Computing and Artificial Intelligence 2024, 2(1), 1388.
https://doi.org/10.59400/cai.v2i1.1388

Article

Utilizing emotion recognition technology to enhance user experience in real-time

Yuanyuan Xu[1,*], Yin-Shan Lin[2], Xiaofan Zhou[3], Xinyang Shan[1]

[1] College of Design and Innovation, Tongji University, Shanghai 200092, China
[2] Khoury College of Computer Science, Northeastern University, Boston, MA 02115, United States
[3] Department of Computer and Information of Science and Engineering, University of Florida, Gainesville, FL 32611, United States
*** Corresponding author:** Yuanyuan Xu, ecusttethys@foxmail.com

CITATION

Xu Y, Lin YS, Zhou X, Shan X. Utilizing emotion recognition technology to enhance user experience in real-time. Computing and Artificial Intelligence. 2024; 2(1): 1388. https://doi.org/10.59400/cai.v2i1.1388

ARTICLE INFO

Received: 17 April 2024
Accepted: 30 May 2024
Available online: 14 June 2024

Abstract: In recent years, advancements in human-computer interaction (HCI) have led to the emergence of emotion recognition technology as a crucial tool for enhancing user engagement and satisfaction. This study investigates the application of emotion recognition technology in real-time environments to monitor and respond to users' emotional states, creating more personalized and intuitive interactions. The research employs convolutional neural networks (CNN) and long short-term memory networks (LSTM) to analyze facial expressions and voice emotions. The experimental design includes an experimental group that uses an emotion recognition system, which dynamically adjusts learning content based on detected emotional states, and a control group that uses a traditional online learning platform. The results show that real-time emotion monitoring and dynamic content adjustments significantly improve user experiences, with the experimental group demonstrating better engagement, learning outcomes, and overall satisfaction. Quantitative results indicate that the emotion recognition system reduced task completion time by 14.3%, lowered error rates by 50%, and increased user satisfaction by 18.4%. These findings highlight the potential of emotion recognition technology to enhance user experiences. However, challenges such as the complexity of multimodal data integration, real-time processing capabilities, and privacy and data security issues remain. Addressing these challenges is crucial for the successful implementation and widespread adoption of this technology. The paper concludes that emotion recognition technology, by providing personalized and adaptive interactions, holds significant promise for improving user experience and offers valuable insights for future research and practical applications.

Keywords: emotion recognition; user experience; human-computer interaction

1. Introduction

In recent years, human-computer interaction (HCI) has experienced rapid advancements, driven by significant technological progress. Among these advancements, emotion recognition technology has emerged as a crucial tool for enhancing user experience [1,2]. This technology leverages sophisticated algorithms to detect and interpret human emotions through various data inputs, such as facial expressions, voice intonations, and physiological signals. The ability to understand and respond to users' emotional states in real-time enables systems to provide more intuitive, engaging, and personalized interactions [3,4].

Emotion recognition technology is increasingly being integrated into a wide range of applications. In online education, for example, this technology can monitor students' emotional states and adjust teaching content dynamically to maintain engagement and improve learning outcomes [5]. In healthcare, emotion recognition can help in monitoring patients' mental health, providing timely interventions and personalized care [6]. In customer service, it can enhance user satisfaction by allowing systems to respond empathetically to users' emotions [7]. In entertainment and gaming, emotion recognition can adjust content based on players' emotional responses, creating more immersive and enjoyable experiences [8].

Despite the significant potential of emotion recognition technology, several challenges remain. Integrating multiple data sources (e.g., facial expressions, voice, physiological signals) to achieve accurate and robust emotion recognition is complex and requires sophisticated models and algorithms [2]. Ensuring real-time processing capabilities while maintaining high accuracy is another significant challenge. Additionally, issues related to privacy and data security must be addressed to ensure users' trust and compliance with regulations [9].

