The Application and Analysis of Emotion Recognition Based on Modern Technology

Lanxin Bi

Rosedale Global High School, 800 Markham, ON, L3R 6G2, Canada

Abstract. This article provides a comprehensive analysis of various emotion recognition methods, focusing on speech emotion recognition, facial expression recognition, and physiological signal emotion recognition. The primary aim is to evaluate the advantages and disadvantages of these methods, offering insights into selecting the most appropriate approach for different application scenarios. The study involves collecting and analysing experimental data, exploring their respective strengths and limitations, and proposing potential solutions to enhance their effectiveness. Speech emotion recognition is effective but sensitive to noise and speaker variability, while facial expression recognition excels under controlled conditions but struggles with changes in lighting and angles. Physiological signal recognition offers deep insights into internal emotional states but requires complex signal processing and is vulnerable to external interferences. Despite the growing application of emotion recognition technology across various fields, including healthcare, traffic safety, and security, there remain significant challenges related to accuracy, robustness, and privacy. This study highlights the need for continued research to improve these technologies, particularly in enhancing their robustness and adaptability. The findings provide valuable guidance for researchers and practitioners seeking to optimize emotion recognition systems for diverse real-world applications.

1 Introduction

Concomitant effects on human feelings and illness. That's why it gets more and more attention to identifying people's feelings. A technique to recognize and detect human emotional state in computer systems is called emotional knowledge technology. It uses various data sets, such as speech, biological signs, body mass and facial images, as the basis for determining the individual's emotional state. The design of an artificial testimony to the knowledge of human emotions is today one of the hottest research materials. Human-computer interaction (HCI) techniques are widely used to understand human feelings. In everyday life, people often use a lot of isolated and non-verbal methods to transfer emotions, so the technology to recognize emotions needs to recognize these fine feelings and to accurately take people's feelings [1].

Corresponding author: huochunlian@ldy.edu.rs

Emotions are an important part of human life and play an independent role in understanding and understanding things [2-4]. The development of environmental diagnostics and its application in society requires technologies that can identify potential customer needs and select suitable solutions for customers. Automatic sensory evaluation is also very important in areas such as education, marketing and entertainment. The appearance of human emotions plays an important role both in everyday life and at work. The definition of "feelings" first appeared in the works of William James, a basic American psychology. He wrote in the book that there are feelings about changes in the body in humans and all feelings are monitored by other changes in the body, such as facial image and visual activity [5]. Over the past thirty years, people have continuously designed numerous methods for analysing human emotions; From the manual methods such as questionnaires carefully designed by psychologists before the advent of artificial intelligence, to the methods gradually involving computers with the development of technology. Nowadays, emotion recognition through computers is applied in many different places. For example, emotion recognition through physiological signals is currently being used to create smart offices and smart homes. In addition, facial detection methods and speech recognition technology are widely used today [6]. The importance of correctly identifying emotions has been recognized by researchers in many different fields, and in recent years, emotion recognition research has mainly been applied in fields such as psychology, clinical therapy, and emotion calculation [7].

This article analyses and compares the principles, techniques, benefits and limitations of three main emotion recognition technologies, namely speech emotion recognition, facial expression recognition, and physiological emotion signal recognition. Voice emotion recognition recognizes emotions by analysing the pitch, speed, volume and cadence of voice signals. Although this technology offers high accuracy, it is strongly influenced by noise and individual differences between speakers. Facial expression recognition perceives emotions by analysing facial expressions, with a focus on changes in eyebrows, eyes and mouth. Although this method is used in front lighting and expression recognition, the accuracy of detection decreases with changes in visual and emotional signals. Emotional states are estimated by electrograms. This technology requires extensive collection of emotional data [1]. Emotion recognition technology improves accuracy and intensity by integrating multiple sources of information. The future development of this field should focus on improving strength, adaptability, and availability, and exploring multimodal transport integration technologies to enhance overall performance. With the continuous development of artificial intelligence and sensory technology, emotion recognition will be more widely applied in all fields.

This article first describes the three methods of identifying emotions in chapter 1, and then gives a detailed description of the processes related to their models, traits and emotional recognition in chapter 2. Chapter 3 analyses and compares the advantages and disadvantages of these technologies and discusses existing challenges and possible solutions. Finally, Chapter 4 gives a comprehensive overview of the main points discussed in the previous chapters.

2 Methodology

2.1 Dataset description and preprocessing

Dataset plays a crucial role in data-driven learning as they can enhance the performance and strength of models. Different datasets contain different types of signals. Based on these different categories of signals, sensor recognition datasets are divided into speech (text, audio), physiological, multimodal, and visual (video and facial expression) datasets. The

main datasets for emotion recognition include SMI glasses for eye tracking, Aware dataset, ESI neural scanning system, DEAP dataset for emotion analysis using physiological signals, and Shanghai Jiao Tong University seed dataset. In the ESI neural scanning system and SMI dataset, data related to various emotions that appear in individual experimental tests and the levels that emotions can reach can be found. The Aware database is an emotion analysis dataset that contains comments on three different domains of applications: performance, gaming, and social media. Long term memory networks (LSTM), DEAP datasets, and SEED datasets were used to train convolutional neural networks (CNN) for sensory recognition and processing [8].

