IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 27 January 2024, accepted 17 February 2024, date of publication 26 February 2024, date of current version 6 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3370431

RESEARCH ARTICLE

Emotion Recognition and Management in the Tourism Industry During Emergency Events Using Improved Convolutional Neural Network

HAIXIA QI[1] AND ZHIFEI HAN[2]
[1]Department of Management, Northeast Petroleum University, Qinhuangdao 066004, China
[2]Hebei Institute of International Business and Economics, Qinhuangdao 066311, China

Corresponding author: Zhifei Han (henry18751280@126.com)

This work was supported by the Guided Innovation Fund of Northeast Petroleum University, in 2020, ''Research on the application of emotional labor strategy in hotel industry'', under Grant 2020 YDQ-12.

ABSTRACT The purpose is to optimize emotion recognition capabilities in the tourism industry during emergency events and enhance management efficiency. Hence, based on the principles of emergency events and emotion recognition, this work has outlined the fundamental emotion recognition process. Additionally, an emotion recognition model is constructed based on a Convolutional Neural Network (CNN), and the process of extracting emotions using a three-dimensional (3D) CNN is proposed. Finally, an attention mechanism is employed to optimize the 3D CNN model, and a comparative analysis of accuracy and precision is conducted using the Bimodal Face and Body Gesture Database (FABO). The research findings reveal that the optimized 3D CNN model exhibits lower error rates in recognizing anxiety emotions, with only 3 samples going unrecognized. Its overall recognition performance is superior to other models. There are variations in the recognition accuracy among different models, but in general, the optimized 3D CNN model performs relatively well across various datasets, achieving recognition accuracies of 83.92%, 82.33%, and 88.81%. Compared to other models, the optimized 3D CNN model demonstrates higher precision in recognizing different emotions, particularly excelling in identifying anger, disgust, and happiness, with precision rates of 97%, 91%, and 94%, respectively. This work has improved the accuracy and efficiency of emotion recognition, providing more intelligent and effective support for emergency event management in the tourism industry.

INDEX TERMS Convolutional neural network, tourism industry, emergency events, emotion recognition management, attention mechanism.

I. INTRODUCTION

A. RESEARCH BACKGROUND AND MOTIVATIONS

With the rapid advancement of the tourism industry, the impact of emergencies on the tourist experience and the industry's reputation is increasingly significant. In these unexpected events, the emotional reactions of tourists are a crucial variable. It affects the resolution of the events and may have long-term effects on the image of the tourism destination and future business [1], [2], [3]. Therefore, accurate and rapid identification and management of emotions in the tourism industry during emergencies have become crucial.

In recent years, with the rapid advancement of artificial intelligence technology, the Convolutional Neural Network (CNN) has achieved remarkable success in the field of image recognition. However, despite the excellent performance of CNN in the image domain, it still faces a series of challenges in handling complex emotion recognition tasks [4]. Emotions involve multidimensional and dynamic psychological states, including semantics, intonation, and facial expressions. Traditional CNNs mainly focus on capturing spatial features of images, which may result in insufficient

The associate editor coordinating the review of this manuscript and approving it for publication was Abdel-Hamid Soliman.

capture of the multidimensional expression of emotions, leading to a decrease in recognition accuracy [5]. Additionally, emotion recognition tasks involve temporal dynamics and context correlation, and traditional CNNs are relatively weak in handling time-series data and global context modeling [6]. Compared to traditional CNNs, the three-dimensional (3D) CNN has more powerful capabilities, allowing for a more comprehensive capture of spatiotemporal information and better adaptation to the multidimensional expression of emotions [7]. The introduction of this network structure enables the model to more effectively understand the temporal and spatial correlations in emotional information, thus accurately capturing and expressing complex emotional states [8]. By introducing the third dimension, namely the time dimension, the 3D CNN can better handle sequential data, providing robust technical support to enhance the flexibility and accuracy of the model [9], [10].

B. RESEARCH OBJECTIVES

This work first establishes the basic emotion recognition process based on emergencies and emotion identification principles. Subsequently, an emotion recognition model is constructed based on CNN, and a 3D CNN approach for extracting emotions is proposed. Finally, an attention mechanism is employed to optimize the 3D CNN model, and the optimized model is compared with common emotion recognition models such as Support Vector Machine (SVM), Recurrent Neural Network (RNN), traditional CNN, and Long Short-Term Memory (LSTM). The research innovation lies in successfully incorporating an attention mechanism into the 3D CNN model. Through the optimization of the basic structure of the 3D CNN, the output of feature maps is effectively enhanced. This innovative design allows the model to exhibit higher accuracy and precision in emotion recognition tasks. Careful adjustments to the model's structure enable the finer and more targeted capture of emotional features. It significantly reduces error rates in emotion recognition for the optimized 3D CNN model and enhances the model performance.

II. LITERATURE REVIEW

The emotion recognition model is a type of model based on machine learning and artificial intelligence technologies, used for identifying and analyzing human emotions. This model typically undergoes training using large annotated datasets to learn how to extract emotional features from text, audio, images, or videos. Common emotion recognition models include rule-based, machine learning, and deep learning-based models [11], [12].