This paper explores how emotion recognition technology can be utilized to monitor users' emotional states in real-time and subsequently enhance user satisfaction and engagement through dynamic adjustments to the interface and content. By integrating facial expression and voice emotion analysis, this study aims to develop a comprehensive multimodal emotion recognition system and evaluate its effectiveness in an experimental setting. The research focuses on addressing the following key questions:

1) How can emotion recognition technology be effectively integrated into an online learning platform?
2) What are the impacts of real-time emotion monitoring and dynamic content adjustments on user satisfaction and engagement?
3) How do users perceive the use of emotion recognition technology in enhancing their interaction experience?

To answer these questions, this study employs a detailed experimental design involving an experimental group and a control group. The experimental group uses an online learning platform integrated with an emotion recognition system, while the control group uses a traditional online learning platform. The results of this study aim to provide insights into the practical applications of emotion recognition technology and its potential to enhance user experience across various domains.

2. Literature review

2.1. Methodologies in emotion recognition

Emotion recognition technology encompasses several methodologies, including computer vision, voice analysis, and physiological signal analysis [3,10].

Computer vision techniques identify emotions by analyzing facial expressions. These methods often utilize convolutional neural networks (CNNs) to detect and classify facial features associated with different emotions [11]. CNNs are particularly effective in recognizing complex facial expressions as they can capture subtle variations in facial features. For example, studies have shown that CNNs can differentiate between similar emotions, such as joy and excitement, by analyzing minute differences in facial muscle movements [12]. Keypoint detection is another common approach, where specific points on the face, such as the corners of the eyes and mouth, are tracked to interpret emotions. Techniques like facial action coding systems (FACS) are also employed to systematically categorize facial movements and their corresponding emotions [13].

Voice analysis methods recognize emotions by examining acoustic features such as pitch, rhythm, and volume. Hidden Markov models (HMMs) and long short-term memory networks (LSTMs) are commonly used algorithms in this domain [14]. HMMs model the temporal dynamics of speech, making them suitable for sequential data like voice recordings [15]. They excel at capturing short-term dependencies and transitions between different states of speech. LSTMs, on the other hand, are adept at capturing long-term dependencies in sequential data, making them effective for analyzing the nuances in speech that correspond to different emotional states. LSTMs can maintain information over extended periods, allowing for a more comprehensive analysis of how emotions evolve during a conversation [16].

Physiological signal analysis involves monitoring physiological signals such as heart rate, skin conductivity, and brain activity to detect emotions. Techniques like electroencephalography (EEG) and galvanic skin response (GSR) are often used. EEG measures electrical activity in the brain, providing insights into emotional states through brain wave patterns [17,18]. GSR measures the electrical conductance of the skin, which varies with sweat gland activity influenced by emotional arousal. These methods provide a direct measurement of the physiological responses associated with different emotional states, offering a complementary perspective to facial and vocal analysis. Combining these physiological signals with computer vision and voice analysis can enhance the accuracy and robustness of emotion recognition systems [19].

2.2. Applications and challenges in emotion recognition

Emotion recognition technology has a wide range of applications across various domains, significantly enhancing user experience and interaction quality.

In the context of online education, emotion recognition technology can monitor students' emotional states to adjust teaching content dynamically [5,20]. Real-time emotion monitoring can improve student engagement and learning outcomes by tailoring the educational experience to the emotional needs of the students. For instance, if a student appears frustrated, the system can offer additional support or simplify the content to alleviate frustration and enhance learning efficiency [5].

In healthcare, emotion recognition technology plays a crucial role, particularly in mental health monitoring and treatment [6]. By continuously monitoring patients' emotional states, healthcare providers can offer timely interventions and personalized care. For example, emotion recognition can help detect early signs of depression or anxiety, enabling proactive mental health management [21]. Continuous monitoring can also provide valuable data for understanding long-term emotional trends and their correlation with treatment outcomes.

In customer service, emotion recognition technology can enhance user satisfaction by allowing systems to respond empathetically to users' emotions. Automated customer service agents can adjust their responses based on the detected emotional state of the user, providing a more personalized and satisfying interaction experience. For example, an empathetic response to a frustrated customer can de-escalate tension and improve the overall customer experience [7].

In the entertainment and gaming industry, emotion recognition can create more immersive experiences by adjusting content based on players' emotional responses. For instance, a game can become more challenging if the player is detected to be bored or less intense if the player appears stressed, thereby maintaining optimal engagement levels. Emotion-driven adjustments can enhance the gaming experience by making it more adaptive and personalized [8].

