

Research Article

Analysis of Concentration in English Education Learning Based on CNN Model

Kun He¹ and Kun Gao² 

¹*Xi'an Siyuan University, Shaanxi, Xi'an 710038, China*

²*Xi'an Jiaotong University, Shaanxi, Xi'an 710048, China*

Correspondence should be addressed to Kun Gao; gaokun@mail.xjtu.edu.cn

Received 28 May 2022; Revised 14 June 2022; Accepted 18 June 2022; Published 8 July 2022

Academic Editor: Lianhui Li

Copyright © 2022 Kun He and Kun Gao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Obtaining the learning behavior of terminal learners solves the problem of inability to understand students' concentration in class. When learners learn English, they can obtain learning concentration to understand learners' concentration, in order to analyze the influence coefficient of various learning behavior data on learning concentration. This paper will design and implement a terminal data acquisition tool to collect the device perception information in the learning environment of learners, and capture the learner's touch screen operation data based on the virtual simulation experiment, and then use the improved neural network to process the collected terminal sensor data for learning behavior. We identify and obtain the learner's learning activity state, and finally fit the learner's behavioral data weight through a linear regression equation, monitor the learner's learning state, and explore the influencing factors of learning concentration.

1. Introduction

In recent years, with the rapid changes in the era of intelligent information and the rapid development of the technological society, the trend of intelligentization has accelerated, and people's demand for intelligent life has gradually increased. Automated driving technology has entered the field of transportation, intelligent voice robots have entered the service industry, and intelligent sweeping robots have entered the field of home furnishing. In the context of education, it is on the agenda to develop intelligent applications suitable for English classroom education.

Concentration, also known as concentration, refers to the continuous and persistent listening state in the classroom and refers to the psychological state of a person when he is concentrating on a certain thing or activity. Due to their young age, primary and secondary school students do not have a strong sense of self-management, and their energy is easily dispersed by other things, resulting in insufficient energy invested in a relatively important thing. For example, when students are listening to a class, part of the time will be

in a "slippage" state, which leads to many students' knowledge breakpoints, which directly affects the quality of the classroom. In addition, when the number of students in the class is large, it is difficult for English teachers to make real-time judgments on everyone's concentration while focusing on lectures. The low concentration of students leads to poor completion of homework after class, which will greatly increase the need for English teachers after class workload, and it is impossible to truly teach students in accordance with their aptitude. Therefore, the monitoring of students' concentration in English classroom is one of the most needed research directions. The traditional English classroom quality analysis is carried out by arranging lectures in the classroom, after-class questionnaires, etc., which is highly subjective, exposing the shortcomings of classroom real-time monitoring and analysis, and cannot meet people's expectations for high-quality English classroom teaching requirements and expectations. Therefore, it is most important to introduce intelligent education into the classroom, assist teachers in intelligent classroom management, and understand the listening status of each student anytime

and anywhere, so as to supervise English classrooms in a timely manner, and customize personalized English teaching plans in a timely manner after class one of the tasks. Research on a concentration monitoring and analysis system that can be applied to English classrooms has strong practicability and market space, and can also promote the development of English education in the direction of intelligence, comprehensiveness, and diversification, and has broad application prospects [1–10].

2. Related Work

Most scholars have done a lot of research on one aspect of mindfulness, attention, and put forward many insightful points, which provide a lot of thought reference for this research. To sum up, the relevant research on concentration is mainly carried out from the following three aspects: the first is the role of attention. It mainly conducts in-depth research on the phenomenon of inattentiveness in children, adolescents, athletes, etc.; the second is the research on the detection methods and evaluation standards of attention; and the third is the research on the cultivation of attention. The research on attention has a complete set of detection methods and evaluation standards. Although it is only a part of attention, it also provides many valuable references for the study of attention. Then, with the improvement of attention, the level of information technology, the research on the intelligent management, concentration monitoring, and analysis system of primary and middle school students' classrooms are moving toward the intelligent mode. In the 1850s, foreign scholar Aryamov made a statistical analysis of the attention of primary school students through the method of observation and recording, and believed that the attention span of primary school students aged 7–10 was about 20 minutes. This method does not have very good objectivity, has a strong subjective color, and also requires the observer to have the relevant ability. Later, the classroom teaching evaluation method appeared mainly by arranging different teachers to randomly listen to the class and record the lecture notes. This method cannot specifically reflect the concentration of each student during class. In recent years, the methods of monitoring students' classroom concentration have been studied in a more objective and convenient direction. The main methods of concentration monitoring include facial microexpression recognition; recognition and analysis of head posture information; and collection of physiological data such as brain waves (EEG), manual filling of questionnaires, and behavioral recognition of video images. However, the above methods have some problems more or less. Although the classroom expression evaluation system can identify most expressions of students in class, there will be certain errors due to the insufficient amount of existing data. A relatively complete classroom expression database is needed to train the model, and there is no such classroom expression database in the public libraries at home and abroad, and it cannot be completed in a short time if the collection is carried out. The above research shows that relying on a single dimension to monitor students' classroom concentration and analyze the results obtained is not

accurate enough and unreliable. The advent of the era of intelligence has prompted the development of traditional classroom management in the direction of technology and informatization, and manual collection methods have gradually been replaced by automated operations. And at this stage, there is not much research on the concentration monitoring and analysis system of primary and secondary school classrooms that combines deep learning and multimodal integration, and there is a big gap, which is worth studying [11–16].

