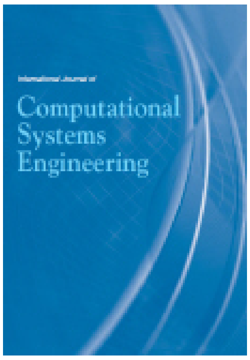
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



**International Journal of Computational Systems Engineering**

ISSN online: 2046-3405 - ISSN print: 2046-3391
https://www.inderscience.com/ijcsyse

# A machine learning framework for academic teaching and learning based on emotional reactions

Yizhu Wang

**DOI:** 10.1504/IJCSYSE.2027.10066344

**Article History:**

| | |
|---|---|
| Received: | 08 December 2023 |
| Last revised: | 11 January 2024 |
| Accepted: | 17 January 2024 |
| Published online: | 24 June 2025 |

# A machine learning framework for academic teaching and learning based on emotional reactions

Yizhu Wang

Chongqing Industry Polytechnic College,
ChongQing, 401120, China
Email: WNgh_9166@outlook.com

**Abstract:** Education suffers from the most significant weakness, which is that teachers are unable to observe their students' learning and, as a consequence, are unable to determine the degree to which their pupils are concentrating on the activities they are being taught. The present model offers a solution to the aforementioned challenge. The courses can be made more difficult by utilising our algorithm's better concentration prediction, which allows us to offer more hard options. By contributing to the expansion of educational theory and practice, this work makes a contribution. The purpose of this paper is to extract facial expression characteristics by utilising a convolutional neural network and manual features from a multi-visual bag-of-words model and support vector machine (SVM) for emotion classification. This is accomplished through the utilisation of a multi-convolutional network-based facial expression identification approach.

**Keywords**: artificial intelligence algorithms; teaching and learning; expression recognition; feature extraction recognition; convolutional neural network; CNN; loss and accuracy.

**Reference** to this paper should be made as follows: Wang, Y. (2025) 'A machine learning framework for academic teaching and learning based on emotional reactions', *Int. J. Computational Systems Engineering*, Vol. 9, No. 11, pp.1–8.

**Biographical notes:** Yizhu Wang is an Associate Professor, and earned her Master's in Pedagogy from the Faculty of Psychology at Southwest University, where she specialised in Psychology. Currently, she serves as the Director of the Mental Health Center at Chongqing Industry Polytechnic College. From 2014 to 2023, she was the Principal Investigator for ten research projects, including one major municipal project and one key municipal project. She has published four papers in core journals and authored an academic monograph on college students' mental health.

## 1 Introduction

Face-to-face offline classrooms are having difficulty meeting the knowledge needs of students in today's rapidly changing environment (Hueso-Romero et al., 2021). This is mostly due to the factors of time and space constraints. According to the findings of certain research, the expansion of network technology and the modification of computer software and hardware would lead to an increase in the number of people going to school. Higher numbers of students will enrol in classes (Zhai, 2016).

In spite of its widespread use, education is facing a number of difficulties. There is a problem with teachers being unable to monitor their students' attentiveness when they are in class (Ying, 2022). The teacher and the student are present in a traditional classroom setting, and the student is able to ask questions. With the podium, the instructor is able to swiftly assess the work of the students, evaluate the level of attention in the class, and correct any mistakes. In order to determine whether or not students are interested in taking a digital course, several authors recommend using deep learning. As a method for gathering behavioural data, this approach makes use of responses to course materials and participation in class. By using this approach, teachers and academics are able to better understand the learning patterns and requirements of students, as well as predict student participation. It is possible that this method will assist educators in developing classroom routines that are both more engaging and instructional (Bhardwaj et al., 2021). A deep face spatiotemporal network is thought to be able to measure the level of interest shown by students. The use of facial expressions allows this instrument to measure the level of pupil involvement. By using this strategy, teachers can better understand the learning needs of their pupils. Lesson planning could be aided by this. It can also assist teachers in meeting the needs of each individual student, which enhances performance (Liao et al., 2021). There are numerous classes that do not provide students with sufficient learning independence, according to the data. Because of this, there is a broad range of enrolment and completion rates for courses (Chen, 2016). E-learning can be managed more effectively by professors with the assistance of interactive simulation models (Boumiza et al., 2018), which can recognise and analyse the voice and body language of pupils.

If teachers want their students to be successful, they need to be aware of how driven their pupils are to learn on their own (Santos et al., 2018). Despite the fact that the learning motivation for students is obvious, this is the case.

