

Multi-Feature Fusion for E-Learning Based Student Concentration Analysis

Changjian Zhou¹, He Jia², Jing Xing¹, and Yunfu Liang^{1,*}

Abstract

Since the outbreak of COVID-19, the hybrid teaching style, combining online and offline methods, has evolved into a normal pedagogical approach. In offline classrooms, teachers may pay attention to the state of students and observe whether they are listening attentively, to adjust the teaching process in time. However, in the E-learning environment, teachers are hindered by their inability to access students' states in time. Particularly, it is challenging to find out whether students are distracted in class. Although there are various existing student concentration analysis models, the metrics, such as convenience and accuracy, of these models often fail to meet the expected requirements. To address these obstacles, a multi-feature fusion method is proposed for E-learning-based student concentration analysis in this work. In this study the 300 questionnaires were collected and seven factor features were summarized. To establish the experimental dataset, 2,000 video clips were acquired, and each one was labelled with one of the five-class concentration degree scores. Finally, the ResNet-50 deep learning model with multilayer perceptron layer was employed for training and fine-tuning. Experimental results demonstrated that the proposed method achieves 0.88 accuracy, outperforming the existing state-of-the-art concentration analysis methods. The proposed method is designed to detect distracted students and to provide reference for teachers to adjust E-learning arrangements, which is of great application value.

Keywords

Artificial Intelligence, Concentration Analysis, Educational Technology, Multi-Feature Fusion, Online Learning

1. Introduction

As the coronavirus disease 2019 (COVID-19) pandemic has spreading all over the world, E-learning has become one of the most prevalent teaching approaches [1]. However, it seems that people are not fully prepared to deal with this transformation. Different from offline teaching styles, E-learning faces various challenges [2]. Due to lack of body language and eye contact in E-learning classroom, it is difficult for teachers to obtain students' statuses in real time like offline teaching environment. Particularly, teachers cannot detect whether students are listening carefully in time, and it is difficult to obtain objective feedback on the effect of online learning. To date, numerous studies have made great efforts to address these problems such as computer vision-based student concentration analysis approach [3], 3D estimated points-based method [4], micro-expression recognition algorithm [5], and so on. These works greatly improve our understanding regarding students' fatigue detection or distraction in offline learning.

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received June 21, 2022; first revision November 28, 2022; second revision January 5, 2023; third revision February 17, 2023; accepted May 4, 2023.

*Corresponding Author: Yunfu Liang (yfliang@neau.edu.cn)

¹ Department of Modern Educational Technology, Northeast Agricultural University, Harbin, China (zhouchangjian@neau.edu.cn, xingjing@neau.edu.cn, yfliang@neau.edu.cn)

² School of Electrical and Information, Northeast Agricultural University, Harbin, China (jiahe@neau.edu.cn)

However, in the E-learning environment, the existing methods suffer from the following limitations. Firstly, there are limited research achievements focusing on student concentration analysis through E-learning, therefore, the teachers struggling to obtain students' class status to adjust teaching progress in time. Secondly, most of the existing E-learning concentration analysis methods rely on the Internet of Things (IoT) sensor signals such as electroencephalogram (EEG), electromyogram, electrocardiogram, electrooculogram, and other indicators. These signal detection methods need to be equipped with professional equipment, which lacks operability in the process of home online teaching under the background of normalization of epidemic prevention and control. Finally, most students' concentration analysis methods are based on existing conventional machine learning models, or a single feature, which cannot reflect students' distracted state comprehensively.

To tackle these challenges, the multi-feature-based student concentration analysis method was proposed in this study. The deep convolutional network was adopted for face detection and seven factors that affecting concentration such as eye closure duration, eye closure times, yawns times, nods times, nod duration, number of turns and tuning time are investigated. To determine the weights of these factors, this work collected 2,000 video clips in E-learning classes, and extracted seven factor scores for each video clip. Finally, each video clip was labelled by concentration score $X = [0,1,2,3,4]$, where a higher score indicates stronger concentration. Thus, the dataset containing 2,000 labeled instances was established, each instance contains the score of the factor features; the total dataset was categorized into five classes, such as worse concentration, poor concentration, concentration, good concentration, and better concentration. The state-of-the-art deep learning model such as the state-of-the-art ResNet-50 deep learning model was employed for training, and the trained model was utilized for students' concentration prediction.