3. Relevant Theories and Technical Methods

3.1. Learning Behavior. Nowadays, with the widespread application of Internet devices such as networks and mobile terminals in all aspects of modern people's daily life, online learning has a positive impact on the progress of education in the field of higher education in my country. It not only breaks through the traditional teaching method and classroom model but also considered as a creative innovation of educational science and technology. By collecting and obtaining data about learners' learning behavior during online learning, it is possible to better understand and effectively optimize learning and the place where it occurs, discover information rules in learning behavior data, and apply them to realize its value. The process of learning behavior data acquisition includes the following three aspects, namely, data collection, data processing, and data analysis and presentation. The content of learning behavior data acquisition is shown in Figure 1 [17].

3.2. English Education Learning Behavior Concentration Recognition Technology

3.2.1. Classroom Attention Behavior Recognition Based on Long Short-Term Memory Network Model. Recurrent neural network (RNN) is a type of directed graph that combines the data connections between each neural node in the order of a corresponding time point, and RNN can be widely used in real-time processing. The input data of the corresponding spatial time series are extracted, and the network structure of the corresponding time-sequential objects of the output data with time-dependent characteristics is extracted. In the recurrent neural network, each node is connected to the next layer of neuron nodes through a one-way connection. During the network iteration process, each neuron contains its own previous information, and its output is affected by the previous neuron. Impact on its expanded unit is shown in Figure 2.

The structure of the RNN unit and its open structure in time are shown in Figure 2. In the figure, $x = (x_0, x_1, x_2, \dots, x_t)$ represents the input sequence of the recurrent neural network, and $y = (y_0, y_1, y_2, \dots, y_t)$ represents the output sequence of the recurrent neural network, and $h = (h_0, h_1, h_2, \dots, h_t)$ represents the hidden state sequence of the recurrent neural network. The output state of each hidden layer of the recurrent neural network depends not only on the sequence t and x of the previous hidden input and output state layer of the current input neuron but also on the

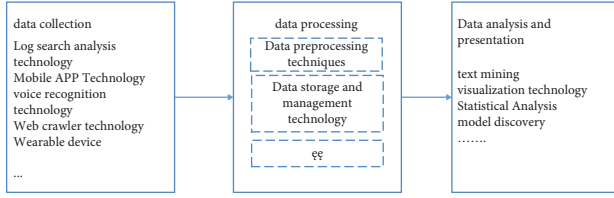


FIGURE 1: Learning behavior data acquisition process.

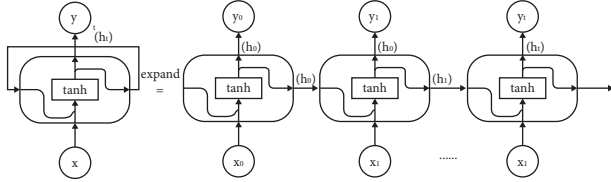


FIGURE 2: RNN unit expansion structure.

sequence of each hidden output state layer of the input neuron at the current moment h , and the expression of the calculation formula of the output layer at time t is as follows:

$$\begin{aligned} h_t &= \tanh(w_{xh}x_t + w_{hh}h_{t-1} + b_h), \\ y_t &= w_{hy}h_t + b_y. \end{aligned} \quad (1)$$

Due to the limited amount of sequence historical data information that can be saved by samples in the hidden state of the recurrent neural network, problems such as gradient explosion or gradient disappearance may occur in the process of processing this long-term sequence of samples. Problems such as gradient explosion and disappearance will cause the recurrent neural network to be unstable and unable to converge to the optimal result, so the long short-term memory network (LSTM) with a strong feedback mechanism effectively solves the drawbacks of long-term dependence in the recurrent neural network. As a variant of the recurrent neural network, LSTM is composed of a self-connected memory data storage unit and three “gates” used to control the memory data information storage. The internal structure of the LSTM unit is shown in Figure 3 [18].

3.2.2. Recognition of English Education Learning Behavior Based on Convolutional Neural Network Model. A convolutional neural network system includes several layers of convolution, pooling, and fully connected layers. There is a huge technical difference between the point-to-layer structure of a convolutional neural network and the point-to-layer structure of a fully connected neural network. Each layer of neurons in a fully connected neural network must be arranged in a one-dimensional order, while each layer of a convolutional neural network must be arranged in a three-dimensional order. Compared with the fully connected neural network, the convolutional neural network has many basic characteristics such as multilayer local data connection, weight data sharing, and next-layer sampling. Local connection means that each neuron is no longer connected to all neurons in the previous layer, and weight sharing means that a group of neurons can share weights at the same

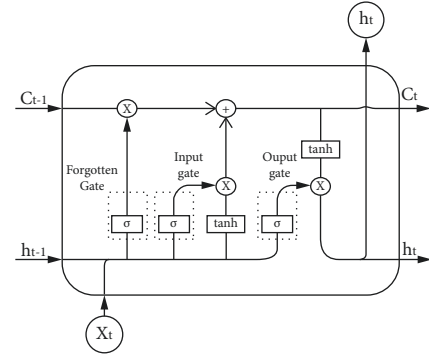


FIGURE 3: LSTM unit internal structure diagram.

position, not each connected neuron has its own positional weights. The downsampling method can greatly reduce the number of samples between each layer through the pooling layer, further greatly reduce the number of each parameter, and also greatly improve the robustness of the model. A schematic diagram of a convolutional neural network is shown in Figure 4.