Rule-based emotion recognition models usually rely on prior knowledge and rule sets to infer emotions; however, this approach is limited by the accuracy and generalization ability of the rules [13]. Machine learning-based models learn patterns for extracting emotional features from data through training algorithms, among which SVM and Random Forests are widely adopted. While these models enhance emotion recognition accuracy to some extent, they may still be constrained by feature representation and generalization issues [14]. In recent years, deep learning-based emotion recognition models have made significant progress. Models like RNN, CNN, and LSTM have demonstrated superior performance in handling multimodal data and capturing complex emotional features [15]. For instance, Song et al. proposed a signal emotion recognition model. By integrating the efficient channel attention module and improving the combination of CNN and Gated Recurrent Units (GRU), the model achieved more comprehensive feature extraction, and significantly enhanced emotion recognition accuracy on the DEAP dataset [16]. Zhang et al. introduced a deep learning-based attention fusion model for EEG emotion recognition. They effectively improved emotion classification performance by extracting differential entropy features, obtaining spatial features through a convolutional encoder, introducing a frequency band attention mechanism, and using LSTM to extract temporal features [17]. Patnaik utilized complex Mel Frequency Cepstral Coefficients as representative features and combined them with deep temporal models for speech emotion recognition, achieving high accuracy on the RAVDESS database [18]. Al-Dujaili and Ebrahimi-Moghadam argued that emotion recognition encompassed multiple modalities, including speech, text, images, and videos [19]. Goel et al. proposed a CNN backbone structure suitable for speech emotion recognition tasks, enhancing the quality of features extracted from log-mel spectrograms by preserving both deep and shallow features [20].

In summary, the development of emotion recognition models has evolved from rule-based and machine learning-based to deep learning-based approaches in previous research. Deep learning has demonstrated significant advantages, especially its capability to handle multimodal data and complex emotional features. However, limitations still exist concerning feature representation, generalization, and the lack of comprehensive integration in studies focusing on emotion recognition across different modalities.

III. RESEARCH MODEL

A. TOURIST ATTRACTION EMERGENCIES

Tourist attraction emergencies refer to sudden and unpredictable events that may significantly impact tourists, staff, and the operation of the tourist attraction itself, occurring within the tourist attraction area [21], [22]. Figure 1 illustrates the specific characteristics of emergencies in tourist attractions.

Crisis management theory is a theoretical framework that involves the strategies and methods organizations, societies, or individuals employ to deal with crises, disasters, or emergencies. The 4R Crisis Management Theory divides crisis management into four stages: Reduction, Readiness, Reaction, and Recovery. It emphasizes reducing the probability of a crisis, being prepared, responding promptly, and recovering in an organized manner. Modern management theories,

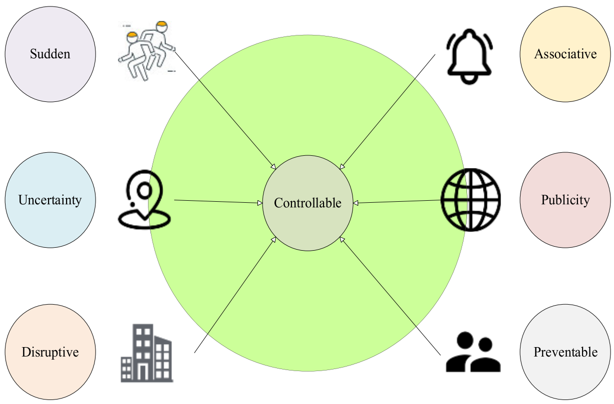
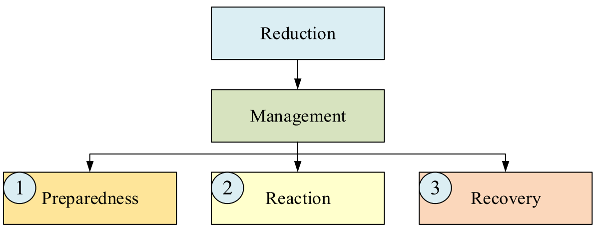


FIGURE 1. Specific characteristics of emergencies in tourist attractions.

drawing from systems theory and information theory, provide concepts for the management systems and information systems in tourist attractions, contributing to improving the effectiveness and adaptability of crisis management [23], [24]. Figure 2 illustrates the structure of the 4R Crisis Management Theory and the classification of modern management theories.



(a) 4R Crisis Management

Classify	Interpretations
System theory	As a tourist attraction, while doing a good job in scenic area marketing, ecological protection and other work, the first thing to do is to ensure the work of tourism safety services. The object contains tourists, the staff of the scenic spot, the facilities and equipment inside the scenic spot, the natural environment and so on.
Information Theory	The information system of the whole scenic area is whether the safety management of the scenic area is in place, and the timeliness of the collection and processing of tourist information of the tourist attractions is the key to the quality supervision of the safety management of the tourist attractions plays an important role.

(b) Modern Management Theory

FIGURE 2. 4R crisis management theory structure and classification of modern management theories.

