Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p26447-cd

A CNN-based cognitive emotion recognition method

Zicheng Song

School of reliability and system engineering, Beihang University, China. E-mail: songzicheng0817@buaa.edu.cn

Jinlong Zhao

China North Vehicle Research Institute, China. E-mail: zhaojinlong817@sohu.com

Dong Zhou

School of reliability and system engineering, Beihang University, China. E-mail: zhoudong@buaa.edu.cn

Xu An

School of reliability and system engineering, Beihang University, China. E-mail: by2414101@buaa.edu.cn

Songliang Shuai

School of reliability and system engineering, Beihang University, China. E-mail: by2414219@buaa.edu.cn

Ziyue Guo

School of Computer Science and Engineering, Beihang University, China. E-mail: guoziyue@buaa.edu.cn

As human error remains a primary cause of failures in increasingly complex systems, this study addresses the role of cognitive emotions—which reflect underlying cognitive states—in influencing human performance. To mitigate such errors, we propose a deep learning-based framework for the objective recognition of cognitive emotions from facial expressions. The approach involves three key steps: first, refining traditional emotion classification by introducing a taxonomy of cognitive emotions; second, extracting facial landmarks using a Dlib-based model to capture emotion-relevant features; and third, training a Convolutional Neural Network (CNN) to classify cognitive emotional states. A case study demonstrates that the model can effectively identify cognitive emotions during task execution. By enabling real-time emotional monitoring and operator alerts, this work offers a practical pathway to enhance human reliability and reduce error rates in safety-critical environments. The findings underscore the potential of emotion-aware systems to support human performance in complex operational settings.

Keywords: human error, cognitive emotion, emotion recognition, affective computing, facial expression, convolutional neural network.

1. Introduction

In recent years, rapid advances in technology have steadily improved the reliability of equipment through better design, materials, and management practices. Meanwhile, as human-machine interaction becomes increasingly close, human error is now a primary cause of industrial accidents(Philippart M 2022). According to an investigation by Made et al., human errors account for 90% of industrial accidents and 99% of unexpected losses, excluding natural disasters(M. Made and R.S. Taufik 2018). Bridges et al. further suggest that, with thorough investigation, premature equipment failures can also be attributed to human error(W. Bridges and R. Tew 2010). From a psychological viewpoint, humans act as key information-processing components within technological systems. Human error occurs when a person's mental activities or behaviors do not achieve the intended outcome, and this cannot be attributed to external factors(REASON J 1990). In short, human error refers to behavioral mistakes during task execution resulting from cognitive biases.

However, directly identifying human cognition is challenging. It is easier to measure cognitive emotions, which are closely related to cognition. Therefore, using computers to automatically recognize human emotional states, known as affective computing, has become a popular and influential research topic(R.W. Picard 2000). For example, Nayak et al. developed an emotion recognition human-computer interaction framework to assist users in emotion management(S. Nayak and B. Nagesh 2021); Massaccesi et al. analyzed the effects of drugs on emotions to detect mental disorders(C. Massaccesi and S. Korb 2022); Pantano employed facial expression recognition to assess consumer evaluation of retail services(E. Pantano 2020); Li et al. identified potential criminal intent based on facial expression recognition to prevent urban crime(Z. Li and T. Zhang 2021).

Among the various types of data used in affective computing, facial expression is the most widely applied visual data. Firstly, studies have shown that facial expressions are key to human emotional expression and contain richer information compared to audio or text data(J.L. Tracy and D. Randles 2015). Secondly, compared to physiological signals, the process of collecting visual data has a smaller impact on task execution(M. Saneiro and O.C. Santos 2014). Thirdly, only a camera is needed to collect visual data for emotion recognition, resulting in low costs. In earlier reasearches, visual data have been proven capable of recognizing emotions with high accuracy(A. Alelaiwi 2019). Due to its advantages and ease of application, visual emotion recognition has been widely studied in fields such as healthcare and industrial production(T.S. Ashwin and R.M.R. Guddeti 2020).