When attempting to identify engaged learners, Zhang et al. (2019) utilised two different types of student behaviour data. The use of the mouse and student expressions are included in this. In this article, the authors discuss an innovative approach to locating students who are interested. The data collected from the classroom camera and mouse (student activities) is utilised. By classifying the data from the pictures, we were able to create two sets of classifiers for training and testing. The article (Liu et al., 2019) discusses ways to improve the recognition of facial expressions in real time. As a result of light and other circumstances, features in rapid photographs may be distorted. Problem with the technology that is now available. If a person's face is altered, it may be impossible to determine what feelings they are experiencing. The visual effects of high-speed capture are investigated by the authors in Mizumoto and Zimmermann (1982). Find a solution to this problem so that the system can function correctly and identify quickly. The technique does not make advantage of the output of current. For recognition purposes, it instead takes the average of the previous image.

There is a suggestion for a GAN-based model in Zhang et al. (2020). Numerous benefits are associated with this paradigm. It is possible for a computer to generate an identity-preserving face from an input face and facial markers by following the intended attitude and expression of the face. This helps to maintain the appearance of the face. Identification can be distinguished from changes in posture and expression through the use of face landmark form geometry. Thirdly, an improved FER work training can be achieved through the creation of face images depicting a wide range of emotions and events.

Face expression recognition is a concept that is proposed by Liang et al. (2021). This network is able to depict faces more accurately since it focuses on facial activity units that combine elements from the neighbourhood. In order to acquire knowledge about spatial and temporal features, video-based identification necessitates the utilisation of an attention mechanism and a temporal stream branch. It makes recognising process easier.

CNN improved the abilities of authors to recognise facial expressions (Fang et al., 2022). FER can benefit from the use of ghost-based CNNs. The ghost-module design benefits from the use of easy linear transformations because it has fewer parameters and creates more feature maps. When it comes to classifying facial expressions, GCNN is assisted by ghost modules.

For the purpose of rapidly determining age, gender, and mood based on facial gestures and photographs of faces, the researchers propose employing a deep learning technique that is based on a convolutional neural network (CNN) model (Khattak et al., 2022).

Attentiveness in the classroom was evaluated by Happy and Routray (2020) by comparing the facial expressions of students who grasped the material and those who did not. Tsiatsos and Demetriadis (2020) estimate the class concentration based on the distance between the face traits. The first findings of this study only identified three different classroom learning states: focused, paying attention, and not caring. This is despite the fact that the recognition rate was quite high.

A mood assessment was conducted by Chen et al. (Huang et al., 2007) using the facial expressions, head position, eye gazing, and physiological indications of the students in the classroom. It is merged with this data. In order to further develop this research method, further data, experimentation, and data processing are required.

As a result of the fact that several academics have taken diverse approaches to defining classroom focus, there is no definitive definition. As a result of the researches described above, researchers have either extracted a single characteristic of students in order to determine their level of focus in the classroom or have combined a number of unique elements.

1 A large number of research have focused on the processing of learners' classroom data using classical machine learning algorithms. These techniques are more complicated in terms of feature extraction, unstable in terms of feature information, and slightly less effective in terms of recognition than deep learning.

2 There has been no research that has examined the accuracy of recognition between various methods that use the learning focus as a guideline and combine single and multiple features.

3 The current research disregards the realities of these students. It is possible that the process of feature extraction will be challenging in classrooms because students may not be looking directly at the camera, their hair or clothing may obscure their faces, or natural light shadows may be present. On the other hand, teachers and students are in distinct surroundings, which makes it more probable that they will engage in conduct that is disguised. On the other hand, the fact that the relationship between the teacher and the student in the classroom is distinct makes it more likely that the student will conceal their behaviour.

Through the use of facial expressions and head posture characteristics, the paper makes a contribution by determining the levels of attentiveness that students exhibit in the classroom. During the course of the lessons, the objective is to evaluate the amount of attention exhibited by the students by analysing their facial expressions and head posture. Constructing the classroom face image dataset, labelling it with concentration level, and then classifying it with emotion labelling were the steps that were taken. An investigation of the connection between feeling and level of focus was carried out. In the following section, we will show the accuracy prediction and loss function for the proposed concentration recognition model for the classroom scenario.

## 2 Proposed model

### 2.1 Face image dataset

It is possible that the existing classroom face image datasets are not suitable for the research that is described in this paper. Because of inconsistencies in actual scenes, differences in expression features, and a lack of head pose data, these datasets may not be acceptable. The authors of this work did a collection of classroom concentration data from a number of different people in a real classroom setting. They then used that information to identify a dataset of images of students' faces.

