Real-time Classroom Student Status Analysis System Based on Intelligent Behavior and Speech Recognition

Hongji Xu[a,*], Yonghui Yu[b], Renzhuo Wang[c], Xinya Li[d]

School of Information Science and Engineering, Shandong University, 72 Binhai Road, Qingdao, China

[a] hongjixu@sdu.edu.cn, [b] yuyonghui2000@163.com, [c] wangrenzhuowd@163.com, [d] lixinya0226@163.com

*Corresponding author

Keywords: Human Activity Recognition, Speech Recognition, Face Recognition, Classroom Evaluation

Abstract: The learning status of students in the classroom is a crucial basis for evaluating the quality of teaching and learning. Therefore, conducting real-time analyses of students' classroom statuses and providing timely reminders for those in negative states can significantly improve the instruction. This paper proposes a real-time student status analysis system in the classroom that utilizes intelligent behavior and speech recognition techniques based on the multimodal data analysis. This scheme offers a solution to the issue of students' lack of attention and low learning outcomes in certain classroom. It can also generate a status report of the classroom which is very valuable for educators seeking to improve the quality of instruction and modify students' learning outcomes.

1. Introduction

Classroom evaluation has a high practical value in assessing the quality of teaching and learning within educational institutions. In order to achieve effective classroom teaching, it is imperative to conduct real-time analysis of students' learning statuses. This enables timely interventions to be made to address any negative trends that may arise, thereby ensuring the effectiveness of students' classroom learning.

In the background of educational informationization, the real-time classroom monitoring based on various artificial intelligence (AI) techniques is an effective way to improve the quality of instruction in the classroom. In order to improve the teaching quality and increase the learning efficiency of students, researchers have started to use deep learning algorithms to identify classroom conditions, assess student learning status in the classroom, optimize teaching methods, and modify teaching strategies, which can recognize the learning status in natural learning contexts and has the benefits of being non-contact, non-intrusive, and easily scaleable [1]. Researchers have achieved numerous accomplishments in this field. For example, Xing Y. et al. [2] used the betapose based gesture key point recognition method and the data augmentation technique to recognize students' behaviors in the classroom, such as dozing, peeking, passing notes, and looking around. The algorithm can be well applied to multi-person scenarios to solve the occlusion problem. Zhao M. et al. [3] suggested a classroom video stream that would identify student behavior in the teaching environment and provide teacher with timely feedback by utilizing image and video processing technology, convolutional neural network algorithm, and application interface programming technology. Wang Y. et al. [4] proposed a strategy for improving algorithms, which incorporates a coordinated attention mechanism into the ghost module to enhance detection accuracy. In order to study the classroom environment, Wang G. et al. [5] created a deep learning network that can count the number of students as well as identify their physical characteristics and facial expressions. In addition, many researchers have designed the systems of behavior recognition to evaluate student behaviors in the classroom [6-7].

At present, some studies only propose algorithms for behavior recognition or speech recognition, without systematic design, which cannot provide teachers with intuitive reports on student learning status in the classroom. Many studies are based on image or video detection, which cannot realize the diverse needs of real-time detection. To address these shortcomings, this paper proposes a real-time

  DOI: 10.25236/ichess.2023.001

status analysis system for students in the classroom based on intelligent behavior and speech recognition. Teachers can adjust the teaching style or teaching rhythm by using this system and improve the quality of teaching.

2. Design for Real-time Student Status Analysis System

2.1. System Framework

The real-time status analysis system for students is based on intelligent recognition of behavior and voice interaction. It relies on hardware devices such as voice collectors, computers, and classroom monitoring systems. This system has eight modules as shown in Figure 1, which contain a multimodal information sensing and acquisition module, a multimodal data intelligence recognition module, a classroom context knowledge base module, a multimodal data analysis module, a student classroom status report output module, a student classroom status judgment module, a student classroom status indication module, and a student classroom status adjustment module. Among them, the multimodal data analysis module is the most important.

