

# *Analysis of Spatiotemporal Characteristics of Student Concentration Based on Emotion Evolution*

Tieming Xiang, Hongjian Ji\*, Jing Sheng

*Xiamen University of Technology, Xiamen, China*

*\*Corresponding author*

**Keywords:** Emotion recognition; spatiotemporal characteristics; virus transmission theory; teaching; computer vision

**Abstract:** Detecting the concentration of students in the classroom can help teachers quickly understand the participation and activity of students. However, the concentration of students has complex spatiotemporal distribution and evolution laws, which is challenging to identify and quantify. This paper proposes a novel student concentration evaluation method based on emotional evolution and virus transmission, which analyzes the spatiotemporal characteristics of concentration. The research contents are as follows: (1) A visual emotion classification method based on deep learning algorithm is developed to identify and quantify the emotion changes of each student. (2) On the basis of quantification results of emotion, the concentration index model with introducing the theory of virus transmission is established and further used to explore the spread of student concentration in spatiotemporal dimensions. (3) The Wilcoxon rank sum test (RST) is used to verify the difference of the results calculated by concentration index model in different semesters, and the reliability of the model can be reflected by the Pearson correlation coefficient between the centroid of the spatiotemporal distribution of concentration and final exam results. The experiments of 64 offline courses have been carried out in a same class for two semesters, and the results show that the concentration of student in the spatial dimension can be affected by negative and positive emotions from different regions, while in the temporal dimension, the high concentration level will decrease with increase of course time, and the generation speed of this phenomenon will be further exacerbated after coupling the spatial factors.

## 1. Introduction

In the current research field of concentration evaluation, many researchers use sensor signal of Electroencephalogram (EEG) to detect the degree of concentration [1]. Although this method has been proven to be effective, it requires the subject to wear a fixed monitoring instrument. For actual demand scenarios, it is unrealistic to realize the application of this method without any psychological resistance and additional costs. In order to achieve the purpose of non-intrusive concentration detection, image recognition technology is employed to classify the apparent properties which is used as key data features for concentration analysis, such as eye movement, gesture, head posture and the associated emotions [2,3]. This kind of method is also an important

part of the mainstream biological information acquisition system.

Emotion is an organic collection of a variety of apparent properties and maintains a strong correlation with concentration, and different combinations of apparent properties can be used to express specific emotions [4]. In the field of education, this research achievement has been successfully applied to the intelligent tutoring system, which enhances student concentration by observing the emotion changes in real time and taking individualized one-to-one counseling activities [5]. In addition, refining the combination of apparent features can obtain a more accurate emotion classification state, which can be further integrated with machine learning algorithms, such as the TAN Bayesian classifier reported by Singh [6], to characterize the concentration for study. The above researches show the feasibility of using the results of emotion classification to enhancing concentration and analyzing its influencing factors, however, the inconsistent apparent properties used for emotion classification make the mapping relationship between concentration and factors diverse, which is difficult to be applied in synchronous teaching environment with single acquisition equipment [7]. To solve this problem, D'mello et al. divided student emotions into confusion, happiness, frustration and boredom through facial expression correlation test [8], the conclusion of which also provides a reference for subsequent evaluation of the relationship between emotion and concentration. The unified mapping relationship between apparent property and emotion classification can achieve stable emotion recognition model after coupling computer vision algorithm and machine learning algorithm, such as Convolutional Neural Networks [9], Optical Flow Method [10], Local Binary Pattern [11], Long Short Memory [12] and OpenFace Framework [13]. This kind of model takes the items of facial expression feature and emotion classification as the input parameters in the model training process. While meeting the needs of non-intrusive detection, it can quantify the level of emotion and provide a reference data for the association and mining work of concentration.

The flow of emotion in the spatiotemporal dimension can be theoretically used to analyze the evolution laws of student concentration and the influence mechanism of its transmission process [14]. The existing research mainly focuses on the form of investigation report and expert evaluation [15], which is feasible for a single student, but for a wide range of group forms, it is difficult to obtain a continuity conclusion on the strength and direction of concentration. The emotion recognition method developed in combination with computing technology is an objective and efficient method to recognize the concentration, such as emotion and concentration statistical system based on LabVIEW [16], and real-time analysis system of facial expression and concentration [17]. Although such methods can achieve the purpose of instantaneous evaluation of individual concentration, the influence factors are inclined to be static [18], which seems not applicable to the situation that both emotion and concentration are dynamically transmitted and interacted in the spatiotemporal dimension, especially in the classroom that pays attention to information transmission and communication.[19] Recently, some studies began to realize that the existence of interaction phenomenon is the main factor affecting the change of concentration, and started to analyze the correlation between concentration and its dynamic evolution law, but there is still a lack of systematic summary [20].

In the present paper, spatiotemporal characteristics based on emotion evolution and virus transmission have been for the first time applied for evaluation of student concentration. Emotion evolution based on facial expression can be used to recognize quantitative concentration in complex interactive environment, which avoid the generation of psychological resistance. Besides, the introduction of virus transmission makes emotion data flow and affect each other in the spatiotemporal dimension, which makes it possible to the dynamic evaluation of concentration. The rest of this paper is organized as follows. In Sect.2, the concentration index model integrated by emotion evolution and virus transmission are explained in detail. Section 3 describes the impact of

the changes of student concentration and its influence factors in spatiotemporal dimension. Sect.4 outlines conclusions.

## 2. Method

The concentration index model with spatiotemporal characteristics is established. The emotion classification results are used as parameters input the model, and the results calculated by the model are further verified by the Wilcoxon rank sum test and Pearson correlation analysis. The related methods are described in detail in this section.

