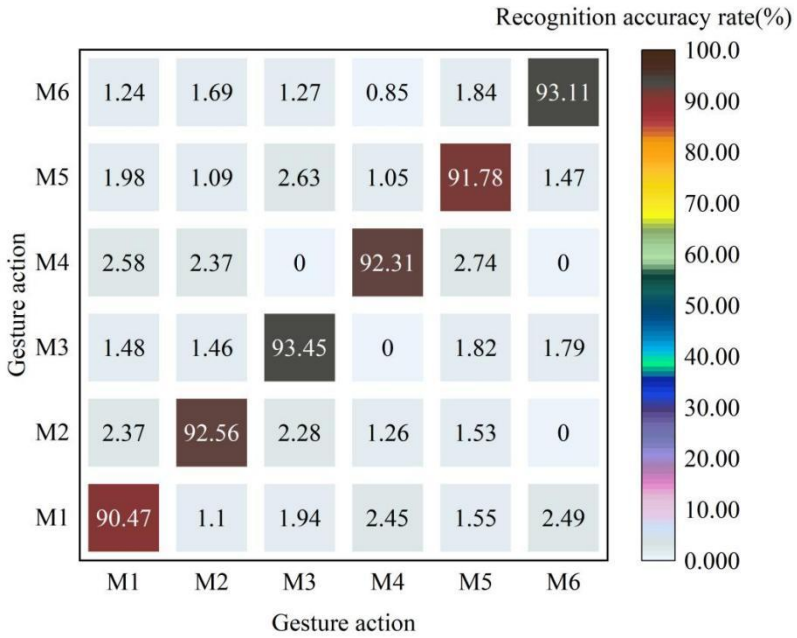


Figure 7: Speech emotion recognition hybrid matrix

Table 2: Student voice emotion parameters

Emotion	Precision (%)	Recall (%)	F1-score (%)
M1	90.47	96.49	95.62
M2	92.56	92.01	91.61
M3	93.45	98.04	97.31
M4	92.31	92.25	93.41
M5	91.78	93.51	95.66
M6	93.11	99.87	98.32

Combining Figure 7 and Table 2, the overall accuracy of the speech emotion recognition model designed in this paper reaches 90.00% or higher, demonstrating relatively precise and balanced recognition capabilities. This model displays the most strong recognition abilities toward (M3) anger and (M6) happiness, achieving accuracy values of 93.45% and 93.11%, separately, with recall values of 98.04% and 99.87%. The corresponding F1 point values are 97.31 percent and 98.32 percent. This performance originates from the situation that anger and happiness display the most obvious emotional changes in sound expression, therefore making them comparatively more easy to recognize.

Classroom Scenarios and Early Childhood Development in the Context of AI Tools

We selected as research objects the E1 class and E2 class of Kindergarten E, each of which has 35 children. E1, which acts as the experimental group, utilized gesture and speech emotion recognition models together with situational teaching; E2, which is the control group, only adopted situational teaching method. After five months of teaching intervening work, children's self-control was carried out measurement from four dimensions: self-consciousness, persistence, delayed satisfaction and impulse controlling, through three evaluation tables: teacher grading, parent grading and situational performance marking, so as to check the

model’s supporting function in teaching.

Self-Control Performance Among Groups Before the Experiment

Table 3 showcases the comparison of the outcomes of the homogeneity test for the overall self - control scores and the four aspects between the experimental group and the control group before the experiment. Statistical analysis of teacher questionnaires indicates that the control group (5.33) exhibited slightly higher mean scores than the experimental group (5.26) across all four dimensions and overall self-control. Furthermore, the two classes showed no significant differences in (D1) Conscientiousness ($P=0.92>0.05$), (D2) Persistence ($P=0.85>0.05$), (D3) Self-Delayed Gratification ($P=0.83>0.05$), and (D4) Impulse Control ($P=0.82>0.05$). There was no statistically notable disparity detected in the overall self - regulation capacity ($P = 0.61 > 0.05$).