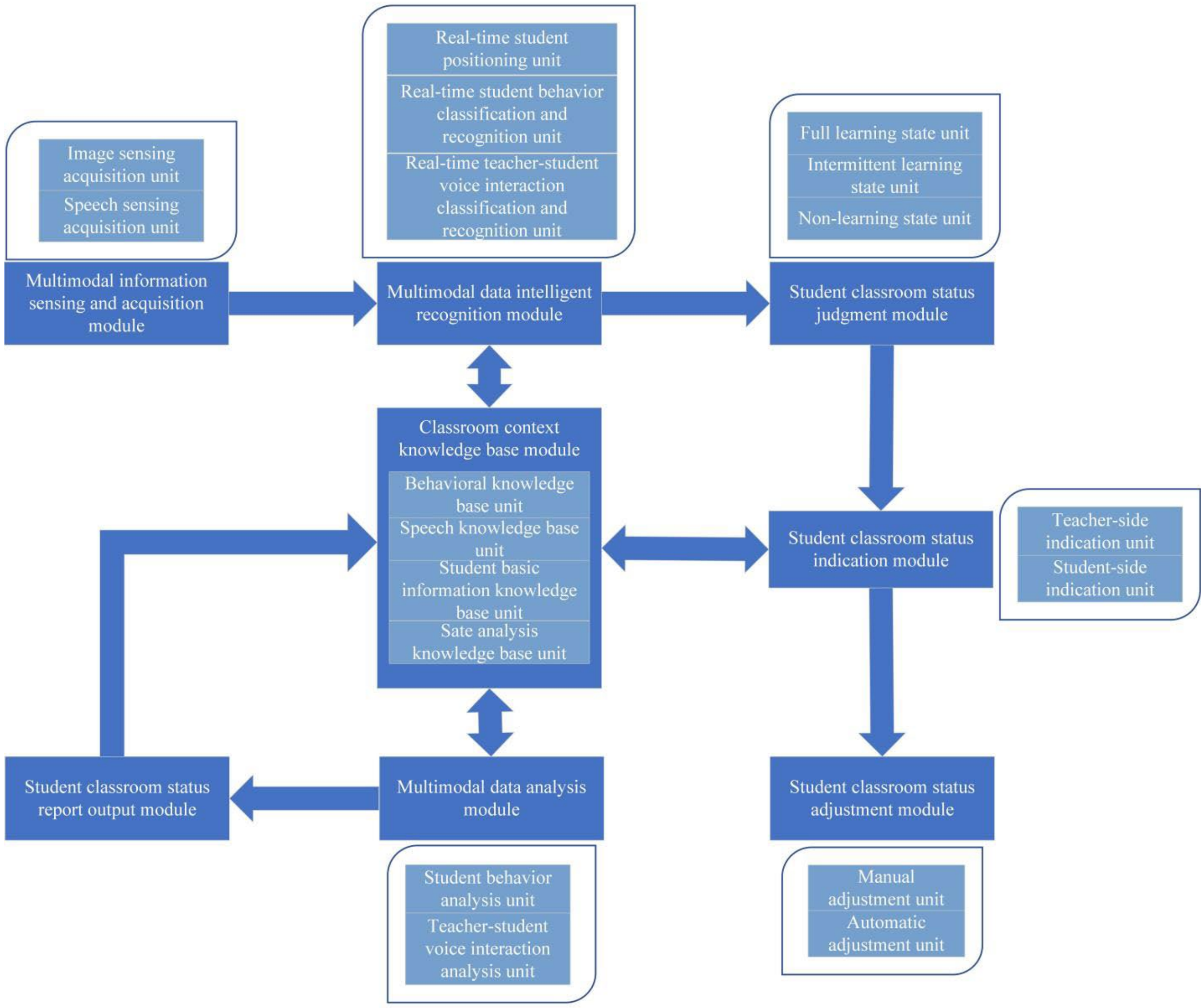

Figure 1 Framework of real-time classroom student status analysis system.

