Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/67/20240671

Medical image intelligent diagnosis system based on facial emotion recognition and convolutional neural network

Qian Leng[1,3,*,†], Lai Peng[2,†]

[1]University of Maryland Eastern Shore, MD, USA
[2]Northeastern University, OH, USA

[3]ql150@georgetown.edu
*corresponding author
[†]These authors also contributed equally to this work

Abstract. The facial emotional information of patients is one of the important indicators of their health status. By combining facial emotion recognition technology, medical imaging intelligent diagnostic systems can more comprehensively understand the psychological and physiological status of patients. This study first selected medical imaging datasets from the publicly available Medical Machine Learning Image Standard database and performed necessary preprocessing, including data cleaning, enhancement, and standardization. Next, a CNN model is constructed to extract key features from medical images and facial expression images through structures such as convolutional layers, activation functions, pooling layers, and fully connected layers. In addition, the system also integrates facial emotion recognition module and medical image analysis module, improving the comprehensiveness of diagnosis through feature fusion. Finally, the diagnostic results are displayed through a user interaction interface. Through comparative experiments, this study verified that the proposed CNN model outperforms traditional LSTM and SVM models in terms of accuracy, recall, and precision. The CNN model has shown significant advantages in medical imaging intelligent diagnosis systems, especially achieving a recall rate of 99%, demonstrating extremely high disease recognition ability. The results of this study not only provide new research directions for the field of intelligent diagnosis in medical imaging, but also lay a solid foundation for future clinical applications.

Keywords: Facial Emotion Recognition, Convolutional Neural Networks, Intelligent Diagnosis of Medical Imaging, Cost Effectiveness.

1. Introduction

The facial emotional information of patients also contains important clues to their health status. Emotion recognition technology can help doctors better understand the psychological state of patients, thereby providing more comprehensive medical services. Therefore, combining facial emotion recognition with CNN in medical imaging intelligent diagnosis systems has important research value and application prospects.

This article proposes a novel framework for intelligent diagnosis of medical images, which integrates facial emotion recognition technology and convolutional neural networks, aiming to improve the analytical ability and diagnostic comprehensiveness of medical images. This study designs a series of experiments to verify the effectiveness of the proposed system and evaluates its performance using actual medical datasets.

The article first introduces the background and research significance of medical imaging intelligent diagnosis system, and then elaborates on the proposed system framework and key technologies in detail. Next, this article demonstrates the experimental design, dataset preparation, and model training process, as well as how to combine facial emotion recognition and CNN to improve diagnostic performance. The article also discusses the experimental results and the performance of the system in practical applications, and finally summarizes the entire article and looks forward to future research directions.

2. Related Work

Numerous researchers are exploring the application of artificial intelligence (AI) technology in medical imaging diagnosis, continuously promoting innovation and development in this field. Based on the rapid development of deep learning and deep neural network technology in the field of computer vision, Zhang Kaixuan conducted research on improving the diagnostic efficiency of colorectal cancer using medical imaging artistic intelligence technology [1]. Based on the principle of intelligent medical imaging assisted diagnosis system and the analysis of its potential and advantages in specific applications, Sun Zhenhu explored its application value in improving the early diagnosis and treatment effect of lung cancer, identified the shortcomings of this technology, and further improved the technology and standards to ensure the reliability and safety of its application effect [2]. Jiang Liu compared the effectiveness of ultrasound assisted thyroid artificial intelligence (AI) diagnostic system with ultrasound physicians of different years in diagnosing medullary thyroid carcinoma (MTC), using papillary thyroid carcinoma (PTC) as a control [3]. Pan Yu selected 25000 cervical liquid based thin layer cell smears and classified cervical cells according to The Bethesda System 2014. He used "AI assisted+manual review" and "TIS assisted+manual review" to review the smears, and compared the application value of artificial intelligence (AI) assisted diagnostic systems and computer assisted review systems (TIS) in cervical cancer screening [4]. Chen Jianhua evaluated the effectiveness of AI technology in assisting undergraduate interns in teaching pulmonary nodules by comparing the theoretical scores, report writing time, nodule detection rate, and diagnostic report scores of students in the AI assisted teaching group and the traditional Picture Archiving and Communication System teaching group after internships [5].