Parent questionnaire results indicated that the control group's mean overall self-control ability (5.04) was slightly higher than the experimental group (4.93), with $P=0.22>0.05$, showing no statistically significant difference. No significant differences were observed across all four dimensions: (D1) Consciousness ($P=0.37>0.05$), (D2) Persistence ($P=0.35>0.05$), (D3) Self-Delayed Gratification ($P=0.35>0.05$), and (D4) Impulse Inhibition ($P=0.28>0.05$).

In the situational performance assessment, the control group (5.40) demonstrated a slightly higher overall self-control ability than the experimental group (5.49), though this difference was not statistically significant ($P=0.45>0.05$). Similarly, no statistically significant differences were found across the four dimensions ($P=0.64, 0.42, 0.72, 0.44>0.05$).

In summary, the overall self-control ability levels of students in the experimental and control groups were homogeneous prior to the experiment, with no significant differences. This strong comparability makes them suitable for subsequent research experiments.

Table 3: Comparison of self-control ability between two classes before the experiment

Dimensionality	Evaluation subject	Class E1(M±SD)	Class E2(M±SD)	P
D1	A1	4.81±0.44	4.40±0.18	0.92
	A2	4.49±0.3	4.57±0.28	0.37
	A3	5.60±0.81	5.22±0.35	0.64
D2	A1	5.51±0.21	6.44±0.59	0.85
	A2	4.60±0.13	4.71±0.14	0.35
	A3	5.65±0.32	5.07±0.07	0.42
D3	A1	5.79±0.26	4.44±0.36	0.83
	A2	4.97±0.49	5.13±0.19	0.35
	A3	5.67±0.48	6.09±0.25	0.72
D4	A1	4.94±0.61	6.03±0.39	0.82
	A2	5.66±0.11	5.73±0.18	0.28
	A3	4.68±0.86	5.59±0.29	0.44
Total score of self-control ability	A1	5.26±0.13	5.33±0.38	0.61
	A2	4.93±0.32	5.04±0.48	0.22
	A3	5.40±0.42	5.49±0.41	0.45

Self-Control Performance Among Groups After the Experiment

Table 4 showcases the post - experiment assessment of self - regulation capabilities between the experimental and control groups, as well as the outcomes of dimensional disparity tests. The results from the teacher questionnaire suggest a notable overall disparity in self - regulation capabilities between the experimental group (6.49) and the control group (6.00) after the experiment ($P = 0.04871 < 0.05$). Significant differences were observed in two of the four dimensions: (D1) Self-Awareness ($P=0.02512<0.05$) and (D4) Impulse Inhibition ($P=0.04721<0.05$).

After organizing, analyzing, and comparing parental questionnaire results, the experimental group (7.51) and control group (5.46) showed extremely significant differences in overall self-control ability post-intervention ($P=0.00000$). Specifically, extremely significant differences were found across three dimensions: (D1) Consciousness ($P=0.00000$), (D2) Persistence ($P=0.00000$), (D3) Self-Delayed Gratification ($P=0.00000$). A statistically significant difference was also observed in (D4) Impulse Inhibition ($P=0.04633<0.05$).

The findings of the situational experiment demonstrate that subsequent to the experiment, there was a highly significant disparity ($P = 0.00000$) in the overall self - regulation capabilities between the experimental group (scoring 7.42) and the control group (scoring 5.82). Specifically, extremely significant differences were found in three dimensions: (D1) Consciousness ($P=0.00000$), (D3) Self-Delayed Gratification ($P=0.00000$), and (D4) Impulse Inhibition ($P=0.00000$). A statistically significant difference was also observed in (D2) Perseverance ($P=0.01278<0.05$).

In summary, under the model-assisted development of the scenario-based teaching approach in this study, children's self-control abilities demonstrated clear and effective development and enhancement. Compared to the single scenario-based teaching approach, this model received consistent recognition from teachers, parents, and practitioners, particularly in the (D1) Consciousness dimension ($P=0.00000$).

