Teaching Quality Evaluation with Integrated Student Emotions and Content Analysis via Machine Learning and Fuzzy Logic

Md Shariful Islam[1] [0009−0009−1120−6278], Bemnet Wondimagegnehu Mersha[2] [0000−0002−1000−8984], Yaping Dai[3] [0000−0001−8795−5333], Kaoru Hirota[4] [0000−0001−5347−6182], Zheng Zhou[5], Wei Dai[6], and Yumin Lin[7]

[1] School of Automation, Beijing Institute of Technology, Beijing, 100081, China email: sharif@bit.edu.cn
[2] School of Automation, Beijing Institute of Technology, Beijing, 100081, China email: MershaBemnetW@bit.edu.cn
[3] Beijing Institute of Technology Zhuhai, Tangjiawan, Zhuhai City, 519088, Guangdong, China email: daiyaping@bit.edu.cn
[4] School of Automation, Beijing Institute of Technology, Beijing, 100081, China email: hirota@bit.edu.cn
[5] School of Automation, Beijing Institute of Technology, Beijing, 100081, China email: 3220220886@bit.edu.cn
[6] River Security Technology Co., LTD, Shanghai 200336, China email:daiwei@gmail.com
[7] River Security Technology Co., LTD, Shanghai 200336, China email:ymlin@riversecurity.com

Abstract. Online and distance learning have recently been increasingly used to enhance smart education. Traditional teaching quality evaluation methods can not be directly used for online or distance learning platforms. To solve this problem, this research proposes an integrated approach that uses fuzzy logic and machine learning. The proposed method's main contribution is combining student engagement with teacher content evaluation for teaching quality evaluation. The emotional engagement is tested on the FER2013 dataset and achieved an average accuracy of 65%. Three experimental scenarios are presented for the proposed fuzzy logic and machine learning integrated method to evaluate teaching quality based on student emotion and teaching content. In these scenarios, more than 94.33% values are assigned as High teaching quality, 65.33% as good teaching quality, and 11% as poor quality.

Keywords: Teaching quality evaluation · Emotion recognition · Fuzzy Logic · Machine Learning · Smart Education

1 Introduction

Online classroom and distance learning platforms are on the rise, with institutions adopting them to train employees and cater to students, adults, and workers[1]. As online and distance teaching becomes an integral part of the educational landscape, there is a pressing need to ensure its teaching quality. Traditional teaching quality evaluation methods, such as self-reported student feedback and expert direct observation, may not adequately capture the intricacies of effective online teaching. The traditional methods are often subjective and time-consuming. Effective and engaging teaching outside the physical classroom requires robust evaluation mechanisms.

This research proposes a Fuzzy logic and machine learning teaching quality evaluation system using video, audio, and text data analysis to comprehensively evaluate teaching quality in online teaching scenarios (see Figure 1). In the proposed method, the teaching content of the teacher and the student's emotional engagement level are used to evaluate the teaching quality level. The main contribution of the research is the use of three types of data to evaluate teaching quality. A video dataset is used to evaluate the engagement level of the students because high-quality teaching is characterized by its ability to capture student's attention. To evaluate the teacher's teaching content, the teacher's teaching content is captured in audio form and then compared to the text that represents the course syllabus.

Integrating machine learning with fuzzy logic enables the development of systems capable of learning from data while also offering interpretable rule-based reasoning through fuzzy logic. This unique combination allows for the creation of intelligent systems that can process complex data and provide human-understandable explanations for their decisions and actions. The paper is organized as follows: Section Two presents a brief literature review, Section Three presents the proposed overall student engagement evaluation, Section Four presents the proposed teaching content evaluation, Section Five presents the Fuzzy Inference System, Section Six presents the results, and Section Seven concludes the paper.

2 Litrature Review

The evaluation of teaching quality is a cornerstone of educational quality assurance, encompassing a variety of methods, each tailored to different aspects of teaching effectiveness. Historically, these methods have included direct observations, student evaluations, self-assessments, and peer reviews. Each method offers unique insights but also presents specific challenges. For example, student evaluations, while providing immediate feedback on teacher performance from the learner's perspective, can be influenced by students' biases or grades, potentially skewing the data [2]. Peer observations involve teachers assessing each other's performances, which can foster a collaborative environment and professional growth. However, these are often criticized for their potential lack of consistency and subjectivity in assessments. Similarly, self-assessments allow educators to reflect on their teaching practices, though they may lack the objectivity necessary for formal evaluation purposes [3].