The capacity of an individual to concentrate in a certain environment is referred to as their concentration level. It is possible to learn this information by observing the facial expressions, eye contact, and body language of a person. Students' degrees of attentiveness are often divided into seven categories (angry, disgust, fear, happy, neutral, sad, and surprise) through the use of the FER 2013 dataset (FER, 2013), which has been characterised in this article into three levels (high level, low level, and poor level). This is done by the process of segmentation of facial images taken from classrooms.

- High level: The student can be described as happy or neutral because they are engaged and involved in the information that is being presented in class.
- Low level: The student's participation in the lesson is modest, and he or she is not paying sufficient attention (Surprise, Sadness). This is a low level.

The learner has reached the poor level when they are not participating in the class and are not paying any attention (anger, disgust, or fear). This indicates that the student has reached the poor level.

Monitoring the levels of concentration of students allows for the evaluation of their degree of participation in the classroom as well as the identification of those pupils who would benefit from receiving additional support.

### 2.2 Data acquisition and processing

This approach is used to the dataset FER (2013), which is the dataset in question. Photographs of faces that are greyscale and have a resolution of 48 by 48 pixels are included in the data. All of the photographs have been automatically aligned in such a way that the proportion of the frame that is occupied by the face of each individual is approximately the same. The identification of emotion categories such as 'high level', 'low level', and 'poor level' and the formulation of particular labelling criteria are necessary in order to ensure that the labelling of emotion analysis data remains consistent. Data in the form of text is labelled after annotation.

First things first, you should gather a lot of images of your face. In order to comply with the regulations, the images must be accompanied with labels that specify the emotions of the subject. The images should be divided into two sets: one for training purposes and another for testing purposes. Both the training set and the test set will be employed in order to accomplish the goal of constructing the machine learning model. There will be training and evaluation of the model, which will be accomplished via the usage of the training and test sets.

A caption should be included to each picture that is part of the training set. Taking the time to closely associate each shot with the particular emotion that the subject is experiencing is essential. In the process of annotating, contributions may be provided by either human beings or computational learning models. Neither of these tactics is ineffective.

Perform training on the model that the machine learning algorithm uses. Our machine learning model will be trained on the pictures that make up the training set, which will be comprised of images that include annotations. As time goes on, the model will develop the capacity to recognise the different emotions that are communicated via the images. On the test set, the functioning of the model will be tested for the goal of making improvements to it. In order to determine whether or not the model is accurate, we will compare the predicted feelings with the actual feelings that individuals actually find themselves experiencing.

A segmentation of the facial picture collection is presented in Table 1. This is absolutely necessary in order to construct a machine learning model that is capable of reliably identifying emotions based on photographs of faces. Once the annotation is complete, check and verify the label's accuracy. In conclusion, the dataset that has been annotated should be divided into training, validation, and test sets for the purpose of training and evaluating the emotion analysis model. Taking this strategy requires patience and care because emotion analysis models require data that is of a high quality and has been annotated.

**Table 1** Face image dataset segmentation

| *Dataset type* | | *Training set* | *Validation set* | *Test set* | *Total* |
|---|---|---|---|---|---|
| Total number | 1746 | 536 | 536 | 2,418 | |
| Concentration level | High level | 321 | 61 | 61 | 403 |
| | Low level | 392 | 85 | 84 | 501 |
| | Poor level | 468 | 110 | 111 | 454 |

In most cases, the dataset is composed of three distinct components: training, validation, and test (Qiao et al., 2015). The training set is a collection of samples that are used to train the parameters of the neural network (Wang, 2014). The validation set is used to validate the model and compare the performance of each model after the training set (Wang, 2016). The test set is used to evaluate the generalisation ability of the model to predict model performance and to objectively evaluate the neural network.