To measure human cognitive emotions through facial expressions, this paper proposes a computational approach comprising three steps: definition, extraction, and recognition. First, to classify human facial expressions more accurately in tasks, the concept of cognitive emotion is introduced, and the widely adopted basic emotion classification method is adapted into a cognitive emotion classification framework. Second, in the cognitive emotion feature extraction stage, a Haar face classifier is employed to extract facial regions from database images, and a facial landmark localization model based on the Dlib library is used to calibrate cognitive emotion-related feature points. Third, the cognitive emotion recognition stage employs a Convolutional Neural Network (CNN) to perform supervised learning on the processed facial images, identify distinctive features, and accomplish classification. This paper is structured as follows. The Methodology section elaborates on the methods for defining, extracting, and recognizing cognitive emotions. In the Case Study section, CNN-based cognitive emotion recognition is implemented to validate the effectiveness of the proposed cognitive emotion computing method. Finally, the research findings are summarized in the Conclusion.

2. Methodology

2.1. *Definition and classification of emotions*

The definition and classification of emotions is the basis of emotion computation, which lays the foundation for subsequent research on emotion feature extraction and emotion state recognition. It is important to define and classify emotions accurately before performing emotion computation, otherwise accurate emotion computation cannot be accomplished.

2.1.1. *Basic emotions*

Many scholars have conducted researches on the classification of emotions and have summarized a series of conclusions until now. Table 1 summarizes the results of some scholars' research(Ortony A and Turner T J 1990).

Table 1. Research on the definition and classification of emotions

Reference	Basic Emotions
Ekman	Anger, disgust, fear, joy, sadness, surprise
Gray	Rage, terror, anxiety, joy
Panksepp	Expectancy, fear, rage, panic
Plutchik	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
Mc Dougall	Anger, disgust, elation, fear, subjection, tender-emotion, wonder
Arnold	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness

Among these studies, Ekman's emotion classification method, which is least influenced

by factors such as culture and gender, has been widely accepted and has the greatest influence in academic circles. The method classifies human emotions into six basic emotions, including anger, disgust, fear, joy, sadness and surprise(Ekman P. 1971), as shown in Fig. 1.

Fig. 1. Ekman's emotion classification method

The common emotion states in daily life can mostly be expressed by the six basic emotions. However, for some specific task processes, the six basic emotions cannot accurately classify human emotions. Therefore, further researches are needed for the definition and classification of emotions.

2.1.2. *Cognitive emotions*

Modern emotional psychology believes that cognition and emotion are two important aspects of human psychological activities, which are closely related. In the research of the relationship between human cognition and emotion, researchers found that the six basic emotions proposed by Ekman have less influence on human cognitive processes(Graesser A and D'mello S 2007). For example, when a person is in a cognitive process, especially when using a computer teaching system, fear and sadness emotions hardly appear persistently. At the same time, through numerous researches, psychologists found that it is not the common six basic emotions that influence human cognitive processes and task processes, but other emotions such as: boredom(Czisikszentmihalyi M 1990, Miserandino M 1996), confusion(Craig S and Graesser A 2004, Kort B and Reilly R 2001), joy(Fredrickson B L and Branigan C 2005, Silvia P J and Abele A E 2002), frustration(Kort B and Reilly R 2001, Patrick B C and Skinner E A 1993), concentration(Czisikszentmihalyi M 1990), and surprise(Schützwohl A and Borgstedt K 2005). The definition of these emotions show that these are emotions that are strongly associated with cognitive processes. Therefore, these emotions that are closely related to human cognition are defined as cognitive emotion(Baker R S and D'mello S K 2010, D'mello S and Graesser A 2011).

Mcdaniel(Mcdaniel B and D'mello S 2007) and Rodrigo(Rodrigo M M T and Rebolledo-Mendez G 2008) conducted experiments on cognitive emotions, and the results showed that the main cognitive emotions which occur most frequently and for most time are: frustration, boredom, confusion, and delight, as shown in Fig. 2. After that, Baker(Baker R S and D'mello S K 2010) conducted another experiment on the four cognitive emotions above, in which the participants were asked to watch a video introducing the task and then complete a series of simple tasks. The emotion changes of the participants were collected and counted during the whole experiment. The statistical results showed that the proportion of participants who showed emotions as four cognitive emotions throughout the experiment was about 90%, while less than 10% of the six basic emotions appeared.

