Journal of E-Technology

DLINE JOURNALS

Print ISSN: 0976-3503
Online ISSN: 0976-2930

JET2025: 16 (4)
https://doi.org/10.6025/jet/2025/16/4/134-141

A Language Assessment System using Deep Neural Networks and Facial Expression Recognition

Peng Chengcai, Cheng Piaoyun, Wei Xia, Huang Zhaowei
Guangxi Ecological Engineering Vocational and Technical College, Liuzhou, Guangxi 545000. China
ijb6281209@163.com

ABSTRACT

Educational Quality assessment serves as a crucial instrument for enhancing teaching quality and bolstering teaching effectiveness. Conventional college English teaching evaluations are static assessments that fail to reflect students' performance throughout the teaching process accurately. Hence, this paper develops a college English teaching evaluation model founded on deep learning neural networks. It achieves emotion classification via facial recognition of students and integrates this with a standard distribution evaluation model to assess students' attitudes toward English teaching quality. The experimental outcomes reveal that the proposed model significantly enhances the accuracy of emotion recognition and classification rates, effectively mirroring students' attitudes towards English instruction in real world applications.

Keywords: Deep Learning, Neural Network, English Teaching, Teaching Assessment

Received: 30 March 2025, Revised 21 July 2025, Accepted 5 August 2025

1. Introduction

Assessing the quality of teaching is an essential component of educational activities and a vital means to guarantee and foster the ongoing enhancement of the learning process. Effective teaching evaluations can provide valuable feedback for educational decisions, enabling educators to analyze the teaching process based on evaluation findings and to make necessary adjustments and incentives according to actual needs. As educational reform and innovation progress, teaching evaluation practices have been continually refined, making the investigation of scientific, practical, reasonable, and authentic teaching evaluation approaches a prominent topic in educational research. In the digital age, teaching evaluation methods and contents increasingly emphasize diversity and comprehensiveness, expanding beyond traditional questionnaires to encompass various educational data, including audio, text, and images. Conventional teaching evaluation methods predominantly depend on manual assessments, which not only exhibit significant subjectivity and

are time consuming but also lack accuracy and struggle to manage large volumes of teaching evaluation data, inhibiting the practical exploration of correlations and influences among the data.

Higher education English instruction plays a crucial role in developing well rounded social talents. As a subset of language education, English teaching emphasizes its comprehensive nature in accordance with new educational standards. However, assessing teaching effectiveness through traditional mathematical evaluation models in language education poses challenges. On the one hand, scores from English assessments do not adequately and multi dimensionally capture the effectiveness of teaching. On the other hand, courses in English speaking and written communication possess significant flexibility and subjectivity, making it difficult for conventional mathematical models to represent the relationship between these components accurately. To enhance the quality of English instruction, numerous educators are integrating multimedia resources to fulfill teaching objectives in a multimodal fashion, while also creating a student focused teaching approach. Within this framework, traditional survey methods often fail to objectively and accurately reflect teachers' teaching quality and cannot process the vast amount of data generated during instruction quickly. Some researchers have suggested that students' responses in the classroom provide critical information that can validate teaching quality. Consequently, this paper develops a model for recognizing emotions in English classrooms using deep neural networks, which gathers essential feedback on teaching by analyzing students' facial and emotional expressions, thereby facilitating English teaching assessment.