## 3 Facial recognition model for classroom

A variety of emotions are experienced by students while they are in the classroom. These emotions are influenced by a number of factors, including the level of difficulty of the content that is being taught, the pace of the class, and the method that the instructor takes. Albert Mehrabian, a notable psychologist, proposed in 1968 that the expression of emotional information involves facial expressions to the extent of 55%, vocal expressions to the extent of 38%, and linguistic expressions to the extent of 7% (Kern et al., 2003). It would appear from this that the most direct form of emotional expression is demonstrated through changes in facial expressions. When it comes to the field of emotion recognition, the emotion model that was proposed by Ekman and his team is the one that is acknowledged the most frequently. These six core emotion models are stated as follows: happiness, surprise, sorrow, fear, rage, and disgust (Litjens et al., 2017). This paradigm is comprised of six different emotion models. Specifically, this may involve the utilisation of computer vision techniques for the purpose of collecting and analysing this data in order to acquire a more comprehensive comprehension of the emotional states of students and the degree to which they are involved (Ashwin and Guddeti, 2020). The multi-convolutional feature extraction recognition mode, which is one of the CNN approaches, is the foundation of the face detection method that was applied in this article in order to extract face images from video frames. This method was utilised in order to meet the requirements of the article. In order to successfully complete the task at hand, this mode of operation was utilised. In the realm of facial recognition technology, CNNs are a relatively new invention that has emerged in recent years. CNNs are a specialised form of deep learning algorithms that can be taught to recognise a wide range of various items by means of the photographs that they are presented with. It has been proved that CNNs are very good when it comes to identifying people based on their faces from the information that they provide. We used these seven emotions as the final emotion categories for the classroom, which was based on an actual experiment that was carried out in the classroom. This was done after adding a neutral emotion state to the fundamental emotions that were proposed by Ekman et al. This study made reference to the facial action coding system (FACS) developed by Ekman and Friesen et al. (Péron et al., 2011), and it implemented a straightforward facial expression characterisation and classification system for the seven emotions that were seen in the classroom. This research was conducted with the intention of standardising the categorisation of the feelings that can be experienced by students when they are learning in a classroom setting. In addition, the authors develop an intelligent system that is capable of determining the level of student participation in large classrooms by identifying facial expressions. By evaluating the facial expressions of the students, this may involve the application of techniques from artificial intelligence and machine learning in order to identify the amount of engagement and learning that is demonstrated by the pupils (Pabba and Kumar, 2021). A easy examination of facial expressions and matching levels of performance in the classroom was carried out for each of these seven categories of classroom emotions, as can be shown in Table 2. This analysis was carried out for each of these seven categories.

The process of emotion recognition in the classroom involves the following:

1 *Image acquisition*: The first thing that needs to be done is to get pictures of the students' faces. This can be accomplished with the use of a webcam, a standalone camera, or even a smartphone camera.

2 *Facial feature extraction*: The next thing that needs to be done is to pull facial features out of the images. Several other approaches, including as Haar cascades, local binary patterns, and CNNs, are all viable options for accomplishing this task.

3 *Feature normalisation*: After that, the face characteristics are normalised, which eliminates any changes in lighting, position, or expression that may have existed.

4 *Classification*: After being normalised, the face features are categorised according to a variety of moods. Support vector machines (SVMs) are utilised in order to accomplish this task.

### *3.1 Machine learning-based model*

A classification or regression model, output, feature extraction, and feature fusion are the four primary components that make up the multi-feature fusion concentration recognition model's block diagram. These components are essentially the building blocks of the model. Presented here is a block diagram of the classroom emotion recognition system described in Figure 1. In the beginning, several sorts of raw data, including facial expressions and head positions, are processed in order to extract valuable aspects through the process of data extraction. After that, the characteristics that have been gathered from the various types of data are combined or merged together according to the requirements. Next, the combined attributes are input into a classification or regression model, which then gives an accurate estimate of the concentration. This process is repeated until the desired result is achieved. In conclusion, the output of the model is the degree of concentration that has been projected to be present. Additionally, the specific structure of the model, as well as the approaches that are applied for feature extraction and fusion, will be determined by the particulars of the problem and the data that is already available. The particulars of the circumstance will be the determining factor in selecting these aspects.

In light of the fact that the dataset consisting of classroom face photos has already been pre-processed for the purpose of size normalisation, the VGG-13 model is used in this study. The reason behind this is because the dataset includes photographs of people's faces. At the beginning of the procedure, the weights are given a random beginning by being derived from a Gaussian distribution that has a mean of zero and a standard deviation of 0.01. This is the first stage in the process. The standard deviation of this distribution is one hundred and one, whereas the mean of this distribution's distribution is zero. It was first set at $10^{-2}$, and then, over the course of ten epochs, it was decreased by 90% until it reached its final value of $10^{-4}$. The initial learning rate was preset at $10^{-2}$. In conclusion, the DSD training plan was carried out in the way that is described in the paragraphs that follow: In accordance with the following, the network was trained for a total of two hundred iterations:

1 during the intensive training phase, the network was trained for a total of one hundred iterations

2 during the sparse training phase, the network was trained for a total of 50 iterations with the sparsity set to 110%

3 during the second intensive training phase, the network was trained for a total of 50 iterations.