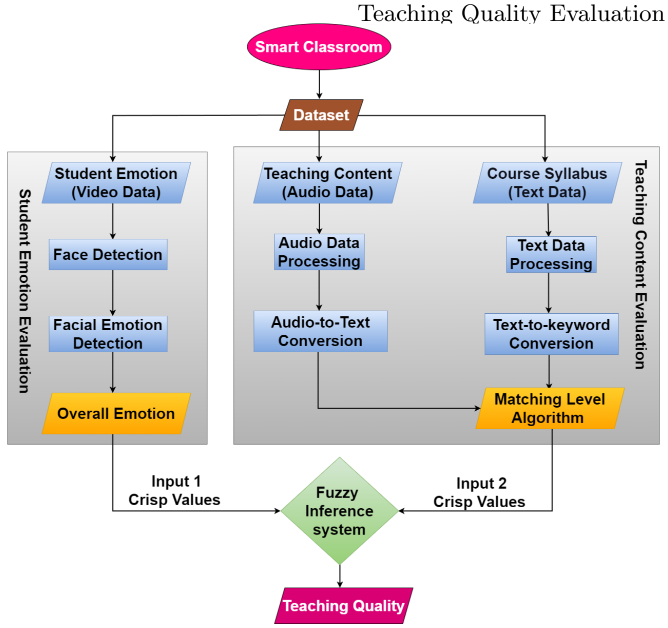


Fig. 1: Proposed Teaching quality evaluation method

The integration of advanced technologies into teaching quality evaluation processes has provided new opportunities for more objective and nuanced assessments. Artificial intelligence (AI) systems, for instance, can analyze verbal and non-verbal communication in the classroom to provide feedback on teacher-student interactions. Machine learning algorithms develop huge amounts of data generated from classroom events to offer insights into the effectiveness of teaching methods and student engagement [4]. Fuzzy Logic-based evaluation in classrooms provides comprehensive, multileveled dynamic feedback in the relevant areas. In the past decade, several researchers have used fuzzy Logic in the field of education to achieve a much finer, more accurate, and fairer assessment of students, paving the way to new opportunities for personalized, adaptive learning experiences[5–7].

Most research neglects to consider the impact of the teacher's teaching content on student engagement. Many researchers often overlook student emotional engagement as a measure of teaching quality. However, excellent teaching involves rich content and high student emotional engagement [8]. In this research, teaching quality is evaluated by assessing the teacher's teaching content and the level of student engagement. Therefore, an evaluation method has been devised that combines teaching content with student emotion assessments.

3 Proposed method for teaching quality evaluation

We used the method shown in Figure 2 to evaluate the teaching quality. The proposed method uses Fuzzy logic with inputs derived from machine learning models. The first input of the Fuzzy logic is used to determine the overall emotion of the students. The second input of the Fuzzy logic is a teaching content evaluation. The teaching content evaluation determines the level of match between the teacher's teaching content and the course syllabus.

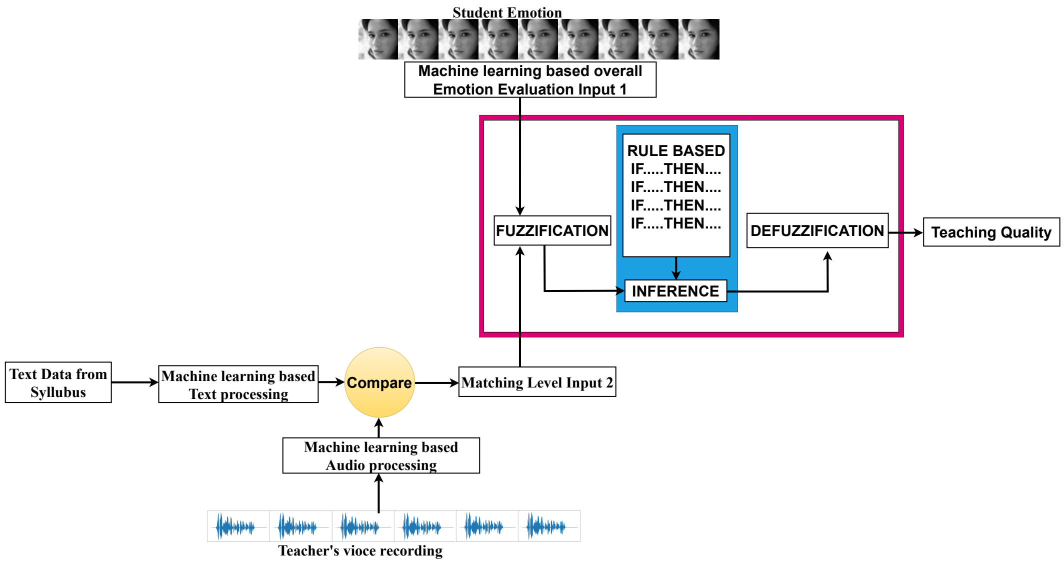


Fig. 2: Teaching Quality Evaluation with Integrated Student Emotions and Content Analysis