2.2. Multimodal Data Analysis Module

The multimodal data analysis module is designed to analyze and evaluate student learning status in the classroom. This module is composed of a student behavior analysis unit and a teacher-student voice interaction analysis unit, which work together to provide comprehensive insights into student performance.

The student behavior analysis unit is designed to evaluate the behaviors of students in the classroom. It accomplishes this by counting the learning behaviors and non-learning behaviors exhibited by each student throughout the teaching process, as well as their durations. The unit analyzes the proportion of different behaviors at different times in the classroom during the learning process and displays the learning status of each student. Students' behaviors can be assessed using the following three metrics.

The effective learning behavior index of individual student is defined as:

$$I_{si} = \frac{T_{si\_1} + T_{si\_2} + T_{si\_3} + \cdots + T_{si\_j} \cdots + T_{si\_N_{bi}}}{T} = \frac{\sum_{j=1}^{N_{bi}} T_{si\_j}}{T} \quad (1)$$

where I_{si} represents the index of effective learning behavior of the ith individual student, T_{si_j} represents the duration of time spent by the jth learning behavior of the ith student in the classroom, and N_{bi} represents the total number of times that learning behavior within the classroom for the ith student.

The average effective learning behavior index of students in the classroom is defined as:

$$I_{se} = \frac{I_{s1} + I_{s2} + I_{s3} + \cdots + I_{si} \cdots + I_{sN_s}}{N_s} = \frac{\sum_{i=1}^{N_s} I_{si}}{N_s} \quad (2)$$

where I_{se} represents the average effective learning behavior index of students in the classroom, I_{si} represents the effective learning behavior index of individual student of the ith student, and N_s represents the total number of students participating in the classroom.

The instantaneous efficiency index of student learning behavior is defined as:

$$E_{sb} = \frac{N_{ss}}{N_s} \quad (3)$$

where E_{sb} represents the instantaneous efficiency index of student learning behavior in the classroom, N_{ss} represents the number of students with learning behavior at a given moment, and N_s represents the total number of students participating in the classroom.

The teacher-student voice interaction analysis unit utilizes the duration and number of teacher-student voice interactions to evaluate the degree of teacher-student voice interaction, using the following three metrics to assess.

The duration of teacher-student voice interaction is defined as:

$$T_{si\_t} = T_{si\_t1} + T_{si\_t2} + T_{si\_t3} + \cdots + T_{si\_tj} + \cdots + T_{si\_tN_{vi}} \quad (4)$$

where T_{si_t} represents the total time of teacher-student voice interaction in the classroom of the ith student, T_{si_tj} represents the duration of the jth voice interaction of the ith student with the teacher, $j = 1, 2, \cdots, N_{vi}$, and N_{vi} represents the number of voice interactions of the ith student with the teacher. It can be seen that the duration of teacher-student voice interaction in the classroom and the index of individual effective learning behaviors obtained by (1) can be combined to analyze the classroom performance of the student.

The total time of teacher-student voice interaction in the classroom is defined as:

$$T_{s\_t} = T_{s1\_t} + T_{s2\_t} + T_{s3\_t} + \cdots + T_{si\_t} + \cdots + T_{sN_s\_t} \quad (5)$$

where T_{s_t} represents the total time of the teacher-student voice interaction in the classroom, T_{si_t} represents the total time of the ith student voice interaction with the teacher, and $i = 1, 2, \cdots, N_s$.

During the calculation, the repeated time is not considered. It can be seen that the total time of the teacher-student voice interaction in the classroom and the average effective learning behavior index of classroom students obtained by (2) can be combined to analyze the learning performance of the whole classroom.

The efficiency of teacher-student voice interaction in the classroom is defined as:

$$E_{st} = \frac{T_{s\_t}}{T} \tag{6}$$

where E_{st} represents the efficiency of teacher-student voice interaction, T_{s_t} represents the total time of classroom teacher-student voice interaction and T represents the duration of the classroom. The efficiency of teacher-student voice interaction in the classroom and the classroom instantaneous student behavioral efficiency obtained by (3) can be combined to analyze the learning efficiency of the whole classroom.