2. Development and Research Status of Teaching Evaluation Methods

Initially, teaching evaluation was predominantly centered on students' examination scores, which assessed their performance. This method of evaluation was not only static but also carried inherent limitations, as it represented overall outcomes based on partial results, leading to substantial biases. Subsequently, researchers recognized that numerous factors affect teaching quality, with test scores merely serving as one tangible representation of that quality. This realization prompted a shift toward multi dimensional evaluation strategies to explore teaching quality more thoroughly. Some scholars have established relevant evaluation metrics, gathered data on teaching activities, processes, student learning outcomes, and feedback, and computed teaching quality scores using weighted or specific mathematical models. However, evaluation indicators differ across disciplines, and the mathematical models employed often fail to express nonlinear relationships, resulting in considerable discrepancies in evaluation outcomes. Furthermore, it has been highlighted that conventional teaching evaluation methods tend to be highly subjective, lacking objectivity and scientific rigor in their findings. In addition, the inherently subjective nature of teaching processes often leads traditional evaluations to overlook critical aspects of the teaching process, making it challenging to identify correlations within the data and resulting in imprecise evaluation outcomes. To bolster the objectivity and scientific validity of teaching evaluation models, some researchers have incorporated Back propagation (BP) neural networks to create specific teaching quality evaluation models. However, the BP neural network is susceptible to becoming trapped in local optima, resulting in elevated error rates in evaluation results. While some researchers have utilized machine learning for teaching quality assessment, their systems of evaluation indicators have varied significantly, leading to inconsistent evaluation outcomes. [8] Scholars also combined deep learning theories to assess teaching quality and effectiveness using deep neural networks [9]. However, most current teaching evaluation models approach evaluation from the perspective of teachers, focusing on the teachers' teaching process and effectiveness, while lacking corresponding evaluation of students' impact.

3. Construction of College English Teaching Evaluation Model Based on Deep Neural Networks

Amid the context of educational reform and innovation, college English instruction has moved beyond the confines of traditional classrooms. The focus has shifted from being teacher centered to student centered, prioritizing personalized growth and democratic principles, while assessment has transitioned from a singular method to a more varied and multi faceted evaluation approach. The assessment of modern college English teaching quality now emphasizes student development and the professional growth of educators. To provide a more objective and scientific representation of teaching quality, this paper develops a college English teaching evaluation model utilizing facial expression recognition and a probability assessment framework. Facial expressions are tangible indicators of human psychological processes. By analyzing facial expressions, essential feature information can be gathered, allowing for insights into an individual's psychological state. Consequently, employing deep neural networks to interpret student expressions in college English classrooms can yield valuable information about students' perceptions of the learning environment and facilitate deeper analysis of their genuine sentiments and reflections regarding English lessons. Deep learning refers to an artificial neural network that achieves intricate function approximation through multiple hidden layers and non-linear network architectures. It enhances the classification of deep level abstract features and the prediction accuracy of the model through extensive training and learning. Moreover, deep neural networks can be trained progressively by layers, which aids in overcoming training challenges. In comparison to other algorithms, deep neural networks exhibit autonomous characteristics, contributing to the enhancement of the model's intelligence and automation to a degree. Based on practical requirements and model specifications, this paper opts to construct a college English teaching evaluation model using a deep Convolutional Neural Network (CNN) and a Deep Belief Network (DBN).

CNN represents a specific kind of feed forward neural network, characterized by a structure composed of numerous neurons with associated weights and biases. Its localized connection method effectively mitigates data redundancy associated with full connections and decreases the number of network parameters, lessening the model's reliance on extensive datasets during training. As such, CNN organizes an image matrix and incorporates three fundamental concepts: local perception, weight sharing, and down sampling. As illustrated in Figure 1, the unique structure of CNN is achieved through local weight sharing. It employs local perception to preserve the correlation information among image pixels, thus facilitating higher dimensional feature extraction and acquiring additional rich data within the image. Subsequently, through weight sharing and down sampling processes, the number of network parameters can be minimized, enhancing the model's robustness and allowing for the sustainable expansion of its depth, which leads to an increase in the number of hidden layers. The two most significant types of layers in CNN are the convolutional layer and the down sampling layer, also referred to as the pooling layer, the convolutional layer is the essential core of the CNN neural network, wherein feature extraction is performed using convolutional kernels.

If the input image is X, the convolution kernel is, and the size is m, the final output feature map is shown in formula (1):