The convolutional layer has the characteristics of realizing weight distribution and local links through the convolutional neural network. The convolutional layer performs feature extraction on the original input. The calculation formula is as follows:

$$a_{i,j} = f\left(\sum_{m=0}^2 \sum_{n=0}^2 w_{m,n}x_{i+m,j+n} + w_b\right). \quad (2)$$

In Formula 2 $w_{m,n}$ represents the weights of the functions in rows m and n , while $x_{i,j}$ -th row represents the elements of rows i, j . The element of row j th column w_b represents the bias term; $f()$ represents the activation function; usually, the model construction selects the ReLU function as the activation function, and the definition of the ReLU function is as follows:

$$f(x) = \max(0, x). \quad (3)$$

Compared with the training process of the fully connected neural network, although the whole training method of the convolutional neural network is more complicated, the main workload and principle of the whole training are the same, mainly using a chain algorithm to obtain the derivation method, to recalculate the partial derivative coefficient of the weight function to the weight of each unit, and then update each weight according to the principle of reverse descent of the weight gradient. The process of training this algorithm requires a reverse weight propagation algorithm.

4. Establishment of Experimental Model of English Education Learning Concentration and Analysis of Experimental Results

4.1. Data Collection

4.1.1. Terminal Data Acquisition Framework. In the whole system, the design of the mobile terminal frame is very

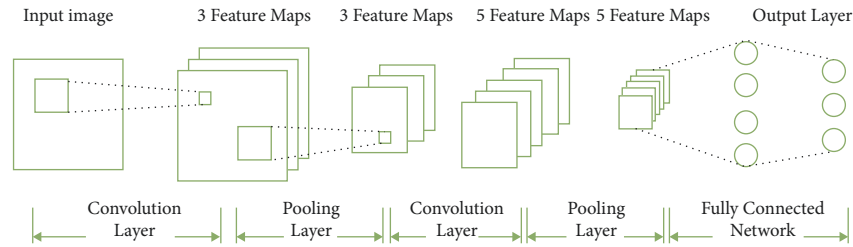


FIGURE 4: CNN neural network architecture.

important, and it is the cornerstone of the whole system. The functions completed in this design are shown in Figure 5. The terminal data, network data, and user usage data of learners' English learning behavior are collected, stored, and uploaded, and the sensor data collected in real time are packaged in API and provided to the focus. Force recognition module is shown in Figure 6.

- (1) Learner behavior data collection module. This module collects learning behavior data of a single terminal learner. The terminal data, network data, and user usage data of mobile terminal learners are collected to facilitate the calling and storage of other modules. The collected data are formed into a unified format "key-value" method in which the key is used as the unique identifier in the system, and the corresponding value saves the learner's learning behavior data in the form of a string.
- (2) Local data storage module. This module receives and stores the collected mobile terminal learning behavior data. The collected data will be directly stored in the local server in the form of a database, providing sample data for subsequent model training.
- (3) API calling module. This module provides the collected mobile terminal learning behavior data to the model in a standard format as sample data or to other models for calling. API (application programming interface) is an interface provided to the application program. After the system call is obtained, the API and the system call jointly complete the data access in the user mode and the kernel mode.
- (4) Terminal behavior data upload module. This module stores the collected learning behavior data of English classroom teaching learners, uploads it, and sends it to the network side. The network side generates summary information according to the data uploaded by the terminal and sends the summary information to the terminal to visualize the results of the collected data.
- (5) Learning concentration recognition module. This module is a research on learning concentration behavior recognition based on sensor data collected on terminal equipment. By building a model based on the LSTM model and integrating the convolutional neural network to improve the low-level processing unit to obtain the local saliency of the time series and the high-level processing unit to obtain the hidden

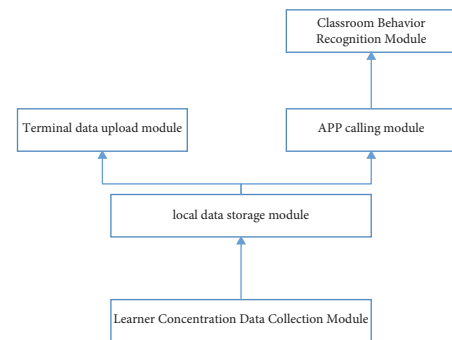


FIGURE 5: Logic diagram of terminal data acquisition framework.

feature of the time series, the classification of the learner's learning behavior and activity state are realized identify. When the terminal data acquisition system performs the startup operation, the application program is initialized, and the data for monitoring the terminal operation of the learner are set. We acquire terminal data information and network data information and return the data to the system; collect user usage data when the learner operation occurs, the corresponding device perception information is successfully captured by the system acquisition module; and store the acquired data in a standard format. We upload the acquired data to the server, and the operation flow of the terminal data acquisition framework is shown in Figure 6.

