J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 2057--2070

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Reading Bimodal Emotion Recognition and Psychological Analysis on the Big Data Blockchain Network Platform

Minghui Shan[1], Xuezhu Liu[2], Dianqing Jiang[2,✉], Yuqi Qiao[3]

[1] *School of Education and Science, Jiamusi University, Jiamusi, Heilongjiang, 154007, China*

[2] *Library, Jiamusi University, Jiamusi, Heilongjiang, 154007, China*

[3] *School of Innovation and Entrepreneurship, Jiamusi University, Jiamusi, Heilongjiang, 154007, China*

ABSTRACT

In order to analyze the reading behavior and its meaning of readers in blockchain online reading platforms, this article conducted research on reading emotion recognition. This article utilized the characteristics of blockchain technology to analyze the reading mode of blockchain internet platforms. By using audio and image bimodal recognition methods, the recognition of readers' reading emotions can be achieved. After feature extraction of speech and facial images, hidden Markov models (HMM) can be used for speech emotion recognition. Support vector machines (SVM) can be used for facial image emotion recognition, and decision level fusion can be used for bimodal emotion recognition. This article obtained the final emotion recognition results to analyze and predict user reading behavior. Analyzing the psychological state of readers based on emotional recognition results can achieve more intelligent reading information push. Experimental results on the effectiveness of reading bimodal emotion recognition showed that the accuracy of reading bimodal emotion recognition based on decision level fusion was much higher than that of single modal emotion recognition. The bimodal method has an average accuracy rate of over 85% in emotion recognition and has a high effect in emotion recognition. Reading bimodal emotion recognition based on audio and image can accurately identify readers' emotions, adjust information push content in a timely manner, and achieve the regulation of readers' emotions, which has high application value.

Keywords: Reading Emotions and Psychology, Bimodal Emotion Recognition, Blockchain Network Reading, Big Data

1. Introduction

Reading is an important way for people to acquire knowledge, convey emotions, and express ideas,

✉ Corresponding author.
E-mail address: tushujiangou@163.com (D. Jiang).
Received 15 April 2024; Revised 05 July 2024; Accepted 25 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-116

as well as an important way of educating people. With the development of the internet, various reading platforms have emerged, and it has become a common phenomenon for people to read through online platforms. The development of online platforms has changed people's reading habits and formed new reading habits, which has led to a reading revolution. At the same time, when conducting online reading, it is possible to analyze readers' emotions, achieve intelligent analysis of readers, and use big data to provide more suitable reading push for readers, which is currently a hot research topic.

Blockchain technology is a distributed ledger technology that uses encryption technology to ensure the integrity and consistency of data. It is considered a disruptive technology for the new information revolution triggered after the internet and mobile internet [1]. Although it is widely used in fields such as social management, its application in the field of reading is relatively limited. Blockchain technology has decentralized, tamper resistant, and traceable features, which can bring new reading experiences [2]. On this basis, the characteristics of blockchain technology can be combined to apply it to reading, providing readers with a safe and reliable reading environment, and analyzing readers' psychological state during reading through blockchain technology.

In the field of audio and video emotion recognition, single mode emotion recognition has achieved good results. In order to fully utilize the emotional information of sound signals and facial expressions, many scholars have conducted extensive research on audio and video bimodal emotion recognition [3]. The integration strategies for audio and video emotion detection are divided into three categories, including feature layer fusion, classification layer fusion, and decision layer fusion [4]. Among them, decision layer fusion is only trained and does not require complete synchronization of audio and video signals, so it is widely used [5]. Bimodal emotion recognition can be applied to reading emotion recognition, with good recognition results.

The application of emerging technologies such as big data and blockchain in various fields is reshaping the socio-economic structure. How to understand and handle the impact of these technologies is a hot topic of general concern in society. This article, represented by blockchain technology, explores the emotional states of users in the reading mode of internet platforms by analyzing the reading mode. This article combines the psychological influencing factors of the psychological analysis research platform and proposes an intelligent reading push method based on user reading emotion analysis. The effectiveness of reading bimodal emotion recognition is verified through experiments.

2. Blockchain Internet Platform Reading Mode

2.1. Publication and Distribution

Copyright is a very important issue in the development of the digital reading industry. This is not only a legal issue, but also reflects the authenticity and reliability of digital content [6]. In the online environment, the impact of the online environment on reading cannot be ignored. Authors generally want to obtain control over copyright through publishing contracts, but publishers can also use other methods to limit these rights, such as limiting printing quantity and pricing [7].

The emergence of blockchain technology provides the possibility to solve the aforementioned problems. Blockchain is a decentralized database that stores all data in a decentralized network, eliminating the need to rely on central power [8]. On this basis, authors can register their works using blockchain technology and track and review them after publication, thereby achieving the protection of their works.

2.2. Reading Management

A decentralized, traceable, and tamper resistant digital reading system constructed using blockchain technology can effectively maintain the copyright relationship between copyright owners and readers. During the reading process, users would be recorded and traced back to the source of the content they read.