3. Workflow for Real-time Classroom Student Status Analysis System

3.1. General Workflow

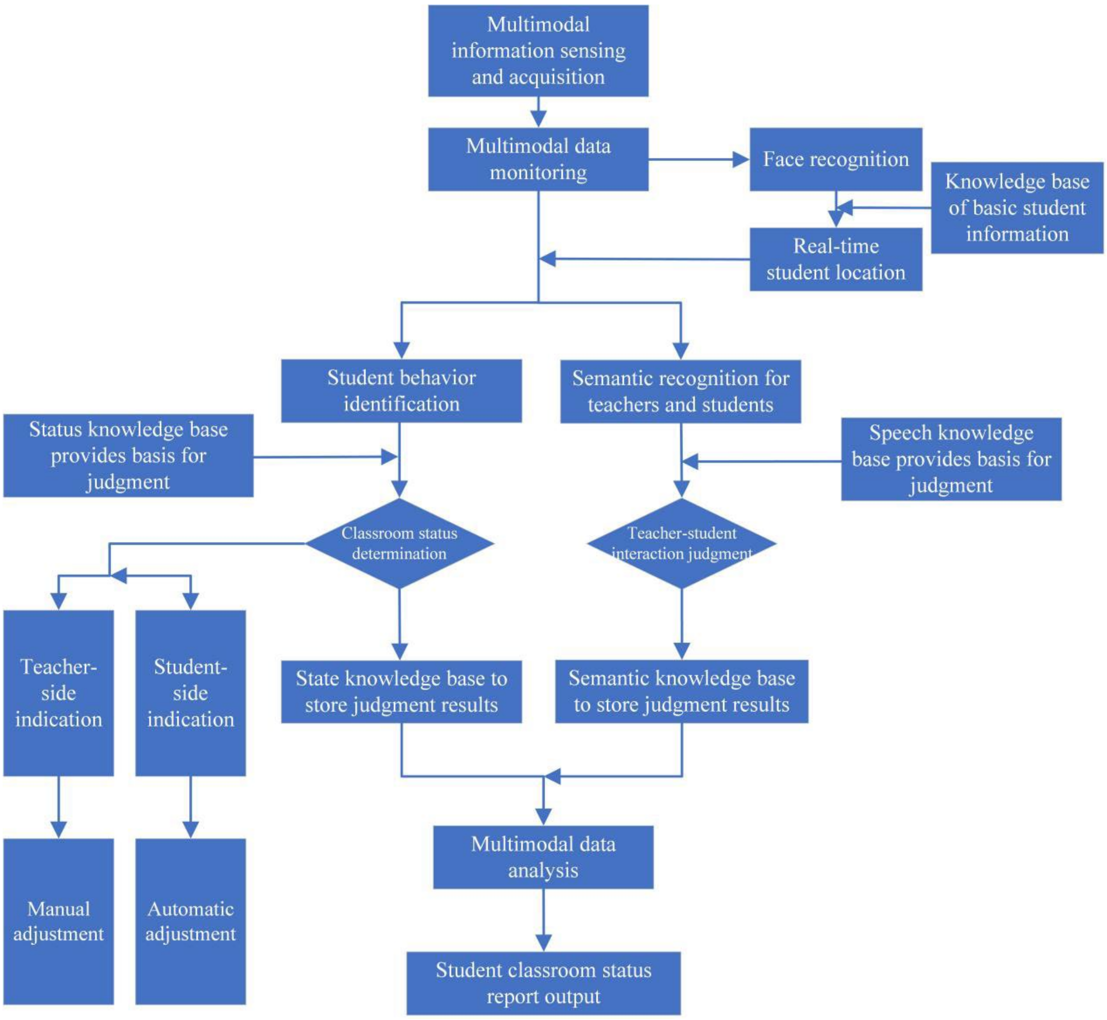

Figure 2 Workflow of the real-time classroom student status analysis system.