The rest of this paper is organized as follows. Section 2 describes the related work. Section 3 details the proposed methodology. Section 4 provides the experiment and results. Section 5 presents the discussion and Section 6 gives the conclusion.

2. Related Work

This study presents a multi-feature fusion method for E-learning-based student concentration analysis, which aiming to achieve a better performance than the existing models. The proposed method integrates three approaches such as the conventional concentration analysis methods, facial landmark localization and 3D head pose estimation. A brief review informing the proposed methodology is given as follows.

2.1 Conventional Concentration Analysis Methods

The conventional methods of student concentration analysis are stated in this section. Dimri et al. [3] proposed a computer vision-based application to identify students who are not concentrating in class. The authors collected the instances containing students who are concentrating or not, and input them into the linear classifier to judge whether students listen carefully. This is very interesting, but it is difficult to make a comprehensive analysis of students' various states in classroom. Lopez Esquivel et al. [4] proposed a new attention deficit detection system, which adopted low-weight machine learning algorithm for remote applications such as online learning. The authors integrated face detection and facial attribute recognition features to predict whether the people were in a distracted state. However, this method adopts

3D estimated points, which are unable to identify the bow or twist state. The micro-expression features are particular interest to researchers in recent years, which are widely used in driver fatigue detection, polygraph test and psychological research. In addition, there are also numerous achievements in the field of students' expression recognition. Pei and Shan [5] developed a micro-expression recognition algorithm for student concentration analysis in offline learning. The authors adopted convolutional neural networks to detect face optical-flow, and extracted micro-expression features. The algorithm was verified to be reasonable in real-world teaching classroom. Chiu et al. [6] analyzed the link between facial micro-expression states and learning, and proposed a facial recognition model to predict the likelihood of student conceptual change. Zheng [7] proposed a machine learning based face fatigue expression recognition. The author collected and fused the facial feature point data to establish an optimal facial feature point dataset. Finally, the fatigue characteristic analysis and modeling were carried out. The single core support vector machine was employed to construct the fatigue detection model. Yang et al. [8] designed a real-time monitoring system of human fatigue state based on face recognition technology. In this work, human eyes were further recognized based on face recognition technology, and the human fatigue state was judged by the changes of human eye contour. These methods can be used for real-time and accurate monitoring students' fatigue state in class. In addition, with the development of IoT and sensor technology, there are various sensor-based methods for student distraction analysis proposed. Yao et al. [9] proposed a prefrontal single channel EEG based method to detect fatigue state. The recognition accuracy of fatigue section of laboratory data based on the above method was 96.80%, and the false positive rate was 3.35%. Huang et al. [10] proposed an infrared image-based fatigue state recognition method. In this work, the AdaBoost algorithm was used to locate the human eye region, and combined with the characteristics of "bright pupil effect". The mesh method was used to calculate the closure of the infrared image after binarization and edge detection, and the human eye closure was obtained. Finally, according to the closure calculation results, the double thresholds were set and combined with PERCLOS criterion to judge the human eye feature state.

2.2 Facial Landmark Localization

The key task of concentration analysis is to detect facial location. Facial landmark localization is one of the most commonly used methods [11], which is also known as face key point detection, detection or face alignment, and locates the key areas of the face, including eyebrows, eyes, nose, mouth, face contour, etc. In this study, the 68 points facial landmarks localization method is employed [12]. As shown in Fig. 1, the five-stage facial landmark localization system consists of 51 inner points and 17 contour points. The inner points are trained by three-level convolutional neural networks, which estimate six regions such as eyes, eyebrows, mouth and nose. The last stage is trained to refine the facial component landmarks, and adds the contour points for final output.