Under this system, users can purchase and download content using cryptocurrency. For certain paid e-books, anime, etc., blockchain can be used to achieve this [9]. In addition, it can also provide readers with functions such as “sharing” and “reading”, allowing them to share their reading materials online.

2.3. Reading Platform

Blockchain technology can provide a new way for electronic reading without the need to create new or central databases. The platform would adopt a distributed, highly autonomous approach and be managed through intelligent contracts. The system can also provide corresponding services to other platforms. On this platform, readers can directly browse to their favorite articles. In addition, the platform can also enable two-way communication between users. This not only helps users make the right choices during the reading process, but also maintains the copyright relationship between authors and readers.

2.4. Book Search

Although electronic reading has been going on for a considerable period of time, people are still accustomed to reading through traditional paper books. This is an ancient form of reading that predates the popularity of the internet. Although this reading method has its irreplaceable advantages, it also has certain drawbacks. Its biggest drawback is that it cannot store too much information. There are tens of thousands of books on the Internet, making it difficult for users to find the books they need in a short period of time.

With the progress of network technology, this problem is gradually being solved. There are many books on the Internet that users can find and read, and the idea of using blockchain technology to achieve a digital library has been proposed. Therefore, it is necessary to introduce blockchain technology in digital libraries to improve book retrieval efficiency. In addition, using blockchain technology can effectively solve the problems of large data volume and slow transmission speed in traditional databases [10].

2.5. Digital Library

How to ensure the security of information resources is the main problem faced by digital libraries. In the online environment, the number of e-books on the internet is increasing day by day. Therefore, how to protect the copyright of books in the online environment is a very important issue. Based on blockchain technology, new methods of intellectual property protection can be created.

For example, this article utilizes blockchain technology to build a decentralized digital library system. During the process of reading e-books, users can directly upload the e-book to the system based on the smart contract, and can authenticate the origin, copyright, etc., of the e-book. In addition, blockchain technology can be utilized to enable each digital resource to have its own encryption key, making all books in the digital library genuine.

2.6. E-commerce

A major feature of blockchain technology is decentralization, which would change the basic architecture of e-commerce and online reading. Credit issue is a thorny issue in e-commerce. In online transactions, it is difficult to conduct online transactions due to mutual distrust between both parties. Blockchain technology ensures the accuracy of digital information, and in addition, the system can effectively reduce errors and fraud during payments.

3. Bimodal Emotion Recognition during the Reading Process

This article conducts emotion recognition on the speech and facial images collected during the reading process. The single-mode and bimodal emotion recognition processes are shown in Figure 1.

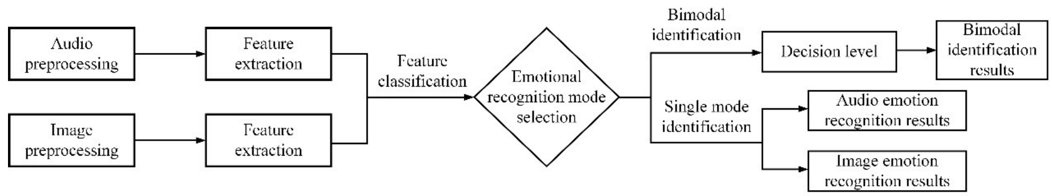


Fig. 1. Emotional recognition process

3.1. Speech Emotion Recognition

When conducting speech recognition research, it is common to use phonological features, spectral related features, sound quality features, etc., or to combine these features [11]. These features are usually performed in terms of time and frequency for time-domain analysis of speech signals. However, the time-domain waveform of speech signals is more susceptible to external environmental noise, and the spectrum has good robustness to noise, and can clearly represent acoustic features such as resonance peak parameters and gene cycle parameters of speech. Therefore, this article would conduct research on speech emotion recognition methods based on spectrograms to better reflect the time-frequency characteristics of speech emotions.

By analyzing the continuous spectrum of speech signals, a two-dimensional graph that can express three-dimensional information can be obtained. As a visual language expression, sound spectrum is a time-frequency distribution of language energy with time-frequency characteristics on a two-dimensional plane, which can connect the time domain with the frequency domain. It can directly reflect the time-frequency characteristics of language and has important application value in related fields such as speech recognition [12].

There are two types of spectrum: narrowband and broadband. The narrowband spectrum has a small bandwidth and a large time bandwidth, which has high spectral resolution ability. It can effectively distinguish the basic frequency band and various harmonics of the signal, but its time resolution ability is poor, presenting a “horizontal line” in the spectrum. Correspondingly, the broadband spectrum has a bandwidth of about 300 Hz and higher time resolution, which can clearly display the frequency of resonance peaks and the energy accumulation area of clear consonants. It presents a “vertical line” in the spectrum, which has the problem of low spectral resolution.

3.1.1. Generation of Spectrograms. According to the characteristics of short-term stability and randomness of speech signal, the methods of framing, windowing and short-time Fourier transform are used to obtain the energy spectral density of speech signal.

1) Framing. As sound is emitted with the movement of the vocal tract, its resonant peak characteristics also change over time. Due to the fact that speech signals can maintain their features unchanged and remain relatively stable within the range of 10 to 30 ms under short-term stability, it is necessary to frame them.