Despite the advancements and potential applications, several challenges remain in the implementation of emotion recognition technology.

Integrating multiple data sources, such as facial expressions, voice, and physiological signals, to achieve accurate and robust emotion recognition is complex. Each modality provides unique information, and effectively combining these sources requires sophisticated models and algorithms [22,23]. Multimodal fusion techniques must address issues such as data synchronization, feature alignment, and the handling of missing or noisy data [24].

Ensuring real-time processing capabilities while maintaining high accuracy is a significant challenge. Emotion recognition systems must process data quickly to provide timely responses, which can be computationally intensive, especially when dealing with multiple data streams simultaneously [25]. Optimizing algorithms for speed without compromising accuracy requires advanced computational techniques and hardware acceleration.

Issues related to privacy and data security must be addressed to ensure users' trust and compliance with regulations. Emotion recognition involves collecting sensitive data, and safeguarding this information is crucial to prevent misuse and protect user privacy [26]. Implementing robust encryption, secure data storage, and strict access controls are essential measures to ensure the confidentiality and integrity of emotional data.

In summary, while emotion recognition technology holds immense potential to enhance user experience across various domains, addressing these methodological and implementation challenges is crucial for its successful deployment. Continued research and innovation in this field will pave the way for more effective and secure emotion recognition systems.

3. Methodology

3.1. Experimental design

To ensure a comprehensive and effective experimental design, this study involves an experimental group and a control group. The experimental group uses an online learning platform integrated with an emotion recognition system that can monitor participants' emotional states in real-time and dynamically adjust learning content and interfaces based on emotional changes. In contrast, the control group uses a traditional online learning platform without emotion recognition features, with learning content and interface remaining unchanged.

Participants include 30 students of different ages, genders, and backgrounds to ensure diversity and representativeness, the demographic distribution of the participants is shown in **Table 1**. Random grouping is used to divide participants into experimental and control groups, with 15 participants in each group. The experiment is set on an online learning platform that provides the following three main tasks: watching instructional videos, participating in discussions, and completing quizzes.

Table 1. Participant demographics.

Demographic attribute	Experimental group	Control group
Total participants	15	15
Gender		
Male	8	7
Female	7	8
Age		
18–24	5	5
25–30	7	6
31–35	3	4
Educational background		
Undergraduate	8	9
Graduate	7	6

3.2. Task descriptions

During the task of watching instructional videos, a camera is used to collect real-time facial expression image data from participants. The system analyzes participants' facial expressions in real-time using a CNN model to recognize emotional states such as happiness, confusion, and boredom. Based on participants' emotional states, the system dynamically adjusts video content and playback speed for the experimental group. During the task of participating in discussions, a microphone is used to collect real-time voice data from participants. The system analyzes participants' voice emotions in real-time using an LSTM model to recognize emotional states such as excitement, calmness, and anxiety. Based on participants' emotional states, the system provides interaction prompts and guidance for the experimental group.

In the task of completing quizzes, the system records participants' answer times, answer sequences, and answer accuracy. It monitors participants' facial expressions and voice emotions in real-time, analyzing emotional changes for the experimental group. The system provides instant feedback and assistance based on participants' emotional states for the experimental group.

The data collection includes facial expression data using a camera and analyzing it with a CNN model, voice emotion data using a microphone and analyzing it with an LSTM model, and user behavior data recording all participant operations on the platform.

3.3. Model selection and implementation

The study utilizes a combination of convolutional neural networks (CNN) and long short-term memory (LSTM) networks to build a multimodal emotion recognition system. The CNN model, designed for facial expression recognition, includes multiple convolutional and pooling layers to extract image features, followed by fully connected layers and a SoftMax classifier for the final classification. The architecture of the CNN model consists of an input layer that processes 48×48 grayscale images, three convolutional layers with 32, 64, and 128 filters respectively, each followed by 2 × 2 max pooling layers, and a fully connected layer with 64 neurons, concluding with a SoftMax layer that outputs seven emotion categories. The CNN model is trained using the Adam optimizer and cross-entropy loss function, leveraging the FER-2013 dataset, which contains facial images representing various emotions. The training process spans 50 epochs and incorporates data augmentation techniques to enhance generalization (**Table 2**).