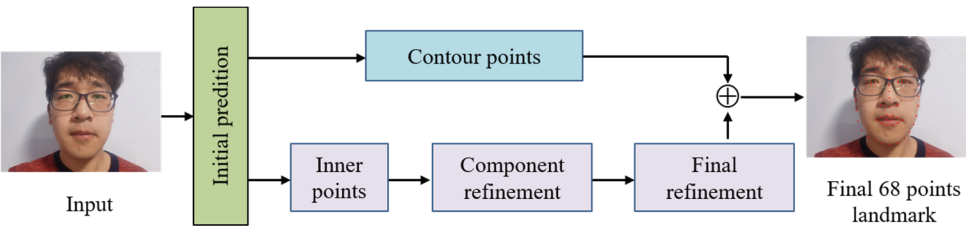


Fig. 1. Five-stage facial landmark localization system.

2.3 Head Pose Estimation

Nodding or turning the head is the important indicator to reflect students' distraction in class. The classic head pose estimation method combines classification-regression approaches [13] with UniPose [14]. The combined classification and regression introduced deep residual network as backbone and the extracted features are input into three independent full connection layers for prediction. The outputs of the three full connected layers are fed into the softmax layer with cross-entropy loss and mean square error (MSE) loss sequentially for estimation as shown in Fig. 2. The UniPose combines the functions of context segmentation and joint localization, which can estimate human posture with high accuracy in a single stage without relying on statistical post-processing methods. The main component of UniPose is waterfall atrous spatial pooling, which combines the cascade methods of hole convolution with the larger field of view obtained by the parallel structure of hole space pyramid pooling module. As the large field of view and multi-scale method are introduced, the method can predict the position of joints by context information.

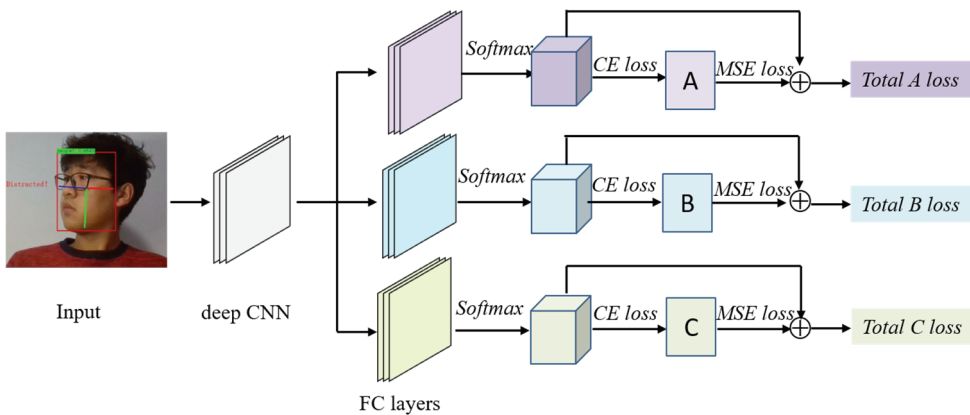


Fig. 2. A 3D head pose estimation method.

3. Proposed Method

This work was approved by the ethical committee of Northeast Agricultural University and obtained the knowledge and consent of all participants. In this work, a multi-feature fusion for E-learning based student concentration analysis method is proposed for student concentration analysis. The architecture is exhibited in Fig. 3.

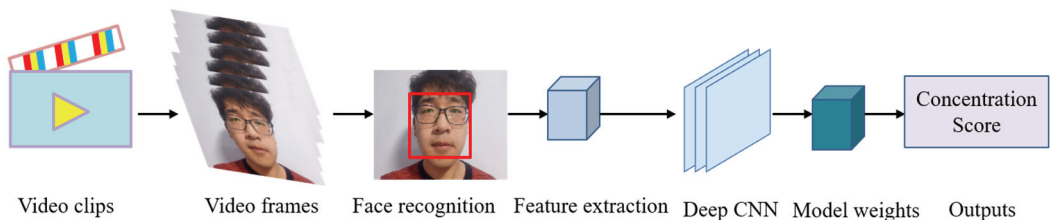


Fig. 3. The architecture of proposed method.