2) Add window. Usually, the windowed speech signal is obtained by multiplying the window function w (n) with the speech signal X(n). Common window functions include rectangular windows, Hanning windows, Hamming windows, Gaussian windows, etc.

The main lobe of a rectangular window is relatively concentrated and operates quickly, but its side lobe amplitude is large, which can easily cause high-frequency interference and spectrum leakage during conversion. Without considering the amplitude, a rectangular window can be selected. Compared with rectangular windows, Hamming windows have wider main lobes and smaller side lobe amplitudes. It solves the problem of spectrum leakage to some extent, while maintaining more details of speech waveform while low-pass filtering, which meets the requirements of time-frequency analysis.

However, due to the increase in the main lobe of the Hamming window, the signal bandwidth it can obtain increases, resulting in a decrease in its spectral resolution. Like the Hamming window, the Hanning window is also a very practical window function, with the only difference being the weight. In speech signal recognition, Hamming window is used for window processing, and the length of the window can be calculated from the length of the window function.

$$n = round\left(44 * fs / 1000\right) \tag{1}$$

Among them, fs is the sampling frequency of the voice. In the Hamming window, the form of calling a function is as follows:

$$w = ham\,min\,g\left(n\right) \tag{2}$$

Each frame after windowing can be considered as a short-term stable signal.

3) Short time Fourier transform. The DTFT (Discrete Time Fourier Transform) formula for signal X(n) is:

$$X_n\left(e^{jw}\right) = \sum_{m=-\infty}^{\infty} x\left(m\right) \cdot w\left(n-w\right) e^{-jwm} \tag{3}$$

Here, $X_n\left(e^{jw}\right)$ is a function of w and n. If $w = 2\pi kj / N$ is set, the DTFT expression for signal X(n) can be transformed into:

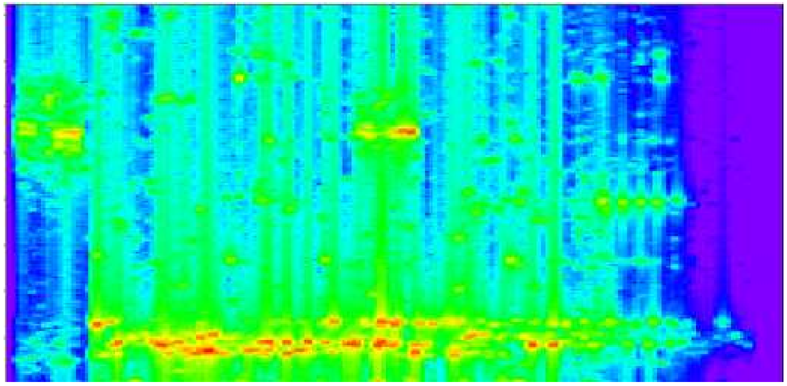
$$X_n\left(K\right) = X_n\left(e^{2\pi kj/N}\right) = \sum_{m=-\infty}^{\infty} X_n\left(m\right) \cdot w\left(n-m\right) e^{-2\pi kj/N} \tag{4}$$

According to the definition of the power spectrum function, it can be seen that the relationship between the short-term power spectrum and the short-term Fourier transform is:

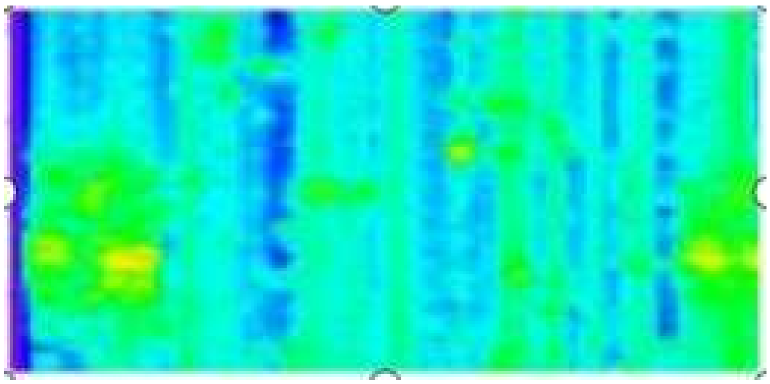
$$S_n\left(e^{jw}\right) = X_n\left(e^{jw}\right) \cdot X_n\left(e^{jw}\right)^* = \left|X_n\left(e^{jw}\right)\right|^2 \tag{5}$$

That is to say, the short-term power spectrum is equal to the square of the amplitude of the short-term Fourier transform, with time n being the abscissa and W being the ordinate. The value of $S_n\left(e^{jw}\right)$ represents the gray value of points $\left(x, w\right)$, and the generated two-dimensional image is a spectrogram of sound.

3.1.2. Preprocessing of Spectrograms. Due to the fact that the generated spectral boundary information is not suitable for speech emotion recognition, in order to remove useless information, it is necessary to segment the spectrogram generated by the program. Secondly, in order to facilitate training, it is necessary to scale the input signal spectrum into a 256x256 image. The preprocessing results of the spectrogram are shown in Figure 2.