2.2 Proposed approach

In this chapter, author will provide a detailed introduction to the main methods of emotion recognition, including speech emotion recognition, facial expression recognition, and physiological signal emotion recognition, from the aspects of technical characteristics, concepts, principles, overall technical processes, and models. In each section, an introduction to the model will be given first, and then relevant data tables will be listed as references.

2.2.1 Speech emotion recognition

Speech emotion recognition technology is basically divided into two stages: extraction and classification of speech features. Speech used for emotion recognition is essentially achieved through the use of changes in audio signals. Speech emotion recognition can be applied in many different scenarios and is currently widely used in speech recognition, customer service, and call centres, as shown in the Table 1. The research centre proposed a two-step operation method based on CNN [9]. Firstly, a thick autoencoder (SAE) with reconstruction separation is used to extract local invariant features (LIF). Then, LIF performs feature input. Another method is to use a semi-CNN in the first step, and then use a retraction folding network to extract features [10].

2.2.2 Facial expression recognition

Systems inspired by biological nervous system control are mainly composed of neural networks. Many machine learning algorithms can support the collaborative work of these machines, so they are not algorithms, but frameworks [2]. The method discussed in this article proposes a neural network approach that evaluates seven different expressions of human disgust, surprise, sadness, happiness, neutrality, fear, and anger through two basic steps. Radial basis functions (RBF) and multi-layer sensor networks (MLP) play important roles in classification. Another method proposes a system that uses rectangles to shrink and expand, extracting facial features from the contours of the eyes, eyebrows, and mouth. You can use fewer features or shorten recognition time compared to other systems. During the recognition process, two interior angles are used as positioning points for the machine to find contours [11]. Comparison of Models is shown in the Table 2.

Table 1. Comparison of Emotion recognition techniques using speech signal

S. No	Method	Database	Recognition Rate	Specifications	Year
1	Artificial nerve Network testing Gender dependence database	Self-generated	Female 65.5 Male 72.055	Discrete Wavelet Transform (DWT)	2009
2	Anchor model	FAU-AIBO	44.19	Cosine based anchor model distance	2013
3	CNN	Sava Emotion Database DES MES	71.8 60.4 57.8	LIF, SAE, SDFA	2014
4	Semi-CNN	Sava Emotion Database DES MES	73.6 79.9	A New Classification Function - Compressed Convolutional Neural Network Learning Candidate Features	2014
5	Mixed deep nerve Network hiding markov model （DNN-HMM）	Berlin Emotion Database	77.92	DNN-HMMs with discriminative learning based unsupervised learning,	2015
6	Fuzz y C means	Self-generated	Female 73.7 Male -63	Pitch, Interface duration, energy	2003
7	Bayesian hypothesis testing	Self-generated	Stress classification best fits the characteristics of TEO	HMM	2003

2.2.3 Physiological signal emotion recognition

Recently, this paper has started developing and using machines to detect emotions generated by human heart activity. Digital filters can facilitate signal decomposition, such as silver conversion integration or discrete microwave conversion (DWT) EMD [5]. These signals can be used to express the five human emotions: disgust, surprise, fear, joy, and sadness. Another method developed by MIT researchers is a wireless signal transmission device called EQ Radio, in which the signal is first reflected from the body and then input into an interpreter [8]. ECG signal-based models are shown in the Table 3.

Table 2. Comparison of Model Based Techniques for Facial Emotion Recognition

S. No	Method	Specifications	Recognition rate (in percentage)	Dataset	Year
1	artificial neural network	MLP and RBF Networks	73	JAFFE	2002
2	Support Vector Machine	Facial feature tracker support vector machine classifier	86	Self- Generated	2003
4	Constructive feedforward neural network	Two dimensional discrete cosine transform (DCT)	93.75	JAFFE	2004
5	Combining Wavelet Transform with Neural Networks to Send Tables	Wavelet transform, Karhunen Loe transform	98.5	Self- generated	2008
6	Using BU-3DFE for 3D facial recognition	Algorithm and particle swarm optimization algorithm (ACO and PSO respectively)	92.3	Self- generated	2008
7	3D deformable model	Candide-3 facial model, tree classifier	87	JAFFE	2009

Table 3. Comparison of Emotion Recognition techniques using ECG signal

S. No	Name	Recognition Rate	Dataset	Year	Specifications
1	Discrete Fourier transform	57.5	Self-generated	2014	Feature extraction of FFT
2	Emotion recognition based on wireless signals	52	Self-generated	2013	Detection of wireless signals and transmission
3	Supervise dimensionality reduction	87	Mahnob-HCI database	2016	NCA,LDA and MCML

3 Result and discussion

For facial recognition technology, author compared the models and feature methods in research. As shown in Figure 1, the model using static microwave transformation has the highest efficiency, followed by the double random forest classifier.