3.1 DAN based Facial Landmark Localization

The precondition of facial landmark localization is locating face region accurately. However, when inputting a whole-body image, the facial landmark localization system is difficult to locate the face region. To solve this bottleneck problem, the deep alignment network (DAN) [15] is introduced to detect face region in this study. The stages of training DAN are progressive; the procedure is iterative stage by stage until the validation error no longer converges. The face region alignment algorithm minimizes the loss function L , which can be represented as (1).

$$L = \min \frac{\|T_t^{-1}(T_t(S_{t-1}) + \Delta S_t) - S^*\|}{d}, \quad (1)$$

where S^* is the true vector, T_t is the transform shape of stage t and d is the distance of S^* and its pupils. In this study, the workflow of DAN based facial landmark localization can be exhibited in Fig. 4.

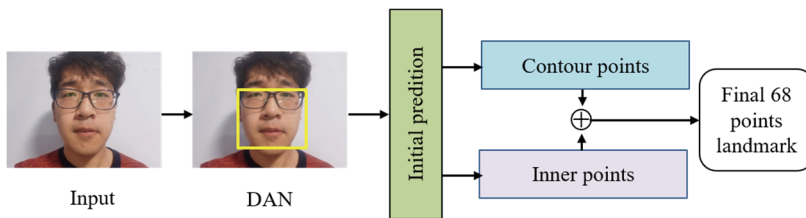


Fig. 4. DAN based Facial landmark localization.

3.2 Concentration Analysis Factors

There are numerous factors that affect the concentration analysis effect in E-learning environment. In order to grasp the key factors accurately, we designed an online questionnaire to collect the main factors in this work. Seven highest ranked items were selected from more than 300 participants for analysis and detailed as follows.

3.2.1 Ocular features

It is said that “the eyes are the windows of the soul,” eye features are one of the important indicators to reflect students’ degree of concentration. In this work, six key points were utilized from DAN based facial landmark localization as demonstrated in Fig. 5. We number the 68 points of the face calibration to facilitate subsequent experiments. $P_i(x, y)$ ($1 \leq i \leq 68$) denotes the coordinates of the i th point. The eye closure parameter ε is demonstrated in Eq. (2).

$$\varepsilon = \frac{\|P_2 - P_6\| + \|P_3 - P_5\|}{2 \times \|P_1 - P_4\|}. \quad (2)$$

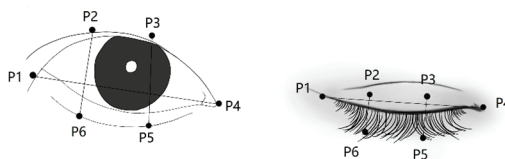


Fig. 5. Six key points of eye features.

In Eq. (2), how to set the parameter ε is a key stage, it needs a lot of practical experience. In this work, we analyzed the video clips of participants, and determined the eye closure state. Fig. 6 shows the diagram of participants' average eye state curve. In this work, we set ε as 0.25, when $\varepsilon < 0.25$, it is considered as eye closure state. When the ε value of the two adjacent video frames falls below the threshold 0.25, we think the person is nodding and record the length of time, and when $\varepsilon > 0.25$, stop timing. The algorithm is detailed in Algorithm 1.

Algorithm 1. The ocular features extraction algorithm

- (1) Input video frames.
 - (2) Calculate the parameter value of eye closure ε according formula (2).
 - (3) Determine eye closure threshold θ as in Fig. 6.
 - (4) When $\varepsilon < \theta$, $\varepsilon_num = \varepsilon_num + 1$, ε_time start timing until $\varepsilon \geq \theta$.
 - (5) Input the next video frame until the end of the video, go to (1).
 - (6) Output ε_num , ε_time .
-

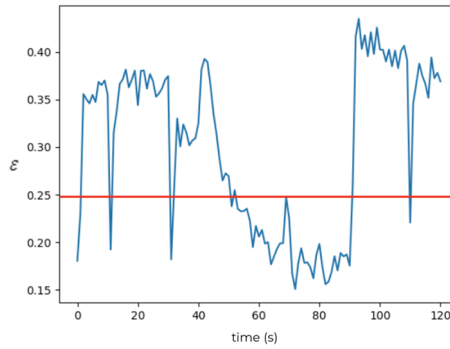


Fig. 6. Eye state curve diagram.