(a) Original spectrogram



(b) Preprocessed spectrogram

Fig. 2. Preprocessing results of spectrogram

There are significant differences in the frequency spectrum generated by speech signals under different emotional states, and the energy gathering sites are also different. When a person is angry or happy, their emotions fluctuate greatly, and in the spectrogram, they exhibit a higher energy and more dark parts. As emotions change, speech presents different forms of expression, forming corresponding spectrograms, which can be used to identify the speaker's emotions and grasp their emotional changes. On this basis, corresponding spectral features are extracted for different emotions.

3.1.3. Speech Emotion Classification. HMM is a widely used statistical analysis model in the field of pattern recognition, used for speech emotion recognition.

HMM is described by two state sets, three probability matrices, and a total of five elements, which means that HMM can be expressed by $\lambda = (N, M, \pi, A, B)$. The meaning of specific parameters in the HMM model is as follows:

N: The number of hidden states in the Markov model, where N states are represented as $S_1, S_2, \ldots, S_N$ and the Markov state at time t is q_t.

M: It represents the number of observation values corresponding to each state, and the value of the observation sequence can be expressed as $V_1, V_2, \ldots, V_M$, and the observation value at time t is called O_t.

The possibility matrix of the initial state, $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$.

A: The possibility matrix of an implicit state transition, $A = (a_{ij})_{N \times N}$.

B: The probability transformation matrix of the observed state, for discrete HMM, HMM,

$B = (b_{jk})_{N \times M}$, b_{jk} is the output probability S_j of the observed symbol k in the state.

In order to achieve speech emotion recognition based on HMM, it is necessary to train the HMM model and select 7 emotion states. Therefore, the Gaussian mixture number of the model is 7, which is $M = 7$.

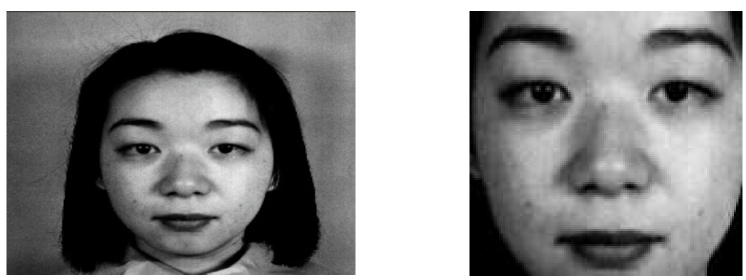
This method divides the data in the corpus into two categories: one is used for training HMM models; Another group is used for cognitive experiments. By analyzing features such as speech speed, fundamental frequency, energy, and resonance peaks, the above features are combined into speech features to obtain more valuable speech features, and based on this, speech emotion recognition is achieved. Through training, each emotion has a corresponding HMM model, resulting in a total of 7 emotion models.

3.2. Facial Expression and Emotion Recognition

Facial expression recognition is essentially the recognition of the position and movement of facial muscles [13]. Facial expression recognition first requires obtaining images through cameras, photos, videos, and other methods, followed by facial detection and image preprocessing, and then facial recognition through feature extraction and classification. Face recognition is used to extract parts of the face, and image preprocessing is used to remove noise and background noise from the image.

3.2.1. Face detection, Localization, and Preprocessing:

1) Face detection. Face detection is the primary step in facial expression recognition, and facial regions need to be extracted from the JAFFE (The Japanese Female Facial Expression) database [14]. This article uses OpenCV (Open Source Computer Vision Library) for facial region detection, and the results before and after facial detection are shown in Figure 3.

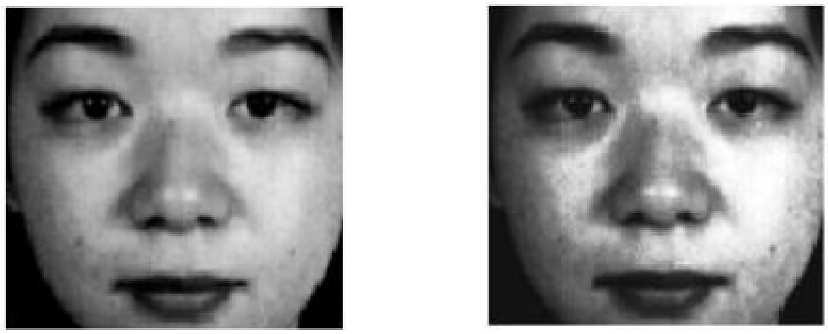


(a) Original image (b) Image after face detection

Fig. 3. Results before and after facial detection

2) Facial expression preprocessing. In the actual process, image acquisition is inevitably affected by interference such as light, head shaking, noise, and so on [15]. The grayscale of most natural images is concentrated in a small range, resulting in blurring of certain details. In order to improve the contrast of the image and improve the definition of the image, the commonly used method is the histogram equalization method.

This paper uses a histogram equalization method to compare the original image with the equalized image, as shown in Figure 4.