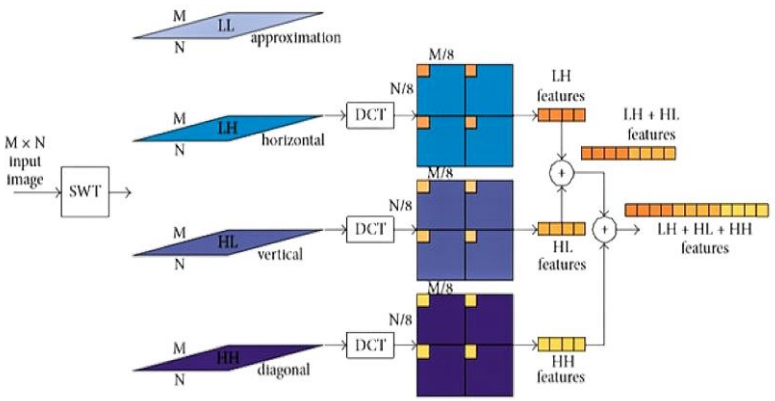

Fig.1. Feature Extraction [12].

Among the three types of emotion recognition, emotions are recognized through speech signals. Research has found that HMM and N-D HMM models have the highest accuracy, while anchor models have the lowest. Researchers have pointed out that over time, the accuracy of using voice signals to detect emotions has significantly improved. For physiological signals, this study analysed two aspects of EEG and ECG signals for emotion recognition. The research shows that these signals require a lot of preprocessing, which makes feature extraction more difficult in the study. Therefore, these are the least used methods in emotion recognition, as shown in Figure 2. The wireless equalizer has the highest accuracy, while most other methods have an accuracy of about 50-60%.

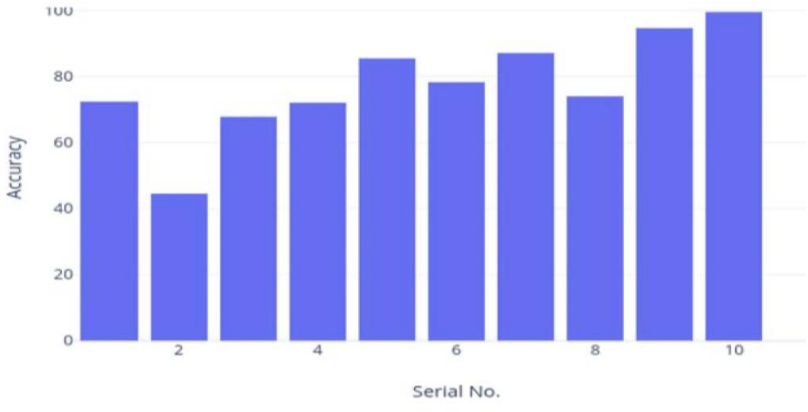

Fig. 2. Recognition rate depiction of Speech based ER [1].

3.1 Discussion

The advantage of speech recognition lies in its naturalness. In human communication, language and emotion are closely linked. Emotional speech recognition technology can naturally capture and analyse emotional information in speech, making human-computer interaction more natural and smoother. It can also make judgments based on different emotional expressions and language characteristics, provide personalized recognition and services according to each person's characteristics, and improve the accuracy and satisfaction of interaction. Its disadvantage is that it cannot accurately perceive people's complex and

diverse emotional expressions influenced by various factors such as cultural background and language habits, making it difficult to accurately identify people's emotional states and disorders. In real-world scenarios, ambient noise and interference can affect the quality of speech signals and the extraction of emotional information. For example, background noise, echoes, etc. This interferes with the analysis and recognition of voice signals. Speech emotion recognition technology has broad application prospects in several areas. In the field of intelligent customer service, by identifying customer emotions, companies can provide more personalized services, better understand each customer's emotional state, and provide more appropriate assistance solutions. In the field of medicine, speech emotion recognition technology can help doctors better analyse patients' mental states and gain a better understanding of their emotions and mental states.

The advantage of facial expression recognition is that it intuitively reflects a person's emotional state and is a widely researched and applied method. This method guides a person's emotions, such as happiness, sadness, anger, etc. The application scenarios of smiling face recognition include but are not limited to: gaming experience: by analysing players' facial expressions, games can adjust the difficulty level in real time to adapt to players' emotional states, thereby improving the gaming experience. Market research: Real time capture of customer reactions and emotions, helping companies better understand customer feedback and optimize products or services. The main disadvantage of recognizing emotions through facial expressions is the slow speed of recognizing emotions in natural environments. Most existing facial recognition technologies have been tested in experimental environments, which greatly reduces the speed of facial recognition under natural conditions, affected by factors such as head position deviation, light changes, shadows, and motion blur. This still makes facial recognition a complex problem under natural conditions.