### 2.1. Emotion Classification

Emotion classification is realized by establishing the correlation between facial expression features of student and the emotion state. The high-definition camera captures facial expressions and forms a image set  $V = \{I_1, I_2, \dots, I_i, \dots, I_n\}$ , where  $I_i$  is the  $i$ -th frame image in video  $V$ , and the  $n$  is the total number of frames per second. As the initial input of emotion classification method, image set  $V$  provides data source for face detection and emotion classification, and the flow chart of this process is shown in Figure 1.

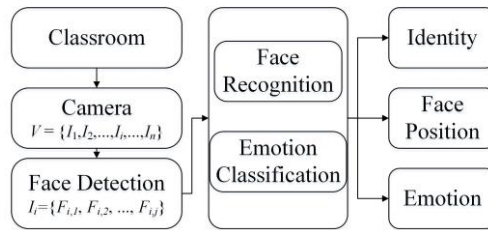


Figure 1: The workflow of visual emotion recognition system.

Face recognition: The deep learning algorithm of Multi-Task Cascaded Convolutional Networks (MTCNN) is used to detect faces in image set  $V$ , which has been successfully applied in the systems of student attendance and school security [21]. Compared with other depth learning algorithm, the MTCNN reduces the number of convolution kernel and increases the depth of each network through a three-layer convolution network to achieve fast and accurate face detection and alignment. Since the algorithm is only used for data acquisition, the algorithm can be constructed reference to the article report of MTCNN [22] to avoid repetitive discussion. The face image set  $F$  and the related feature vector  $f$  can be obtained by inputting the image  $I_i$  to MTCNN. The key feature points of the face, such as eyes, tip of nose, corner of mouth, eyebrows and other contour points, are located by calculating vector  $f$ , and the face alignment is completed by affine transformation. The alignment results are compared with the faces in the training model [23] to achieve the purpose of face recognition.

Emotion classification: The forms of emotion in the cognitive range are generally divided into classified emotion state and dimensional emotion space, which correspond to discrete and continuous information expression respectively [24]. In the classified emotion states, emotions are composed of a series of states. These states will appear on the face and naturally form a specific expression on person without professional training. In the dimensional emotion space, emotion is usually defined as a point in  $n$ -dimensional space, which is similar to the Pleasure-Arousal-Dominance (PAD) space [25] or the Valence-Arousal (VA) space [26]. Although the dimensional emotion space is flexible in describing the continuous evolution of emotion, it is often unable to intuitively explain the instantaneous emotional state, which undoubtedly increases the difficulty of real time analysis. The interpretability of discrete emotional state is beneficial to modeling, which can be defined as happiness, surprise, neutrality, sadness, disgust, fear and anger.

The parameters of convolutional neural network Res-Net50 [27] in deep learning algorithm are adjusted and compared with FER2013 data set [28], which contains 28709 emotion training sets and 3589 test sets. The artificial accuracy of the test is  $65 \pm 5\%$ , while the accuracy of the training model in this paper is 70.2%. It is considered that the algorithm can provide intermediate variables for the subsequent emotion evolution analysis. When face image set  $F_i$  are input into the algorithm, the set  $S$  of different emotions will be output, which includes the probability  $p_e(t)$  of emotion at time  $t$ . The emotion recognition result with the highest accuracy probability is the expression recognition result. The relationship between emotion with different weights and dummy variables is shown in Table 1.

Table 1: The weights for seven facial expression

Facial expression	Happy	Surprise	Neutral	Angry	Sad	Fear	Disgust
$e$	1	2	3	4	5	6	7
$W_e$	7	6	5	4	3	2	1

## 2.2. Spatiotemporal Concentration Index Model

The classification of concentration is actually the clustering result of different types of emotions in high-order space, which needs the combination of weighted emotion calculation and threshold division [7]. Previous studies have regarded the problem of concentration recognition as a binary classification problem, and divided it into positive and negative states. Hence, the expression of concentration score and concentration state can be expressed as (Li et al., 2020)[7].

$$g(t_m) = \frac{\sum_{e=1}^k (W_e p_e(t_m))}{(W_{e,max} - W_{e,min})} \quad (1)$$

$$C(S_x, t) = \begin{cases} P & g(t_m) > \alpha \\ N & g(t_m) < \alpha \end{cases} \quad (2)$$

Where  $g(t_m)$  represents the concentration score at  $t_m$  moment,  $C(S_x, t)$  represents the concentration state of student  $x$  at time  $t$ ,  $p_e(t_m)$  represents the probability of obtaining emotion  $e$  at  $t_m$ ,  $k$  represents the number of emotions actually judged,  $W_{e,max}$  and  $W_{e,min}$  represent the maximum weight and minimum weight of each emotion at  $t_m$  moment respectively,  $P$  and  $N$  represent positive and negative states respectively,  $\alpha$  represents the threshold of the partition.

In practical observation, we find that the change of concentration has gradient change characteristics within a certain time range. Obviously, it is difficult to characterize this phenomenon by the two classification methods of Eq.(1) and Eq.(2). In order to characterize the transition of concentration from one state to another, it is necessary to subdivide the concentration state and convert the transition interval between each state into low stability states. The low stability states are easy to influence and change each other, while far away from this state is the high stability states. Hence, the average of concentration scores  $g'(t_m)$  in a certain period of time  $t$  is used to represent the time transition of state, and each state is divided into high and low steady state. The Eq.(1) and Eq.(2) can be improved as follows

$$g'(t_m) = \frac{\left( \sum_{t=t_m-\Delta t}^{t_m} g(t) \right)}{count(g(t_m))} \quad (3)$$

$$C'(S_x, t) = \begin{cases} HP & g'(t_m) > \alpha_1 \\ LP & \alpha_2 < g'(t_m) < \alpha_1 \\ LN & \alpha_3 < g'(t_m) < \alpha_2 \\ HN & g'(t_m) < \alpha_3 \end{cases} \quad (4)$$

Where,  $C'(S_x(t))$  is the subdivided concentration state.  $\Delta t$  is the time interval.  $LP$  and  $LN$  are the low positive and negative states.  $HP$  and  $HN$  are the high positive and negative states.