(a) Original image (b) Histogram equalization image

Fig. 4. Image processing result of histogram equalization method

3.2.2. Expression Feature Extraction. The uniform LBP (Local binary pattern) model has important application value in face recognition, age assessment, emotion recognition, object detection, medical image analysis, and other fields. In response to the complexity issue of the algorithm in this system, this article uses a LBP algorithm based on a consistent model for facial expression feature extraction. The specific steps are as follows:

1) It is possible to construct a textured area of 3×3, where the critical value is the central pixel of the textured area. In this article, the 8 surrounding pixels are compared with the critical value. If the value is greater than the critical pixel value, the adjacent area is set to 1; If this value is less than this critical value, the adjacent area is set to 0.

2) In the 3×3 range, the numerical value generated by 8 pixels constitutes an 8-bit binary number in a clockwise direction. This article statistics the number of jumps from 0 to 1 or 1 to 0 in 8-bit binary numbers. If the number of jumps is less than 2, then the decimal number corresponding to this binary number is the LBP value of the 3×3 neighbor center; When the number of jumps exceeds 2, this paper sets P=8 and the LBP value to P+1=9.

3) This article obtains the LBP of the entire image by scanning each pixel, and combines these LBPs in a certain order to form the LBP of the entire image.

3.2.3. SVM Based Facial Expression Classification. SVM aims to improve the generalization ability of classification and regression, and has high prediction accuracy in untrained samples. It is an inference method from data to distribution. Support vector machine can effectively deal with the problem of small samples, and can also effectively avoid the curse of dimensionality. After comparing various classification algorithms, this paper uses SVM to learn and classify facial expressions.

1) Constructing multi class SVM classifiers. This article adopts a one-on-one approach to construct an SVM classifier, constructing a classifier between every two classes. This article constructs n (n-1)/2 classifiers for n-class problems, and uses voting to make decisions. This article classifies the identified samples to obtain the highest score as the attribution classification of the identified samples.

2) Training and Classification of SVM. This article selected 7 facial expressions, and the main features of these 7 expressions are listed in Table 1.

Table 1. Seven facial expression features

Expression	Facial features
Calm	Facial muscles are in a relaxed state, evenly distributed and without significant changes, and maintaining facial expressions for the longest time.
Anger	Eyes widened, pupils narrowed, eyebrows pressed down, nostrils unconsciously dilated, and breathing became thicker.
Sad	The facial muscles will droop overall, the eyebrows will be slightly tight and pressed down, and the eyelids will droop.
Happy	Raise the corners of the mouth and make the eyes smaller, while also raising the corners of the eyes.
Hate	Eyebrows pressed down, eyes small. Mild disgust may cause the corners of the mouth to slightly open and a bit of teeth to leak out. In cases of extreme disgust, the eyes and mouth may become tightly closed.
Fear	Facial expression muscles will become tense, eyebrows and eyelids will also rise, and pupils will first become larger and then gradually smaller.
Surprised	Eyes wide open, pupils dilated, upper eyelids and eyebrows raised, and mouth wide open.

During the training phase, the samples are grouped according to facial posture and divided into three groups, including the left side of the face, the right side of the face, and the front of the face. This article extracts the features of all images and configures multiple classifiers for each group. In multi-level learning of SVM, each group contains n (n-1)/2 SVM classifiers, storing all the information required for multiple levels. In the classification stage, this article uses a trained SVM classifier to classify unlabeled expressions to determine the category.

3.3. Bimodal Emotion Recognition Method Based on Decision Level Fusion

This article uses SVM algorithm to recognize and classify emotions in facial expression images. By preprocessing the speech data, this paper extracted acoustic parameters from the speech data and used the HMM model to classify the speech data, obtaining the corresponding speech data. This article synthesizes these two methods according to certain rules, and the final result is the identified emotional state.

In facial expression recognition, this article assumes that the length of a paragraph is t and there are n images during this period. Therefore, the recognition of facial expressions during this period can be as follows:

$$W\left(N_i, M_j\right) = \begin{cases} 1, SVM(N_i) = M_j \\ 0, SVM(N_i) \neq M_j \end{cases} \tag{6}$$

$$P\left(M_j \mid O_i\right) = \frac{\sum_{i=1}^{n} W\left(N_i, M_j\right)}{n} \tag{7}$$

N_i represents the i-th image, M_j represents the jth emotional state, and $W\left(N_i, M_j\right)$ is used to indicate whether the image i is the jth emotional state after classification in the SVM classifier. If the i-th emotional state image is related to the jth emotional state, then $W\left(N_i, M_j\right)$ is 1, otherwise it is 0.

$\sum_{i=1}^{n} W\left(N_i, M_j\right)$ represents the classification of all facial expressions related to the jth emotional state in SVM. $P\left(M_j \mid O_i\right)$ shows the proportion of images related to the jth emotional state within time t in SVM classification, compared to the total number of SVM images n.

In order to recognize speech emotions, the input languages of 7 HMM models can be calculated, and the probability of confirming that they belong to the corresponding emotional state can be recorded as $P\left(S_j\right)$.

This article fuses the emotion recognition result $P(S_j)$ of speech and the recognition result $P(M_j \mid O_i)$ of facial expressions during this period according to certain rules, and the emotion state with the highest probability value in the final result is its belonging emotion state. When fusing, the recognition results are fused using summation rules.