B. EMOTION RECOGNITION

Emotion refers to the subjective experiences and physiological changes that humans or other organisms undergo at a specific moment, typically manifested as various emotional states such as joy, anger, sadness, and surprise. Emotions are commonly categorized into multiple fundamental classes: anger, joy, sadness, and surprise [25], [26]. Figure 3 illustrates a two-dimensional emotion model.

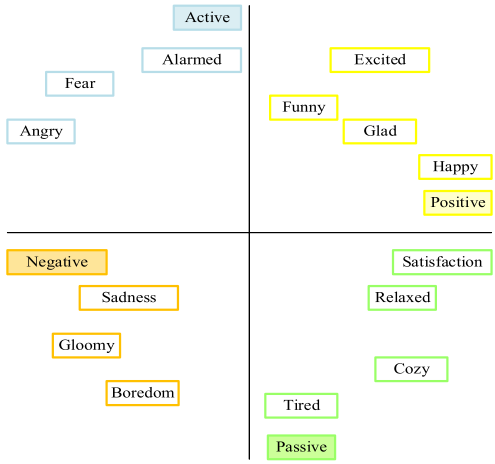


FIGURE 3. Two-dimensional model of emotion.

Emotion recognition is a technology that accurately identifies and understands an individual's current emotional state through the analysis of various signals such as facial expressions, speech, text, and physiological features like heart rate and skin conductance. Through techniques from computer vision, natural language processing, and bio-signal processing, emotion recognition systems can infer emotional categories such as joy, anger, sadness, and more. These systems play a role in various application areas, including social media analysis, user experience improvement, and mental health monitoring [27], [28]. Figure 4 outlines the basic steps in the emotion recognition process.

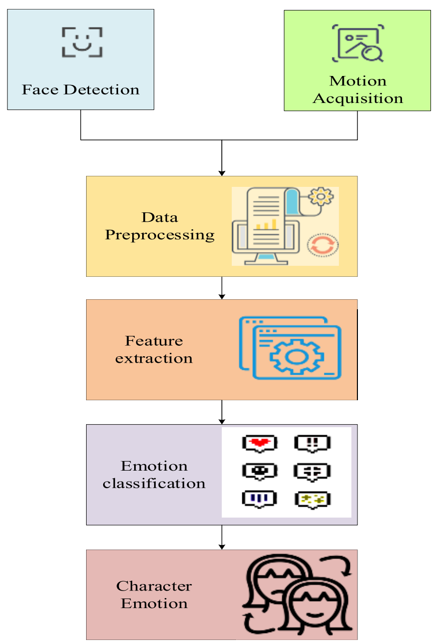


FIGURE 4. Basic process steps for emotion recognition.

C. CONVOLUTIONAL NEURAL NETWORK

CNN is a type of deep learning model with a core structure that includes convolutional layers for sliding operations with convolutional kernels to extract features from input data, pooling layers to reduce the size of feature maps and computational complexity, activation functions for introducing non-linearity, and fully connected layers for the final layer connection to perform classification or regression [29].

3D CNN is a deep learning model specifically designed for processing 3D data, such as videos or time series [30]. It performs convolution operations across three dimensions (width, height, time), emphasizing spatiotemporal information extraction, making it particularly suitable for handling 3D data like videos or time series. Its unique feature lies in its ability to simultaneously capture spatial and temporal features. By sliding convolutional kernels across multiple dimensions, the model effectively adapts to spatiotemporal relationships, enabling it to comprehensively understand complex dynamic data. Operations like parameter sharing and pooling help reduce network parameters and computational complexity, providing powerful tools for spatiotemporal modeling in tasks like video analysis and action recognition [31]. Figure 5 displays the structures of CNN and 3D CNN.

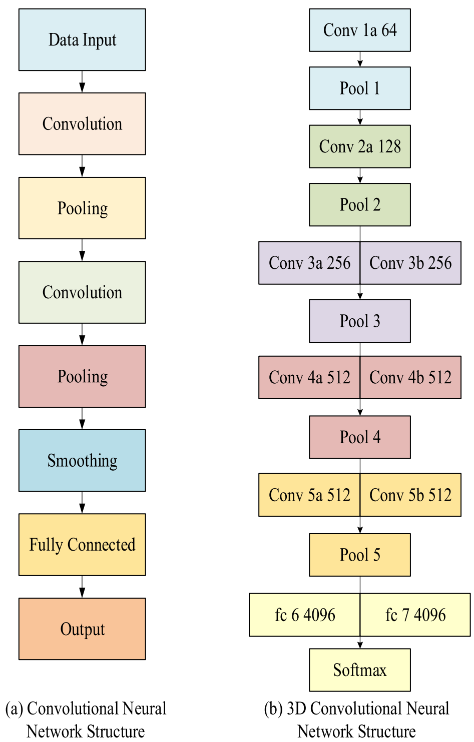


FIGURE 5. Convolutional neural network structure and 3D convolutional neural network structure.

The input data is denoted as X, the convolutional kernel as K, and the output of the convolution operation as Y. For each position (i, j, k), the output $Y(i, j, k)$ can be calculated using the following specific equation:

$$Y(i,j,k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{P=0}^{P-1} X(i+m, j+n, k+p) \times K(m,n,p) + b \tag{1}$$

m, n, p are the width, height, and depth of the convolutional kernel, respectively. $X(i+m, j+n, k+p)$ represents the value of the input data at the corresponding position. $K(m,n,p)$ is the weight of the convolutional kernel at the corresponding position. The process involves element-level multiplication with the local region of the input data, and the results are summed to obtain the final output. This process is implemented through loops across the three dimensions [32], [33].