$$Y_j(j \in p*q) = f(\sum_{i \in m*m} X_i * K_i + b) \quad (1)$$

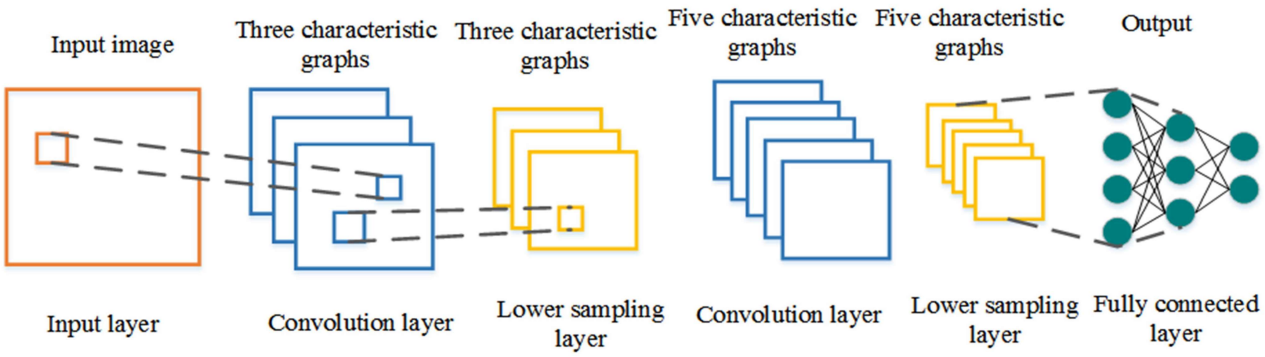

Figure 1. Simple structure diagram of a Convolutional neural network

The offset amount in the equation is expressed as b, and the final output image size is expressed as p_*q.

The pooling layer is to remove redundant information from the feature maps extracted through convolutional layers, preserve the most important information, and achieve the goal of feature map dimensionality reduction. The pooling layer generally includes maximum pooling, average pooling, and sum pooling, with the first two being the most commonly used types of pooling.

Assuming that the j_{th} output feature map in the layer is represented as a, after passing through the pooling layer, as shown in formula (2):

$$a_j^L = f(down(a_j^{L-1}) + b_j^L \qquad (2)$$

The offset in the equation is expressed as b, and the pooling function is down. The CNN Activation function is shown in Formula (3):

$$relu(x) = \max(0, x) \qquad (3)$$

The Loss function is shown in Formula (4):

$$loss(Y, \hat{Y}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \qquad (4)$$

In this model, a multi-task Convolutional neural network is used, which is based on the multi task cascade face detection framework to complete face detection and key point location at the same time. It is composed of three network structures, P-Net, R-Net and O-net. P-Net is a fully convolutional network that primarily determines whether faces are present in each 12 * 12 range area. If there are faces, the boxes containing the faces will be regressed to obtain the corresponding regions in the original image. R-Net is a pure Convolutional neural network, which is mainly used to improve the candidate window. O-net is a pure Convolutional neural network in the third level network, which will perform Bilinear interpolation steps on the possible face candidate window, and perform more refined extraction operations on face detection and key points.

Deep belief network is a probabilistic Generative model, which includes multiple restrictive Boltzmann machine machines and a top level algorithm that can be used for discrimination. During the training process, the deep belief network mainly optimizes multiple hidden layers through greedy algorithms. Then it combines reverse fine-tuning algorithms to achieve the best state of model training. The restrictive Boltzmann machine is usually the introductory module that can reflect the relationship between random variables. It includes the visible layer and the hidden layer. There is connectivity between them, but there is no connectivity within the layer. The energy function expression between visible and hidden layer nodes is shown in (5):

$$E(V,H|\theta)=-\sum_{i=1}^{n}a_i v_i-\sum_{j=1}^{m}b_j h_j-\sum_{i=1}^{n}\sum_{j=1}^{m}v_i w_{ij}h_j \tag{5}$$

Among them, the unit state with sequence number i in the visible layer is denoted as v, the unit state with sequence number in the hidden layer is denoted as h, the connection weight between the two is denoted as w, and the offset values of the two are denoted as a and b, respectively.

By combining the above formula, the joint probability distribution can be obtained, as shown in (6):

$$p(v,h|\theta)=\frac{e^{-E(V,H|\theta)}}{Z(\theta)} \tag{6}$$

In the model, the energy summation that may occur in the visible layer and hidden layer node sets is represented as Z.