If the speech emotion recognition result is $P(M_j \mid S)$ and the facial expression recognition result is $P(M_j \mid L)$, the probability assigned to each type of emotion state is determined by the following sum rule:

$$\begin{cases} P_j = a \cdot P(M_j \mid S) + b \cdot P(M_j \mid L) \\ \qquad 1 \le j \le 6 \\ \qquad a + b = 1 \end{cases} \tag{8}$$

In the formula, a and b represent the weights of speech emotion classification results and facial expression recognition results, respectively, while P_j represents the probability of being assigned to emotional state j after the fusion of speech emotion recognition results and facial expression recognition results. This article selects the maximum value of the results from P_1, P_2, P_3, P_4, P_5, P_6, and the corresponding emotional state is the bimodal recognition result after the fusion of facial expression recognition and speech emotion recognition.

4. Online Platform Reading Recommendation Method Based on Reader Emotions

During the reading process, this article analyzes the emotions of readers, thereby analyzing their psychological state and timely capturing their negative emotional trends, based on their current psychological state. This can recommend a book suitable for them to ensure their mental health and prevent their mental illness.

The blockchain online reading platform would conduct emotional analysis on readers and obtain their emotional portraits. This paper then uses big data information push technology according to their emotional orientation. This can push books that have the best effect on them, such as self-help books, inspirational books, philosophy books, and so on. Among books of the same type, one can carefully select the most complete and authoritative classic work.

During the reading process, through emotional communication with the author, readers can eliminate negative emotions such as irritability, depression, and anxiety that are suppressed in their own hearts from their subconscious, thereby discovering their problems and finding solutions, and restoring a calm state of mind. In addition, social media is also a primary way of emotional communication and information acquisition with readers.

5. Experiment on Reading Bimodal Emotion Recognition and Psychological Analysis

This study was divided into three groups, with the first group using only speech recognition mode, the second group using only face recognition mode, and the third group using bimodal recognition mode. This article is classified four times, and data for each emotion is randomly divided into four groups. All four groups of data are tested to obtain the classification accuracy of the four classification experiments, and the average of the four classification accuracy is obtained as the final accuracy.

5.1. Speech Emotion Recognition

The recognition accuracy of emotion recognition using only speech modality is shown in Figure 5.

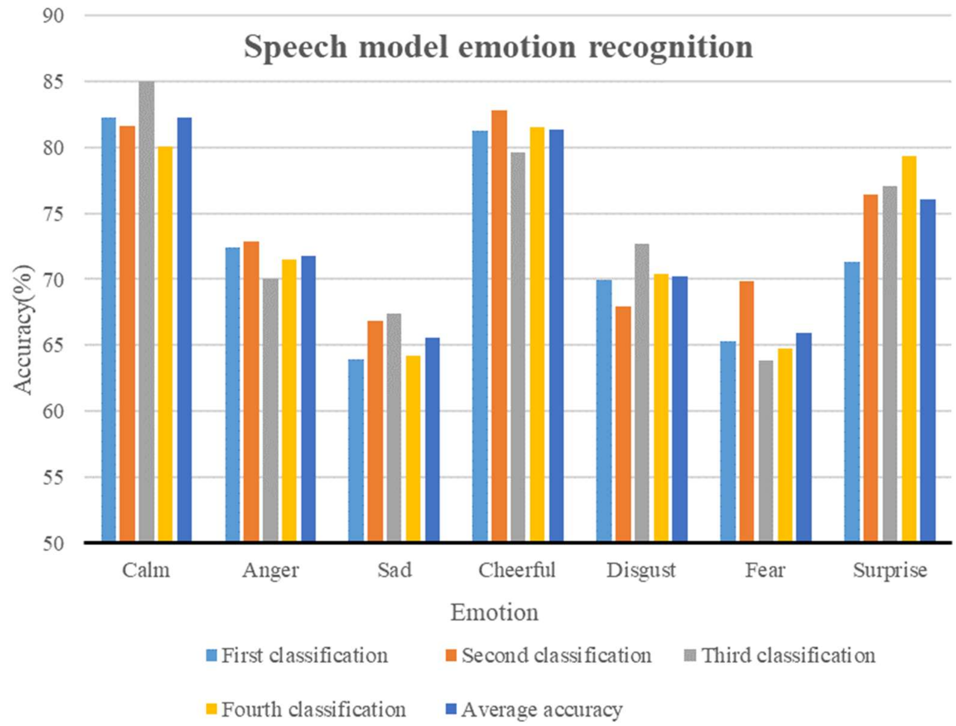


Fig. 5. Recognition accuracy of emotion recognition using only speech modality

Figure 5 shows the accuracy of seven emotions recognition when only speech recognition is performed, for seven emotions such as calmness, anger, sadness, joy, disgust, fear, and surprise. The recognition accuracy of calm and happy emotions is relatively high, with an average recognition accuracy of over 80%. For other emotions, the recognition accuracy is relatively low. When only using speech recognition, the average recognition accuracy for seven emotions is between 65% -83%, with a lower overall recognition accuracy and a significant difference in recognition accuracy between different emotions.