3.1 Fuzzy logic Input 1: Proposed machine learning based overall emotion recognition

The methodology for detecting the overall emotion from video footage involves facial detection, emotion classification and classroom overall emotion determination. The steps are described as follows:

(1) **Facial Detection:** For each frame f_i in the video, the facial region R_{f_i} is detected using the Haar Cascade Classifier, which is represented as

$$R_{f_i} = \text{DetectFaces}(f_i) \tag{1}$$

where "DetectFaces$(\cdot)$" is the face detection function applied to frame f_i.

(2) **Emotion Classification:** Each detected facial region R_{f_i} is then classified into one of the emotion categories $E = \{e_1 = \text{Anger}, e_2 = \text{Disgust}, e_3 = \text{Fear}, e_4 = \text{Sad}, e_5 = \text{Neutral}, e_6 = \text{Happy}, e_7 = \text{Surprise}\}$ using the emotion detection model. The CNN model shown in Figure 3 is used to capture the emotion from the video data. The video data is changed to several frames, and the CNN model is used to capture the emotions of the students.

The prediction for each face P_{f_i} is given by:

$$P_{f_i} = \text{EmotionModel}(R_{f_i}) \tag{2}$$

where "EmotionModel($\cdot$)" represents the emotion classification CNN model.

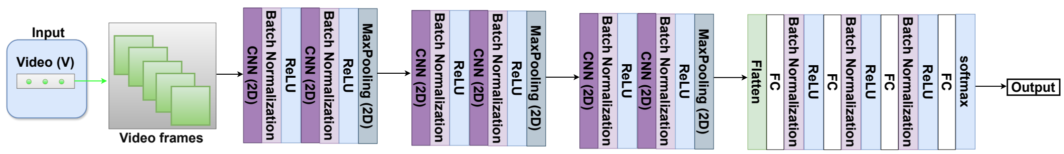


Fig. 3: The CNN model

(3) **The overall emotion $E_{\textbf{overall}}$:** is determined by identifying the emotion category with the maximum frequency of occurrence across all frames:

$$E_{\text{overall}} = \arg\max_{e \in E} \sum_{i=1}^{N} \delta(P_{f_i}, e_i) \tag{3}$$

$$\delta(P_{f_i}, e_i) = \begin{cases} 1, & \text{if } P_{f_i} = e_i \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where N is the total number of frames, and $\delta(P_{f_i}, e)$ is an indicator function that equals 1 if there is a predicted emotion or zero if there is no emotion predicted.

3.2 Fuzzy logic Input 2: Proposed teaching content evaluation method

The proposed teaching content evaluation method evaluates the alignment of the teacher's content with the course syllabus. To do this, the keywords from the course syllabus are extracted using text-to-keyword extraction. Next, the teacher's audio is processed and converted into text. Finally, the processed teacher's content and the keywords from the course syllabus are compared using the matching level algorithm to determine the level of alignment between the teacher's content and the course syllabus.

Teacher teaching Audio File-to-Text Conversion

The audio recording from the teacher teaching content is converted to text using the following steps:

(1) **Noise Reduction:** Initial preprocessing reduces ambient noise in audio recordings for clearer verbal content. This is achieved by

$$Y(t) = X(t) - \lambda \cdot N(t); \tag{5}$$

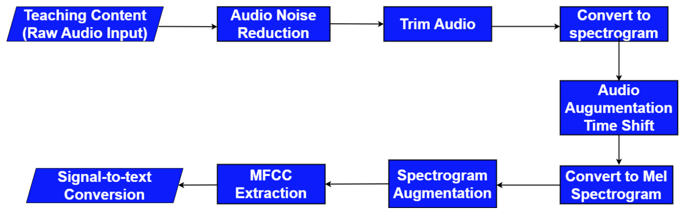


Fig. 4: Audio to text data processing

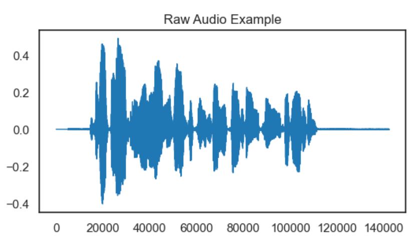
where $Y(t)$ represents the noise-attenuated signal, $X(t)$ the original signal, $N(t)$ the estimated noise, and λ the coefficient moderating the noise deduction.