Table 2. Participant demographics.

Layer	Details
Input layer	48 × 48 grayscale images
Conv layer 1	32 filters, 3 × 3 kernel, ReLU
Max pooling 1	2 × 2 pool size
Conv layer 2	64 filters, 3 × 3 kernel, ReLU
Max pooling 2	2 × 2 pool size
Conv layer 3	128 filters, 3 × 3 kernel, ReLU
Max pooling 3	2 × 2 pool size
Fully connected	64 neurons, ReLU
Output layer	SoftMax, 7 emotion categories

For voice emotion recognition, the LSTM model is employed. This model comprises multiple LSTM layers to process the time series features of voice signals, along with fully connected layers for final classification. The LSTM model's architecture includes an input layer for 40-dimensional Mel-Frequency Cepstral Coefficients (MFCC) features, two LSTM layers with 128 and 64 units respectively, and a fully connected layer with 32 neurons, culminating in a SoftMax layer with seven neurons corresponding to the emotion categories. The LSTM model is also trained using the Adam optimizer and sparse cross-entropy loss function, using the RAVDESS dataset, which includes voice recordings of various emotions. The training process involves 50 epochs with 20% of the data allocated for validation (**Table 3**).

Table 3. Participant demographics.

Layer	Details
Input layer	40-dimensional MFCC features
LSTM layer 1	128 units, tanh
LSTM layer 2	64 units, tanh
Fully connected	32 neurons, ReLU
Output layer	SoftMax, 7 emotion categories

To integrate the results from both facial expression and voice emotion recognition, a multimodal fusion approach is employed. The outputs of the CNN and LSTM models are concatenated and passed through a fully connected layer for the final emotion classification. The fusion model is trained and evaluated using an independent validation set. Performance metrics, including accuracy, recall, and F1 score, demonstrate the superior performance of the multimodal fusion model over single-modal models (**Table 4**).

Table 4. Performance metrics comparison.

Emotion category	Accuracy (CNN)	Accuracy (LSTM)	Accuracy (Fusion)	Recall (CNN)	Recall (LSTM)	Recall (Fusion)	F1 score (CNN)
Happy	0.88	0.89	0.92	0.84	0.86	0.88	0.86
Confused	0.85	0.86	0.89	0.81	0.82	0.85	0.83
Bored	0.80	0.83	0.85	0.76	0.79	0.82	0.78
Calm	0.90	0.93	0.95	0.87	0.90	0.93	0.88
Excited	0.86	0.88	0.90	0.83	0.85	0.87	0.84
Anxious	0.82	0.84	0.87	0.78	0.80	0.83	0.80
Overall	0.85	0.87	0.90	0.80	0.84	0.86	0.82

4. Results

This section provides a detailed analysis of how improved emotion classification results translate into enhanced task performance and user satisfaction. The critical link between enhanced emotion classification algorithms and task performance improvements is thoroughly discussed.

4.1. Quantitative and qualitative analysis results

The integration of advanced CNN and LSTM models allowed for accurate real-time detection of participants' emotional states. The improved classification accuracy (90% for the fusion model) enabled the system to make precise adjustments to the content, thereby enhancing the overall learning experience. This improved emotion recognition directly impacted task performance and user satisfaction in several ways (**Table 5**).

Table 5. Quantitative analysis results.

Indicator	Experimental group (Mean ± SD)	Control group (Mean ± SD)	Statistical significance (*p*-value)
Task completion time (seconds)	300 ± 45	350 ± 50	<0.05
Error rate (%)	5 ± 2	10 ± 3	<0.01
Quiz score (out of 100)	85 ± 8	75 ± 10	<0.05
User satisfaction rating (1–5)	4.5 ± 0.5	3.8 ± 0.6	<0.01

To evaluate task completion time, we measured the duration each participant took to complete the assigned tasks. The results indicated that participants in the experimental group completed their tasks significantly faster than those in the control group. Specifically, the average task completion time for the experimental group was 300 ± 45 s, whereas the control group took an average of 350 ± 50 s. This suggests that the emotion recognition system's real-time adjustments to the learning content helped maintain participant engagement and efficiency.