4. College English Teaching Evaluation Model Based on Deep Neural Networks

Students' expressions in English teaching classrooms can be categorized into three states: positive, neutral, and negative. The positive state includes expressions of pleasure and focus, the neutral state includes calm expressions, and the negative state includes expressions of fatigue and boredom. Each state is defined by its corresponding salient features. When students find the English teaching content interesting or pay serious attention, they will exhibit focused expression. When students understand the English teaching content and actively participate, they usually show a joyful expression. When students have doubts or don't understand the English teaching content, they generally display a puzzled expression. When students feel that the English teaching content is beyond their cognitive scope and level, they may show expressions of fatigue and boredom. Each expression state corresponds to specific salient features on students' faces, and the deep neural network model can evaluate college English teaching based on the results of student facial expression recognition.

To verify the effectiveness of the deep neural network based college English teaching evaluation model in recognizing and classifying basic facial expressions, this paper selected four other models for basic facial expression classification. The results are shown in Figure 2. The results in the figure indicate that the GBDN model can classify happy, surprised, sad, and fearful facial expressions very well, with classification rates above 0.9 for each. The Ada+BUs model only achieved a classification rate above 0.9 for happy expressions. The *Ada+ibdn* model achieved classification rates above 0.9 for happy and angry expressions. The BDBN model, except for surprised expressions, achieved classification rates above 0.9 for other expressions. In this paper's model, the classification rates for all six expressions are above 0.92, significantly higher than the classification rates of the other four models. Among the six expressions, the classification rate for surprised expressions is relatively low,

mainly because surprised expressions have some similarity with angry and fearful expressions, making them relatively harder to discriminate.

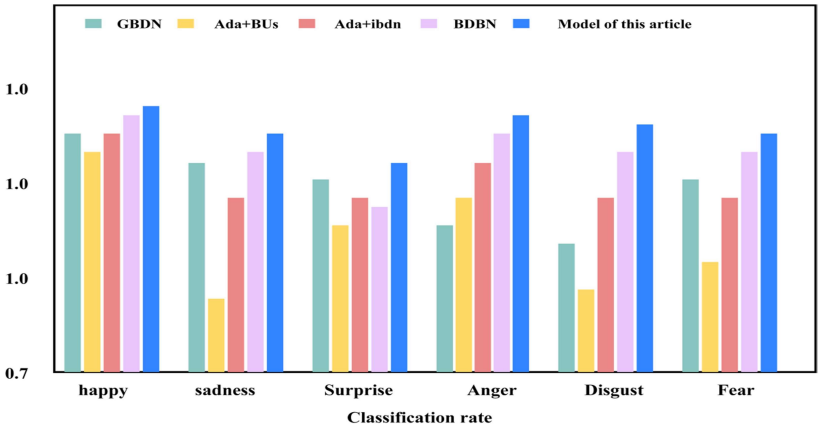

Figure 2. Comparison results of basic facial expression classification rates for five models

To further confirm the effectiveness of the college English teaching evaluation model utilizing deep neural networks, this research randomly selected a college English classroom. It employed the regular distribution evaluation model to assess the English instruction. The findings are illustrated in Figure 3. The posterior probability and the two parameters of the evaluation results show a positive correlation with the likelihood of true values. Thus, from the figure's results, it is evident that when the posterior probability reaches its peak, the parameter value lies within the range of [0.605-0.607], indicating that the overall emotional response of students in this classroom is quite favorable. This parameter represents the spread of the posterior probability, suggesting that a higher value of this parameter corresponds to greater data dispersion, and conversely, a lower value indicates data concentration. The parameter is 0.0406, signifying that the emotional index of students in this classroom is highly concentrated around the average, implying that most students view the English teaching content positively, and the teaching quality is commendable. Based on the evaluation findings and feedback, educators can focus more on students at both extremes of the normal distribution and adjust their teaching strategies to engage a broader range of students, ultimately enhancing the quality of English instruction.