The VGG-13 network was trained to its maximum capacity over a total of two hundred cycles of training.

The purpose of this article is to offer a technique that makes an attempt to locally adjust the performance of the training system to the characteristics of the training set in each area of the input space. A demonstration of this method may be found in this article. In order to understand how the local learning algorithm works, the following is a list of the essential processes that are involved in its operation:

1 the selection of a small number of training samples that are located in close proximity to a particular test set

2 the training of a classifier that makes use of only these samples

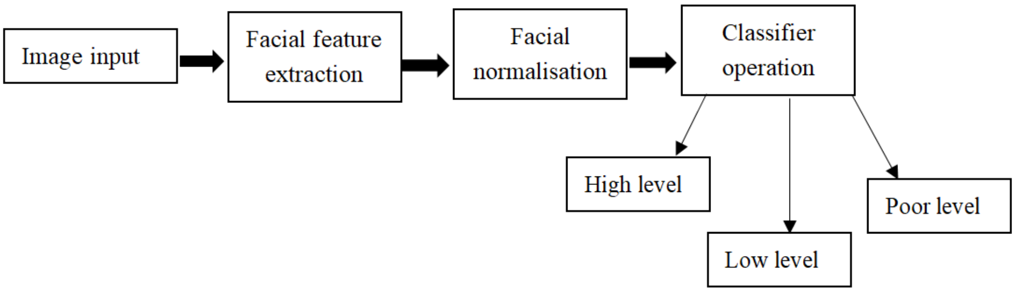
3 the application of the trained classifier to the problem of predicting the class labels utilised by the test set.

For the purpose of addressing the local classification issue that has been explored throughout this study, the SVM classifier is used. It is possible for the conventional SVM to take on a nonlinear shape when it is used within the context of the local learning approach. The difficult categorisation issue is broken down into a number of smaller problems that are simpler to deal with via this approach, which makes the problem easier to solve. Additionally, it chooses those instances that are most similar to the assessment samples, which helps to limit the number of alternatives that are included within the training set. This is done in order to simplify the process.

**Table 2** Categorisation of emotions in classroom

| *Learning about emotions* | *Facial expression characteristics* | *Corresponding classroom performance* |
|---|---|---|
| Happy | High cheekbones, contracted orbicularis oculi, curled upper lip, and the beginnings of a nasolabial furrow indicate middle age. | Content perhaps because they enjoy the class or have finally grasped the material smiles of pure joy because they adore their professor or have finally mastered the material |
| Surprise | Lips protrude, mouth gapes, eyebrows raise, eyes bulge. | Expresses astonishment at the listener's sudden comprehension or extraordinary grasp of the material being taught. |
| Sad | The eyes become smaller, the eyebrows are lifted, and the corners of the lips are pulled down. | Disappointment in oneself or one's instructor, resulting in sadness |
| Fear | Eyes bulging, mouth agape, lips drawn back, brows knitted | Acts frightened because the lesson is too complex, questions from the speaker are intimidating, etc. |
| Angry | The person's nostrils constricted, mouth contracted, eyebrows raised, and eyes expanded. | Discontent with the instructor or the content of the lesson are two common sources of anger in the classroom. |
| Disgust | Lips outstretched, mouth open, lower lip down. | Dislike for either the instructor's teaching method or the subject matter of the class |
| Neutral | Untroubled demeanour; hardly much has changed | Feelings that arise spontaneously as a result of listening intently or absorbing the material being presented. |

**Figure 1** Block diagram of classroom emotion recognition



## 4 Experimental results and analysis

Within the scope of this study, we evaluate the outcomes of the tests that were carried out in order to categorise the prediction of emotions in the classroom into seven distinct categories. However, despite the fact that the results are positive, they also imply that there is scope for progress in this specific sector. However, there is one aspect of the tale that may require some improvement in terms of the amount of realism with which the sentiments of repulsion and awe are represented in the narrative. These emotions are often conveyed in a way that is not overt, which makes it more difficult to notice them when they are there. Another aspect of the model that needs to be improved is its flexibility in the face of fresh data. This is one of the many aspects that need to be improved. It is vital to run tests on the model using data acquired from different school districts and from a range of nations in order to guarantee that it can be used to a bigger population. This enables the model to be applied to a larger population. It is feasible to carry out these tests in order to determine whether or not the model can be employed for a population that may be considered more extended. Overall, the findings of this research indicate that measuring feelings in a school setting is a task that can be accomplished. This is the conclusion that can be drawn from the findings. The construction of models that are capable of accurately and dependably anticipating the feelings that students are experiencing when they are physically present in a classroom setting is not an impossible task; nonetheless, this undertaking will require a significant amount of study and development.