In addition, Zhou S K first introduced the characteristics of medical imaging, emphasized the clinical needs and technical challenges of medical imaging, and described how the emerging trend of deep learning can solve these problems [6]. Jussupow E used induction to understand how artificial intelligence advice affected the decision-making process of doctors, introduced five decision-making patterns, and developed a medical diagnosis decision enhancement process model with artificial intelligence advice [7]. Abdulkareem K H introduced three main diagnostic scenarios for Corona Virus Disease 2019, such as diagnosis based on raw and standardized datasets and diagnosis based on feature selection. Compared with benchmark studies, his proposed SVM model achieved the most significant diagnostic performance (up to 95%) [8]. Turkoglu M used the AlexNet Convolutional Neural Network architecture based on pre-trained CNN, which employed transfer learning methods. Then he used Support Vector Machine (SVM) method to classify the effective features selected from all layers of the architecture using relief feature selection algorithm [9]. Chowdary M K processed emotion recognition using transfer learning methods, using pre-trained networks from Residual Network 50, Visual Geometry Group 19, Inception Network Version 3, and Mobile Convolutional Neural Networks [10]. Although the above research has made significant progress in the field of medical imaging intelligent diagnosis, its theoretical research is not yet in-depth enough. This article proposes a medical imaging intelligent diagnosis system based on facial emotion recognition and CNN. The system not only focuses on the recognition of disease markers, but also innovatively combines facial emotion analysis to provide more comprehensive patient information and help patients further recover.

DOI: 10.54254/2755-2721/67/20240671

3. Method

3.1. Dataset Preparation and Preprocessing

The role of medical imaging datasets in intelligent medical imaging diagnosis systems cannot be ignored [11-12]. Firstly, it is necessary to select these datasets, which are usually sourced from the publicly available Medical Machine Learning Image Standard database. This is a medical image dataset published by researchers from institutions such as Shanghai Jiao Tong University, which includes 2D and 3D biomedical image data. The 2D dataset contains 708069 images, while the 3D dataset contains 9998 images. These datasets cover data scales ranging from 100 to 100000.

The sample must cover a wide range of disease types and include samples of normal conditions, which is crucial for training models to recognize different disease features. Meanwhile, each sample should have accurate annotation information, including disease type, location, and severity, which is crucial for subsequent model training and validation, as shown in Table 1:

Table 1. Sample annotated data

Sample ID	Imaging Type	Disease Category	Location	Severity	Date of Annotation	Annotator
001	CT	Pneumonia	Right Lung	Moderate	2022-01-15	Dr. Zhang
002	MRI	Brain Tumor	Cerebellum	Mild	2022-01-20	Dr. Li
003	X-ray	Fracture	Left Arm	Severe	2022-02-05	Dr. Wang
004	CT	Tuberculosis	Both Lungs	Mild	2022-02-10	Dr. Zhao
005	MRI	Stroke	Brainstem	Moderate	2022-02-15	Dr. Chen
006	X-ray	Arthritis	Right Knee	Severe	2022-03-01	Dr. Zhou

Table 1 provides a detailed sample information set for the medical imaging intelligent diagnosis system. Through this data, the system can learn how to recognize the imaging features of different diseases, and thus provide faster and more accurate auxiliary diagnostic services in practical applications.

Emotion recognition selects specialized emotion recognition datasets that contain facial expression images of various basic emotion categories, such as happiness, sadness, anger, and surprise. Each facial expression image needs to be labeled with a corresponding emotion category for use in training emotion recognition models.

In the data preprocessing stage, data cleaning is first performed to remove noise from medical diagnostic images, improve image quality, and remove irrelevant information such as background and non facial regions from facial expression images [13-14]. The core idea of using Gaussian filters for denoising is to use Gaussian distribution functions as weights to perform convolution operations on the image, in order to smooth the image and reduce noise:

$$f(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

$f(x,y,\sigma)$ is a Gaussian function, σ is the standard deviation of a Gaussian distribution, and x and y are coordinates on the image plane. e is the base of natural logarithms, approximately equal to 2.718285, π is pi, a mathematical constant, approximately equal to 3.14159.

Rotating, flipping, scaling, and adjusting brightness data augmentation techniques can increase the diversity of the dataset, improve the model's generalization ability, and adaptability to environmental changes. Rotation transformation can be achieved through rotation matrix. For two-dimensional images, the rotation matrix R of rotation angle θ is defined as:

$$R(\theta)=\begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} \tag{2}$$

Feature extraction is an important step in the preprocessing process, which requires extracting key features of medical images, such as shape, texture, and edges, as well as key features of facial expressions, such as the position and shape of eyes, eyebrows, mouth, and the movement of facial muscles.