The benefits of emotion recognition physiological signals mainly include their authenticity: physiological signals are generated by the activity of the human autonomic nervous system, are not controlled or disguised by human will, and can objectively reflect the physical and mental state of human activities. Diversity: Physiological signals include electroencephalography, electrocardiography, pulse, respiration, skin temperature, electromyography, skin conductance, and many other types. The comprehensive analysis of these signals can improve the accuracy and reliability of emotion recognition. The disadvantages of emotion recognition based on physiological signals mainly include the complexity of signal processing, sound interference, and difficulty in expressing emotions. The processing of physiological signals is relatively complex and requires the use of several denoising algorithms to deal with noise in the signal. Different physiological signals (such as electrocardiogram, electromyographic signals, etc.) have different noise characteristics, so specific noise reduction methods need to be developed for each signal. This increases the complexity and workload of the research. In addition, physiological signals are also affected by external sound interference, such as pseudo motion, background drift, etc. In the field of road safety, emotion recognition technology can improve brake response by detecting the driver's emotional state (such as anger or fatigue) and warning the driver of steering wheel vibration and warning sounds, thus reducing traffic accidents caused by driver negligence and anger. In the medical field, emotion recognition technology can help healthcare professionals assess patients' emotional states, such as by detecting injuries to mechanics, user experiences in video games, and so on.

4 Conclusion

Emotions play a crucial role in human life and significant progress has been made in recognizing emotions through facial expressions, speech signals and physiological signals. Facial Emotion Recognition is mainly based on feature-based techniques such as Gabor

wavelets and local direction maps, as well as model-based methods such as neural networks and supporting vector machines. Speech recognition of emotions uses prosodic features, mathematical models, and neural networks to recognize emotions. In physiological signals, emotions are identified by breaking them down into smaller components and then extracting and classifying them, with techniques such as empirical model decomposition (EMD) playing a key role. Despite the potential of emotion recognition technology, it faces significant challenges, including technological limitations and privacy issues. Many current methods derive emotional states from external cues such as facial expressions, but there is little scientific evidence to support the accuracy of these conclusions. In addition, the widespread use of emotion recognition technology has raised serious privacy concerns as it can lead to unauthorized recording and analysis of a person's emotional state. To meet these challenges, it is crucial to establish ethical standards and strengthen data protection regulations, which will help emotion recognition technology develop responsibly within legal and ethical frameworks.

References

1. A. Konar, A. Halder, A. Chakraborty, Introduction to emotion recognition. Emotion Recognition: A Pattern Analysis Approach, 1-45 (2015)
2. D. J. Vandervoort, The importance of emotional intelligence in higher education. Current Psychology: Development, Learning, Personality, Social, 25(1), 4–7 (2006)
3. C. Scotty, G. Arthur, S. Jeremiah, Affect and learning: An exploratory look into the role of affect in learning with Auto Tutor, Journal of Educational Media, 29(3), 241-250 (2004)
4. M. Nussbaum, Emotions as Judgments of Value and Importance. Series in affective science, 183-199 (2004)
5. A. Saxena, A. Khanna, D. Gupta, Emotion recognition and detection methods: A comprehensive survey. Journal of Artificial Intelligence and Systems, 2(1), 53-79 (2020)
6. M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan, Real Time Face Detection and Facial Expression Recognition: Development and A lications to Human Computer Interaction, 2003 Conference on Computer Vision and Pattern Recognition, 53-53 (2003)
7. R. Guo, H. Guo, L. Wang, Development and application of emotion recognition technology—a systematic literature review. BMC psychology, 12(1), 95 (2024)
8. Y. Cai, X. Li, J. Li, Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review. Sensors, 23(5), 2455 (2023)
9. Q. Mao, M. Dong, Z. Huang and Y. Zhan, Learning Salient Features for Speech Emotion Recognition Using Conational Neural Networks, in IEEE Transactions on Multimedia, 16, 2203-2213 (2014)
10. Z. Huang, M. Mao, Q. R. Zhan, Speech Emotion Recognition Using CNN. MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia, 801-804 (2014)
11. S.C. Tai and K. C. Chung, Automatic facial expression recognition system using Neural Networks, TENCON 2007 - 2007 IEEE Region 10 Conference, 1-4 (2007)
12. Q. Y. Huma, Facial Expression Recognition Using Stationary Wavelet Transform Features. Mathematical Problems in Engineering, 1-9 (2017)