4.1.2. Terminal Data Acquisition Development Platform and Tools. In this study, the Android system was selected as the development platform, and a set of learner terminal data acquisition system was implemented. Based on common technology architecture (MVC), Android Studio is used as the main development tool. Android Studio is an application editor developed by Google and specially developed for Android developers. The Android Studio development tool supports the flexible construction of the operating system based on Gradle and supports the automatic construction of multiple versions and the simultaneous generation of multiple APK files. Based on the above advantages, this paper uses Android Studio to develop and design the learner terminal data acquisition system.

4.1.3. Realization of Terminal Data Acquisition. The terminal data acquisition system frame includes three parts,

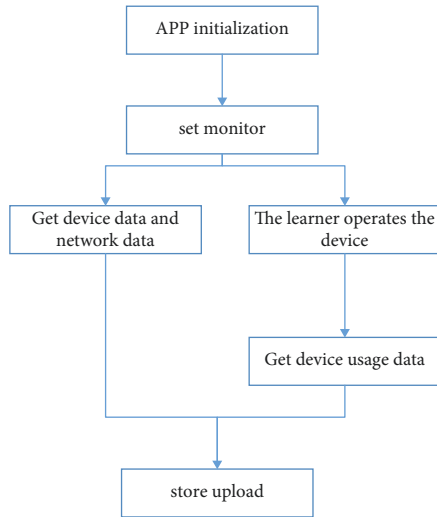


FIGURE 6: Operation flow chart of terminal data acquisition framework.

namely, terminal data, network data, and user usage data acquisition. The framework takes the sense controller class as the core, which combines the contents of device data (terminal data), network data (network data), and equipment usage data (user usage data) to work together. The core control class structure is shown in Figure 7.

- (1) The realization of terminal equipment information acquisition is shown in Figure 7. The terminal device information acquisition is based on the digital serial number and DEVICE_ID dialing provided by the Android system to the mobile phone developer for acquiring the designated mobile terminal device name symbol.
- (2) The realization of terminal power information acquisition, an information module about mobile phone power supply, is designed, and the mobile phone power supply is obtained in this way so that the learning behavior of learners on the mobile terminal can be better monitored and analyzed.
- (3) As shown in Figure 8, the application “data-Acquisition,” “rediobutton,” “qqttest,” and “Phone” are displayed as the list of APP usage in the virtual machine. The purpose of obtaining the usage of the APP is to obtain whether the learner has switched to other applications during English learning. During the learning process, multitasking is performed in parallel, which affects the concentration of learning.
- (4) The realization of unlocking the screen event acquisition is shown in Figure 9. The purpose of acquiring the unlocking screen event is to analyze the behaviors of the English learners during the learning process. By analyzing the unlocking screen state, the learner-related learning information is obtained.

4.2. Human Behavior Recognition. In this study, a new neural network is designed: a two-layer convolutional neural

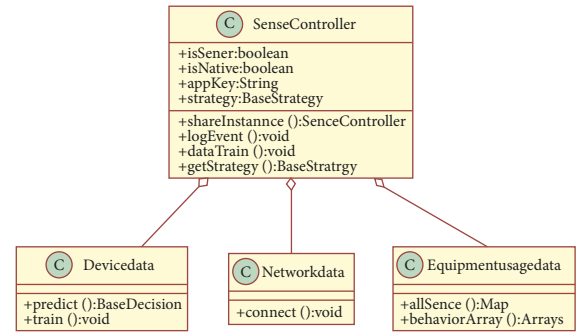


FIGURE 7: Core control class structure diagram.

network is integrated into the long short-term memory network. It only needs to send the preprocessed data to the network for training and does not need to manually extract the feature set.

The convolutional neural network plays a great role in identifying the concentration of English education learners. The low-level processing unit can obtain the local saliency of the time series, and the high-level processing unit can obtain the hidden features of the time series. In the face of time-series sensor data, convolutional neural networks need to perform convolution and pooling calculations along the time dimension, stack time-series sensor data into three-dimensional arrays, and extract time-series feature sets through convolution pooling. The long short-term memory network is a recurrent network that simulates the temporal correlation in time series problems through memory units, which can effectively obtain the temporal dependence of features.