Figure 2 shows the workflow of the proposed real-time classroom student status analysis system. It can be seen that the work consists of four parts: data acquisition, data preprocessing, classroom status judgment and data result analysis. The four modules work in collaboration, enabling real-time analysis and indication reminders for classroom students' statuses, and generating the final teaching report. Among these four modules, the classroom status judgment module is the most important one.

3.2. Classroom Status Judgment

3.2.1. Student Behavior Recognition and Teacher-student Semantic Recognition

Classroom status judgment mainly involves the recognition of student behavior recognition and the semantic recognition of teacher-student.

Real-time classification and identification of student behavior is based on image data acquisition, which provides data for the subsequent analysis of student behavior. Some intelligent information processing techniques would be used here, for instance, the features of student behavior can be obtained from the scale-invariant feature transform (SIFT) features and directional gradient histogram (HOG) features of the students' classroom behaviors information obtained in the two-dimensional image after data preprocessing.

The synchronously collected voice data is uploaded to the database of servers for storage. Semantic recognition technology is used to automatically identify the voice interaction signals of teachers and students, and then record the time points and durations of the identified teacher-student voice interactions. This type of semantic recognition can distinguish various types of voice interactions, including learning voice interactions and non-learning voice interactions.

3.2.2. Classroom Status Judgment and Teacher-student Interaction Judgment

The classroom status of a student is determined based on his behaviors by utilizing the classifier, which would also combine with the state context knowledge base to determine what kind of learning status it is.

It is recognized as the full learning state, if the student is with a learning behavior within a specified duration (such as 3 minutes) by means of the data collection and the classification recognition; It is recognized as the intermittent learning state if the student is with both a learning behavior and a non-learning behavior within a specified duration; It is determined to be the non-learning behavior if the student is with a non-learning behavior within a specified duration.

Meanwhile, the teacher-student voice interaction status can be determined by utilizing the voice data in the classroom, combined with the judgment basis of teacher-student voice interaction in the voice knowledge base unit. If it is determined to be teacher-student voice interaction, it will be stored in the system, and if it is not, it will be ignored.

3.2.3. Student Classroom Status Indication and Storage

The detected students' behaviors are carried out to indicate the students' classroom statuses in the teacher indication end and the student indication ends, according to the judgments of students' classroom statuses and teacher-student interaction statuses. At the same time, the realized judgment results of students' classroom statuses and teacher-student voice interactions are respectively uploaded and stored in the state context knowledge base and voice knowledge base to provide data for subsequent multimodal data analysis.

The teacher's instruction end may rely on a touchable liquid crystal display (LCD) screen that indicates specific information. The LCD monitor may display the name, learning status, identification (ID) photo as well as the real-time image frame of the student at a certain seat. When the corresponding name is indicated in green, yellow, or red, it indicates that the student is in the full learning state, the intermittent learning state, and the non-learning state, respectively. The student indicator end is a light emitting diode (LED) indicator that will be mounted on the student's desk, and displays red, yellow, and green, just like the teacher's instruction end and can serve as a reminder.

4. Experimental Analysis of the Proposed System

Two experiments are carried out based on the system designed in this paper.

For experiment 1, the number of students in the classroom is 30 and the duration of this class is 50 minutes. Upon analysis through the proposed system, it is determined that out of the 30 students, 19 are in full learning state, 8 are in intermittent learning state and 3 are in non-learning state. For experiment 2, the relevant settings of it are kept consistent with those of experiment 1. Upon analysis, it is determined that out of the 30 students, 11 are in full learning state, 6 are in intermittent learning state and 13 are in non-learning state.

Figure 3 shows the learning behavior durations of the students in the classroom, where the horizontal axis represents the student number in both experiments, while the vertical axis represents the learning durations of each student in the classroom. In Figure 4, the horizontal axis represents the student number, while the vertical axis represents the voice interaction durations between each student and the teacher. By using these data, we can calculate the average effective learning behavior index of students, the total time of teacher-student voice interaction, and the efficiency of teacher-student voice interaction in the classroom.