Conceivably, the eye closure duration is obtained, according to experience, the larger eye closure duration, the lower degree of concentration. The ordinate of Fig. 6 represents the value of ε , which depicts the size of the eye opening. The numerical unit is the Euler distance between different points on the image.

The PERCLOS [16] in this work is employed to determine whether one enters fatigue state. It is defined as the time occupied by a certain proportion of eye closure in a unit time (generally 15%), which is considered to enter the state of fatigue. In this work, when the proportion of eye closure in a certain time interval exceeds 15%, the person is considered to be in a dozing state with his eyes closed. The Boolean parameter ε_tag is used to label whether one is in dozing state or not, and parameter ε_time is used to record eye closure duration.

3.2.2 Mouth features

The yawn times can be adopted as an indicator to judge students' degree of concentration. According to one's behavior in daily life, the difference between normal speaking and yawning is the size of the mouth. The eight points were utilized from DAN based facial landmark localization as demonstrated in Fig. 7. It is possible to judge whether one is yawning by the mouth length width ratio quantitative indicators as shown in Fig. 8. The parameter τ is defined as the aspect ratio of mouth which demonstrated in (3):

$$\tau = \frac{\|P_{51} - P_{59}\| + \|P_{52} - P_{58}\| + \|P_{53} - P_{57}\|}{3 \times \|P_{55} - P_{49}\|}. \quad (3)$$

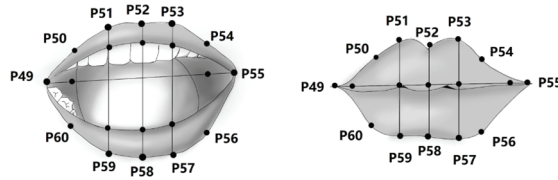


Fig. 7. Eight key points of mouth features.

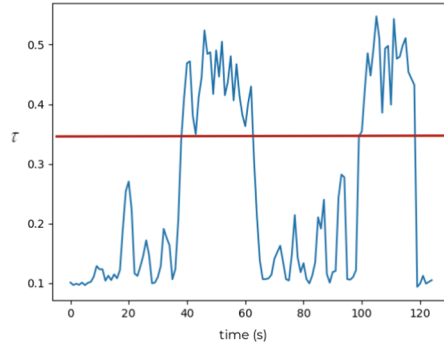


Fig. 8. The aspect ratio of the mouth.

3.2.3 Head features

The head pose estimation algorithm is employed to collect the key factors which used to determine students' concentration in E-learning class. In this work, the features such as nodding duration, nodding times, turning duration, and turning times are four key features to judge whether one is distracted or not. The torque angle parameter ψ is used to record whether one is in a turning state. When the angle parameter value $\psi \in [-15^\circ, 15^\circ]$, we think one is in a normal listening state, otherwise one is distracted, and the torque duration parameter ψ_time starts timing. Similarly, the nodding times parameter φ and nodding duration parameter φ_time are defined.

Together, seven parameters such as eye closure duration, eye closure times, yawn times, nodding duration, nodding times, turning duration, and turning times were extracted from each video clip. A total of 2,000 examples were collected in this work, they were curated into five classes such as worse concentration, poor concentration, concentration, good concentration, and better concentration. All the instances were split into three parts of training/validation/testing as 70%/15%/15%.

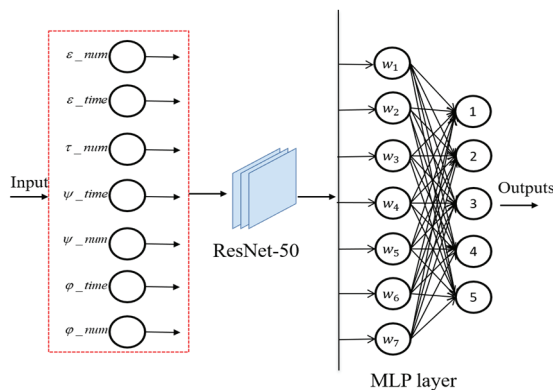


Fig. 9. The deep learning model for concentration analysis.