There is a classification report of classroom sentiment prediction based on a multi-CNN that is shown in Table 3 for each of the seven categories. The classification report for the three levels that were identified is presented in Table 4. For the purpose of classifying the emotions that were communicated by pupils in the classroom, SVM (Wang, 2016) was applied. Using a multi-CNN, Figure 2 illustrates the accuracy of emotion recognition in the classroom. Additionally, the loss function is provided in this figure. The loss is the sum of the changes in loss that have occurred across all networks. The accuracy of the validation set model reached a stable state after the 150th epoch, and the recognition accuracy of the model reached 83.64% with this stability. Additionally, the validation set model proved to be accurate 83.64% of the time.

A comparison of the suggested method with the work that has already been done is presented in Table 5. A classroom emotion identification system that is based on a multi-CNN is evaluated with regard to its accuracy and variance in the loss function. The multi-CNN was able to obtain an accuracy of prediction of 83.14% when applied to the test set. It was discovered that the number of emotions that were successfully detected ranged from 15 for sad emotions to 138 for neutral emotions. The highest number of emotions that were correctly identified was for neutral emotions. The recognition of melancholy was still the weakest, with an F1 score of 0.600, which indicates that the qualities of sadness were not visible enough for the individuals who participated in the test.

**Table 3** Classification of classroom emotion prediction for seven categories

| *Classroom mood* | *Happy* | *Surprised* | *Neutral* | *Fearful* | *Angry* | *Sad* | *Disgusted* |
|---|---|---|---|---|---|---|---|
| Accuracy | 0.896 | 0.729 | 0.869 | 0.865 | 0.777 | 0.765 | 0.727 |
| Recall | 0.878 | 0.552 | 0.958 | 0.874 | 0.810 | 0.510 | 0.858 |
| F1 | 0.887 | 0.628 | 0.911 | 0.869 | 0.793 | 0.612 | 0.843 |

**Table 4** Classification for three levels identified

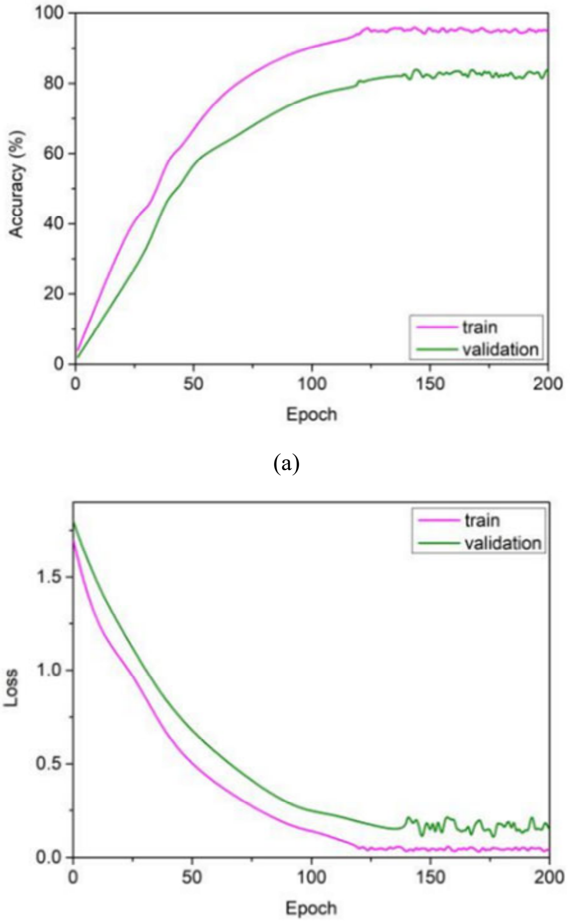
| *Classroom concentration* | *High level* | *Low level* | *Poor level* |
|---|---|---|---|
| Accuracy | 0.896429 | 0.964286 | 0.946939 |
| Recall | 0.832653 | 0.830612 | 0.821429 |
| F1 | 0.816327 | 0.793878 | 0.832653 |