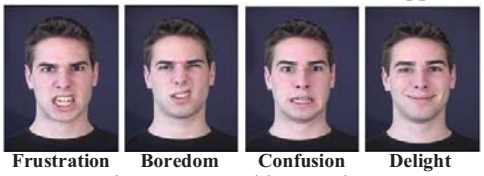

Fig. 2. Four cognitive emotions

Therefore, this paper will focus on the four cognitive emotions mentioned above that appear frequently during the experiment and can deeply influence human cognition. Subsequent recognition of cognitive emotion states will classify facial expressions into these four types to provide objective measures of cognitive emotions during the task and to analyze human reliability.

2.2. *Recognition of cognitive emotions*

In recent years, cognitive emotion classifiers are developed in order to objectively identify human cognitive emotions. Since facial expressions play a decisive role in human emotion expressions, cognitive emotion classifiers mainly obtain results by analyzing human facial expressions. These classifiers have a variety of applications, such as assisting in micro-expression researches, monitoring the emotions of task workers in real time, helping doctors to determine the mental status of patients. Deep learning is a commonly used method for the development of cognitive emotion classifiers.

2.2.1. *Advantages of CNN in image processing*

There are several types of layers in the neural network, and different types of layers can be used to process different types of data. For example, Dense layers are usually used to process vector data; LSTM layers are usually used to process sequence data; and Conv2D layers are usually used to process image data. This is determined by the characteristics of the different types of layers.

Convolutional neural network (CNN) is the neural network with mainly convolutional layers. CNN runs with its convolutional kernels translating sequentially over the image, which means that the convolutional layers learn the characteristics of the partial image. It makes CNN has the following two important characteristics:

- CNN' s learning is translation invariant. A feature learned at one location in an image can be recognized at another location. Therefore, CNN can have high efficiency in processing image data, which not only shortens the learning time, but also reduces the number of samples needed for learning.
- CNN's learning is spatially hierarchical. For multiple consecutive convolutional layers, the first layer learns smaller partial features such as edges, and each subsequent layer learns features that are combined from the characteristics learned in the previous layer. It means that when there are enough convolutional layers in the CNN, it can effectively learn extremely complex image data.

The above two characteristics make CNN has obvious advantages over other types of layers in the image processing field. Therefore, this paper uses CNN for deep learning of human face images to learn the pattern between human facial expressions and expressed emotions, and achieve an objective measure of human cognitive emotions.

2.2.2. *Construction of CNN model*

The CNN model is constructed by first building a linear stacking model Sequential, then adding input layers to the model, the input layer is a convolutional layer, setting the number of neurons, the size of the convolutional kernel, defining the activation function, and setting the input shape of the input image.

After that, hidden layer, which is created by Conv2D, MaxPooling2D, Dense and other neural network layers, is added to the model, and the hyperparameters of each layer are set. Dropout layer is added at the same time, so that some neurons are randomly abandoned in the neural network during each training process to avoid over fitting.

Finally, output layer is added to the model, and the number of neurons in the output layer should be the same as the classification dimension. After construction, the model is trained using the back propagation algorithm. Specifically, use compile function to set the loss function, optimizer and metrics, and then set the training data parameters, training period epochs and batch size.

2.3. *Face image processing method*

2.3.1. *Facial emotion feature points calibration*

Extracting facial expression information from images and then conducting cognitive emotion recognition requires locating the emotion feature points in the target images. Emotion feature points are defined as a series of specific points that can be recognized by computers and used to distinguish different regions of face images. Emotion feature points are mainly distributed around eyebrows, eyes, nose and mouth, etc.

This paper uses the face emotion feature point calibration model based on dlib database, namely "shape_predictor_68_face_landmarks", which is used to calibrate 68 emotion feature points, including 17 face contour feature points, 10 eyebrow feature points, 12 eye feature points, 9 nose feature points and 20 mouth feature points, as shown in Fig. 3.

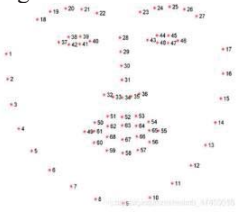

Fig. 3. 68 facial emotional feature points

2.3.2. *Facial emotion feature point extraction*

The facial emotion feature point extraction method is used to further reduce the complexity of image information. The coordinates of 68 emotion feature points are obtained from the output of emotion feature point calibration model, and a zero matrix of appropriate size is

created, each point and the surrounding 3*3 pixel points are assigned 255, and then the matrix is transformed into gray scale image. The result after treatment is that except the 68 emotion feature points are pure white, the rest of the image is pure black, Then the points are connected according to the arrangement pattern, and the image containing only emotion feature points information can be obtained.