(2) **Normalization:** Ensuring consistent audio volume in recordings is crucial for efficient data processing. This is achieved via:

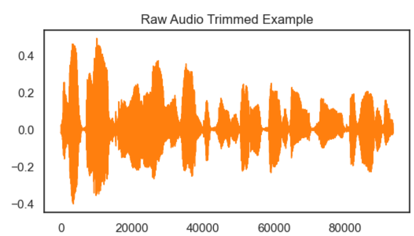
$$X(t) = \frac{Y(t)}{\max(|Y(t)|)} \cdot C, \tag{6}$$

where, $X(t)$ denotes the normalized signal, $Y(t)$ the original signal, with C as the normalization constant.

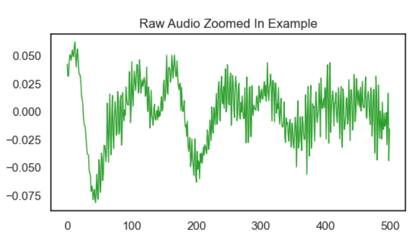
(3) **Spectrogram:** A spectrogram shows a sound signal's frequency spectrum over time. It is created using the Short-Time Fourier Transform (STFT), which breaks down the sound signal into its frequency components across brief, overlapping intervals. The mathematical framework underpinning the



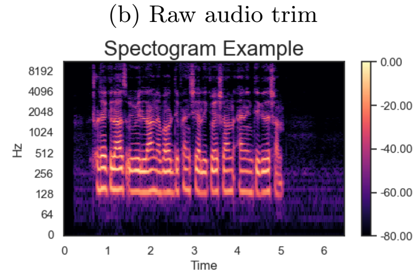
(a) Raw audio



(b) Raw audio trim



(c) Raw audio zoom



(d) Audio spectrogram

Fig. 5: Audio Processing

STFT is as

$$\text{STFT}\{X(t)\}(f,\tau) = \int_{-\infty}^{\infty} X(t) \cdot w(t-\tau) \cdot e^{-j2\pi f t} \, dt, \tag{7}$$

where, $X(t)$ signifies the sound signal under analysis, $w(t-\tau)$ denotes a window function centered at time τ, and f stands for frequency. The STFT produces a complex function representing the magnitude and phase of the signal for each frequency and time, which is then displayed as a spectrogram.

(4) **Mel-Spectrogram:**Spectrograms visually represent sound frequency, intensity, and pitch over time but don't fully account for how humans perceive sound. The mel-spectrogram addresses this by using the mel scale, which aligns better with the human ear response to different frequencies. This scale compresses higher frequencies and expands lower ones to mimic human hearing sensitivity. The mathematical transformation from frequency (f) in Hertz to the mel scale (m) employs

$$m = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \tag{8}$$

Text-to-keyword Extraction

To extract keywords from the syllabus, the method shown in Figure 6 is used. The process begins with text processing, where the text is converted to lowercase and punctuations are removed for consistency. Then, tokenization is used to convert the text to tokens for easier analysis. The next stage involves normalization, usually achieved through lemmatization, to reduce words to their root forms for consistency and accuracy in handling various word forms. Following this, stop words, such as "we," "will," "a," "is," "that," "of," "or," and "be" are removed.

After modifying the text, the keywords are extracted using the TF-IDF (Term Frequency-Inverse Document Frequency) method. TF-IDF assigns scores to each term based on its frequency within the text relative to a larger corpus, highlighting the most important terms. Finally, the terms with the highest TF-IDF scores are taken as keywords. This approach ensures that the retrieved keywords are relevant and useful.

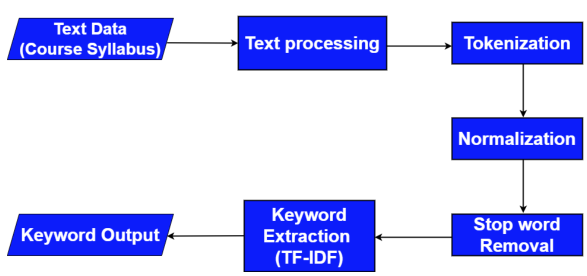


Fig. 6: Text to keywords conversion

Matching algorithm

The algorithm uses the Audio File-to-Text and Text-to-keyword extraction methods described above (see Figure 7). The output of the Matching algorithm is a percentage value that shows how much the teacher's teaching content is related to the course syllabus learning material.