Table 4: Comparison of self-control ability between two classes after the experiment

Dimensionality	Evaluation subject	Class E1(M±SD)	Class E2(M±SD)	P
D1	A1	6.27±0.69	5.91±0.65	0.02512
	A2	7.88±0.87	5.04±0.43	0.00000
	A3	7.4±0.72	5.9±0.58	0.00000
D2	A1	6.59±0.64	6.01±0.56	0.11346
	A2	7.77±0.81	5.52±0.51	0.00000
	A3	7.21±0.82	6.91±0.67	0.01278
D3	A1	6.49±0.63	6.53±0.44	0.13969
	A2	7.54±0.77	5.39±0.65	0.00000
	A3	7.59±0.84	5.28±0.67	0.00000
D4	A1	6.59±0.85	5.56±0.54	0.04721
	A2	6.84±0.65	5.88±0.57	0.04633
	A3	7.47±0.67	5.19±0.43	0.00000
Total score of self-control ability	A1	6.49±0.68	6.00±0.68	0.04871
	A2	7.51±0.76	5.46±0.58	0.00000
	A3	7.42±0.72	5.82±0.59	0.00000

Self-Control Performance in the Experimental Group Before and After the Experiment

Table 5 showcases the self - regulation performance of the experimental group both prior to and subsequent to the experiment. It also includes the outcomes of the difference tests carried out across multiple dimensions. Analysis of teacher questionnaire results indicates that post-intervention (6.49), the experimental group's overall self-control ability (5.26) showed a significant improvement compared to pre-intervention ($P=0.04872<0.05$). Statistically significant differences were observed across all four dimensions, with the (D1) Self-Awareness dimension exhibiting an extremely significant difference ($P=0.00000$).

In the parent questionnaire results, except for the (D4) Impulse Inhibition dimension, the experimental group children demonstrated extremely significant differences ($P=0.00000$) in overall self-control ability and in the (D1) Self-Awareness, (D2) Perseverance, and (D3) Self-Delayed Gratification dimensions after the experiment compared to before.

The situational experiment results showed that after the intervention, the experimental group children's overall self-control ability (7.42) demonstrated a significant improvement compared to before the intervention (5.40) ($P=0.00092<0.001$). Statistically significant differences were observed across all four dimensions, with particularly pronounced differences in (D2) Perseverance, (D3) Self-Delayed Gratification, and (D4) Impulse Inhibition ($P=0.00000$).

Data analysis makes clear that under the help of the model which this study has designed, the built scenarios—in both structure and content aspect—better match the physical and psychological development demands of children, therefore significantly pushing forward the all-round development of their self-control capabilities.

Table 5: Comparison of self-control ability before and after the experiment in Class E1

Dimensionality	Evaluation subject	Before the experiment (M±SD)	After the experiment (M±SD)	P
D1	A1	4.81±0.44	6.27±0.69	0.00000
	A2	4.49±0.3	7.88±0.87	0.00000
	A3	5.60±0.81	7.4±0.72	0.00023
D2	A1	5.51±0.21	6.59±0.64	0.03562
	A2	4.60±0.13	7.77±0.81	0.00000
	A3	5.65±0.32	7.21±0.82	0.00000
D3	A1	5.79±0.26	6.49±0.63	0.00465
	A2	4.97±0.49	7.54±0.77	0.00000
	A3	5.67±0.48	7.59±0.84	0.00000
D4	A1	4.94±0.61	6.59±0.85	0.00963
	A2	5.66±0.11	6.84±0.65	0.04321
	A3	4.68±0.86	7.47±0.67	0.00000
Total score of self-control ability	A1	5.26±0.13	6.49±0.68	0.04872
	A2	4.93±0.32	7.51±0.76	0.00000
	A3	5.40±0.42	7.42±0.72	0.00092

5 Conclusion

For solving the problem that the current situation teaching methods do not match the early childhood development, this thesis puts forward a gesture identifying model which is based on an improved YOLOv7 network and a speech emotion identifying model that is strengthened by using voice augmentation. These models give help to the constructing, the refining, and the adapting of classroom scenes, so that they can better satisfy the physical and psychological growth demands of young children.

The gesture identification model that is based on the improved YOLOv7 network has attained stable accuracy of 80.00% or higher after 30 iterative processes on the experimental data set, with loss numerical values controlled between 0.2 and 1.4, hence it demonstrates both accuracy and effectiveness.The speech emotion recognition model enhanced by voice augmentation demonstrated highly reliable feasibility in practical applications, achieving recognition accuracy rates of 90.00% or higher for children's pronunciation and equally high rates for speech emotion recognition.