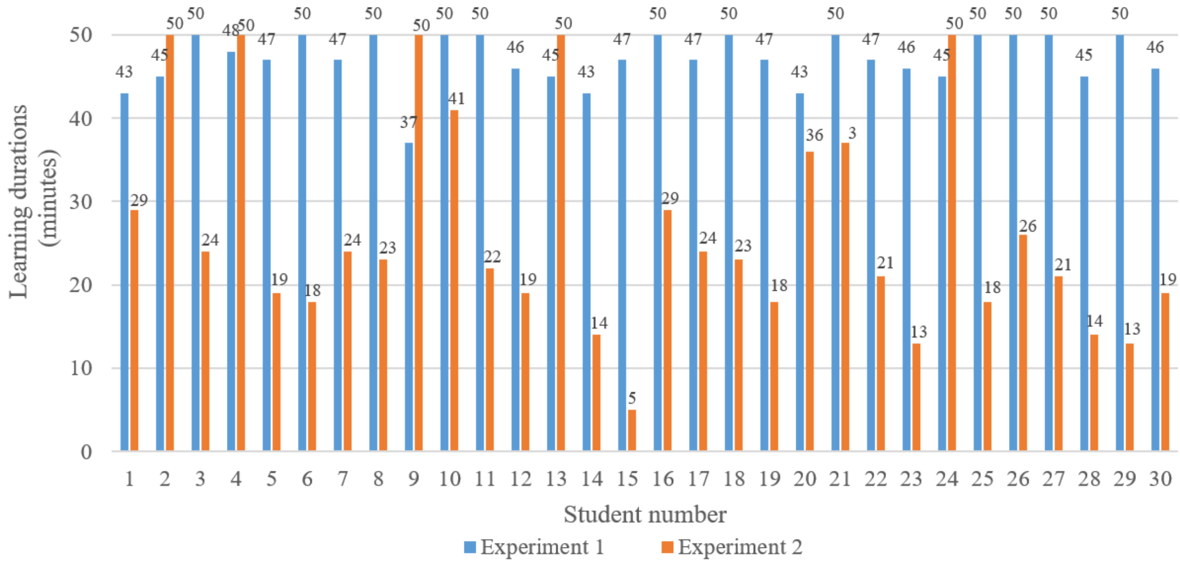

Figure 3 The learning behavior durations of the students in the classroom.

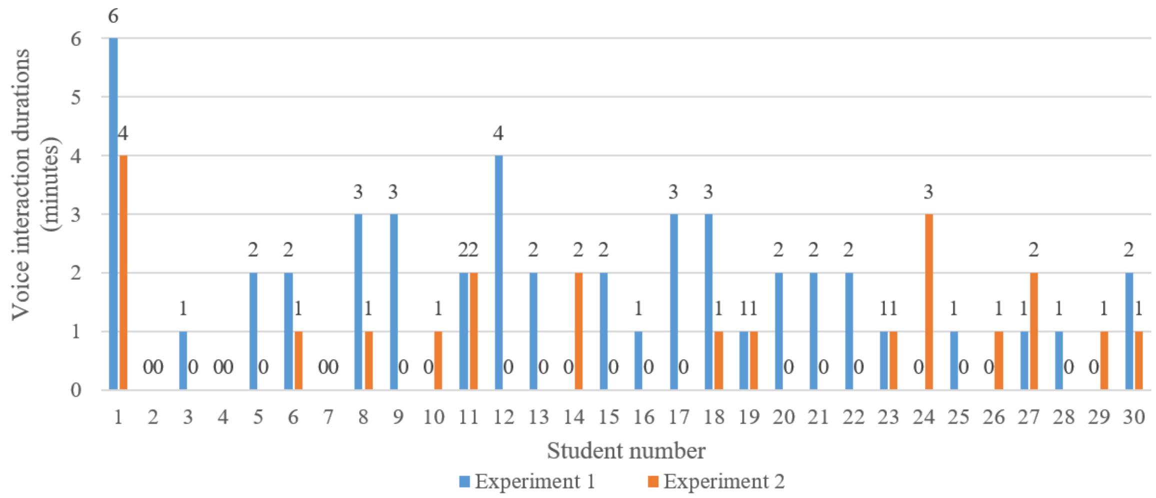

Figure 4 The voice interaction durations of the students in the classroom.