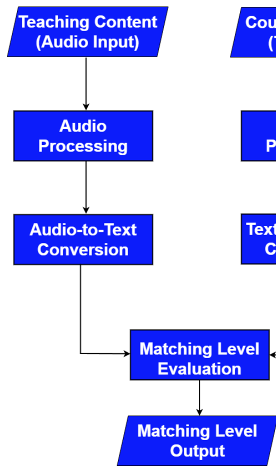


Fig. 7: Matching output between audio and text data

(1) **Speech-to-Text Conversion:** The audio file is converted into a textual transcription T, using the Audio File-to-Text $AFT(\cdot)$ conversion method:

$$T = AFT(\text{audio\_file}). \tag{9}$$

(2) **Keyword Extraction:** Keywords are retrieved from the text T extracted from audio teacher recording and the course syllabus learning materials L using Text-to-keyword extraction function $TKW(\cdot)$:

$$K_T = TKW(T), \quad K_L = TKW(L). \tag{10}$$

(3) **Match Level Computation:** The match is calculated as the proportion of keywords from the transcription (K_T) to the keywords from the course syllabus learning materials (syllabus) (K_L):

$$\text{Match\_Level} = \left(\frac{|K_T \cap K_L|}{|K_L|} \right) \times 100\% \tag{11}$$

Based on the computed match level, the final categorization into fuzzy sets: Match_level = {"no_match," "little_match," "match," "good_match," "very_good_match"} are determined using

a) **No Match:** Match_Level = 0%, indicating no keywords were found in the text.
b) **Little Match:** 0% < Match_Level ≤ 25%, a minimal presence of keywords.

c) **Match:** 25% < Match_Level ≤ 50%, a moderate level of keyword presence.
d) **Good Match:** 50% < Match_Level ≤ 75%, a majority of the keywords are present.
e) **Very Good Match:** Match_Level > 75%, an extensive presence of keywords indicating a high relevance of the text to the keyword set.

3.3 Research Components

The research components of the proposed method is shown in Table 1.

Table 1: Teaching Evaluation System: Research Components

Component	Function	Method	Accuracy	System Robustness
Camera & Audio Sensor	Track teacher and student behaviors	Video and audio data capture in real-time	High	Dependent on sensor quality and placement
Emotion Training	Train emotion detection models	Machine learning with FER2013	Moderate-High	Sensitive to data variability
Emotion Analysis	Obtain overall emotion in the classroom	Aggregate emotional data analysis	Moderate	Robust to fluctuations in data
Machine Evaluation System	Collect and process teacher's audio	Speech-to-text conversion and text-to-keyword extraction	High	Requires clear audio signals
Text and Audio Matching	Evaluate the alignment between text data and audio	Comparison algorithms	Moderate	Sensitive to discrepancies in data
Fuzzy Evaluation System	Evaluate teaching content quality	Fuzzy logic based on inputs from emotion and text matching	Varies	Adaptive to input variations

4 The Fuzzy logic design

In this section, the membership functions and the Fuzzy rules are described in detail.

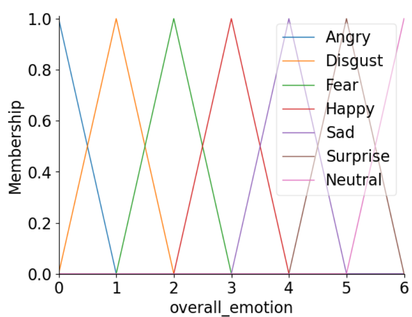
4.1 Design of membership functions

(1) **Overall emotion**: For the input variable "overall emotion," which is derived from facial emotion recognition, the membership functions are designed to

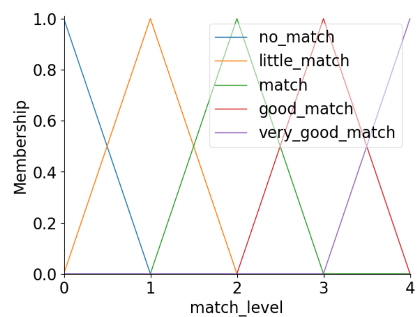
categorize the predominant emotional ambiance of a classroom. Each emotion category is assigned a specific value on the scale of 0 to 6, corresponding to the output of the emotion detection algorithm. The MF for each emotion is assigned a value from 0 to 1, depending on the overall detected emotion, as shown in Figure 8a.

(2) **Match level:** For the input variable $match\_level$, which assesses the alignment between teaching content and learning objectives, the membership functions need to reflect the continuum from no alignment to perfect alignment. Triangular MFs are used to represent the degree of match because of their simplicity and ease of interpretation. For Match_level = $\{M\_L_1 =$ "no_match", $M\_L_2 =$ "little_match", $M\_L_3 =$ "match", $M\_L_4 =$ "good_match", $M\_L_5 =$ "very_good_match"$\}$ as shown in Figure 8b.