5.2. Facial Emotion Recognition

The recognition accuracy of emotion recognition using only facial modality is shown in Figure 6.

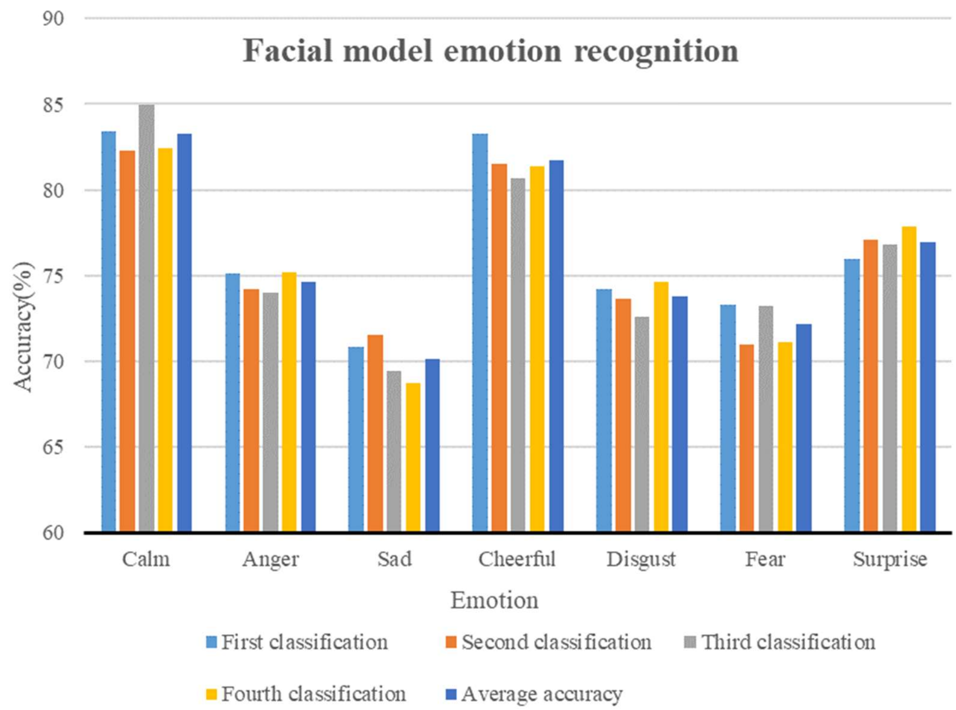


Fig. 6. Recognition accuracy of emotion recognition using only facial modality

Figure 6 shows the accuracy of recognizing 7 emotions when only facial recognition is performed. For the recognition results of 7 emotions, calm and happy have the highest recognition accuracy, and the average recognition accuracy is also above 80%. The recognition accuracy of surprise is second, above 75%, while the recognition accuracy of anger, sadness, disgust, and fear is lower. Compared with emotion recognition results based solely on speech recognition, emotion recognition accuracy based solely on facial recognition has been improved to a certain extent. The average recognition accuracy of the seven emotions is over 70%, but there is still no high emotion recognition effect, and the overall recognition accuracy is below 85%.

5.3. Dual Modal Emotion Recognition

The recognition accuracy of using bimodal emotion recognition is shown in Figure 7.