Error rate was another critical metric, particularly for the quizzes. Participants in the experimental group had a notably lower error rate, averaging 5% ± 2%, compared to the control group's 10% ± 3%. This reduction in errors indicates that the emotion recognition system could identify and mitigate participants' anxiety and confusion during quizzes, providing calming and supportive feedback that contributed to more accurate responses.

Quiz scores further highlighted the benefits of the emotion recognition system. The experimental group achieved higher scores, averaging 85 ± 8 out of 100, while the control group averaged 75 ± 10. This performance gap underscores the effectiveness of real-time emotional monitoring and content adjustment in enhancing participants' understanding and retention of the material.

User satisfaction was assessed through a questionnaire survey, utilizing a Likert scale to gauge participants' satisfaction with their learning experience. The survey covered various aspects of the learning experience, including the relevance and practicality of the content, the effectiveness of interaction discussions, the friendliness of the system interface, and the timeliness of emotional feedback.

The results from the satisfaction ratings revealed that the experimental group had significantly higher satisfaction levels compared to the control group. The experimental group's average satisfaction rating was 4.5 ± 0.5 on a 5-point scale, indicating a high level of satisfaction. In contrast, the control group had an average satisfaction rating of 3.8 ± 0.6. The higher satisfaction ratings in the experimental group suggest that the emotion recognition system's ability to provide personalized learning experiences and timely emotional feedback contributed to a more positive learning experience overall.

In summary, the quantitative analysis demonstrates that the emotion recognition system significantly improved task completion performance and user satisfaction. The system's real-time emotional monitoring and content adjustments led to faster task completion times, lower error rates, higher quiz scores, and greater overall satisfaction among participants in the experimental group compared to those in the control group.

Table 6. Qualitative analysis results.

Theme	Experimental group feedback example	Control group feedback example
Overall learning experience	"The system's timely content adjustments kept me focused."	"Sometimes the video content felt monotonous."
Satisfaction with interaction discussions	"The discussion prompts and guidance were very helpful."	"It's hard to stay active in discussions, lacking motivation."
Emotional changes during quizzes	"The system's feedback during quizzes reduced my anxiety."	"I felt a lot of pressure during quizzes without much help."
Views on the emotion recognition system	"The system's understanding of my emotions and adjustments were great."	"Without the emotion recognition system, the learning process sometimes felt dull."
Suggestions for future improvements	"More interactive content and richer learning resources would be great."	"I hope the platform can be more personalized, providing content adjustments based on my emotions."

The qualitative analysis of the experimental and control groups provides insights into participants' subjective experiences with the learning platform (**Table 6**). Feedback was collected and organized into key themes, highlighting differences in the overall learning experience, satisfaction with interaction discussions, emotional changes during quizzes, views on the emotion recognition system, and suggestions for future improvements.

Participants in the experimental group frequently mentioned the benefits of the emotion recognition system's real-time adjustments. For example, one participant stated that the system's timely content adjustments kept them focused, indicating that the adaptive nature of the platform helped maintain their engagement. In contrast, a participant from the control group remarked that sometimes the video content felt monotonous, suggesting that the lack of dynamic adjustments led to a less engaging experience.

The experimental group found the discussion prompts and guidance provided by the system to be very beneficial. One participant noted that the discussion prompts and guidance were very helpful, indicating enhanced interaction and communication. On the other hand, a control group participant expressed difficulty staying motivated, saying it was hard to stay active in discussions, lacking motivation.

During quizzes, participants in the experimental group reported reduced anxiety due to the system's feedback. One participant shared that the system's feedback during quizzes reduced their anxiety, highlighting the supportive role of the emotion recognition system. Conversely, a control group participant felt pressured, stating that they felt a lot of pressure during quizzes without much help.

Participants in the experimental group appreciated the system's ability to understand and respond to their emotions. One participant mentioned that the system's understanding of their emotions and adjustments were great, reflecting satisfaction with the personalized experience. In contrast, a control group participant noted that without the emotion recognition system, the learning process sometimes felt dull, indicating a preference for the enhanced interaction provided by the system.