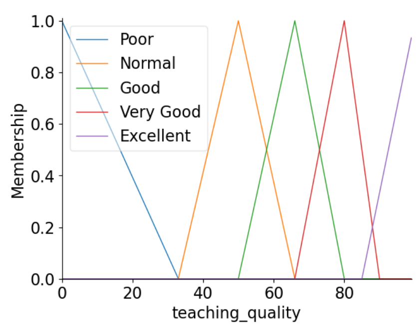
(3) **Teaching quality:** Designing membership functions for the output variable, "teaching quality," defined within the range of 0 to 100, with membership functions categorizing this range into different levels of teaching effectiveness: Poor, Normal, Good, Very Good, and Excellent (see Figure 8c).



(a) Input 1: overall emotion



(b) Input 2: matching level



(c) Fuzzy output teaching quality

Fig. 8: Membership Functions

a) **Poor Teaching Quality:** The "Poor" teaching quality is represented by a triangular membership function, capturing the lowest end of the teaching quality spectrum. This function peaks at a low score and increases as the score moves away from this low value.
b) **Normal Teaching Quality:** "Normal" teaching quality is represented by a triangular membership function that intersects with both "Poor" and "Good" at its boundaries.
c) **Good Teaching Quality** The "Good" category is also represented by a triangular membership function designed to capture the mid-range of teaching quality.
d) **Very Good Teaching Quality** For teaching quality that is better than "Good" but not quite "Excellent," the "Very Good" category is defined.
e) **Excellent Teaching Quality**Finally, "Excellent" teaching quality captures the highest tier of effectiveness, with a membership function that increases towards the upper limit of the scale.

4.2 Rule base construction

Expert knowledge is used to define the rules for teaching quality evaluation based on the overall_emotion value and the match_level value. See Table 2 for a partial list of rules.

Table 2: Sample fuzzy rules for teaching quality evaluation based on overall emotion and match level

No.	Input 1 (overall_emotion)	Input 2 (match_level)	Output (teaching_quality)
1	Happy	Very Good Match	Excellent
2	Angry	Match	Normal
3	Disgust	Good Match	Good
4	Neutral	Very Good Match	Very Good
⋮	⋮	⋮	⋮

5 Results and analysis

5.1 The Emotional classification result and analysis

The emotional classification model was tested using the 2013 Facial Emotion Recognition series dataset, which is publicly available through the Kaggle platform. It contains 35,000 48×48 grayscale pictures of faces having a range of feelings.

The key parameters for the model were a learning rate of 0.01, a batch size of 64, and an 80% training and 20% testing dataset split. An early stopping mechanism was used to avoid overfitting. The training was conducted using the FER2013 dataset with 35 epochs. The model was trained ten times, showing an accuracy ranging from 0.65 to 0.73 (see Figure 9). The confusion matrix indicated fairly good classification, with most categories above 60%.