We tested the subdivided concentration state  $C'(S_x(t))$  in a same class and show the results in Figure 2. The recognition ability of Eq.(3) and Eq.(4) in concentration is verified, more importantly, the time transition effect of subdivided  $LP$  and  $LN$  means that there is a certain transmission law between  $HP$  and  $HN$ . For example, from Figure 2(a) to Figure 2(f), the slight change of  $HP$  and  $HN$  in quantity is accompanied by the significant gradient change between  $LN$  and  $LP$  in time. The change process of states is difficult to be explained mathematically.

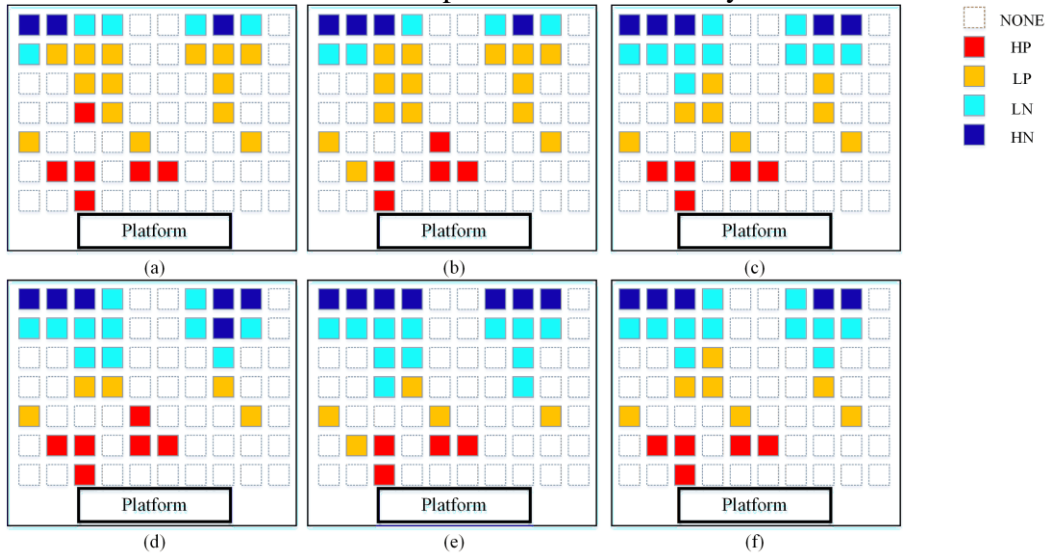


Figure 2: Changes of concentration state in a same class.

Due to the lack of quantitative expression of transmission law, the state change process is difficult to be explained mathematically. In order to solve this problem, the questionnaire [7] is adopted to collect other influence factors. The questionnaire results show that attention of surveyed student is usually affected by the facial expression, voice, action of students in adjacent seats, while students far away from the surveyed students are not mentioned in the questionnaire. Combining the questionnaire results with Figure 2, it can be found that there is an obvious dividing line between  $HP$  and  $HN$ . The high steady states are difficult to approach each other, and can only rely on the low steady state to carry out gradient transmission in space. Assuming that  $HN$  is a harmful emotion and  $HP$  is a healthy emotion. The process of transmission between them through  $LP$  and  $LN$  shows a similar process with the virus transmission in the spatiotemporal dimension, which also provides a theoretical basis for the introduction of virus transmission theory.

There are three common virus transmission models: single population model [29], compound population model [30] and micro individual model [31] [32]. For this kind of model is mainly aimed at the mobile scene, it cannot be used directly in a relatively certain classroom environment. However, the common factors contained in the virus transmission model can be used for reference, which are the immune population ( $IP$ ) and susceptible population ( $SP$ ).  $IP$  is analogous to  $HP$  and  $HN$  in the concentration state, while  $SP$  corresponds to  $LP$  and  $LN$ . Based on the virus transmission theory, the transmission scenario within the effective range  $D$  is established to simulate the diffusion

and evolution of concentration among student groups, as shown in Figure 3.

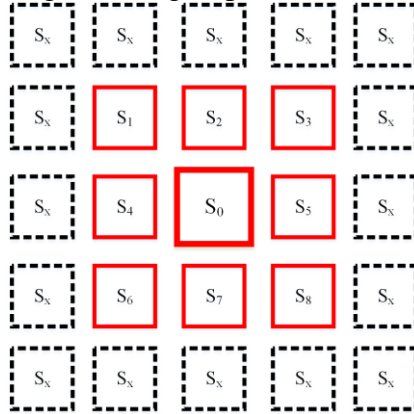


Figure 3: Maximum affected range of S0.