4.2.1. Data Collection and Preprocessing. In order to obtain a form that is more conducive to learner feature extraction, the data should be preprocessed before feature extraction and model recognition. In order to fully mine the focus state information of English education learners, the data standardization is carried out by using the standard descriptive method, and the data are preprocessed by the sliding window technology of the sensor, and the data collected every second are used as a sample; that is, the sensor samples the learning behavior of the learner at a frequency of 50 Hz. When it is completed, all the collected data are sampled in a certain time dimension. When dividing, every 128 consecutive samples are taken as 1 sample. The distance between each sample and the starting position is 64 samples as a point, and each active window is described by 561 features. This study uses two-dimensional convolution operation to process time series, obtains the local dependency of sensor data time dimension, uses two-dimensional convolution to capture the spatial characteristics of English education learning behaviors, and obtains the two-dimensional matrix of learning action input according to the operation size, and the calculation formula of its dimension N is as follows:

$$N = \sqrt{m * 2 + \text{cell}(\sqrt{m})^2}, \quad (4)$$

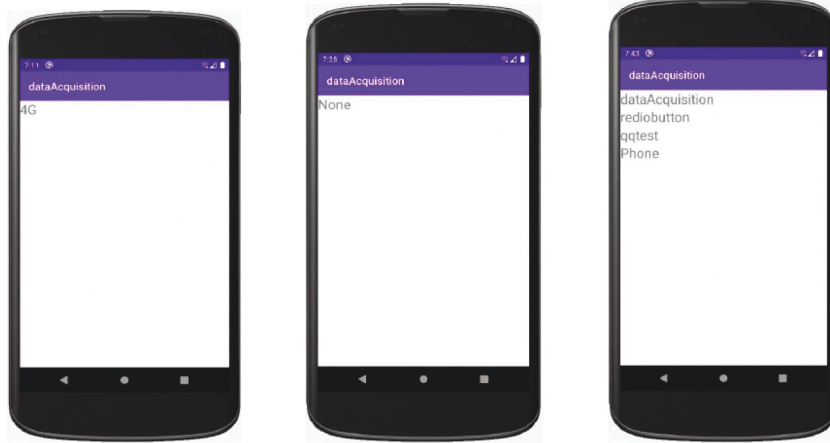


FIGURE 8: Access network type data.

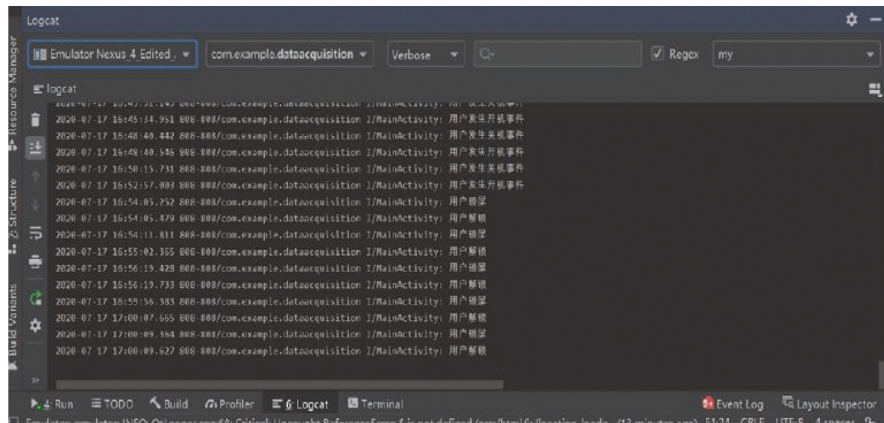


FIGURE 9: English learner learning behavior data.

where m is the number of collected data samples, and $\text{cell}()$ is rounded up. The processed time-series data are input into the neural network for training, and the algorithm model is obtained. The output layer in the model is the softmax layer, and its output is a probability distribution, so the labels of the sample data also need to appear in the form of probability distributions.

4.2.2. Learner Learning Concentration Recognition Model.

In this paper, the method of learner focus recognition is mainly based on the segmentation of various learning behaviors, and each learning behavior has a unique and refined action and category label.

Attention learning behavior recognition model takes human behavior, that is, multi-action sequence information as the input of the model, in which the multi-action output is used to predict the human behavior category of the model. The model mainly uses convolutional neural network and long short-term memory. The combined form of the network is used for human behavior recognition. The model is defined as the sequential Keras model, and a CNN model is defined with the Keras deep learning library. The network structure framework is shown in Figure 10. It mainly includes two one-dimensional convolutional layers, a dropout

layer and a pooling layer. The training of the convolutional neural network model is very fast in which the dropout layer can slow down the learning process and make the final model better, and the pooling layer shortens the learned features to $1/4$, so that it retains the most important elements. To better enable the model to learn features from the input data, we wrap the entire CNN model in a time distributed layer, allowing the same CNN model to be read in each of the four subsequences of the window. After the convolutional and pooling layers, the learned features are unrolled into a long vector, the extracted features are flattened and fed to the single-hidden layer LSTM model for reading, and the CNN-LSTM model will be in chunks. We read subsequences of the main sequence, allow the LSTM to interpret and extract features from each block, define dropout layers to reduce the model's overfitting to the training data, extract its own features before final classification of the activity, and finally, in turn, classify through fully connected layers and softmax layers.