Participants in the experimental group suggested further enhancements, such as more interactive content and richer learning resources. This feedback points to a desire for even more engaging and varied content. Similarly, a control group participant expressed a need for personalization, stating that they hope the platform can be more personalized, providing content adjustments based on their emotions.

4.2. Linking enhanced emotion classification to performance improvement

The key to the improved task performance and user satisfaction lies in the sophisticated emotion classification algorithms employed. The CNN and LSTM models provided high accuracy in detecting subtle emotional cues from facial expressions and voice intonations. These accurate classifications allowed the system to personalize content by recognizing when a participant was confused or bored and adjusting the content to better match the participant's emotional state, thereby maintaining engagement and improving comprehension. During discussions and quizzes, the system's ability to detect anxiety or excitement allowed it to provide immediate and relevant feedback, reducing stress and enhancing performance. The dynamic adjustments based on emotional states kept participants engaged throughout the tasks, preventing disengagement and promoting a more immersive learning experience.

The experimental results support the research hypothesis that the emotion recognition system significantly improves users' learning experience and satisfaction. The enhanced emotion classification algorithms enabled precise real-time adjustments to the content, which directly translated into better task performance and higher user satisfaction. The system's ability to monitor and respond to users' emotional states in real-time and provide personalized interactions highlights the potential of emotion recognition technology in various applications, including online education, healthcare, customer service, and entertainment.

5. Discussion

5.1. Innovation and problem solving

This study presents a novel approach to enhancing user experience through the integration of advanced emotion recognition technology. By leveraging sophisticated CNN and LSTM models, we achieved high accuracy in detecting emotional states in real-time, enabling dynamic adjustments to content and interactions based on users' emotions. Our research addresses several critical challenges in the field of emotion recognition and user experience. Specifically, our work introduces a comprehensive multimodal emotion recognition system that combines facial expression analysis with voice emotion recognition. The innovative use of CNN and LSTM models allows for the accurate detection of subtle emotional cues, leading to precise content adjustments. This integration addresses the challenges of real-time emotional monitoring and personalized user experience, enhancing engagement and satisfaction, particularly in online education where it significantly improves learning outcomes.

5.2. Advancing the field of emotion recognition and user experience

Our findings contribute to the advancement of emotion recognition technology and its application in improving user experiences. The successful integration of CNN and LSTM models demonstrates the potential of multimodal approaches in achieving higher accuracy and robustness in emotion detection. This research pushes the boundaries of current methodologies and provides a framework for future studies to build upon. By showcasing the effectiveness of combining different neural network models, our study sets a new standard for future research in this area. The practical implications of our emotion recognition system are profound, as it can improve user experiences across various domains, including online education, healthcare, customer service, and entertainment, by providing real-time, personalized feedback.

5.3. Application to real-world scenarios

The insights gained from our research have significant practical applications. In online education, our system can be integrated into learning platforms to provide adaptive learning experiences, helping students stay engaged and improve their academic performance. In healthcare, the system can be used to monitor patients' emotional states, providing timely interventions and personalized care. Customer service can benefit from the system's ability to detect and respond to customers' emotions, leading to more empathetic and effective interactions. In entertainment, the system can enhance user engagement by adapting content based on real-time emotional feedback. Overall, our research demonstrates the transformative potential of emotion recognition technology in enhancing user experiences across various domains, addressing the challenges of real-time emotional monitoring and personalized content delivery, and paving the way for future innovations in this field.

6. Conclusion

This paper studies the application of emotion recognition technology in enhancing user experience, and designs and implements an experiment to verify the effect of emotion recognition systems on an online learning platform. The experimental results show that emotion recognition technology can significantly improve users' learning experience and satisfaction, supporting the effectiveness and practicality of emotion recognition. By comparing the experimental and control groups, it was found that the emotion recognition system has significant advantages in real-time monitoring of users' emotional states and dynamically adjusting learning content and interfaces. Specifically, the experimental group participants performed significantly better than the control group in task completion time, error rate, and quiz scores. The experimental group participants also scored higher in all dimensions of user satisfaction, especially in the relevance and practicality of the content, the effectiveness of interaction discussions, the friendliness of the system interface, and the timeliness of emotional feedback. These results verify the potential of emotion recognition technology in enhancing user experience, indicating that emotion recognition can achieve personalized learning experiences, thus improving learning effectiveness and user satisfaction.