The model takes student  $S_0$  as the center, and the solid border is the maximum affected range  $D$  of the student.  $\sigma$  represents the intensity at which students in a high steady state are affected,  $\delta$  represents the intensity at which students in a low steady state are affected. The impact intensity comes from the virus transmission model SIRS [33], in which,  $\sigma$  is used to describe the transition rate from non susceptible type to susceptible type, which is usually set to 0.01-0.6, and  $\delta$  is used to describe the transformation rate from susceptible type to non susceptible type, which is usually set as 0.6-1. For the concentration of students, the influencing factors  $\sigma$  is less than the value of the real virus index and can be set within 0.01-0.3. The value of Influence factor  $\delta$  is set between 0.3-1. The concentration index model  $C_I(S_x(t))$  considers the dynamic transmission and interaction of different students' concentration in the space-time dimension, as shown in Eq.(5).

$$C_I(S_x(t)) = \begin{cases} (g(t) \pm \sigma_x g'(t))T_l & \text{for } (C(S_x(t)) = HP, C(S_x(t)) = HN) \\ (g(t) \pm \delta_x g'(t))T_l & \text{for } (C(S_x(t)) = LP, C(S_x(t)) = LN) \\ 0 & \text{for } C_I(S_x(t)) < 0 \end{cases} \quad (5)$$

Where  $C_I(S_x(t))$  is the concentration index of Student  $x$  in time  $t$ ,  $T_l$  is normalization parameters.

### 2.3. Verification Method

It is assumed that learning environment for student is stable for a certain period of time, which is the prerequisite for the analysis of spatiotemporal characteristics of concentration. The experiment is conducted in 2 semesters, and the results of spatiotemporal concentration index model are divided into two groups independent sample,  $n_1$  and  $n_2$ , according to semesters. The accuracy of the model is verified by the sample difference and correlation test between independent samples and semester grades.

Common test methods for testing the sample difference include T-test [34], Wilcoxon rank sum test [35] (RST) and Wilcoxon signed rank test [36]. Among them, the T-test needs the sample to meet the condition of normal distribution and homogeneity of variance, and the Wilcoxon signed rank test ignores the difference between the observed value and the data center position. Considering that the concentration index in the spatiotemporal dimension needs to be quantified by the cluster center, the RST is used as the test method, which is shown in Eq.(6).



$$Z = \frac{T - \frac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 \times n_2 (n_1 + n_2 + 1)}{12}}} \quad (6)$$

where,  $T$  is the rank sum of smaller samples,  $Z$  is the test statistic. The significance level  $P$  is 0.05. When  $p < 0.05$ , the assumption that the calculation results in two semesters are independent samples is rejected, which means there is no significant difference in the concentration index between two semesters. This is obviously inconsistent with the reality and shows that the model is inaccurate.

The center point of spatiotemporal concentration index is found through clustering algorithm, which is used to quantify the student concentration in a certain period of time. In order to reflect the practical law of the positive correlation between concentration and course score, the accuracy of the model needs to be tested by the Pearson correlation between the center point and the course score. Therefore, the spatiotemporal concentration model can be considered to be feasible for practical application when meeting the conditions of rank sum test and correlation analysis at the same time.

### 3. Results

#### 3.1. Experiment Setup

The experiment is conducted in 2 semesters with a total of 64 lessons. Unified classrooms are adopted and students are advised to reduce the number of seat exchanges between classes. The HD camera is used as the acquisition equipment of facial features, the parameter of which is  $1920 \times 1080$  resolution and 2 megapixel. The acquisition frame rate is set to be 3 frames per second to avoid a large number of duplicate images in set  $V$ . The same class with a total of 28 students is selected to participate in the experiment, and before the experiment, we will inform all students and get their consent. There is no deliberate interference and arrangement in the experiment environment, which is consistent with the actual situation of normal class. The experiment environment is shown in Figure 4.

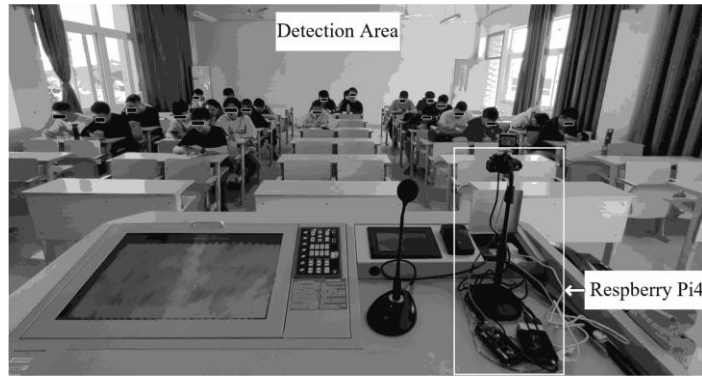


Figure 4: Classroom environment

#### 3.2. Analysis of Spatiotemporal Characteristics of Student Concentration

According to the flow chart in Figure 1, each image  $I_i$  in the image set  $V$  captured by HD camera is input into the module of Face recognition to recognize the  $j$ -th student corresponding to the face image set  $F_{i,j}$ . The  $F_{i,j}$  is sorted by time series and input into the module of Emotion classification, and the subdivided concentration states  $C'(S_x(t))$  are obtained by Eq.(4).

In the spatial dimension, the classroom is divided into 6 areas, which corresponds to the free choice behavior of student based on the attention inertia before class. The division area of the

classroom is shown in Figure 5. The students identified in Figure 6 are marked in turn according to the divided areas, and the selection frequency of students in each area is counted, which is shown in Figure 7. There are obvious spatial differences in the selection frequency of divided areas. In the front area of the classroom (close to the blackboard), such as Area1, Area2 and Area3, students who have selected these areas more than 50 times rarely appear in Area4, Area5 and Area6 in the back, and vice versa. This shows that the concentration inertia of some students in LP and LN states is more difficult to change than that of other students, which also reflects the polarization of student performance.