In the application experiment at E Kindergarten, the proposed model tool assisted in constructing the overall self-control ability level of preschoolers in the experimental group within classroom scenarios. This method not only got sustained approval from teachers (6.49) and guardians (7.51) but also showed a notable enhancement when compared with pre-experiment results ($P=0.04872>0.05$).

Funding

This work was supported.by:

Liaoning Education Society 15th Five-Year Plan Project 2026 (XH2026988): Research on Data-Driven Personalized Reading Intervention for Young Children

About the Author

Ge Cui have the birth in 1976, at Shanghe County which belongs to Shandong Province, of China. I have gotten a master diploma in the engineering of software from Dalian University of Technology. I at the present time hold a work position at the Teacher Education Center of Anshan Normal University.My main major are teacher education and information technology-assisted instruction.

References

[1] Haslip, M. J., & Gullo, D. F. (2018). The changing landscape of early childhood education: Implications for policy and practice. Early Childhood Education Journal, 46(3), 249-264.

[2] Bakken, L., Brown, N., & Downing, B. (2017). Early childhood education: The long-term benefits. Journal of research in Childhood Education, 31(2), 255-269.

[3] Qi, X., & Melhuish, E. C. (2017). Early childhood education and care in China: History, current trends and challenges. Early Years, 37(3), 268-284.

[4] Li, H., Yang, W., & Chen, J. J. (2016). From ‘Cinderella’to ‘Beloved Princess’: The evolution of early childhood education policy in China. International Journal of Child Care and Education Policy, 10(1)

[5] Hong, X., Luo, L., & Cui, F. (2013). Investigating regional disparities of preschool education development with cluster analysis in mainland China. International Journal of Child Care and Education Policy, 7(1), 67-80.

[6] Huang, J., & Xiong, C. (2021). Quality preschool education in China: development connotation and realization approach. Journal of East China Normal University (Educational Sciences), 39(3), 33.

[7] Pan, Y., Wang, X., & Li, L. (2018). Early childhood education and development in China. In International handbook of early childhood education (pp. 599-622). Dordrecht: Springer Netherlands.

[8] Su, Y., Rao, N., Sun, J., & Zhang, L. (2021). Preschool quality and child development in China. Early Childhood Research Quarterly, 56, 15-26.

[9] Li, K., Zhang, P., Hu, B. Y., Burchinal, M. R., Fan, X., & Qin, J. (2019). Testing the ‘thresholds’ of preschool education quality on child outcomes in China. Early Childhood Research Quarterly, 47, 445-456.

[10] Lu, J. (2024, July). Research on the Development of Teaching Models in Preschool Education in China. In 5th International Conference on Language, Art and Cultural Exchange (ICLACE 2024) (pp. 102-108). Atlantis Press.

[11] Ngumbi, E. (2022). Challenges Facing Early Childhood Teachers and Possible Solutions. Education & Child Development, 1(1).

[12] Devi, J. S., Sreedhar, M. B., Arulprakash, P., Kazi, K., & Radhakrishnan, R. (2022). A path towards child-centric Artificial Intelligence based Education. International Journal of Early Childhood, 14(3), 9915-9922.

[13] Su, J., & Yang, W. (2024). Artificial Intelligence (AI) literacy in early childhood education: An intervention study in Hong Kong. Interactive Learning Environments, 32(9), 5494-5508.

[14] Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. IEEE access, 8, 75264-75278.

[15] Lamanauskas, V. (2025). Pre-service preschool and primary school teachers' position on artificial intelligence: Aspects of benefits and impact in the future. Gamtamokslinis ugdymas bendrojo ugdymo mokykloje, 31(1), 24-35.

[16] Wang, H. (2022, May). The application of interactive artificial intelligence in the intelligent integration of infants playing teaching aids. In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 30-33). IEEE.

[17] Vidović, E. (2024). Applying the Situational Approach in Foreign Language Communication with Children in Preschool Institutions: An Overview. DHS-Društvene i humanističke studije: časopis Filozofskog fakulteta u Tuzli, 26(26), 1215-1230.