3.3 Deep Learning Model based on Multi-Feature Instances

In this work, the state-of-the-art deep learning approaches such as ResNet-50, DenseNet121, VGG-16, Inception-ResNet-v2, EfficientNet-B3 were employed for training concentration analysis model, and the ResNet-50 backbone-based architecture achieved the best performance. As illustrated in Fig. 9, each training instance with the seven features values and the pre-set label is fed into the deep learning model, the optimization function is Adam, and the binary cross-entropy loss function is adopted for monitoring the model running state of overfitting or vanishing gradient.

4. Results and Analysis

In order to verify the effectiveness of the proposed method, we selected 400 E-learning videos for experiment. Five clips were randomly intercepted from each video, and each clip is up to 60 seconds. Together, a total of 2,000 video clips data sets was established. All video clips were manually labelled with one of the five concentration scores. The operation effect videos are available as supplementary materials.

To comprehensively compare the effect of the proposed method, this work selected the recent existing concentration analysis models such as facial emotion [2], student emotion recognition system [17], EEG [18], classroom action [19], and facial expressions [20]. The performance of the selected models and the proposed method is depicted in Table 1. From the experimental results, it can be found that the proposed method achieved the best performance by adopting multiple features. The existing state-of-the-art models usually adopt only one or two features of them, which is difficult to reflect the students' concentration in E-learning class comprehensively. In addition, the proposed method investigated key factor features from 300 questionnaires, obtaining the first-hand and reliable data for training deep learning models. The classic deep learning models such as ResNet-50 were employed for training, which improved the accuracy of concentration state recognition largely.

Table 1. Comparison with the recent methods

Models	Adopted methods	Accuracy
Facial emotion [2]	Facial emotion analysis	0.62–0.65
SERS [17]	Emotion recognition	0.70–0.73
EEG [18]	EEG signals and machine learning	0.77–0.81
Classroom action [19]	Audio-visual media in online learning	0.80–0.85
Facial expressions [20]	Facial expressions	0.77–0.84
Proposed method	Multi-feature and deep learning	0.83–0.88

The bold font indicates the best performance in each test.

The proposed approach extracts the parameter feature information of each frame of the given video to be tested, and compares the concentration evaluation grade given by the model with the manual human evaluation results. Compared with the other methods, the advantages of the proposed method in this work are as follows: 1) it performs better in the accuracy of evaluation, that is, whether the results predicted by the paper method are consistent with the results of human evaluation. 2) This research proposes a better method with faster running speed to extract more comprehensive features through computer vision technology and predict through depth learning methods. The image information is simplified into sequence information to improve efficiency. 3) This work has established a relatively complete analysis process of students' concentration in classroom learning and visualization of practical applications, rather than just studying relevant methods.

5. Discussion

5.1 Research Motivation

The hybrid teaching mode combining online and offline will become a normal learning method in the future [1]. In offline teaching environment, teachers can easily interact with students' expressions and eyes, thus, the teaching progress can be adjusted in time according to the students' classroom performance. However, compared with offline teaching, E-learning encounters more difficulties and challenges, such as network, computer operation skills and some unknown interference factors. Teachers need to face the screen and focus on slides, there is not enough time or energy to pay attention to the dynamics of students. The proposed method can be used as an auxiliary tool for E-learning, which helps teachers understand students' E-learning status more objectively. It can also provide data support for teaching management institutions.

5.2 Why Not Utilizing Training Video Clips to Train Model?

First of all, massive video data processing occupies large computing resources, which may affect the fluency of E-learning. In addition, it is a huge task to label each extracted frame of video, and it is difficult to fully reflect the information of video clips. Finally, there is lack of large data training set for concentration prediction model training.