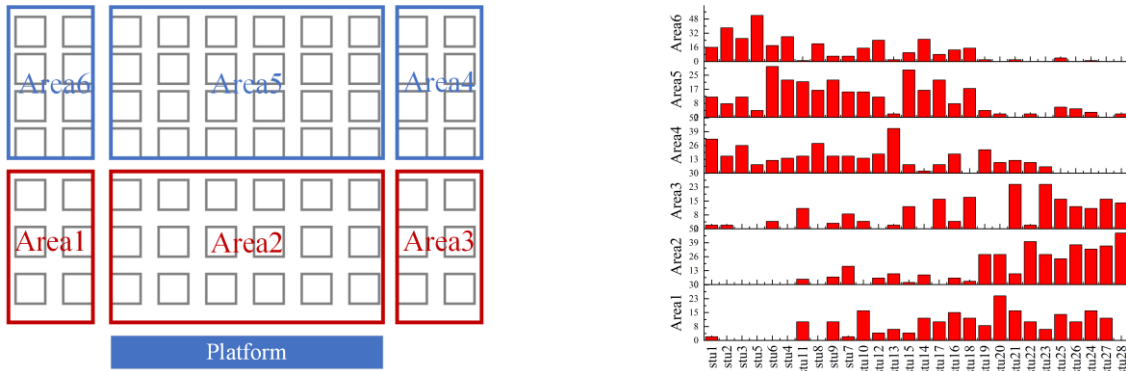
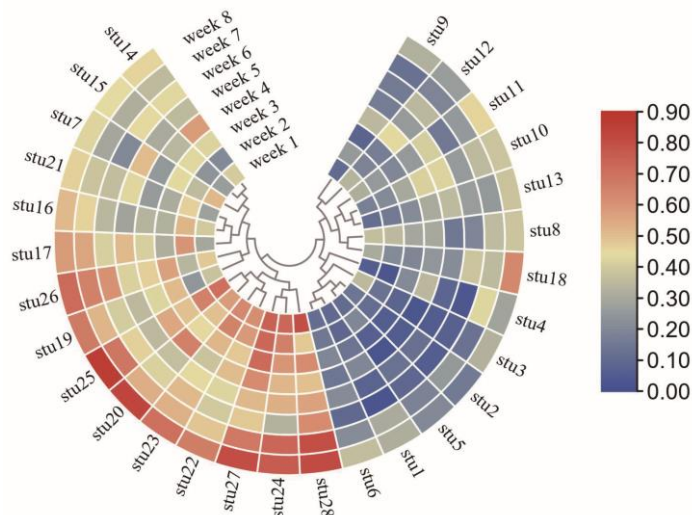


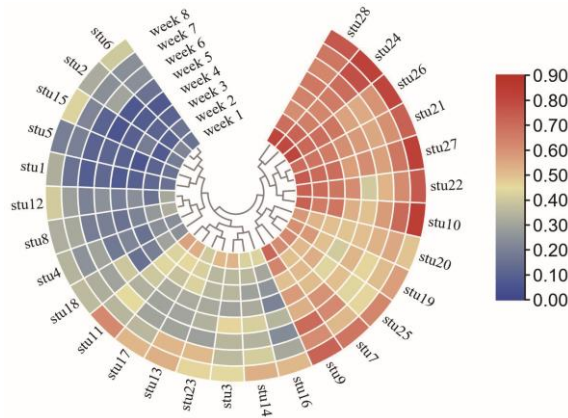
Figure 5: Division of six areas in the class. Figure 6: Selection frequency of students in each area.

Cluster analysis is carried out on the state score  $g'(t_m)$  corresponding to the subdivided concentration states  $C'(S_x(t))$  in Figure 4. The cluster results are shown in Figure 7. The concentration scores of students are clearly divided into two major categories with the color mark of 0.4, which reflects the high steady state and low steady state for concentration. As the threshold of concentration score  $g'(t_m)$  clustering, the color mark can provide threshold in Eq.(4), where  $\alpha_2 = 0.4$ . The clustering results also confirm the significance of the existence of the binary classification state and its threshold  $\alpha$  in Eq.(1). In each main category, two secondary categories are divided by color thresholds of 0.7 and 0.15, which correspond to the threshold of  $\alpha_1$  and  $\alpha_3$  respectively in Eq.(4). Combined with the results of area selection in Figure 6, students in each area are classified into the category of concentration state, and the spatial distribution of concentration state can be shown in Table 2.



a) Average concentration score of each student per week in the first semester





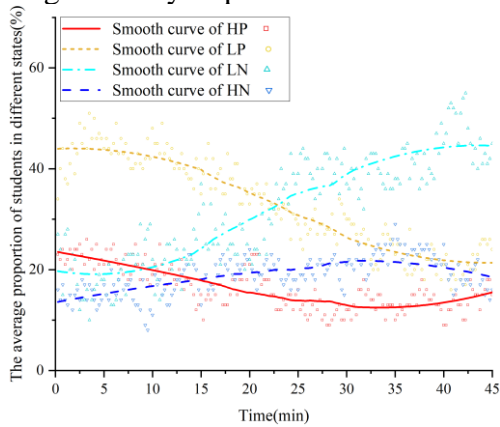
b) Average concentration score of each student per week in the second semester

Figure 7: Average concentration score.