[18] Li, D. (2024). An interactive teaching evaluation system for preschool education in universities based on machine learning algorithm. Computers in human behavior, 157, 108211.

[19] Dogmus, Z., Erdem, E., & Patoglu, V. (2014). ReAct!: An interactive educational tool for AI planning for robotics. IEEE Transactions on Education, 58(1), 15-24.

[20] Zhou, C., Kuang, D., Liu, J., Yang, H., Zhang, Z., Mackworth, A., & Poole, D. (2020, April). AISpace2: an interactive visualization tool for learning and teaching artificial intelligence. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 09, pp. 13436-13443).

[21] Ramu, M. M., Shaik, N., Arulprakash, P., Jha, S. K., & Nagesh, M. P. (2022). Study on potential AI applications in childhood education. International Journal of Early Childhood, 14(03), 2022.

[22] Zhai, H. (2021, June). The application of VR technology in preschool education professional teaching. In 2021 2nd International Conference on Artificial Intelligence and Education (ICAIE) (pp. 319-323). IEEE.

[23] Li, Q. (2022). A study on mobile resources for language education of preschool children based on wireless network technology in artificial intelligence context. Computational and Mathematical Methods in Medicine, 2022(1), 6206394.

[24] Weng, X., Ye, H., Dai, Y., & Ng, O. L. (2024). Integrating artificial intelligence and computational thinking in educational contexts: A systematic review of instructional design and student learning outcomes. Journal of Educational Computing Research, 62(6), 1420-1450.

[25] Yang, W. (2022). Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation. Comput. Educ. Artif. Intell., 3, 100061.

[26] Kölemen, E. B., & Yıldırım, B. (2025). A new era in early childhood education (ECE): Teachers' opinions on the application of artificial intelligence. Education and Information Technologies, 1-42.

[27] Fikri, Y., & Rhalma, M. (2024). Artificial Intelligence (AI) in early childhood education (ECE): Do effects and interactions matter?. International Journal of Religion: Politics, Sociology, Culture, 5(11), 7536-7545.

[28] Latorre-Medina, M. J., & Abdelmaula-Mesaud, S. (2025). Artificial intelligence applied to early childhood education: A focus for educational research?. Contemporary Issues in Early Childhood, 26(1), 140-153.

[29] Qayyum, A., Bukahri, M., Zulfiqar, P., & Ramzan, M. (2024). Balancing artificial intelligence and human insight in early childhood education: Implications for child development. Social Science Review Archives, 2(2), 1520-1536.

[30] Prentzas, J. (2013). Artificial intelligence methods in early childhood education. In Artificial intelligence, evolutionary computing and metaheuristics: In the footsteps of Alan Turing (pp. 169-199). Berlin, Heidelberg: Springer Berlin Heidelberg.

[31] Durrani, R., Iqbal, A., & Akram, H. (2024). Artificial intelligence (AI) in early childhood education, exploring challenges, opportunities and future directions: A scoping review. Qlantic Journal of Social Sciences, 5(2), 411-423.

[32] Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. Computers and Education: Artificial Intelligence, 3, 100049.

[33] Yi, H., Liu, T., & Lan, G. (2024). The key artificial intelligence technologies in early childhood education: a review. Artificial Intelligence Review, 57(1), 12.

[34] Aslan, S., Durham, L. M., Alyuz, N., Okur, E., Sharma, S., Savur, C., & Nachman, L. (2024). Immersive multi-modal pedagogical conversational artificial intelligence for early childhood education: An exploratory case study in the wild. Computers and Education: Artificial Intelligence, 6, 100220.

[35] Purwandari, N., & Nasution, T. (2025). Development Of Interactive Learning Applications Based On Artificial Intelligence To Improve Early Childhood Literacy Skills At RA Darussalam 009. International Journal of Science, Technology & Management, 6(3), 561-566.

[36] Li, X., & Taber, K. S. (2022). The future of interaction: Augmented reality, holography and artificial intelligence in early childhood science education. In STEM, robotics, mobile apps in early childhood and primary education: Technology to promote teaching and learning (pp. 415-442). Singapore: Springer Nature Singapore.