D. CONSTRUCTION OF EMOTION RECOGNITION MODEL BASED ON THREE-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORK

This work initially constructs a foundational 3D CNN emotion recognition model. For video processing, a frame sequence of 16 frames per second is captured, and this continuous sequence of video frames is input into the 3D CNN. Through 3D convolutional operations, the obtained feature dimensions are typically high. In order to reduce feature dimensions while retaining crucial information, 3D pooling operations are employed, and the method of max pooling is adopted. This 3D max pooling helps reduce network computational complexity, enhance network invariance, and improve the model's generalization ability. This foundational model serves as the basis for subsequent experiments and optimizations, with the potential to achieve good performance in emotion recognition tasks.

Figure 6 illustrates the process of extracting information using 3D CNN.

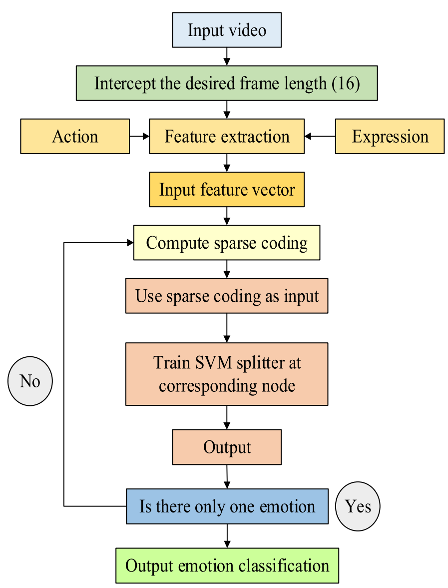


FIGURE 6. Three-dimensional convolutional neural network extraction information process.

Max pooling is a commonly used pooling method employed to reduce the dimensions of feature maps. In the 3D context, max pooling operates across three dimensions—width, height, and depth/time [34], [35], [36], [37].

The primary purpose of max pooling is to retain the most significant features within each window, thereby reducing the dimensions of the feature map while preserving essential information. This aids in reducing the computational complexity of the model while enhancing its robustness to translation and partial occlusion [38].

The calculation for 3D pooling output is as follows [39]:

$$M_{\alpha\beta\gamma} = MAX_{0\leq n\leq \mathrm{A},0\leq m\leq \mathrm{B},0\leq j\leq \Gamma}\left(I_{\alpha Q_1+n,\beta Q_2+m,\gamma Q_3+j}\right) \tag{2}$$

Here, I represents the 3D input vector, and Q_1, , Q_2, Q_3 are the sampling strides along the coordinate $\alpha\beta\gamma$. The product of A, B, and Γ corresponds to the size of the sampling region.

In addition, this work also selects common models such as SVM, RNN, CNN (2D CNN), and LSTM for comparative analysis [40], [41]. The chosen evaluation indicators for comparison are accuracy and precision, with the specific calculation equation as follows:

$$Accuracy = TP + TN / (TP + TN + FP + FN) \tag{3}$$

$$Precision = TP / (TP + FP) \tag{4}$$

TP represents the number of samples correctly predicted as the positive class, TN is the number of samples correctly predicted as the negative class, FP is the number of samples actually belonging to the negative class but wrongly predicted as the positive class, and FN is the number of samples actually belonging to the positive class but wrongly predicted as the negative class.

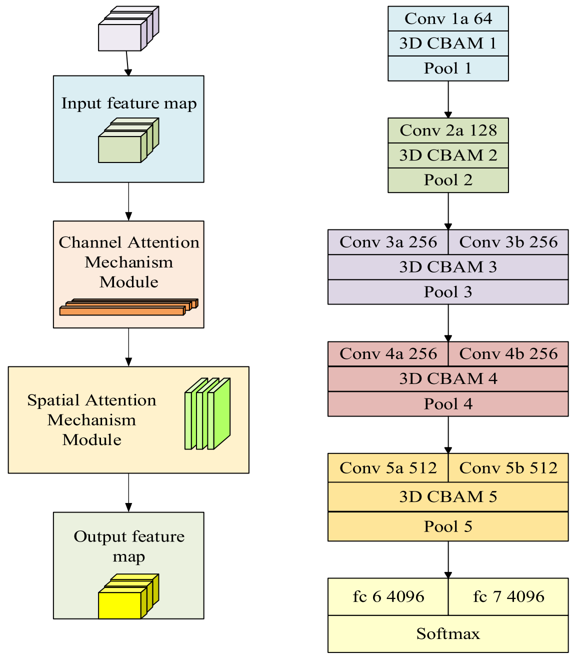
The accuracy and precision indicators have values ranging from 0 to 1, with higher values indicating better performance. Accuracy measures the proportion of correctly classified samples among all samples, while precision measures how many of the samples predicted by the model as the positive class are truly positive.