In terms of practical implications, emotion recognition systems hold substantial value across various domains. In online education, these systems can monitor students' emotional states in real-time, dynamically adjusting teaching content and interaction methods to improve student engagement and learning outcomes. For massive open online courses (MOOCs), emotion recognition technology can help teachers understand students' learning states, adjust teaching strategies promptly, and provide personalized teaching services. In remote medical care, emotion recognition systems can monitor patients' emotional states, helping doctors better understand patients' emotions and psychological states to provide more targeted treatment plans. This is particularly useful in psychological counseling and mental health fields, where real-time understanding of patients' emotional changes can facilitate timely psychological intervention and support.

In the realm of intelligent customer service, emotion recognition technology can enhance the responsiveness and personalization of customer service robots by understanding users' emotional states in real-time. This leads to improved user satisfaction and service quality, especially in handling customer complaints and after-sales service where timely identification and response to user dissatisfaction can prevent conflict escalation. Furthermore, in human-computer interaction contexts such as smart homes and intelligent assistants, emotion recognition technology can enhance emotional understanding, enabling systems to adjust interaction methods based on users' emotional states, thereby providing more humanized services. In gaming and entertainment, emotion recognition technology can dynamically adjust game difficulty and plot direction based on players' emotional states, enhancing the gaming experience and player satisfaction.

For user experience design, emotion recognition technology enables systems to monitor users' emotional states in real-time and dynamically adjust content and interfaces based on emotional changes, providing personalized user experiences. Designers should leverage this technology to develop personalized interactive content and interfaces, thus improving user satisfaction. Moreover, systems should incorporate emotional feedback mechanisms to identify users' negative emotions, such as confusion, anxiety, and boredom, and provide appropriate feedback and assistance to alleviate these emotions and improve the quality of user experience. It is also crucial to prioritize user emotional data protection, ensuring that the collection, storage, and use of emotional data comply with legal regulations and ethical norms to protect users' privacy rights and data security.

Designers should consider applying multimodal emotion recognition technology in user experience design, combining facial expressions, voice, body language, and physiological signals to improve the accuracy and robustness of emotion recognition. This approach provides a more comprehensive and accurate emotional understanding and response. In summary, emotion recognition technology has significant practical application value and broad prospects for enhancing user experience. Designers should fully utilize this technology to develop innovative applications and services, thereby improving user satisfaction and experience quality.

Author contributions: Conceptualization, YX and YSL; methodology, YX, XZ and XS; software, YX and YSL; validation, YX, XZ and XS; formal analysis, YX, XZ and XS; investigation, YSL; resources, YSL; data curation, YX, XZ and XS; writing—original draft preparation, YX, XZ and XS; writing—review and editing, YSL; visualization, YX, XZ and XS; supervision, YSL; project administration, YSL; funding acquisition, YSL. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Chen CM, Wang HP. Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. Library & Information Science Research. 2011; 33(3): 244-255. doi: 10.1016/j.lisr.2010.09.010
2. Lim JZ, Mountstephens J, Teo J. Emotion recognition using eye-tracking: taxonomy, review and current challenges. Sensors. 2020; 20(8): 2384. doi: 10.3390/s20082384
3. Dzedzickis A, Kaklauskas A, Bucinskas V. Human emotion recognition: Review of sensors and methods. Sensors. 2020; 20(3): 592. doi: 10.3390/s20030592
4. Abdullah SMSA, Ameen SYA, Sadeeq MAM, et al. Multimodal emotion recognition using deep learning. Journal of Applied Science and Technology Trends. 2021; 2(01): 73-79. doi: 10.38094/jastt20291
5. Savchenko AV, Savchenko LV, Makarov I. Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. IEEE Transactions on Affective Computing. 2022; 13(4): 2132-2143. doi: 10.1109/taffc.2022.3188390
6. Hasnul MA, Aziz NAA, Alelyani S, et al. Electrocardiogram-based emotion recognition systems and their applications in healthcare—A review. Sensors. 2021; 21(15): 5015. doi: 10.3390/s21155015
7. Saxena A, Khanna A, Gupta D. Emotion recognition and detection methods: A comprehensive survey. Journal of Artificial Intelligence and Systems. 2020; 2(1): 53-79. doi: 10.33969/ais.2020.21005
8. Silva GM, Souto JJS, Fernandes TP, et al. Interventions with serious games and entertainment games in autism spectrum