5.3 Limitations

Although the proposed method greatly solves the headache problem of teachers in E-learning, there are still some obstacles to be solved. Firstly, this method occupies certain bandwidth resources to transmit video data to the computing server, and it affects the fluency of E-learning on terminals with tight network resources. Secondly, the successful deployment of the system requires the support of high-performance computing platform, the large amount of video data only can be processed in this way. Finally, the successful implementation of this method requires some computer skills in order to adjust the parameters in time according to the actual situation.

6. Conclusion

In this study, a novel method of students' concentration analysis in E-learning environment was proposed. The novel approach detects the face region of students, and identifies the components of facial components using DAN-based facial landmark localization method. Through questionnaire survey, seven factors such as eye closure duration, eye closure times, yawn times, nodding duration, nodding times, turning duration, and turning times are investigated and expressed by quantifiable parameters. A total of 2,000 video clips are collected and labelled as a long sequence feature parameters for training and testing deep learning models, and the trained model with high accuracy is obtained. Experimental results indicated that the proposed method outperforms the existing state-of-the-art concentration analysis methods, which proves the effectiveness of the proposed method. Furthermore, it is known that the COVID-19 is changing people's lifestyle, which promotes the rapid development of online teaching in the field of education. However, the experience of E-learning is still inferior to offline learning, the proposed method in this study devotes to bridge the digital gaps between online and offline learning by artificial intelligence technology, and contributing to the development of modern educational technology.

Conflict of Interest

The authors declare that they have no competing interests.

Funding

This study was supported by the 2022 Heilongjiang Higher Education Reform Research Project (Grant NO. SJGY20220178).

References

- [1] Q. Ai, T. Yang, H. Wang, and J. Mao, "Unbiased learning to rank: online or offline?," *ACM Transactions on Information Systems (TOIS)*, vol. 39, no. 2, article no. 21, 2021. <https://doi.org/10.1145/3439861>
- [2] P. Sharma, M. Esengonul, S. R. Khanal, T. T. Khanal, V. Filipe, and M. J. Reis, "Student concentration evaluation index in an e-learning context using facial emotion analysis," in *Technology and Innovation in Learning, Teaching and Education (TECH-EDU 2018)*. Cham, Switzerland: Springer, 2019, pp. 529-538. https://doi.org/10.1007/978-3-030-20954-4_40
- [3] A. Dimri, D. Chaudhary, S. Shrivastava, S. Maurya, A. Chauhan, and S. Kumar, "Identifying distracted students using computer vision and statistical methods," in *Proceedings of 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2021, pp. 791-797.
- [4] A. A. Lopez Esquivel, M. Gonzalez-Mendoza, L. Chang, and A. Marin-Hernandez, "Real time distraction detection by facial attributes recognition," in *Advances in Computational Intelligence (MICAI 2021)*. Cham, Switzerland: Springer, 2021, pp. 265-276. https://doi.org/10.1007/978-3-030-89817-5_20
- [5] J. Pei and P. Shan, "A micro-expression recognition algorithm for students in classroom learning based on convolutional neural network," *Traitement du Signal*, vol. 36, no. 6, pp. 557-563, 2019. <https://doi.org/10.18280/ts.360611>
- [6] M. H. Chiu, H. L. Liaw, Y. R. Yu, and C. C. Chou, "Facial micro-expression states as an indicator for conceptual change in students' understanding of air pressure and boiling points," *British Journal of Educational Technology*, vol. 50, no. 1, pp. 469-480, 2019. <https://doi.org/10.1111/bjet.12597>
- [7] D. Zheng, "Face fatigue expression recognition based on machine learning," Ph.D. dissertation, Beijing University of Posts and telecommunications, Beijing, China, 2019.
- [8] Y. Yang, C. Sheng, L. Zhu, W. Wang, and J. Hu, "Design of fatigue monitoring system based on face recognition technology," *Instrument Technology*, vol. 2020, no. 8, pp. 5-6+39, 2020. <https://doi.org/10.19432/j.cnki.issn1006-2394.2020.08.002>
- [9] J. Yao, K. Lu, X. Ma, and G. Cheng, "Research on fatigue detection method based on EEG and EEG," *Electronic Design Engineering*, vol. 28, no. 6, pp. 115-120, 2020. <https://doi.org/10.14022/j.issn1674-6236.2020.06.025>
- [10] B. Huang, Q. Luo, H. Wang, W. Yan, and H. Su, "Fatigue state recognition method of infrared image," *Computer Measurement and Control*, vol. 25, no. 7, pp. 230-234, 2017. <https://doi.org/10.16526/j.cnki.11-4762/tp.2017.07.057>
- [11] N. Wang, X. Gao, D. Tao, H. Yang, and X. Li, "Facial feature point detection: a comprehensive survey," *Neurocomputing*, vol. 275, pp. 50-65, 2018. <https://doi.org/10.1016/j.neucom.2017.05.013>
- [12] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Extensive facial landmark localization with coarse-to-fine convolutional network cascade," in *Proceedings of 2013 IEEE International Conference on Computer Vision Workshops*, Sydney, Australia, 2013, pp. 386-391. <https://doi.org/10.1109/ICCVW.2013.58>
- [13] N. Ruiz, E. Chong, and J. M. Rehg, "Fine-grained head pose estimation without keypoints," in *Proceedings*