E. OPTIMISING A 3D CONVOLUTIONAL NEURAL NETWORK EMOTION RECOGNITION MODEL

The attention mechanism is a technique applied in deep learning to learn attention weights for each input element, enabling the model to selectively focus on important parts of the input. This mechanism enhances the model's focus on different positions or regions when processing sequence data or images, thereby improving its performance and generalization ability.

In optimizing the 3D CNN model, an attention mechanism is employed to enhance its performance. First, the input 3D feature map is processed using a channel attention mechanism to generate a channel attention feature map. This is then used as input for the spatial attention mechanism. In the spatial attention mechanism, max-pooling and average-pooling operations are performed on the channel attention feature map, and the features are merged and dimensionally reduced through convolution. This results in the generation of a spatial feature map. Multiplying the spatial feature map with the original 3D feature map yields the final feature map incorporating the attention mechanism. This optimized feature map is then input into the 3D CNN model to enhance the accuracy and effectiveness of emotion recognition tasks. By introducing the attention mechanism, the model can focus more selectively on important features, thus optimizing emotion recognition performance.

Figure 7 shows the basic structural positions of the attention mechanism and the optimized Attention Mechanism-3D CNN (AM-3D CNN).



(a) Attention Mechanism Basic Structure Location

(b) Optimised AM-3D CNN structure

FIGURE 7. Attention mechanism basic structure location and optimized AM-3D CNN structure.

IV. EXPERIMENTAL DESIGN AND PERFORMANCE EVALUATION

A. DATASETS COLLECTION

This work utilizes data from the Bimodal Face and Body Gesture Database (FABO) dataset. FABO is a comprehensive database that integrates facial expressions and body gesture information, encompassing a multimodal dataset with both facial expressions and body gestures.

Six emotions are chosen for the emotion recognition experiments: anger, anxiety, boredom, disgust, happiness, and uncertainty. In order to conduct the experiments, three different datasets from the FABO dataset are employed: the facial expression video dataset (D1), the body movement video dataset (D2), and a combined dataset containing both facial expressions and body movements (D3). The datasets are split into training, validation, and test sets in a 3:1:1 ratio.

B. EXPERIMENTAL ENVIRONMENT AND PARAMETERS SETTING

The experimental setup for this work is based on an Intel Core i5-10200H processor, 16GB of RAM, and the Windows 10 operating system. Python 3.6.8 is used for model construction and testing. Each sample video in every dataset is captured as a frame sequence containing 16 frames with dimensions of $3 \times 16 \times 128 \times 171$. During training, the frame sequence dimensions are cropped to $3 \times 16 \times 112 \times 112$.

The model training utilizes GPU, with a maximum of 30,000 iterations, a batch size of 32, an initial learning rate of 0.0001, and a weight decay of 0.00005. Stochastic gradient descent is chosen as the optimizer, with a momentum decay of 0.9 and a step size of 5000 for each iteration. In order to prevent overfitting, L2 regularization is applied to each convolutional layer, and dropout is introduced after the fully connected layer with a dropout rate of 0.5. The optimal model is searched by gradually decreasing the learning rate.

C. PERFORMANCE EVALUATION

Figure 8 presents the number of unrecognized samples of different emotions across various models.

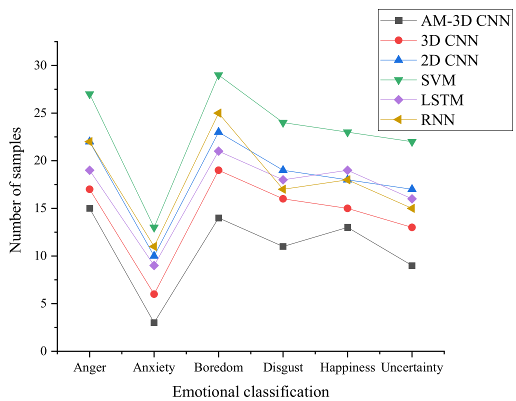


FIGURE 8. Results on the number of unidentified samples for different models for different emotions.

The research results indicate that in the AM-3D CNN model, the number of unrecognized samples for the emotions is as follows: 15 for anger, 3 for anxiety, 14 for boredom, 11 for disgust, 13 for happiness, and 9 for uncertainty. The 3D CNN performs well in recognizing anger, anxiety, and happiness emotions, with relatively low numbers of unrecognized samples, while it exhibits higher numbers of unrecognized samples for boredom, disgust, and uncertainty emotions. The 2D CNN model performs well in recognizing boredom and disgust emotions but has relatively high numbers of unrecognized samples for anger, anxiety, and uncertainty emotions. The SVM model performs relatively well in recognizing anger, anxiety, and boredom emotions but has a higher number of unrecognized samples for uncertainty emotions. The LSTM model performs well in recognizing anxiety and boredom emotions but has relatively high numbers of unrecognized samples for anger, disgust, and uncertainty emotions. The RNN model performs well in recognizing boredom emotions but has relatively high numbers of unrecognized samples for anger, anxiety, disgust, and uncertainty emotions.

Figure 9 displays the emotion recognition accuracy results for different models on various datasets.