4.2.3. Learning Behavior Recognition Model Training. In the process of fitting the CNN-LSTM model, no shuffle operation is performed on the sequence data, but the window of the input data is randomly adjusted during training, so the

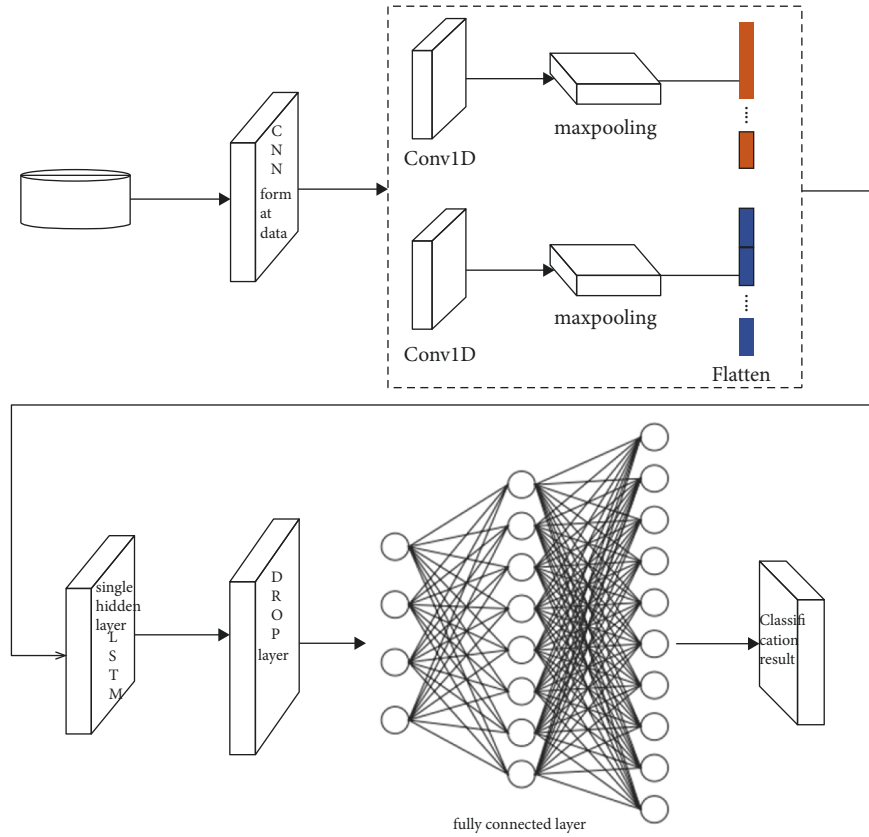


FIGURE 10: Building a neural network structure framework.

log information is not output to the standard output stream during model training. The number of iterations of the training model is 15, and the learning efficiency of the model is set to 0.1 to improve the generalization ability of the model. The Adam optimizer is used as the stochastic gradient descent algorithm to iteratively update the network weights for training data, providing a sparse stochastic gradient weight descent optimization algorithm that can effectively deal with the problem of network noise reduction. In the process of training the model, the optimization objective is set as the cross-entropy categorical_crossentropy to describe the accuracy of the model for human action recognition. Cross-entropy is defined as follows:

$$H_{y'} y = - \sum_{i=1}^N y'_i \log y_i. \quad (5)$$

Among them, y'_i and y_i represent the i -th ground-truth sample labels and the i -th predicted sample labels are the number of identified categories. After the weight matrix calculates the cross-entropy loss function, the weight matrix and the offset are corrected in turn. The neural network uses the gradient descent method to solve the minimum value of the loss function, which is defined as follows:

$$x_{\text{new}} = x_{\text{old}} - \varphi \frac{\partial f(x)}{\partial(x)}. \quad (6)$$

Among them, x_{new} and x_{old} represent the modified weight matrix and the uncorrected weight matrix,

respectively, and $f(x)$ represent the learning rate, which is a hyperparameter of the neural network. The data features and labels of the test set are input into the evaluation function run_experiment respectively, and the accuracy of the model used for the test set data is calculated. By creating and evaluating models, debug information and a sample score are printed for each model.

4.2.4. Experimental Results and Analysis. The experimental results mainly compare and analyze the improved CNN-LSTM model proposed in this paper with the machine learning hierarchical hidden Markov model, machine learning algorithm, and LSTM model algorithm, in order to verify the accuracy of the model. First, some data sets used in the experiment, experimental environment, and parameter configuration in other experiments are briefly introduced, and then, the corresponding experiments are designed according to the needs and analyzed with the comparison results.

The data of the experiment come from the machine learning knowledge base of the public website UCI. The HAR data set collects sensor data on smartphones. The data set is recorded from 30 students, who collect six learning behaviors while wearing smartphones around their waists activity. They are dictation, reading, answering, discussion, thinking, and classroom practice. After the sampling work results are detected, all samples and data are divided according to the time dimension. The analysis results are

TABLE 1: UCI data sample distribution table.

	Dictation	Read	Answer	Discuss	Think	Classroom exercises	
Training set	1253	1307	1359	1355	1086	992	7352
Test set	491	537	532	496	471	420	2947
Total	1744	1844	1891	1851	1557	1412	10299

TABLE 2: Experimental parameter settings.

Parameter name	Parameter value
Length of time sample	128
Number of times to train the model	25
Number of batches per training	64
Time step	32
The number of neurons in the hidden layer	100
Learning rate	0.1
Dropout ratio	0.5

2.56 seconds and a 50% overlapping fixed window for sampling. The segmentation of various learning behaviors is that every 128 consecutive samples are taken as a new sample, each sample data are 64 sampling points apart at its start and end positions, and each active window is described by 561 feature methods, and these data are analyzed. After normalization, the data values are between $[-1, 1]$. The entire sample set was redivided and split into two groups, with the behavioral data of 21 students as the training set and the behavioral data of 9 students as the test set. The entire UCI data set has a total of 10,299 data samples, and the specific sample distribution is shown in Table 1 [19–23].