To train the model more efficiently, it is necessary to normalize and standardize the image data, scale the image pixel values to 0-1, and make the data mean 0 and variance 1. The normalization formula is:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{3}$$

x is the original pixel value, x' is the normalized value, and $min(x)$ and $max(x)$ are the minimum and maximum values of all pixel values in the dataset, respectively. Finally, converting the emotion category labels into a unique hot encoding form for the neural network to learn and classify:

$$onehot = \begin{cases} 1 & if\ \ j = k \\ 0 & otherwise \end{cases} \tag{4}$$

j is the category label, k is the true category of the current sample, and $one\ hot$ is the transformed unique hot encoding vector.

During the process, attention should be paid to maintaining the diversity and representativeness of the data, avoiding overfitting or underfitting, in order to improve the accuracy and reliability of medical imaging intelligent diagnostic systems.

3.2. Construction of Convolutional Neural Network Model

The model input layer in the medical imaging intelligent diagnosis system is responsible for receiving preprocessed medical images and facial expression image data, which are the foundation of model learning [15-16]. Next, the convolutional layer utilizes a set of learnable filters to extract local features in the image, which is the core of convolutional neural networks. Through the characteristics of local receptive field weight sharing and sparse connections, it effectively reduces model parameters and improves computational efficiency.

The application of CNN models in medical imaging intelligent diagnosis systems has achieved significant results. The CNN model, through its unique structural design, can effectively extract useful features from medical images and facial expression images, and perform accurate classification and recognition.

In the CNN model, the input layer receives preprocessed image data, which is processed by filters in the convolutional layer. The weights of the convolutional layer are learned and can capture local features of the image. These filters can detect visual elements such as edges, textures, and shapes, providing a foundation for subsequent analysis. As the network hierarchy deepens, the nonlinearity introduced through activation functions enables the model to handle more complex image features.

The pooling layer is used to reduce the spatial size of the feature map, reduce the number of parameters and computational complexity of subsequent layers, while maintaining the recognition ability of important features. This downsampling operation helps the model focus on key information, reduce sensitivity to details, and thus improve the model's generalization ability.

The fully connected layer is located at the end of the network and is responsible for integrating and classifying the previously extracted features. These layers learn advanced combinations between features through fully connected methods, providing support for the final classification task. The output layer usually uses the Softmax function to convert the network's output into a probability distribution, representing the model's prediction probability for each category.

The loss function is used to measure the difference between model predictions and actual labels, while the optimizer minimizes the loss function by adjusting network weights. The setting of learning rate and momentum is crucial for the training effectiveness of the model, as they affect the convergence speed and final performance of the model. The choice of batch size also affects the training efficiency and memory consumption of the model. Regularization techniques and early stopping methods help prevent overfitting and ensure the model's generalization ability on new data.

DOI: 10.54254/2755-2721/67/20240671

This structural design and optimization strategy help to construct CNN models that can be effectively applied to facial emotion recognition and medical image intelligent diagnosis tasks. However, in practical applications, adjustments and optimizations need to be made based on specific problems and the characteristics of the dataset to achieve the best diagnostic results [17-18].

3.3. System Framework Design and Implementation

The overall architecture design of the diagnostic system aims to achieve an efficient process from data input to final diagnostic result output. The system first receives medical image data and facial expression image data through a data input module, which are sourced from the hospital's image database or on-site collection. The data preprocessing module performs necessary cleaning, standardization, and enhancement on these raw data to ensure the quality of the data and the effectiveness of model training.

The facial emotion recognition module utilizes pre trained CNN models to recognize and analyze facial expressions, extract emotion related features, while the medical image analysis module focuses on analyzing medical images, identifying potential disease markers and abnormal features [19-20]. The outputs of these two modules are then fed into the feature fusion module, where emotional and physiological features are effectively combined to provide more comprehensive information for comprehensive analysis.

The fused data is sent to the diagnostic decision module, which uses CNN to analyze features and generate diagnostic reports. Ultimately, the user interaction interface presents the diagnostic results to doctors and patients in an easily understandable form, while providing further medical advice and treatment plans. Through this architecture design, the system not only improves the accuracy and efficiency of medical diagnosis, but also to a certain extent improves the patient's medical experience, especially when combined with emotional recognition features, it can provide doctors with more comprehensive disease information, help them better understand the patient's condition and formulate appropriate treatment plans.