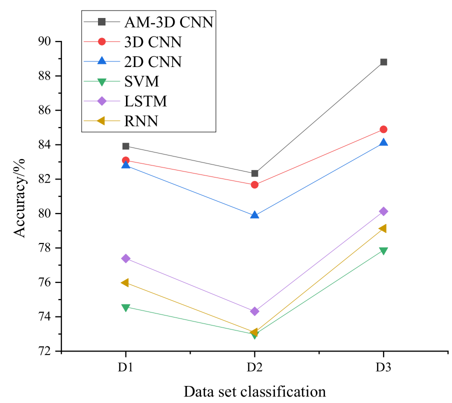


FIGURE 9. Results of emotion recognition accuracy of different models for different datasets.

The research results demonstrate that the AM-3D CNN model performs excellent emotion recognition across different datasets. On the D1 dataset, the accuracy of the AM-3D CNN model reaches 83.92%, slightly surpassing the 3D CNN (83.09%) and 2D CNN (82.79%). On the D2 dataset, the AM-3D CNN maintains its lead with an accuracy of 82.33%, while 3D CNN and 2D CNN have accuracies of 81.67% and 79.88%, respectively. Notably, on the D3 dataset, the AM-3D CNN achieves an outstanding accuracy of 88.81%, significantly outperforming 3D CNN (84.89%) and 2D CNN (84.1%). In comparison to traditional approaches, SVM shows accuracies ranging from 72.98% (D2) to 77.88% (D3), while RNN (both LSTM and RNN) accuracies fall between 73.1% and 80.13%. These results suggest that AM-3D CNN performs better in emotion recognition tasks, particularly when dealing with multimodal datasets (D3).

Figure 10 presents the precision results of different models for various emotions.

The research results indicate that AM-3D CNN performs exceptionally well across all emotion categories. For the emotion of anger, the precision of AM-3D CNN reaches 97%, significantly higher than other models, where 3D CNN and 2D CNN achieve precisions of 89% and 82%, respectively. In the anxiety emotion category, AM-3D CNN maintains a leading position with a precision of 81%, while precisions of other models range from 75% to 80%. In the boredom emotion category, AM-3D CNN achieves a precision of 89%, surpassing other models, with 3D CNN and 2D CNN precisions at 82% and 77%, respectively. For the disgust emotion, AM-3D CNN demonstrates a precision of 91%, markedly higher than other models, where 3D CNN and 2D CNN achieve precisions of 86% and 81%, respectively. In the happiness and uncertainty emotions, AM-3D CNN achieves

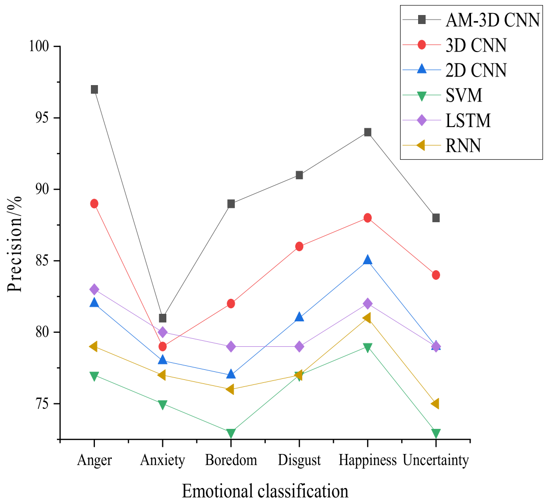


FIGURE 10. **Results of recognition precision rate of different models for different emotions.**

precisions of 94% and 88%, respectively, showing significant superiority over other models.

D. DISCUSSION

This work employs a novel model, AM-3D CNN, based on 3D CNN and attention mechanisms, making full use of multimodal data, including facial expressions and body movements, resulting in significant improvement. Compared to traditional CNN and other models such as SVM, LSTM, RNN, AM-3D CNN demonstrates superior recognition performance across different emotions. In comparison to prior research, this work introduces attention mechanisms, optimizing 3D CNN to better focus on crucial features when processing emotional data, thereby enhancing model accuracy and robustness. The model excels in recognizing emotions such as anger, anxiety, and happiness. It provides a more reliable and efficient solution for areas like affective computing, human-computer interaction, and emotional intelligence systems.

V. CONCLUSION

A. RESEARCH CONTRIBUTION

This work initially establishes the fundamental process for emotion recognition, followed by the construction of a CNN-based emotion recognition model, specifically utilizing 3D CNN. Finally, it optimizes the 3D CNN by introducing an attention mechanism and compares it with common emotion recognition models such as SVM, RNN, CNN, and LSTM. The research findings reveal that the AM-3D CNN model demonstrates relatively superior recognition performance across different emotions. Compared to other models, it exhibits excellence in recognizing emotions like anger, anxiety, and happiness, showcasing the multimodal emotional data's effective fusion and classification performance. Across all datasets, the AM-3D CNN model demonstrates relatively high emotion recognition accuracy, particularly achieving 88.81% on the D3 dataset. It exhibits outstanding performance across different emotion categories, showcasing robust emotion recognition capabilities.