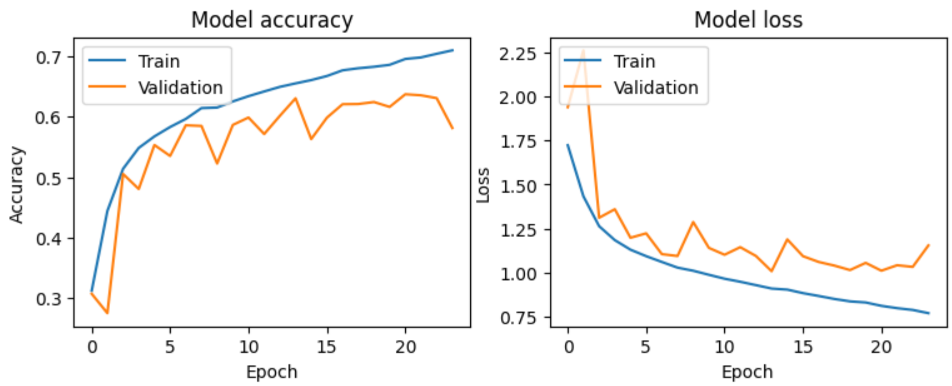


Fig. 9: The CNN model Accuracy and loss on FER2013 dataset

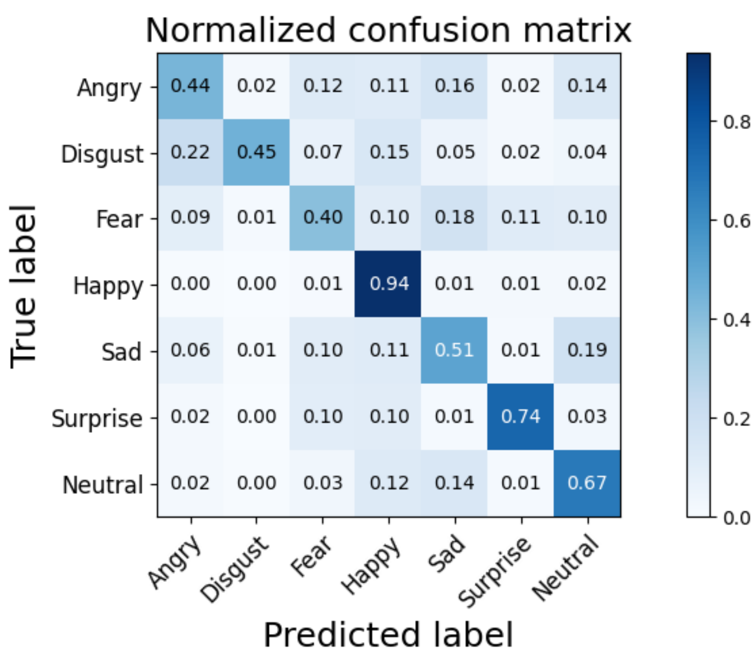


Fig. 10: Confusion Matrix on FER 2013

5.2 The Fuzzy Inference System result and analysis

The fuzzy logic model for evaluating teaching effectiveness was evaluated using the following experimented scenarios.

(1) **Scenario A High Engagement and Content Relevance:**
Inputs: Overall Emotion = Happy, Match Level = very_good_match
Output: Teaching Quality = Excellent (94.33% (see Figure 11))
This scenario reflects a situation where students are generally happy, and the teaching content closely aligns with learning objectives, resulting in an excellent teaching quality assessment.

(2) **Scenario B Moderate Engagement, Adequate Content Relevance:**
Inputs: Overall Emotion = Neutral, Match Level = match
Output: Teaching Quality = Good (65.33%)
Here, the emotional ambiance is neutral, and the content relevance is adequate, leading to a good assessment of teaching quality.

(3) **Scenario C Low Engagement, Poor Content Relevance:**
Inputs: Overall Emotion = Sad, Match Level = no_match
Output: Teaching Quality = Poor (11%)
This scenario represents a case where students are mostly sad, and the content does not meet learning objectives, resulting in a poor teaching quality rating.

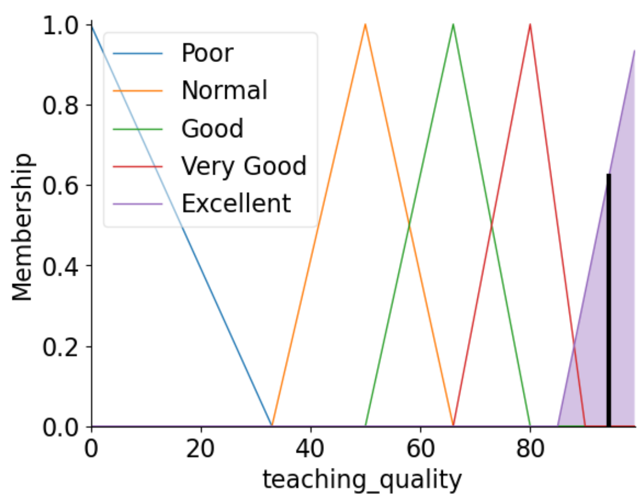


Fig. 11: Experimental scenario for excellent teaching quality

The fuzzy logic model shows that emotional engagement and content relevance are key to teaching effectiveness. Positive emotional engagement, like happiness, strongly correlates with higher evaluations of teaching quality. The alignment between teaching content and learning objectives is crucial for high-quality teaching. The model is sensitive to emotion variations, and teaching

Table 3: Comparison of Fuzzy Logic in Classroom Evaluation

Research Focus	Components Used	Method	Evaluation Focus	Implementation Context
Outcome Based Assessment Using Fuzzy Logic [9]	Calculus class performance data	Fuzzy logic for continuous performance assessment	Student performance in specific subjects	Traditional classroom settings
Fuzzy Logic and Multilevel Analysis-Based Evaluation [10]	Online classroom data, fuzzy logic model	Multilevel fuzzy logic evaluation	Teaching quality in digital environments	Online teaching and digital classrooms
Applying Fuzzy Fault Tree Method [11]	Fault tree analysis, fuzzy logic	Fuzzy fault tree method for reliability	Classroom teaching effectiveness	University classrooms
Learning Evaluation Using Fuzzy Logic [12]	Student monitoring data, fuzzy logic	Fuzzy model for student monitoring	Student engagement and learning progress	Technology-mediated classrooms
Our Project	Cameras, audio sensors, fuzzy logic system	Fuzzy logic with emotion and text matching	Teacher performance, teaching effectiveness	Distance teaching with real-time data analysis

content evaluations match level variations. This underscores the importance of a balanced approach to evaluating and enhancing educational quality.