3. Case Study

3.1. *Get training samples from databases*

This paper uses the public databases Belfast Natural Induced Emotion Database and MMI Facial Expression Database, which contain several video files, each with various expressions shown by the participants, including the four defined cognitive emotions. In order to obtain face images that will be used as samples for deep learning, the videos in the database need to be saved by taking one frame every 0.2s, and the result is shown in Fig. 4.

Fig. 4. Face images saved by frame

In the saved face images, there are large areas that are not relevant to face expression recognition. In order to reduce the scope and difficulty in the recognition, it is necessary to find the region of face in the image, and then only focus on this region in the deep learning. In the face detection field, the Haar face classifier is the most commonly used classifier.

The Haar classifier can output the region of the image where the face is located by inputting an image containing a face, and the output shape is set to a fixed 192*192 to ensure that the shape of the image remains consistent. During this process, incomplete faces and incorrect angles may cause the Haar classifier to fail to recognize faces. The face region is cropped out of the image and saved separately so that only the region can be processed in subsequent operations, as shown in Fig. 5a. and Fig. 5b.

After the image with only faces is obtained, it can be processed according to the method in section II. Input the image into the emotion feature point calibration model, and the coordinates of the 68 points are obtained after calculation, so that the emotion feature points can be calibrated on the image, as shown in Fig. 5c. The face image after calibrating emotion feature points still has some redundant information, if it is directly used as a sample for deep learning, it not only increases the calculation complexity but also adds unnecessary interference, so extract the emotion feature points separately and connect them, where the size of zero matrix is set to 200*200, and the image containing only emotion feature points is obtained, as shown in Fig. 5d. and Fig. 5e. Now the image can be used as the training sample for deep learning, which not only retains most features of the face image but also omits a large amount of irrelevant information. It can greatly improve the accuracy and efficiency of deep learning.

After getting the training samples, all the samples need to be classified. Combining with the cognitive emotion classification method proposed in this paper and the actual situation, it is necessary to define an expressionless state beyond the four cognitive emotions, named Neutral. The emotions of the samples were manually classified and labeled, and the samples were classified into five categories, including: neutral, frustration, boredom, delight, confusion.

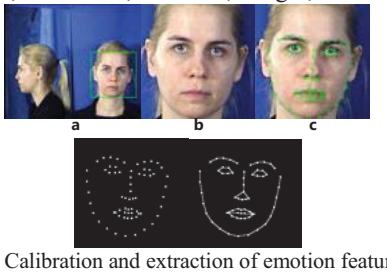

Fig. 5. Calibration and extraction of emotion feature points

3.2. *Deep learning based cognitive emotion recognition*

The deep learning in this paper classifies the samples according to the emotion type, so this is a supervised learning. Supervised learning is a kind of machine learning using samples of defined labels. Specifically, for the samples, the relationship between data and labels is learned

based on the training samples, the hyperparameters in the neural network are adjusted based on the validation samples, which can improve the neural network model's adaptability and prevent from overfitting, and the generalization ability is judged based on the fitting degree of test samples. After processing and classifying the face images in the database, a total of 609 valid samples were obtained, of which the amount and ratio of each cognitive emotion are shown in the Table 2.

Table 2. Amount and Ratio of emotions in the sample set

Emotion Type	Amount	Ratio
Neutral	193	31.69%
Frustration	54	8.87%
Boredom	33	5.42%
Delight	172	28.24%
Confusion	157	25.78%
Total	609	100%

In order to achieve a better learning result, all the samples should be cut into three parts: training samples, validation samples and test samples in the ratio of 7:2:1 at the beginning of training. The amount and ratio of the three samples are shown in Table 3.

Table 3. Amount and Ratio of Training & Validation & Test sets

Set Type	Amount	Ratio
Training Set	429	70.44%
Validation Set	120	19.70%
Test Set	60	9.85%
Total	609	100%

After the samples are processed, the next step is to construct the cognitive emotion recognition model based on CNN. For the input layer, the activation function is set as "relu", and the input data is the emotion feature point image, whose input shape is (200, 200, 1). For the hidden layer, set the ratio of abandoning neurons in the dropout layer in each training iteration to 0.1. For the output layer, because the number of neurons should be the same as the classification dimension, this paper sets it to 5 according to the emotion classification labels, and the activation function selects "softmax" which is often used in multi-classification problems.