- of the *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, UT, USA, 2018, pp. 2074-2083. <https://doi.org/10.1109/CVPRW.2018.00281>
- [14] B. Artacho and A. Savakis, "UniPose: unified human pose estimation in single images and videos," 2020 [Online]. Available: <https://arxiv.org/abs/2001.08095>.
- [15] M. Kowalski, J. Naruniec, and T. Trzcinski, "Deep alignment network: a convolutional neural network for robust face alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, 2017, pp. 88-97. <https://doi.org/10.1109/CVPRW.2017.254>
- [16] D. F. Dinges and R. Grace, "PERCLOS: a valid psychophysiological measure of alertness as assessed by psychomotor vigilance," 1998 [Online]. Available: <https://trid.trb.org/View/498744>.
- [17] L. B. Krithika and G. G. Lakshmi Priya, "Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric," *Procedia Computer Science*, vol. 85, pp. 767-776, 2016. <https://doi.org/10.1016/j.procs.2016.05.264>
- [18] H. Huang, G. Han, F. Xiao, and R. Wang, "An online teaching video evaluation scheme based on EEG signals and machine learning," *Wireless Communications and Mobile Computing*, vol. 2022, article no. 1399202, 2022. <https://doi.org/10.1155/2022/1399202>
- [19] N. Karlina and R. Setiyadi, "The use of audio-visual learning media in improving student concentration in energy materials," *PrimaryEdu: Journal of Primary Education*, vol. 3, no. 1, pp. 17-26, 2019. <https://doi.org/10.22460/pej.v3i1.1229>
- [20] S. Vijayakumar, K. R. Sai, and B. S. Reddy, "Concentration level of a learner using facial expressions on e-learning platform," in *Advances in Data Science and Management*. Singapore: Springer, 2022, pp. 575-584. https://doi.org/10.1007/978-981-16-5685-9_56



Changjian Zhou <https://orcid.org/0000-0002-2094-6405>

He received M.S. degree in Harbin Engineering University in 2012, China. He is currently as a teacher in Northeast Agricultural University. His current research interests include artificial intelligence and computer vision.



He Jia <https://orcid.org/0000-0003-3978-4474>

He is currently pursuing the B.S. degree in college of electrical & information, Northeast Agricultural University. He is also the member of High-Performance Computing and Artificial Intelligence Laboratory. His research interests include image processing and agricultural artificial intelligence.



Jinge Xing <https://orcid.org/0000-0001-8764-0673>

He received B.S. degree in Harbin Engineering University in 1996, China. Now he is as a senior engineer in Northeast Agricultural University. His current research interests include cyberspace security and artificial intelligence.



Yunfu Liang <https://orcid.org/0000-0003-3414-2886>

He received B.S. degree in Northeast Agricultural University, China. Now he is as a teacher in Northeast Agricultural University. His current research interests include cyberspace security and educational informatization.