Through the analysis of the data set, this paper continuously adjusts and tests the parameters, and obtains the optimal relevant parameters for the current training. The experiment is defined as a shallow-level CNN that extracts features from the input data, flattens the feature vector into a single-hidden-layer LSTM model, and adds a dropout layer that reduces the model's overfitting to the training data, and a dropout layer for classification. Softmax layer uses a fully connected layer to interpret the features extracted by the CNN-LSTM hidden layers before using the output layer for prediction. In the practice process of training the model, the method of stochastic gradient descent is used for training, and the optimization objective is set as the cross-entropy categorical_crossentropy to describe the accuracy of the model for classifying human actions. When performing the training operation iteratively, only a small part of the data is used for stochastic gradient descent each time, and 64 items are randomly selected from the training set as mini_batch; that is, the batch size is set to 64, and 64 data windows are inputs into the model before updating the model weights. In the CNN-LSTM model, the settings of parameters and matrix size are learned through model training, so 30% of the data set is selected as the training set to train the model to obtain the optimal parameters, and the remaining data are used as the test set to evaluate the performance of the model. The specific experimental parameter settings are shown in Table 2.

This experiment is a training test of the neural network model built based on the public data set of sensor

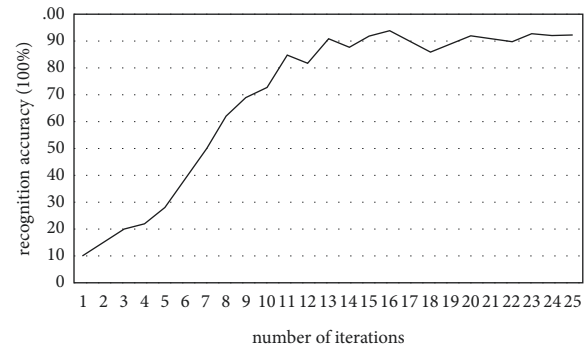


FIGURE 11: Variation curve of recognition accuracy of CNN-LSTM model training set.

data. In order to make the experimental results more accurate, the experiments are carried out under the weight parameter that obtains the highest accurate value, and the experimental results are compared with the hierarchical hidden Markov model, machine learning algorithm, and LSTM model. The recognition accuracy is used as an evaluation index for experimental comparison. The higher the accuracy, the better the performance of the model classifier.

When experimenting with the CNN-LSTM model, we set the experimental parameters. Figure 11 shows the change process of the recognition accuracy of the experimental training set based on the CNN-LSTM model. It can be seen that with the increase of the number of iterations, the recognition accuracy rate rises rapidly and remains above 0.92; the standard deviation is 0.373; and the recognition accuracy is accurate. The rate is about 92.499%.

Through the above experiments, the operation results based on the hierarchical hidden Markov model, machine learning algorithm, LSTM model, and CNN-LSTM model are obtained. The comprehensive recognition accuracy of the above experiments is summarized in Table 3. Compared with the existing research, it is found that the hierarchical hidden Markov model structure saves the storage space of the mobile phone and reduces the computational complexity, but the space complexity is expensive and the data overfitting phenomenon that may occur in the experimental classification process is not carried out. Processing. In the experiment based on machine learning algorithm, the data features selected manually after dimensionality reduction of the data by principal component analysis are still insufficient; when recognizing students' concentration behavior based on the LSTM model, the single-layer LSTM structure is used to process long-term sequences. There is a problem of taking too long to sample. Compared with the above three

TABLE 3: Comparison of experimental results.

Method	Recognition accuracy (%)
Hierarchical hidden Markov model	72.5
Machine learning algorithms	78.2
LSTM model	89.1
CNN-LSTM model	92.5

methods of human behavior recognition, the recognition accuracy of the CNN-LSTM model has increased, which indicates that the model has a higher accuracy rate and better effect in recognizing concentration behaviors in English classroom teaching.

By integrating the convolutional neural network on the basis of the LSTM model, the local saliency of the time series obtained by the low-level processing unit and the hidden features of the time series obtained by the high-level processing unit are improved. On the one hand, the model does not require manual feature extraction, which leads to deviations in the experimental results; on the other hand, the combination of convolutional neural network and long short-term memory network can make the model better learn features from the input data and achieve more effective results. *English learning concentration activity recognition*. Finally, the model is compared with the hierarchical hidden Markov model, machine learning algorithm, and LSTM model, and the models improve the recognition accuracy by nearly 19.99%, 14.29%, and 2.79%, respectively, thus verifying the effectiveness of the model.