disorder: a systematic review. Developmental Neuropsychology. 2021; 46(7): 463-485. 10.1080/87565641.2021.1981905
9. Hassouneh A, Mutawa AM, Murugappan M. Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked. 2020; 20: 100372. doi: 10.1016/j.imu.2020.100372
10. Akçay MB, Oğuz K. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Communication. 2020; 116: 56-76. doi: 10.1016/j.specom.2019.12.001
11. Akhand MAH, Roy S, Siddique N, et al. Facial emotion recognition using transfer learning in the deep CNN. Electronics. 2021; 10(9): 1036. doi: 10.3390/electronics10091036
12. Khaireddin Y, Chen Z. Facial emotion recognition: State of the art performance on FER2013. arXiv. 2021; arXiv:2105.03588
13. Canal FZ, Müller TR, Matias JC, et al. A survey on facial emotion recognition techniques: A state-of-the-art literature review. Information Sciences. 2022; 582: 593-617. doi: 10.1016/j.ins.2021.10.005
14. Lameris H, Mehta S, Henter GE, et al. Prosody-controllable spontaneous TTS with neural HMMs. In: Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2023. pp. 1-5.
15. Syed SA, Rashid M, Hussain S, et al. Comparative analysis of CNN and RNN for voice pathology detection. BioMed Research International. 2021; 2021: 1-8.
16. Rizvi DR, Nissar I, Masood S, et al. An LSTM based deep learning model for voice-based detection of Parkinson’s disease. International Journal of Advanced Science and Technology 2020; 29(8).
17. Gao Z, Dang W, Wang X, et al. Complex networks and deep learning for EEG signal analysis. Cognitive Neurodynamics. 2021; 15(3): 369-388. doi: 10.1007/s11571-020-09626-1
18. Arsalan A, Majid M. Human stress classification during public speaking using physiological signals. Computers in Biology and Medicine. 2021; 133: 104377. doi: 10.1016/j.compbiomed.2021.104377
19. Niu X, Yu Z, Han H, et al. Video-based remote physiological measurement via cross-verified feature disentangling. In: Proceedings of the Computer Vision—ECCV 2020: 16th European Conference; 23–28 August 2020; Glasgow, UK. pp. 295-310.
20. Wang W, Xu K, Niu H, et al. Emotion recognition of students based on facial expressions in online education based on the perspective of computer simulation. Complexity. 2020; 2020: 1-9. doi: 10.1155/2020/4065207
21. Ayata D, Yaslan Y, Kamasak ME. Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems. Journal of Medical and Biological Engineering. 2020; 40(2): 149-157. doi: 10.1007/s40846-019-00505-7
22. Fei Z, Yang E, Li DDU, et al. Deep convolution network based emotion analysis towards mental health care. Neurocomputing. 2020; 388: 212-227. doi: 10.1016/j.neucom.2020.01.034
23. Zhang J, Yin Z, Chen P, et al. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion. 2020; 59: 103-126. doi: 10.1016/j.inffus.2020.01.011
24. Hu J, Liu Y, Zhao J, et al. MMGCN: Multimodal fusion via deep graph convolution network for emotion recognition in conversation. arXiv. 2021; arXiv:2107.06779.
25. Park CY, Cha N, Kang S, et al. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. Scientific Data 2020; 7(1): 293. doi: 10.1038/s41597-020-00630-y
26. Adjabi I, Ouahabi A, Benzaoui A, et al. Past, Present, and Future of Face Recognition: A Review. Electronics. 2020; 9(8): 1188. doi: 10.3390/electronics9081188