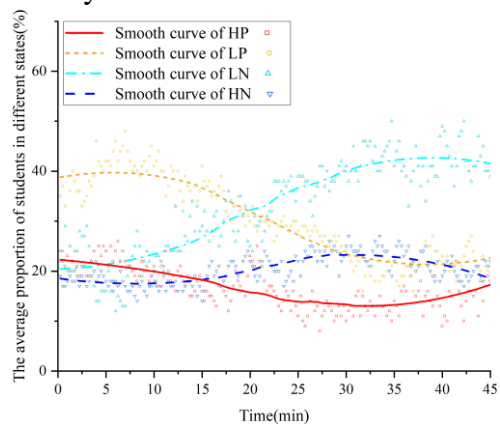
Table 2: The spatial distribution of concentration state

Area	Concentration state High steady state		Concentration state Low steady state	
	HP	HN	LP	LN
Area1, Area2, Area3	Stu28,Stu27,Stu24, Stu26,Stu25,Stu23	-	Stu22,Stu21,Stu20, Stu19,Stu17,Stu16, Stu10	Stu18
Area4, Area5, Area6	-	Stu1,Stu2,Stu3, Stu4,Stu5,Stu6	Stu7,Stu9	Stu14,Stu15,Stu13, Stu12,Stu8,Stu11

In the temporal dimension, the concentration state  $C'(S_x(t))$  of each student is calculated and classified according to the time series. In order to intuitively show the distribution of student concentration in the classroom, the curve of the proportion of concentration state with time is adapted and shown in Figure 8. The number of students in the states of HP and LP decreases gradually with time, while the number of LN and HN states increases. After 17 minutes of class in Figure 5(a) and 20 minutes of class in Figure 5(b), the proportion of students in LN and HN states is gradually greater than that in HP and LP states. The curves in two semesters are similar, which reveals that the concentration state has a certain inertia in the temporal dimension. When the inertia is broken, it is difficult to significantly improve concentration without any intervention.



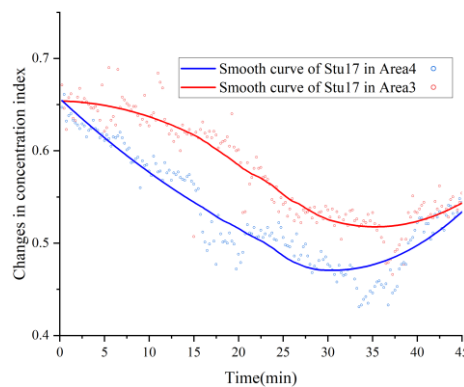
a) Changes in the proportion of students with different concentration states in the 32 lessons in the first semester



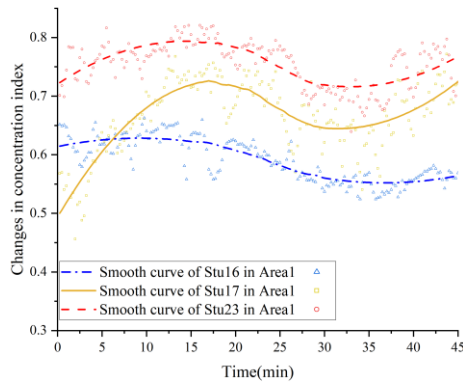
b) Changes in the proportion of students with different concentration states in the 32 lessons in the second semester

Figure 8: Changes in concentration of different students within two semesters.

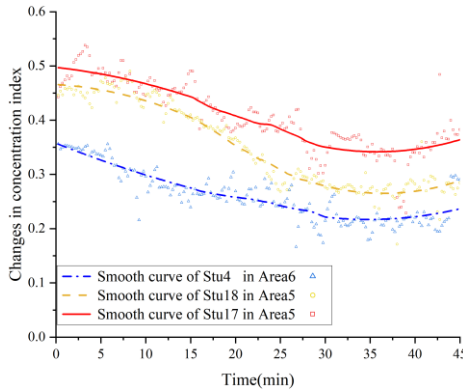
After coupling the spatial factors, the concentration index  $C_I(S_x(t))$  can reflect the influence of concentration flow with time among students in different spatial locations. It is proposed in  $C'(S_x(t))$  that the performance of students in high steady state in class is more stable than that of students in low steady state, so the concentration index of students in low steady state can effectively reflect the impact of broadcasting. In Figure 9, the concentration index of each student decreased significantly after 10-15 minutes of class. In Figure 9(a), stu17 has no other students within the effective influence range D, and the concentration index in both classes decreases significantly with time flow, and the index in region 4 is significantly smaller than that in area3. In Figure 9(b), stu16 with low steady state and stu23 with high steady state are within the effective influence range D of stu17 with low steady state where both of them are in area1. The influence can be described by using the difference of the index. The smaller the difference, the greater the influence. The difference of the index between stu17 and stu23 gradually decreases from 0.19 to about 0.07, and the difference between stu17 and stu23 increases from -0.1 to about 0.11 which is greater than 0.07, so it is determined that the influence of student with high steady state will be greater than that of students with low stable state. Similarly, in Figure 9(c), after 20 minutes, the difference in the concentration index of stu18 with low steady state and stu4 with high steady state rapidly decreases to 0.02, which indicates that stu18 is greatly affected by stu4, and there is a cross regional impact on the concentration of stu18 in area5 and stu4 in area6 during this time period. In Figure 9(d), the exponential difference between stu17 with low steady state and stu8 with low steady state continues to decrease, and the arrival time of the lowest exponential value of stu17 is 10 minutes faster than that of stu17 in Figure 9(a). Therefore, stu8 exacerbates the reduction of stu17 index. To sum up, students with low steady state will be guided by the state of students in high steady state within their effective range D, and over time, such impact can be cross regional, while students in high steady state are not easy to be affected. When there are no high stable students in the range of D, the interaction between low stable students will often accelerate the trend of negative state. In addition, we note that the average index in area1 and area3 is generally higher than that in area4, area5 and area6, which means that students close to the podium are more focused than students away from the podium.