5. Conclusion

This paper obtains learning behavior data based on the application scenario of mobile terminal learning, obtains the learning status device perception information of the learner by implementing a terminal data acquisition tool, captures the touch screen operation of the learner experiment based on the virtual simulation experiment, and then improves the CNN-LSTM by improving the CNN-LSTM. The collected terminal acceleration sensor data are processed for human behavior recognition, and the learning activity status of the learner is obtained. Finally, the influence of the learning behavior of the learner's mobile terminal on the learning concentration is analyzed by fitting the weights of various types of learner behavior data. The acquisition of learning behavior data of mobile terminal learners is the premise of analyzing the influence of learners' online learning concentration factors. The acquisition of learner behavior data provides experimental data sets for subsequent research modules such as intelligent learning guidance and personalized diagnosis.

Data Availability

The data set can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] C. Klonoff David, "New wearable computers move ahead: Google glass and smart wigs," *Journal of diabetes science and technology*, vol. 8, no. 1, 2014.
- [2] N. Takashi, M. Daisuke, I. Hiroyuki, O. Yuki, and K. Ueda, "Application of Augmented Reality (AR) Technology to Locate the Cutaneous Perforator of Anterolateral Thigh Perforator Flap: A Case Report," *Microsurgery*, vol. 44, no. 18, pp. 15–18, 2021.
- [3] C. A. Corneanu, M. O. Simon, J. F. Cohn, and S. E. Guerrero, "Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1548–1568, 2016.
- [4] S. Neha and E. Sharma, Khular and K. Amita, Hybrid meta-heuristic algorithm based deep neural network for face recognition," *Journal of Computational Science*, vol. 30, no. 11, pp. 14-15, 2021.
- [5] T. B. Moeslund, A. Hilton, V. Krüger, and L. Sigal, *Visual Analysis of Humans*, Springer, London, 2011.
- [6] A. Kadhodamohammadi, A. Gangi, M. De Mathelin, and P. Nicolas, "Articulated clinician detection using 3D pictorial structures on RGB-D data," *Medical Image Analysis*, vol. 35, 2017.
- [7] D. Case, *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*, Academia Press, Amsterdam NL, 2002.
- [8] Y. Lee and S. B. Cho, "Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D accelerometer," in *Proceedings of the International Conference on Hybrid Artificial Intelligence Systems*, pp. 460–467, Springer, Wroclaw, Poland, May 2011.
- [9] F. Marton and R. Säljö, "On qualitative differences in learning: I-outcome and process," *British Journal of Educational Psychology*, vol. 46, no. 1, pp. 4–11, 1976.
- [10] J. B. Biggs, "Student Approaches to Learning and Studying," *Australian: Australian Council for Educational*, vol. 1, pp. 8–19, 1987.
- [11] T. W. Smith and S. A. Colby, "Teaching for deep learning," *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, vol. 80, no. 5, pp. 205–210, 2007.
- [12] J. Biggs, Tang, and C. Teaching, *For Quality Learning at university: What the Student Does* pp. 214–220, Open University Press, Berkshire, UK, 3rd edition, 2007.
- [13] H. Tüzün, M. Yılmaz-Soylu, T. Karakuş, Y. İnal, and G. Kızılkaya, "The effects of computer games on primary school students' achievement and motivation in geography learning," *Computers & Education*, vol. 52, no. 1, pp. 68–77, 2009.
- [14] K. Kieser, H. Herbison, W. Waddell, K. Kardos, and I. Innes, "Learning in oral biology: a comparison between deep and surface approaches," *New Zealand Dental Journal*, vol. 102, no. 3, pp. 64–68, 2006.
- [15] H. S. Barrows, "Problem-based learning in medicine and beyond: a brief overview," *New Directions for Teaching and Learning*, vol. 68, pp. 3–12, 1996.
- [16] W. Hung, "The 9-step problem design process for problem-based learning: application of the 3C3R model," *Educational Research Review*, vol. 4, no. 2, pp. 118–141, 2009.
- [17] A. Alrahlah, "How effective the problem-based learning (PBL) in dental education. A critical review," *The Saudi Dental Journal*, vol. 28, no. 4, pp. 155–161, 2016.

- [18] B. Hoffman and D. Ritchie, "Using multimedia to overcome the problems with problem based learning," *Instructional Science*, vol. 25, no. 2, pp. 97–115, 1997.
- [19] H. S. Barrows and M. Robyn, *Tamblyn Problem-Based Learning: An Approach to Medical Education* Springer Publishing Company, Springer, Berlin, Germany, 1980.
- [20] J. R. Savery, "Overview of problem-based learning: definitions and distinctions," *Interdisciplinary Journal of Problem Based Learning*, vol. 1, no. 1, pp. 9–20, 2014.
- [21] J. Biggs and C. Tang, *Teaching for Quality Learning at University*, McGraw-Hill, New York, NY, USA, 3rd edition, 2007.
- [22] P. R. Pintrich, "The dynamic interplay of student motivation and cognition in the college classroom," in *Motivating Students: An Initial Attempt to Operationalize the Curiosity gap Model, Developments in Business Simulation and Experiential Learning*, J. G. Gentry, Ed., pp. 69–75, 2001.
- [23] N. F. Harun, K. M. Yusof, M. Z. Jamaludin, and S. A. H. S. Hassan, "Motivation in problem-based learning implementation," *Procedia-Social and Behavioral Sciences*, vol. 56, pp. 233–242, 2012.