The structural information of the completed cognitive emotion recognition model is shown in Fig. 6a., and the accuracy and loss of the training process are shown in Fig. 6b. It can be observed that the accuracy of the training and validation samples of this CNN model during the training process tends to increase, and the loss tends to decrease. It has a good training effect without overfitting. Therefore, the test set can be used to test it and analyze the accuracy.

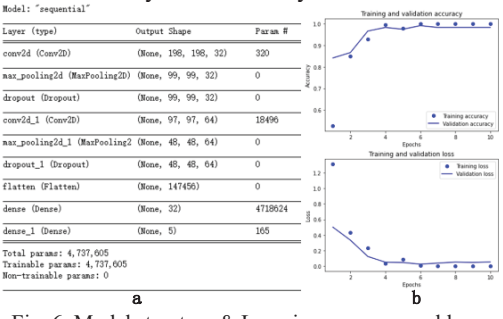

Fig. 6. Model structure & Learning accuracy and loss

3.3. *Result analysis of CNN-based cognitive emotion recognition model*

The test samples are unknown samples that have not been trained by the neural network, based on which the classification and generalization ability of the network model can be analyzed. Considering the prediction randomness of CNN, the average value of the accuracy is calculated to reduce the effect of the randomness of the model after making several predictions on the test samples.

For the above CNN model, use the test set for 50 rounds of prediction, and count the prediction and label of each sample in each round. If they are the same, it indicates that the prediction of the sample is correct. Record the amount of prediction-label of each round, as shown in the Table 4.

Table 4. The amount of prediction-label of each round

Label / Prediction	Neutral	Frustration	Boredom	Delight	Confusion	Total
Neutral	18.8	0.02	0	0.12	0.5	19.44
Frustration	0	5.16	0	0	0.02	5.18
Boredom	0	0.04	2.88	0.02	0.26	3.2

Delight	0.02	0.02	0	17.08	0	17.12
Confusion	0.02	0.08	0.08	0.04	14.84	15.06
Total	18.84	5.32	2.96	17.26	15.62	60

It can be seen from the above table:

- The average prediction accuracy of each test set can be obtained by adding the five diagonal values. The value is 58.76, so the average prediction accuracy is 97.93%. For each emotion, the average prediction accuracy is 99.79%, 96.99%, 97.30%, 98.96% and 95.01%. The accuracy is satisfying and there is no obviously special individual.
- In each group of 60 test samples, the average amount of five cognitive emotions is 18.84, 5.32, 2.96, 17.26 and 15.62, while the theoretical amount obtained by multiplying the total amount of samples by the ratio of each emotion is 19.01, 5.32, 3.25, 16.95 and 15.47. Comparing the two groups of data, it is found that the maximum absolute difference is delight, which is 0.31. The largest relative difference is boredom, which is 8.92%, but the reason for this is that the amount of boredom samples is too small. Generally speaking, the gap between the two groups of data is small, which means that the selection of test samples basically meets the random requirements.

In conclusion, the classification effect of the model on the test set is satisfying, which proves that the cognitive emotion computing method proposed in this paper can accurately judge human cognitive emotion, and provides a feasible method to reduce human error and improve human reliability through cognitive emotion monitoring.

4. Conclusion

The traditional basic emotion classification method cannot well classify the emotion state of human in the task. Therefore, this paper proposes the concept of cognitive emotion and analyzes its classification method. In addition, this paper also builds a cognitive emotion recognition model based on CNN, and verifies the model by using databases. However, there are still some problems. Firstly, almost all the participants in the database are white man, but the differences in facial expressions of different races have an important influence on the classification results, which leads to the low generalization of the model. In the subsequent process of optimizing the database, a large number of facial expression data of other races should be added. Secondly, the research object of this paper only stays at the level of static facial expression, but it can also analyze human cognition from more angles, such as voice, posture, EEG and so on. In the further research, it can consider from these angles.

References

Philippart M(2022). The petroleum industry fails to uncover valuable human error lessons from incidents - Are you making the same mistakes? . Journal of Space Safety Engineering, 9(4): 571-6.