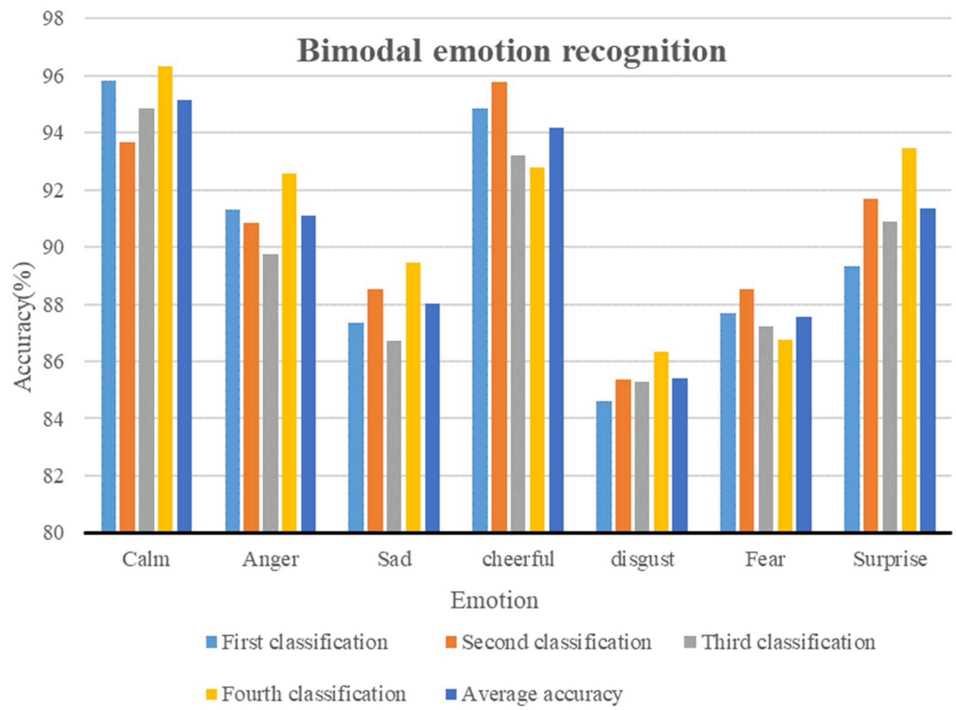


Fig. 7. Recognition accuracy of emotion recognition using bimodal approach

Figure 7 shows the accuracy of seven types of emotion recognition using bimodal recognition that combines facial recognition and emotion recognition. The average accuracy of emotion recognition using bimodal methods is above 85%, which is much higher than the recognition accuracy using only facial modality for emotion recognition and only facial modality for emotion recognition. When using bimodal recognition methods for emotion recognition, although there is still a difference in the recognition accuracy of each emotion. Compared with the results of single mode recognition, the difference is reduced, and each emotion has a higher recognition accuracy.

6. Conclusions

This article mainly explored the application of emerging technologies such as big data, artificial intelligence, and blockchain in reading modes from a technical perspective, and studied and analyzed the reading mode characteristics of blockchain network platforms. In response to the issue of inaccurate recognition results caused by inconsistent single modal emotion recognition results, this article studied the audio video dual modal emotion recognition algorithm. This article extracted and classified emotional features from speech and facial images. Based on the obtained speech emotion recognition results and facial image emotion recognition results, decision level fusion is performed to obtain bimodal emotion recognition results. This article used emotion recognition results for reader psychological analysis, intelligently adjusting information push content, and regulating readers' psychological state. This bimodal emotion recognition technology has high accuracy in emotion recognition, solving the problem of low accuracy in single modal emotion recognition, and has good effectiveness.

References

[1] Alsharari, Nizar. "Integrating blockchain technology with internet of things to efficiency." International Journal of Technology, Innovation and Management (IJTIM) 1.2 (2021): 01-13.

[2] Shen, Bin, Ciwei Dong, and Stefan Minner. "Combating copycats in the supply chain with permissioned

blockchain technology." Production and Operations Management 31.1 (2022): 138-154.

[3] Schoneveld, Liam, Alice Othmani, and Hazem Abdelkawy. "Leveraging recent advances in deep learning for audio-visual emotion recognition." Pattern Recognition Letters 146.6 (2021): 1-7.

[4] Liu, Min, and Jun Tang. "Audio and video bimodal emotion recognition in social networks based on improved alexnet network and attention mechanism." Journal of Information Processing Systems 17.4 (2021): 754-771.

[5] Verkholyak, Oxana, Anastasia Dvoynikova, and Alexey Karpov. "A Bimodal Approach for Speech Emotion Recognition using Audio and Text." J. Internet Serv. Inf. Secur. 11.1 (2021): 80-96.

[6] Zheng Jingjing ,Teng Shuhua ,Li Peirong ,Ou Wei ,Zhou Donghao ,and Ye Jun. "A Novel Video Copyright Protection Scheme Based on Blockchain and Double Watermarking." Security and Communication Networks 2021.13 (2021): 1-16.

[7] Coiro, Julie. "Toward a multifaceted heuristic of digital reading to inform assessment, research, practice, and policy." Reading Research Quarterly 56.1 (2021): 9-31.

[8] Sanka, Abdurrashid Ibrahim, Irfan Muhammad , Huang Ian , Cheung Ray C.C. "A survey of breakthrough in blockchain technology: Adoptions, applications, challenges and future research." Computer Communications 169.3 (2021): 179-201.

[9] Miglani, Arzoo, and Neeraj Kumar. "Blockchain management and machine learning adaptation for IoT environment in 5G and beyond networks: A systematic review." Computer Communications 178.10 (2021): 37-63.

[10] Unalan, Serhan, and Sercan Ozcan. "Democratising systems of innovations based on Blockchain platform technologies." Journal of Enterprise Information Management 33.6 (2020): 1511-1536.

[11] Akcay, Mehmet Berkehan, and Kaya Oguz. "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers." Speech Communication 116 .56(2020): 56-76.

[12] Koduru, Anusha, Hima Bindu Valiveti, and Anil Kumar Budati. "Feature extraction algorithms to improve the speech emotion recognition rate." International Journal of Speech Technology 23.1 (2020): 45-55.

[13] Arora, Malika, and Munish Kumar. "AutoFER: PCA and PSO based automatic facial emotion recognition." Multimedia Tools and Applications 80.9 (2021): 3039-3049.

[14] Hassan, Alia K., and Suhaila N. Mohammed. "A novel facial emotion recognition scheme based on graph mining." Defence Technology 16.5 (2020): 1062-1072.

[15] Holland, Alison C., Garret O Connell, and Isabel Dziobek. "Facial mimicry, empathy, and emotion recognition: A meta-analysis of correlations." Cognition and Emotion 35.1 (2021): 150-168.