Not only are the findings of this study encouraging, but they also imply that anticipating emotions in a classroom setting is a task that can be accomplished, despite the restrictions of the study. The construction of models that are capable of accurately and dependably anticipating the feelings that students are experiencing when they are physically present in a classroom setting is not an impossible task; nonetheless, this undertaking will require a significant amount of study and development. This information can be utilised to enhance the teaching and learning process by providing educators with insights into the feelings that their students are experiencing while they are in the classroom. It is possible to achieve this goal by making use of the data in order to provide teachers with insights into the feelings that their students are experiencing. It is possible to make use of this information in order to enhance instructional strategies, offer assistance to students who are having difficulty, and create a learning atmosphere that is more encouraging and conducive to successful endeavours.

**Table 5** Comparison of proposed method with existing work

| *Author* | *Zhang et al. (2020)* | *Liang et al. (2021)* | *Khattak et al. (2022)* | *Proposed method* |
|---|---|---|---|---|
| Feature chosen | Facial | Facial | Facial | Facial |
| Methodology used | GAN | Multi-scale action unit (AU)-based network | CNN | Multi-labelled for classroom concentration and classroom emotion |
| Performance metric | 92.09% (accuracy) | - | 95.69% (accuracy) | 96.8% (accuracy) |

**Figure 2** Prediction accuracy and loss function, (a) prediction accuracy (b) loss function (see online version for colours)



(a)

(b)

There is a strong probability that both students and instructors will participate in the implementation of the suggested technology. Concerns like privacy, prejudice, and the amount of work that teachers have to do are addressed by the framework, which was created with user requirements in mind. A suitable amount of training and assistance is provided to teachers so that they can properly use the framework. The students are provided with information on the use of their data and are given the ability to alter their privacy settings. Implementation of the framework is carried out in a manner that strengthens and supports the teaching practices that are already in place. Therefore, a framework for machine learning that is based on emotional responses has the potential to transform education by producing a learning experience that is more individualised, engaging, and successful for everyone.

## 5 Conclusions

For the goal of this investigation, facial expressions are used as a means of overseeing the attention of the students. Picture recognition techniques and algorithms are used in order to analyse the facial expressions and head postures of students who are present in the classroom. This is done in order to ascertain the level of attention that the students are displaying. This article develops and multi-labelled the dataset by making use of a classroom face dataset that was self-constructed for the purpose of analysing subjects' levels of attentiveness and mood.

The purpose of this study is to extract facial expression features by utilising a CNN and manual features based on a multi-visual bag-of-words model by employing a multi-convolutional network-based facial expression identification approach. Additionally, the paper uses SVMs for emotion categorisation. At long last, a hierarchy of emotions and concentrations in the classroom has been established. In order to achieve accuracy, concentration identification algorithms that are based on deep learning require a significant amount of annotated data as well as processing resources. It is necessary to have hardware in order to have precise eye movement tracking models. Despite the fact that they provide comprehensive data, multimodal models are difficult to interpret and analyse. Machine learning models have the potential to overfit, necessitating the selection of features manually. The purpose of this research is to offer a multi-feature fusion concentration recognition model that has a stable architecture, inexpensive processing resources, quick results, and good reliability.

## Acknowledgements

This paper is supported by Chongqing Municipal Major Project Funding for Vocational Education Teaching Reform Research, Project Number Z231003.

## References

Ashwin, T.S. and Guddeti, R.M.R. (2020) 'Affective database for e-learning and classroom environments using Indian students' faces, hand gestures and body postures', *Future Generation Computer Systems*, Vol. 108, pp.334–348.

Bhardwaj, P., Gupta, P.K., Panwar, H., Siddiqui, M.K., Morales-Menendez, R. and Bhaik, A. (2021) 'Application of deep learning on student engagement in e-learning environments', *Computers & Electrical Engineering*, Vol. 93, p.107277, ISSN 0045-7906.

Boumiza, S., Bekiarski, A., Pleshkova, S. and Souilem, D. (2018) 'Development of simulation models for interactive audiovisual control of students in the e-learning environment', in *2018 International Conference on High Technology for Sustainable Development (HiTech)*, pp.1–4.

Chen, M. (2016) 'The construction of the model of distance education under the background of 'Internet+'', *Continuing Education*, Vol. 30, No. 8, pp.44–46.

Fang, B., Chen, G. and He, J. (2022) 'Ghost-based convolutional neural network for effective facial expression recognition', in *Proceedings of the 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE)*, Guilin, China, 25–27 February.