B. FUTURE WORKS AND RESEARCH LIMITATIONS

While this work has achieved significant results, there are still some limitations. First, the training and optimization of the model might require a larger dataset to enhance generalization. Moreover, further in-depth research is needed for fine-tuning the attention mechanism. Future research directions could involve expanding the dataset, incorporating additional modalities such as text or audio to improve emotion recognition accuracy, and further optimizing the model architecture. Additionally, considering the impact of complex environments on emotion recognition would contribute to enhancing the research's practicality and applicability.

REFERENCES

[1] D. Grewal, A. Guha, E. Schweiger, S. Ludwig, and M. Wetzels, "How communications by AI-enabled voice assistants impact the customer journey," *J. Service Manag.*, vol. 33, no. 4, pp. 705–720, Jul. 2022.

[2] M. S. Frederiksen, V. Schmied, and C. Overgaard, "Creating coherent perinatal care journeys: An ethnographic study of the role of continuity of care for Danish parents in a vulnerable position," *Women Birth*, vol. 36, no. 1, pp. 117–126, Feb. 2023.

[3] A. Sayer and J. Morgan, "A realist journey through social theory and political economy: An interview with Andrew sayer," *J. Crit. Realism*, vol. 21, no. 4, pp. 434–470, Aug. 2022.

[4] S. Bagherzadeh, K. Maghooli, A. Shalbaf, and A. Maghsoudi, "Emotion recognition using effective connectivity and pre-trained convolutional neural networks in EEG signals," *Cognit. Neurodynamics*, vol. 16, no. 5, pp. 1087–1106, Oct. 2022.

[5] H. N. Vu, M. H. Nguyen, and C. Pham, "Masked face recognition with convolutional neural networks and local binary patterns," *Int. J. Speech Technol.*, vol. 52, no. 5, pp. 5497–5512, Mar. 2022.

[6] W. Zou, D. Zhang, and D.-J. Lee, "A new multi-feature fusion based convolutional neural network for facial expression recognition," *Int. J. Speech Technol.*, vol. 52, no. 3, pp. 2918–2929, Feb. 2022.

[7] M. F. Bashir, A. R. Javed, M. U. Arshad, T. R. Gadekallu, W. Shahzad, and M. O. Beg, "Context-aware emotion detection from low-resource Urdu language using deep neural network," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 5, pp. 1–30, May 2023.

[8] T. Debnath, M. M. Reza, A. Rahman, A. Beheshti, S. S. Band, and H. Alinejad-Rokny, "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity," *Sci. Rep.*, vol. 12, no. 1, p. 6991, Apr. 2022.

[9] A. Pise, H. Vadapalli, and I. Sanders, "Facial emotion recognition using temporal relational network: An application to e-learning," *Multimedia Tools Appl.*, vol. 81, no. 19, pp. 26633–26653, Aug. 2022.

[10] L. Trinh Van, T. Dao Thi Le, T. Le Xuan, and E. Castelli, "Emotional speech recognition using deep neural networks," *Sensors*, vol. 22, no. 4, p. 1414, Feb. 2022.

[11] L. Hansen, Y. Zhang, D. Wolf, K. Sechidis, N. Ladegaard, and R. Fusaroli, "A generalizable speech emotion recognition model reveals depression and remission," *Acta Psychiatrica Scandinavica*, vol. 145, no. 2, pp. 186–199, Feb. 2022.

[12] A.-H. Jo and K.-C. Kwak, "Speech emotion recognition based on two-stream deep learning model using Korean audio information," *Appl. Sci.*, vol. 13, no. 4, p. 2167, Feb. 2023.

[13] Y. Xu, H. Su, G. Ma, and X. Liu, "A novel dual-modal emotion recognition algorithm with fusing hybrid features of audio signal and speech context," *Complex Intell. Syst.*, vol. 9, no. 1, pp. 951–963, Feb. 2023.

[14] T. Xu, W. Dang, J. Wang, and Y. Zhou, "DAGAM: A domain adversarial graph attention model for subject-independent EEG-based emotion recognition," *J. Neural Eng.*, vol. 20, no. 1, Feb. 2023, Art. no. 016022.

[15] A. Iyer, S. S. Das, R. Teotia, S. Maheshwari, and R. R. Sharma, "CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings," *Multimedia Tools Appl.*, vol. 82, no. 4, pp. 4883–4896, Feb. 2023.