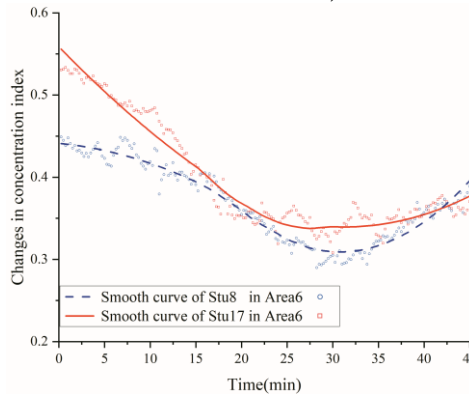
a) Changes in concentration index of stu17 in two lessons



b) Changes in concentration index of stu17, stu16 and stu23 in one lesson



c) Changes in concentration index of stu17, stu18 and stu4 in one lesson



d) Changes in concentration index of stu17 and stu8 in one lesson

Figure 9: The changes in the degree of attention of adjacent students in the classroom.

### 3.3. Verification

The verification of the experiment will be described from two aspects: the stability and reliability of the algorithm. Among them, Wilcoxon rank sum test is used to verify the stability of the algorithm, and the hypothesis is put forward: for the experiment in these two semesters, we assume that there is no obvious difference in students' concentration index in the same environment and the same teacher in different two periods (each student's study habits will not change greatly). For this reason, the significance level we selected is 0.05, that is, the original hypothesis is rejected when the p value is less than 0.05. As is shown in Table 3, there is no significant change in the concentration index of most students. Although the P-value of stu7, stu9, stu19 and stu21 is obviously less than 0.05, we attribute such abnormalities to the imperfection of the algorithm and the change of personal learning state. Except for the students' own factors and the influence of abnormal

interference events, the p-values of other students are in line with the hypothesis, which fully proves the stability of the algorithm. Table 4 gives the Pearson correlation coefficient R of students' weekly average concentration index  $C_i$  and each students' final exam performance  $R_i$ . It can be seen that the R value of each week is greater than 0.75, and most of the weeks are greater than 0.8, which means that the results of the algorithm show a high correlation with students' scores. Among them, the average concentration index in the eighth week is higher than that in other weeks, which is usually understood as that the teacher will explain and summarize the knowledge in the last week, and the concentration is higher than usual. Therefore, it can be determined that the concentration index value can reflect the learning status of students.

Table 3: P value of each student's concentration comparison in two semesters

Student	Stu1	Stu2	Stu3	Stu4	Stu5	Stu6	Stu7
P-value	0.924	0.401	0.327	0.263	0.484	0.344	0.025*
	Stu8	Stu9	Stu10	Stu11	Stu12	Stu13	Stu14
P-value	0.674	0.012*	0.237	0.889	0.208	0.779	0.233
	Stu15	Stu16	Stu17	Stu18	Stu19	Stu20	Stu21
P-value	0.575	0.208	0.216	0.093	0.012*	0.674	0.021*
	Stu22	Stu23	Stu24	Stu25	Stu26	Stu27	Stu28
P-value	0.211	0.208	0.335	0.285	0.069	0.455	0.575

\* indicates that P-value is less than 0.05.

Table 4: Correlation coefficient R value between concentration index and final exam score

Semester 1	Week1	Week2	Week3	Week4	Week5	Week6	Week7	Week8	average
R-value	0.802	0.759	0.834	0.756	0.778	0.868	0.824	0.875	0.917
Semester 2	Week1	Week2	Week3	Week4	Week5	Week6	Week7	Week8	average
R-value	0.811	0.83	0.761	0.784	0.809	0.834	0.851	0.864	0.893

## 4. Conclusions

Aiming at the problem that the concentration of student is difficult to quantify in the classroom, this paper proposes a method for evaluating the concentration of student based on emotional evolution and virus transmission, and the following conclusions are obtained:

(1) In the classroom, the proportion of the number of people who maintain a positive focus state in the total number decreases with the flow of time, and changes significantly after 10 minutes in the classroom, while the number of people who maintain a negative focus state continues to increase, and reaches the highest between 30-35 minutes. This effectively reflects the temporal law of students' learning state.

(2) From the perspective of the whole semester, the spatial distribution of students often presents gradient characteristics. Students with high concentration often choose the area close to the podium, students with low concentration often choose the back area, and other students are distributed between the two types of students. And for students in different areas, their concentration also has a cross-regional impact.

(3) The students can be divided into two categories according to the state of concentration: high steady state and low steady state. Among them, the change of concentration index of students in low steady state is no more than 0.3, while the change of concentration index of students in high steady state is no more than 0.2. Within the effective influence range D of students in low steady state, students in high steady state have a greater impact on them than students in low steady state. At the same time, the time flow across regions will aggravate the spatial gradient distribution of

concentration state.

## Acknowledgment

The authors would like to acknowledge the financial support from Fujian Science and Technology Planning Project (Grant No. 2021H0027).