M. Made, R.S. Taufik, T. Gustiyana（2018）, Fatigue and Human Errors Analysis in Petrochemical and Oil and Gas Plant’s Operation, in: Proc. International Conference on Industrial Engineering and Operations Management, Bandung, Indonesia .

W. Bridges, R. Tew（2010）, Human Factors Elements Missing from Process Safety Management (PSM), in: Process Improvement Institute, 6th Global Congress on Process Safety and the 44th Annual Loss Prevention Symposium, pp. 01－24 .

REASON J（1990）. Human Error. Cambridge, UK: Cambridge University Press

R.W. Picard（2000）, Affective Computing, MIT Press.

S. Nayak, B. Nagesh, A. Routray, M. Sarma(2021), A Human–Computer Interaction framework for emotion recognition through time-series thermal video sequences, Comput. Electr. Eng. 93 107280.

C. Massaccesi, S. Korb, M. Willeit, B.B. Quednow, G. Silani (2022), Effects of the mu-opioid receptor agonist morphine on facial mimicry and emotion recognition, Psychoneuroendocrinology 105801.

E. Pantano(2020), Non-verbal evaluation of retail service encounters through consumers’ facial expressions, Comput. Hum. Behav. 111 106448.

Z. Li, T. Zhang, X. Jing, Y. Wang(2021), Facial expression-based analysis on emotion correlations, hotspots, and potential occurrence of urban crimes, Alex. Eng. J. 60 1411–1420.

J.L. Tracy, D. Randles, C.M. Steckler(2015), The nonverbal communication of emotions, Curr. Opin. Behav. Sci. 3 25–30.

M. Saneiro, O.C. Santos, S. Salmeron-Majadas, J.G. Boticario(2014), Towards emotion detection in educational scenarios from facial expressions and body movements through multimodal approaches, Sci. World J.

A. Alelaiwi(2019), Multimodal patient satisfaction recognition for smart healthcare, IEEE Access 7 174219-174226.

T.S. Ashwin, R.M.R. Guddeti(2020), Impact of inquiry interventions on students in e-learning and classroom environments using affective computing framework, User Model. User-Adapt. Interact. 30 759–801.

Ortony A, Turner T J(1990). What's basic about basic emotions?. Psychological review, 97(3): 315.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. Nebraska Symposium on Motivation, 19, 207–283.

Graesser A, D'mello S, Chipman P, et al(2007). Exploring relationships between affect and learning with AutoTutor// Proc Int Conf AIED.

Czisikszentmihalyi M(1990). Flow-The Psychology of Optimal Experience. Harper & Row.

Miserandino M(1996). Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. Journal of educational psychology, 88(2): 203.

Craig S, Graesser A, Sullins J, et al(2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. Journal of educational media, 29(3): 241-250.

Kort B, Reilly R, Picard R W(2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion// Advanced Learning Technologies, Proceedings IEEE International Conference on. IEEE: 43-46.

Fredrickson B L, Branigan C(2005). Positive emotions broaden the scope of attention and thought‐action repertoires. Cognition & emotion, 19(3): 313-332.

Silvia P J, Abele A E(2002). Can positive affect induce self-focused attention? Methodological and measurement issues. Cognition & emotion, 16(6): 845-853.

Patrick B C, Skinner E A, Connell J P(1993). What motivates children's behavior and emotion? Joint effects of perceived control and autonomy in the academic domain. Journal of Personality and social Psychology, 65(4): 781.

Schützwohl A, Borgstedt K(2005). The processing of affectively valenced stimuli: The role of surprise. Cognition & emotion, 19(4): 583-600.

Baker R S, D'mello S K, Rodrigo M M T, et al(2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. International Journal of Human-Computer Studies, 68(4): 223-241.

D'mello S, Graesser A(2011). The half-life of cognitive-affective states during complex learning. Cognition & emotion, 25(7): 1299-1308.

Mcdaniel B, D'mello S, King B, et al(2007). Facial features for affective state detection in learning environments// Proceedings of the Annual Meeting of the Cognitive Science Society.29.

Rodrigo M M T, Rebolledo-Mendez G, Baker R, et al(2008). The effects of motivational modeling on affect in an intelligent tutoring system// Proceedings of International Conference on Computers in Education.57: 64.