[16] Y. Song, Y. Yin, and P. Xu, "A customized ECA-CRNN model for emotion recognition based on EEG signals," *Electronics*, vol. 12, no. 13, p. 2900, Jul. 2023.
[17] Y. Zhang, Y. Zhang, and S. Wang, "An attention-based hybrid deep learning model for EEG emotion recognition," *Signal, Image Video Process.*, vol. 17, no. 5, pp. 2305–2313, Jul. 2023.
[18] S. Patnaik, "Speech emotion recognition by using complex MFCC and deep sequential model," *Multimedia Tools Appl.*, vol. 82, no. 8, pp. 11897–11922, Mar. 2023.
[19] M. J. Al-Dujaili and A. Ebrahimi-Moghadam, "Speech emotion recognition: A comprehensive survey," *Wireless Pers. Commun.*, vol. 129, no. 4, pp. 2525–2561, Apr. 2023.
[20] D. P. Goel, K. Mahajan, N. D. Nguyen, N. Srinivasan, and C. P. Lim, "Towards an efficient backbone for preserving features in speech emotion recognition: Deep-shallow convolution with recurrent neural network," *Neural Comput. Appl.*, vol. 35, no. 3, pp. 2457–2469, Jan. 2023.
[21] G. Ke, Q.-J. Jiang, A. J. Abualhamayl, and X.-Y. Xu, "Dynamic nonlinear differential investment decision model for scenic spot system with uncertainties and emergencies," *Fractals*, vol. 30, no. 2, Mar. 2022, Art. no. 2240108.
[22] R. Liu, "Intelligence information intervention in emergency management of emergencies: The case of the '7.20 Zhengzhou rainstorm,'" *Frontiers Manag. Sci.*, vol. 2, no. 3, pp. 18–27, Jun. 2023.
[23] J. Fasth, V. Elliot, and A. Styhre, "Crisis management as practice in small-and medium-sized enterprises during the first period of COVID-19," *J. Contingencies Crisis Manag.*, vol. 30, no. 2, pp. 161–170, Jun. 2022.
[24] S. Y. Oh, "From a super spreader of MERS to a super stopper of COVID-19: Explaining the evolution of South Korea's effective crisis management system," *J. Conting Crisis Man.*, vol. 15, no. 2, pp. 250–265. Jun. 2022.
[25] J. C. Richards, "Exploring emotions in language teaching," *RELC J.*, vol. 53, no. 1, pp. 225–239, Apr. 2022.
[26] A. Ortony, "Are all 'basic emotions' emotions? A problem for the (basic) emotions construct," *Perspect. Psychol. Sci.*, vol. 17, no. 1, pp. 41–61, Jan. 2022.
[27] M. K. Chowdary, J. Anitha, and D. J. Hemanth, "Emotion recognition from EEG signals using recurrent neural networks," *Electronics*, vol. 11, no. 15, p. 2387, Jul. 2022.
[28] P. Pandey and K. R. Seeja, "Subject independent emotion recognition from EEG using VMD and deep learning," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1730–1738. Oct. 2022.
[29] M. Krichen, "Convolutional neural networks: A survey," *Computers*, vol. 12, no. 8, p. 151, Jul. 2023.
[30] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: A survey," *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, Mar. 2022.
[31] S. Abbas, Y. Alhwaiti, A. Fatima, M. A. Khan, M. A. Khan, T. M. Ghazal, A. Kanwal, M. Ahmad, and N. Sabri Elmitwally, "Convolutional neural network based intelligent handwritten document recognition," *Comput., Mater. Continua*, vol. 70, no. 3, pp. 4563–4581, 2022.
[32] S. K. Towfek and N. Khodadadi, "Deep convolutional neural network and metaheuristic optimization for disease detection in plant leaves," *J. Intell. Syst. Internet Things*, vol. 10, no. 1, pp. 66–75, 2023.
[33] T. Hur, L. Kim, and D. K. Park, "Quantum convolutional neural network for classical data classification," *Quantum Mach. Intell.*, vol. 4, no. 1, p. 3, Jun. 2022.
[34] A. F. Nurjannah, A. S. D. Kurniasari, Z. Sari, and Y. Azhar, "Pneumonia image classification using CNN with max pooling and average pooling," *J. RESTI*, vol. 6, no. 2, pp. 330–338, Apr. 2022.
[35] A. Zafar, M. Aamir, N. Mohd Nawi, A. Arshad, S. Riaz, A. Alruban, A. K. Dutta, and S. Almotairi, "A comparison of pooling methods for convolutional neural networks," *Appl. Sci.*, vol. 12, no. 17, p. 8643, Aug. 2022.
[36] S. Li, J. Yu, X. He, H. Guo, and X. He, "VoxDMRN: A voxelwise deep max-pooling residual network for bioluminescence tomography reconstruction," *Opt. Lett.*, vol. 47, no. 7, pp. 1729–1732, Apr. 2022.
[37] S.-H. Wang, M. Attique Khan, and Y.-D. Zhang, "VISPNN: VGG-inspired stochastic pooling neural network," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 3081–3097, 2022.
[38] X. Li, X. Feng, H. Fang, N. Yang, G. Yang, Z. Yu, J. Shen, W. Geng, and Y. He, "Classification of multi-year and multi-variety pumpkin seeds using hyperspectral imaging technology and three-dimensional convolutional neural network," *Plant Methods*, vol. 19, no. 1, p. 82, Aug. 2023.
[39] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting," *Ann. Data Sci.*, vol. 10, no. 1, pp. 183–208, Feb. 2023.
[40] C. M. Driskill, J. E. Childs, B. Itmer, J. S. Rajput, and S. Kroener, "Acute vagus nerve stimulation facilitates short term memory and cognitive flexibility in rats," *Brain Sci.*, vol. 12, no. 9, p. 1137, Aug. 2022.
[41] D. Durstewitz, G. Koppe, and M. I. Thurm, "Reconstructing computational system dynamics from neural data with recurrent neural networks," *Nature Rev. Neurosci.*, vol. 24, no. 11, pp. 693–710, Nov. 2023.