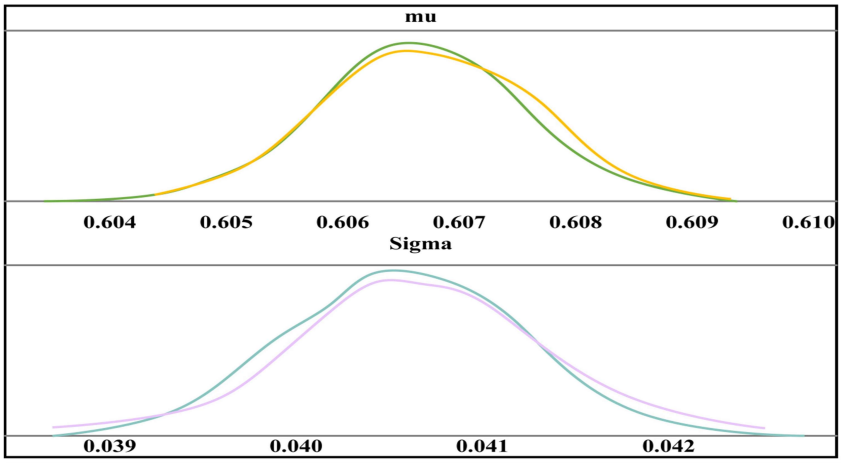

Figure 3. Results of normal distribution parameters' probability distribution for college English teaching classroom data

In summary, the college English teaching evaluation model grounded in deep neural networks can proficiently recognize and categorize students' facial expressions. The classification rates for all expressions surpass those of the other four models and exceed 0.92, offering reliable technical support and scientific data for practical applications. In real world scenarios, the model effectively captures students' genuine emotional reactions to English teaching and analyzes the dispersion of emotional data through parameters, thus achieving a multi-dimensional representation of English teaching quality.

5. Conclusion

With the rapid advancement of educational innovation and reform, college English teaching has transitioned from a teacher centered to a student centered approach, where instructors fully adopt a guiding role. As a result, previous college English teaching evaluation models lack adaptability and fail to provide a comprehensive assessment of English teaching from various dimensions. Consequently, this study integrated deep neural networks to develop a college English teaching evaluation model that identifies and classifies students' facial expressions. It subsequently analyzes students' emotional attitudes towards English teaching content using the normal distribution evaluation model to finalize the assessment of English instruction. The experimental outcomes demonstrate that, in comparison to the other four models, the college English teaching evaluation model based on deep neural networks significantly enhances the classification rates of all facial expressions, with each expression's classification rate exceeding 0.92, thereby improving the model's classification and recognition capabilities. In practical application, this model can accurately detect students' emotional fluctuations, reflect their authentic emotional attitudes through the normal distribution evaluation model, and analyze the dispersion of emotional data, resulting in a comprehensive evaluation of English teaching.

References

[1] Sharma, K., Saxena, A., Kumar, P. (2012). Alignment of DNA sequence using the features of global and local algorithms along with matrices. *Advanced Materials Research*, *403*, 2012–2015.

[2] Eban, E., Schain, M. (2017). Scalable learning of non-decomposable objectives. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (pp. 832–840).

[3] Radzikowski, K., Wang, L., Yoshie, O., others. (2021). Accent modification for speech recognition of non-native speakers using neural style transfer. *EURASIP Journal on Audio, Speech, and Music Processing*, *2021*(1), 1–10.

[4] Dharmawansa, A. D., Nakahira, K. T., Fukumura, Y. (2012). On-line visualization of student facial emotion in virtual e-learning environment. *Kansei Engineering International Journal*, *11*(4), 267–276.

[5] Zhang, Z., Wang, Y., Yang, J. (2021). Text-conditioned transformer for automatic pronunciation error detection. *Speech Communication*, *130*, 55–63.

[6] Wang, Y.xin. (2017). Analysis of the factors influencing the quality of junior high school English teaching based on analytic hierarchy process. *Educational Modernization*, (49), 388–390.

[7] Yang, Y., Li, J. (2015). Research on employment quality evaluation of college graduates based on AHP and BP neural network. *Journal of Chinese Education*, *11*, 148–149.

[8] Krizhevsky, Alex. (2013). ImageNet classification with deep convolutional neural networks. *Journal of Machine Learning Research*, *45*, 12–16.

[9] Hinton, G. E., Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*, 504–507.

[10] Jugo, I., Kovaèic, B., Slavuj, V. (2016). Increasing the adaptivity of an intelligent tutoring system with educational data mining: A system overview. *International Journal of Emerging Technologies in Learning*, *(3)*, 423–425.