4. Results and Discussion

The application of facial recognition and CNN in intelligent diagnostic systems has been achieved, and reliable experimental data is needed to support the specific application effect of CNN. This study aims to comprehensively compare CNN models with other popular machine learning algorithms, including Long Short Term Memory Networks (LSTM) and Support Vector Machines (SVM), through a series of comparative experiments to evaluate their application effectiveness. This study will analyze the specific application performance of each model in depth from three dimensions: accuracy, precision, and recall.

Firstly, it is necessary to prepare a consistent testing environment to ensure consistency of results and avoid interference from external variables, including ensuring consistent hardware configurations, consistent software versions, and standardization of data preprocessing processes. Subsequently, load the trained and optimized CNN model and compare the LSTM and SVM models to ensure that all models are in the best testing state. After defining performance evaluation metrics, testing can begin.

This article will collect 12 sets of data from the Medical Machine Learning Image Standard database for experiments. Through the evaluation indicators of accuracy, precision, and recall, the application effects of these three models in the medical image intelligent diagnosis system based on facial emotion recognition will be analyzed and compared in depth. Accuracy will reflect the overall correct classification ability of the model, while precision will measure the proportion of samples predicted to be truly positive, while recall focuses on the model's ability to recognize all positive samples. By comparing these indicators, this study can better understand the advantages and limitations of each model, providing guidance for future research and practical applications.

4.1. Accuracy

Accuracy is a key indicator for measuring model performance, which directly reflects the overall correctness of the model in classification tasks. The experimental data is shown in Figure 1:

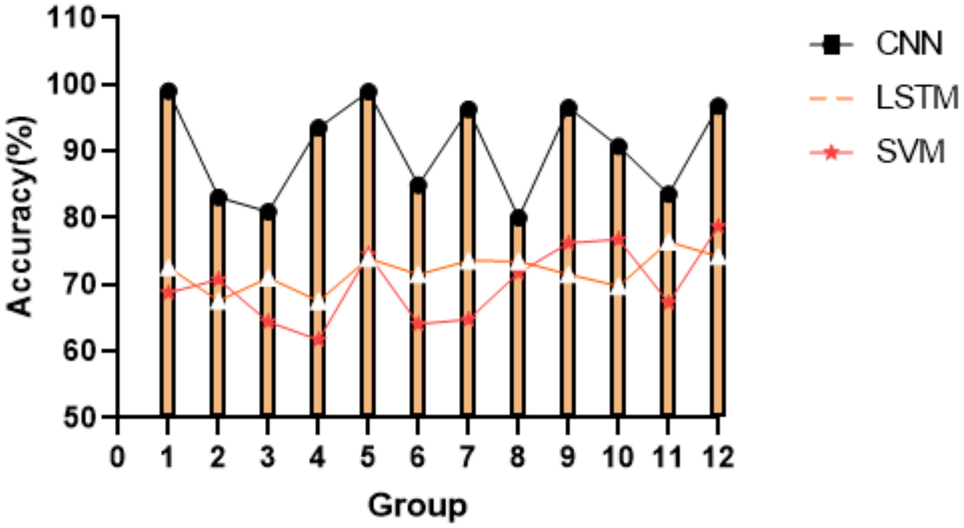

Figure 1. Accuracy Comparison

As shown in Figure 1, CNN has a significant advantage in accuracy in intelligent diagnostic systems, with the highest accuracy reaching 99%, compared to 76.4% for LSTM and 78.7% for SVM. This result indicates that when processing medical image data, CNN can effectively capture complex features in the image and perform accurate classification. This high-performance performance is due to the multi-layer convolution and pooling operations of CNN, which gives it a natural structural advantage in image recognition tasks.

4.2. Recall

Recall rate is a crucial evaluation indicator that measures the model's ability to correctly identify all positive samples. The recall rate directly affects whether the diagnostic system can detect and warn potential health risks in a timely manner, which is of great significance for improving the success rate of early disease detection and treatment. As shown in Figure 2:

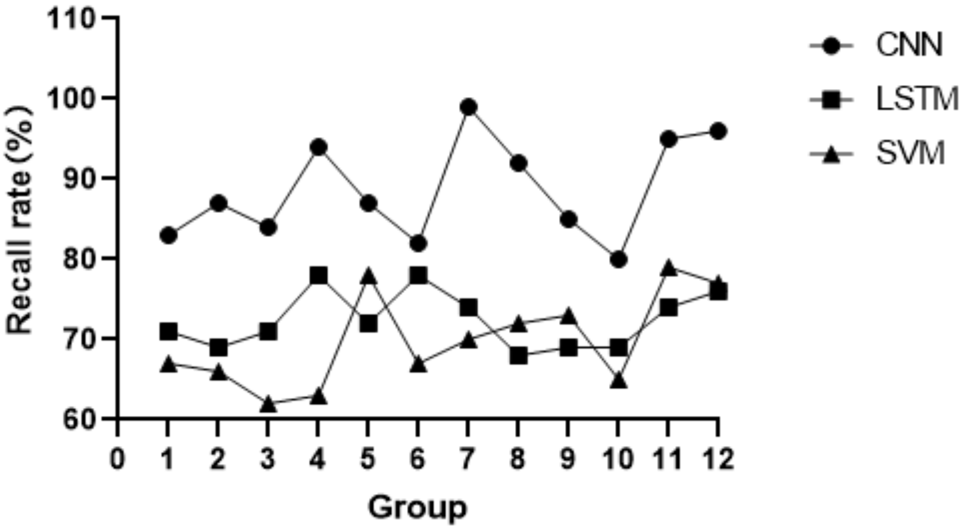

Figure 2. Comparison of recall rates

From the data in Figure 2, it can be seen that the recall rate of CNN is higher than that of LSTM and SVM, with the highest recall rate reaching 99%. In the same situation, LSTM is only 78%, while SVM is higher, but only 79%. This indicates that in medical imaging intelligent diagnosis systems, CNN can effectively identify the vast majority of positive samples, which are actually samples in disease status.

4.3. Accuracy

High accuracy means that the model has a higher proportion of true disease in the identified positive samples, thereby reducing the risk of misdiagnosis. As shown in Figure 3:

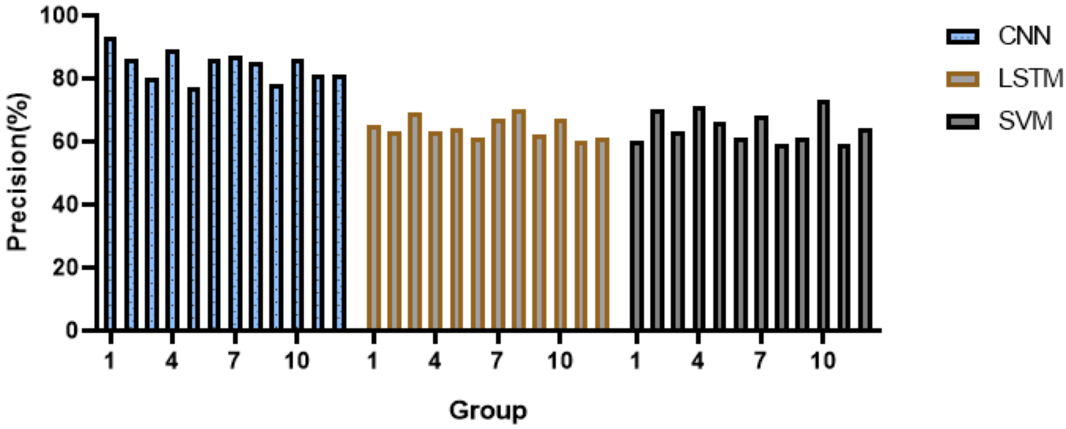

Figure 3. Precision Comparison

As shown in Figure 3, it can be seen that the cost-effectiveness of CNN is significantly higher than that of LSTM and SVM. Among them, CNN has a maximum accuracy of 93% and a minimum accuracy of 77%, while LSTM has a maximum and minimum accuracy of 70% and 60%, and SVM has a maximum and minimum accuracy of 73% and 59%, respectively. There is still a considerable gap compared to CNN. These data indicate that although LSTM and SVM may have advantages in certain specific tasks, from the perspective of accuracy, CNN provides higher accuracy.

5. Conclusion

This study successfully developed and validated a medical imaging intelligent diagnosis system based on the combination of facial emotion recognition and CNN. Through comparative experiments, this study found that the system performs well in key performance indicators such as accuracy, recall, and precision, demonstrating its high efficiency and reliability in identifying diseases. This achievement not only demonstrates the important value of facial emotion information in medical diagnosis, but also demonstrates the potential of deep learning technology in improving diagnostic accuracy and comprehensiveness.