The developed method was compared with other methods (see Table 3). Unlike other methods, the proposed teaching quality evaluation uses student engagement and teaching content assessment through video, audio and text processing. This gives the method a significant advantage over other methods.

6 Conclusion

This study presents a method for evaluating teaching quality in distance learning platforms by combining student engagement levels with the alignment of teaching content to the course syllabus. Using the FER2013 dataset, the emotion recognition component achieved a maximum accuracy of 73%, demonstrating its potential in capturing real-time student engagement during interactive classroom settings. This provides valuable insights into students' emotional responses to different teaching techniques and activities. The content evaluation component, based on keyword extraction and TF-IDF matching, is adaptable to a wide range of subjects, from technical fields to the humanities, where course materials and instructional strategies vary significantly. This flexibility ensures the method's applicability across diverse educational contexts. The combined outputs from the emotion evaluation and content evaluation components were used as inputs for a fuzzy inference model to assess teaching quality. Teaching quality levels were categorized as Excellent, Very Good, Good, Normal, and Poor. The results indicate that this approach shows promise in effectively evaluating teaching quality in distance learning environments.

Bibliography

[1] P. Wannapiroon, P. Nilsook, J. Jitsupa, and S. Chaiyarak, "Digital competences of vocational instructors with synchronous online learning in next normal education," *International Journal of Instruction*, vol. 15, no. 1, pp. 293–310, 2022.

[2] W. T. Liang, "The establishment of supported decision-making model for evaluation teachers professional development effects based on fuzzy theory," *International Journal of Learning and Teaching*, vol. 7, no. 1, 2021.

[3] F. Ma, Z. Zhu, M. Zhou, and W.-T. Pan, "Fuzzy comprehensive evaluation of classroom teaching quality of college teachers," *The International Journal of Electrical Engineering & Education*, 2020.

[4] M. Guruprasad, R. Sridhar, and S. Balasubramanian, "Fuzzy logic as a tool for evaluation of performance appraisal of faculty in higher education institutions," in *SHS web of conferences*, vol. 26, p. 01121, EDP Sciences, 2016.

[5] M. Chkiwa, M. Chkiwa, and F. Achour, "Student knowledge evaluation system: a case of application of fuzzy logic in intelligent education," in *2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA)*, pp. 1–6, IEEE, 2023.

[6] F. TUĞRUL and M. ÇİTİL, "A new perspective on evaluation system in education with intuitionistic fuzzy logic and promethee algorithm," *Journal of Universal Mathematics*, vol. 4, no. 1, pp. 13–24, 2021.

[7] N. Pokrovskaia, Y. Margulyan, Y. Lvin, and A. Bulatetskaia, "Neurotechnologies and fuzzy logic for intellectual capital evaluation in education and business," in *IOP Conference Series: Materials Science and Engineering*, vol. 940, p. 012090, IOP Publishing, 2020.

[8] S. Lavy and E. Naama-Ghanayim, "Why care about caring? linking teachers' caring and sense of meaning at work with students' self-esteem, well-being, and school engagement," *Teaching and Teacher Education*, vol. 91, p. 103046, 2020.

[9] A. Varghese, S. Kolamban, J. P. Sreedhar, and S. Nayaki, "Outcome based assessment using fuzzy logic," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 1, 2017.

[10] Z. Han, "A fuzzy logic and multilevel analysis-based evaluation algorithm for digital teaching quality in colleges and universities," *Scientific Programming*, vol. 2021, pp. 1–7, 2021.

[11] L. Wang, M. Fan, and F. Zhang, "Applying fuzzy fault tree method to evaluate the reliability of college classroom teaching," *Frontiers in Psychology*, vol. 12, p. 593068, 2021.

[12] W. R. Malvezzi, A. B. Mourão, and G. Bressan, "Learning evaluation in classroom mediated by technology model using fuzzy logic at the university of amazonas state," in *2010 IEEE Frontiers in Education Conference (FIE)*, pp. S2C–1, IEEE, 2010.