FER (2013) *Dataset* [online] https://www.kaggle.com/datasets/msambare/fer2013 (accessed 20 August 2023).

Happy, S.L. and Routray, A. (2020) 'Deep learning for facial expression recognition: a step closer to a smartphone that knows your moods', *Computers & Electrical Engineering*, Vol. 81, p.106522.

Huang, D., Quan, Z. and Jia, W. (2007) *An Authentication Method based on Revocable Handwritten Signature*, Hefei Institute of Intelligent Machinery, Chinese Academy of Sciences.

Hueso-Romero, J., Gil-Quintana, J., Hasbun, H. et al. (2021) 'The social and transfer massive open course: post-digital learning', *Future Internet*, Vol. 13, No. 5, p.119.

Kern, N., Schiele, B. and Schmidt, A. (2003) 'Multi-sensor activity context detection for wearable computing: Eusai', *LNCS.*

Khattak, K., Asghar, M.Z., Ali, M. and Batool, U. (2022) 'An efficient deep learning technique for facial emotion recognition', *Multimed. Tools Appl.*, Vol. 81, No. 2, pp.1649–1683.

Liang, L., Lang, C., Li, Y. and Feng, S. (2021) 'Fine-grained facial expression recognition in the wild', *IEEE Trans. Inf. Secur.*, Vol. 16, No. 8, pp.482–494.

Liao, J., Liang, Y. and Pan, J. (2021) 'Deep facial spatiotemporal network for engagement prediction in learning', *Appl. Intell.*, Vol. 51, No. 10, pp.6609–6621.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M. and Sánchez, C.I. (2017) 'A survey on deep learning in medical image analysis', *Medical Image Analysis*, Vol. 42, pp.60–88.

Liu, K-C., Hsu, C-C., Wang, W-Y. and Chiang, H-H. (2019) 'Real-time facial expression recognition based on CNN', in *Proceedings of the 2019 International Conference on System Science and Engineering (ICSSE)*, Dong Hoi, Vietnam, 19–21 July.

Mizumoto, M. and Zimmermann, H.J. (1982) 'Comparison of fuzzy reasoning methods', *Fuzzy Sets & Systems*, Vol. 8, No. 3, pp.253–283.

Pabba, C. and Kumar, P. (2021) 'An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition', *Expert Systems*, No. 1, p.e12839.

Péron, J., El, T.S., Grandjean, D. et al. (2011) 'Major depressive disorder skews the recognition of emotional prosody', *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, Vol. 35, No. 4, p.987.

Qiao, H., Yin, P., Li, R. et al. (2015) 'The significance of the intersection of robotics and neuroscience: thoughts on the future development of intelligent robots', *Journal of the Chinese Academy of Sciences*, No. 6, pp.762–771.

Santos, P.B., Wahle, C.V. and Gurevych, I. (2018) 'Using facial expressions of students for detecting levels of intrinsic motivation', in *2018 IEEE 14th Int. Conf. e-Science*, pp.323–324.

Tsiatsos, T. and Demetriadis, S. (2020) 'Artificial intelligence in education: the application of deep learning to enhance teaching and learning', in *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*, Springer, Cham, pp.381–391.

Wang, T. (2016) 'Hot directions in artificial intelligence and robotics before', *Intelligent Robotics*, No. 4, pp.31–33.

Wang, W. (2014) 'Information push-Thomson Reuters and the CAS Literature and Information Center jointly publish Frontiers of Research 2014 (Psychology and Bioscience section)', *Support Systems.*

Ying, Y. (2022) 'Quantitative analysis of Chinese classroom teaching activity under the background of artificial intelligence', *Education and Information Technologies Article*, April, Vol. 27, pp.11161–11177.

Zhai, X. (2016) *Research on the Influencing Factors of Flipped Classroom Learner Satisfaction and its Mechanism of Action*, University of Science and Technology of China.

Zhang, F., Zhang, T., Mao, Q. and Xu, C. (2020) 'Geometry guided pose-invariant facial expression recognition', *IEEE Trans. Image Process.*, Vol. 29, pp.4445–4460.

Zhang, Z., Li, Z., Liu, H., Cao, T. and Liu, S. (2019) 'Data-driven learning engagement detection via facial expression and mouse behavior recognition technology', *J. Educ. Comput. Res.*, Vol. 152, DOI: https://doi.org/10.1177/0735633119825575.