The successful implementation of the system provides new ideas and methods for future medical imaging diagnosis. By combining advanced image processing techniques and deep learning algorithms, this study can more effectively assist doctors in disease diagnosis and improve the quality of medical services. In addition, this study also provides valuable experience and data support for subsequent research work. In the future, the performance and application scope of the system can be further improved by introducing more data types, optimizing model structures and algorithms.

References

[1] Zhang Kaixuan, Ding Kang. Application progress of deep learning based artificial intelligence technology in medical imaging diagnosis of colorectal cancer [J]. Shandong Medical Journal, 2024,64 (4): 107-110.

[2] Sun Zhenhu, Hu Fangyun, Xu Min, Zhou Jianguo. Clinical application exploration of artificial intelligence medical imaging assisted diagnostic system in lung nodule screening [J]. China Health Industry, 2023,20 (22): 149-151+155.
[3] Jiang Liu, Chen Lei, Zhang Xiaoting, Liu Chang, Liang Zhenwei, Sun Xiuming, Shao Yuhong, Chen Luzheng. Ultrasound assisted artificial intelligence diagnostic system for medullary thyroid carcinoma [J]. Chinese Medical Imaging Technology, 2024,40 (2): 208-211.
[4] Pan Yu, Zhou Donghua, Fan Juhua, Ping Ping, Ouyang Chenjie, Ma Huisi, Xu Xuefen, Liu Lantian. Comparative observation of the application of artificial intelligence assisted diagnostic systems and computer-aided film reading systems in cervical cancer screening [J]. Shandong Medical Journal, 2024,64 (4): 55-57.
[5] Chen Jianhua, Sun Bin, Ge Huiting, Lin Li, Li Jian. Preliminary application of artificial intelligence technology in teaching medical imaging diagnosis of pulmonary nodules [J]. Journal of Fujian Medical University (Social Sciences Edition), 2023,24 (4): 63-67.
[6] Zhou S K, Greenspan H, Davatzikos C, et al. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises[J]. Proceedings of the IEEE, 2021, 109(5): 820-838.
[7] Jussupow E, Spohrer K, Heinzl A, et al. Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence[J]. Information Systems Research, 2021, 32(3): 713-735.
[8] Abdulkareem K H, Mohammed M A, Salim A, et al. Realizing an effective COVID-19 diagnosis system based on machine learning and IOT in smart hospital environment[J]. IEEE Internet of things journal, 2021, 8(21): 15919-15928.
[9] Turkoglu M. COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble[J]. Applied Intelligence, 2021, 51(3): 1213-1226.
[10] Chowdary M K, Nguyen T N, Hemanth D J. Deep learning-based facial emotion recognition for human–computer interaction applications[J]. Neural Computing and Applications, 2023, 35(32): 23311-23328.
[11] Rajpurkar P, Lungren M P. The current and future state of AI interpretation of medical images[J]. New England Journal of Medicine, 2023, 388(21): 1981-1990.
[12] Lebovitz S, Lifshitz-Assaf H, Levina N. To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis[J]. Organization science, 2022, 33(1): 126-148.
[13] Budholiya K, Shrivastava S K, Sharma V. An optimized XGBoost based diagnostic system for effective prediction of heart disease[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(7): 4514-4523.
[14] Rawal N, Stock-Homburg R M. Facial emotion expressions in human–robot interaction: A survey[J]. International Journal of Social Robotics, 2022, 14(7): 1583-1604.
[15] Arora M, Kumar M. AutoFER: PCA and PSO based automatic facial emotion recognition[J]. Multimedia Tools and Applications, 2021, 80(2): 3039-3049.
[16] Li Z, Liu F, Yang W, et al. A survey of convolutional neural networks: analysis, applications, and prospects[J]. IEEE transactions on neural networks and learning systems, 2021, 33(12): 6999-7019.
[17] Lindsay G W. Convolutional neural networks as a model of the visual system: Past, present, and future[J]. Journal of cognitive neuroscience, 2021, 33(10): 2017-2031.
[18] Sarvamangala D R, Kulkarni R V. Convolutional neural networks in medical image understanding: a survey[J]. Evolutionary intelligence, 2022, 15(1): 1-22.
[19] Ilesanmi A E, Ilesanmi T O. Methods for image denoising using convolutional neural network: a review[J]. Complex & Intelligent Systems, 2021, 7(5): 2179-2198.
[20] Habib G, Qureshi S. Optimization and acceleration of convolutional neural networks: A survey[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(7): 4244-4268.