## References

- [1] Lim S., Yeo M., & Yoon G. (2019). Comparison between concentration and immersion based on EEG analysis. *Sensors*, 19, 7, 1669.
- [2] Noroozi F., Corneanu C. Kamińska, D. Sapinski T. Escalera S. & Anbarjafari G. (2018). Survey on emotional body gesture recognition. *IEEE transactions on affective computing*, 12, 2, 505-523.
- [3] Kar P., Chattopadhyay S., & Chakraborty S. (2020). Gestatten: Estimation of User's Attention in Mobile MOOCs From Eye Gaze and Gaze Gesture Tracking. *Proceedings of the ACM on Human-Computer Interaction*, 4(EICS), 1-32.
- [4] Guo X., Zhou J., & Xu T. (2018). Evaluation of teaching effectiveness based on classroom micro-expression recognition. *International Journal of Performability Engineering*, 14, 11, 2877.
- [5] Kulik J. A., & Fletcher J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of educational research*, 86, 1, 42-78.
- [6] Singh A., Karanam S., & Kumar D. (2013). Constructive learning for human-robot interaction. *IEEE Potentials*, 32, 4, 13-19.
- [7] Li J., Shi D., Tumnark P., & Xu H. (2020). A system for real-time intervention in negative emotional contagion in a smart classroom deployed under edge computing service infrastructure. *Peer-to-Peer Networking and Applications*, 13, 1706-1719.
- [8] D' Mello S., Picard, R. W., & Graesser A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, 22, 4, 53-61.
- [9] Gu J. X., Wang Z. H., Kuen J., Ma L. Y., Shahroudy A., Shuai B., Liu T., Wang X. X., Wang G, Cai J. F., & Chen T. H. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, 77, 354-377.
- [10] Gibson J. J. (1951). *The Perception of the Visual World*. *The American Journal of Psychology*, 64, 3, 43-55.
- [11] Lis A. M., Black K. M., Korn H., & Nordin M. (2007). Association between sitting and occupational LBP. *European spine journal*, 16, 2, 283-298.
- [12] Greff K., Srivastava R. K., Koutn k J., Steunebrink B. R., & Schmidhuber J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28, 10, 2222-2232.
- [13] Monkaresi H., Bosch N., Calvo R. A., & D'Mello S. K. (2016). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, 8, 1, 15-28.
- [14] Kim J. K., & Kang Y. (2016). The effects of smart-phone use experience on emotion regulation and attention concentration of young children. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 6, 4, 97-109.
- [15] Olausson D., & Vallmark J. J. (2014). Emotion, concentration and performance in a Swedish female handball team: Development and implementation of the team specific instrument. *24, 11, 33-39*.
- [16] Yao K. C., Huang W. T., Chen T. Y., Wu C. C., & Ho W. S. (2022). Establishing an intelligent emotion analysis system for long-term care application based on LabVIEW. *Sustainability*, 14, 14, 8932.
- [17] Ko B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors*, 18, 2, 401.
- [18] Šuvakov M., Garci A. D., Schweitzer F., & Tadić B. (2012). Agent-based simulations of emotion spreading in online social networks. *arXiv preprint arXiv*, 1205. 6278.
- [19] Larison K. D. (2022). *On Beyond Constructivism: Using Intersubjective Approaches to Promote Learning in the Science Classroom*. *Science & Education*, 31, 1, 213-239.
- [20] Teixeira T., Wedel M., & Pieters R. (2012). Emotion-induced engagement in internet video advertisements. *Journal of marketing research*, 49, 2, 144-159.
- [21] Zhang K., Zhang Z., Li Z., & Qiao Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23, 10, 1499-1503.
- [22] Wang Z., Cui S., Wang X., & Tian J. F. (2022). Research on Facial Expression Capture Based on Two-Stage Neural Network. *Cmc-computers Materials & Coutinua*, 72, 3, 4709-4725.
- [23] Schroff F., Kalenichenko, D. & Philbin J. (2015). Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 815-823.
- [24] Ekman P. (1992). An argument for basic emotions. *Cognition & emotion*, 6, 3-4, 169-200.
- [25] Russell J. A., & Mehrabian A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in*

*Personality*, 11, 3, 273-294.

- [26] Nicolaou M. A., Gunes H., & Pantic M. (2011). Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Transactions on Affective Computing*, 2, 2, 92-105.
- [27] He K., Zhang X., Ren S., & Sun J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770-778.
- [28] Khairuddin Y., & Chen Z. (2021). Facial emotion recognition: State of the art performance on FER2013. *arXiv preprint arXiv*, 2105. 03588.
- [29] Kermack W. O., & McKendrick A. G. (1991). Contributions to the mathematical theory of epidemics--I. 1927. *Bulletin of mathematical biology*, 53, 1-2, 33-55.
- [30] Grenfell B., & Harwood J. (1997). (Meta) population dynamics of infectious diseases. *Trends in ecology & evolution*, 12, 10, 395-399.
- [31] Wallinga J., Edmunds, W. J., & Kretzschmar M. (1999). Perspective: human contact patterns and the spread of airborne infectious diseases. *Trends in Microbiology*, 7, 9, 372-377.
- [32] Liljeros F., Edling C. R., & Amaral L. A. N. (2003). Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and infection*, 5, 2, 189-196.
- [33] Feng L. P., Wang H. B., & Feng S. Q. (2011). Computer Network Virus Propagation Model Based on Biology Principle. *Jisuanji Gongcheng/ Computer Engineering*, 37, 11, 298-321.
- [34] De Winter, J. C. F. (2013). Using the Student's *t*-test with extremely small sample sizes. *Practical Assessment Research and Evaluation*, 18, 1, 10-12.
- [35] Dewan I., & Rao B. L. S. P. (2005). Wilcoxon-signed rank test for associated sequences. *Statistics & probability letters*, 71, 2, 131-142.
- [36] Perolat J., Couso I., Loquin K., & Straussa O. (2015). Generalizing the Wilcoxon rank-sum test for interval data. *International Journal of Approximate Reasoning*, 56, 46, 108-121.