Table 1 Evaluation indices of the classroom.

Evaluation indices	Experiment 1	Experiment 2
Effective learning behavior index of individual student (1st student)	86%	58%
Average effective learning behavior index of students in the classroom	94%	53%
Instantaneous efficiency index of student learning behavior	90%	57%
Duration of teacher-student voice interaction (1st student)	6 minutes	4 minutes
Total time of teacher-student voice interaction in the classroom	47 minutes	22 minutes
Efficiency of teacher-student voice interaction in the classroom	94%	44%

The indices obtained from the two experiments related to students' classroom learning statuses and teacher-student voice interactions are shown in Table 1. The effective learning behavior index of individual student and duration of teacher-student voice interaction takes the first student as an example.

For experiment 1, the effective learning behavior index of most students and the instantaneous efficiency index of student learning behavior are maintained at a high level. The average effective learning behavior index of students in the classroom reaches 94%, which can be observed that the students in this classroom exhibited a high level of concentration throughout the entire classroom. Based on the teacher-student voice interaction duration of each student, it is evident that there are variations in the duration of voice interaction among different students. Nonetheless, the total time of teacher-student voice interaction is 47 minutes and the efficiency of teacher-student voice interaction is 94% which represents the high quality of teacher-student communication and interaction in the classroom. For experiment 2, the average effective learning behavior index of students in the classroom is 53%, which means that the overall learning levels of the students required to be improved. Furthermore, the total time of teacher-student voice interaction is 22 minutes, and the efficiency of teacher-student voice interaction is 44%, which are both lower than those of experiment 1. Based on the analysis, it appears that many students in experiment 2 require prompt reminders from teachers to enhance their learning efficiency.

5. Conclusion

For the needs of high-quality teaching and learning, a real-time status analysis system for students in the classroom based on intelligent behavior and speech recognition is proposed in this paper. This system realizes the data analysis of student's classroom performance based on the combination of behavior and teacher-student voice interaction, making the analysis results more comprehensive. The system can also generate a status report which is very helpful to both students and teachers.

Future research can consider further expanding the types of multimodal data, for instance, emotion is an important factor to achieve a comprehensive perception. Therefore, physiological information such as skin electrical response and heart variability rate collected by smart bracelets can be added, so we can analyze students' classroom behaviors emotionally.

Acknowledgements

This work was financially supported by the 2021 Undergraduate Teaching Reform Research Project of Shandong Province of China (M2021228).

References

[1] Wei, Y., Luo, J., Hu, M., Li, W. and Yao, H. (2022) Online Learning Emotion Recognition Based on Videos. Computer Science, 49, 105-110.

[2] Yan, X., Kuang, Y., Bai, G. and Li, Y. (2023) Student Classroom Behavior Recognition Method Based on Deep Learning. Computer Engineering, 49, 251-258.

[3] Zhao, M., Hu, H. and Liu, H. (2023) Design of Classroom Behavior Monitoring System Based on Target Detection. Technology Innovation and Application, 13, 35-38.

[4] Wang, Y. and Ma, Z. (2023) Research on Classroom Behavior Recognition of Students Based on Deep Learning. Software Engineering, 26, 40-43+62.

[5] Wang, G., Zhang, X. and Zheng, H. (2022) Analysis of Students' Classroom Learning Status Based on Deep Learning. Journal of Higher Education, 8, 1-5.

[6] Liu, X. (2023) Design of Student Classroom Learning Behavior Monitoring System Based on Artificial Intelligence Technology. China New Telecommunications, 25, 65-67.

[7] Hu, F., Wang, X. and Bu, C. (2023) Research on Classroom Deep Learning Performance Prediction Model Based on Behavior Recognition. Chinese Journal of ICT in Education, 29, 108-118.