A Comparative Analysis Between Machine and Deep Learning Models for Facial Emotion Recognition

Genghao Du

Quanzhou NO.5 Middle School, Fujian, 362000, China
100266@yzpc.edu.cn

Abstract. Facial emotion recognition is vital for enhancing human-computer interactions, providing personalized services, and diagnosing mental health conditions. This technology holds promise for improving accessibility and emotional intelligence in various fields, from healthcare to education. In recent years, Artificial Intelligence (AI) has made significant improvements in emotion recognition, particularly in the development of facial emotion classification models. Deep learning dominates this domain, yet conventional machine learning approaches, e.g. Support Vector Machine (SVM), still hold value in applications. Each model has its strengths and weaknesses, necessitating a choice based on specific needs. At the forefront, neural network models like Convolutional Neural Networks (CNNs) have excelled in facial expression recognition, with numerous studies enhancing recognition accuracy through optimized algorithms. However, AI-based emotion recognition still faces challenges of data volume and interpretability. Looking ahead, AI emotion recognition holds broad application prospects in areas like mental health and education, promising to bring more convenience to human society.

Keywords: Emotion Recognition, Support Vector Machine, Convolutional Neural Networks.

1 Introduction

Artificial Intelligence (AI) has become a buzzword in the field of technological innovation in recent years. Its applications span across various domains, driving innovation in healthcare, finance, transportation, and education. It optimizes operations, enhances decision-making, and improves user experiences through personalized interactions. AI's significance lies in its ability to process vast data, automate tasks, and develop intelligent systems that augment human capabilities, leading to breakthroughs in efficiency and new discoveries. With its continuous development facial emotion recognition becomes a hotspot of its application.

Emotion recognition could be exploited for inferring human emotion, via various kinds of signals. It now often refers to AI-based automatic recognition system, which is an important aspect of affective computing. The application of AI in this area has greatly helped research in psychology-related fields such as mental health monitoring.

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_40

At the same time, researchers are actively expanding the practical applications of AI emotion recognition in fields such as market research, education, and security.

Among all data sources, facial images have also become an important physical signal for AI recognition. It offers several advantages over physiological signal-based methods. It is non-invasive and respects privacy, as it does not require any physical contact or sensors attached to the body. This approach is also more natural and less obtrusive, allowing for real-time emotion detection without the need for individuals to wear or interact with monitoring devices. Furthermore, advancements in computer vision and machine learning have strongly advanced the precision of facial emotion recognition. It shapes this setting as a robust alternative for emotion analysis in various settings, from mental health to customer service.

Therefore, this article classifies, organizes, and analyzes the facial emotion classification models used by researchers in recent years, summarizes the advantages and disadvantages of all mentioned models, and aims to provide reference value for subsequent researchers in the selection of decision-making models. In addition, this article will also provide prospects for future practical applications and development.

2 Machine Learning-based Facial Emotion Recognition

Machine learning (ML) emerged in the 1980s as a type of artificial intelligence decision-making. As described by Dr. Danko Nikolic, machine learning aims at equipping a computer with the capacity to automatically operate without explicit programming, via learning from their own experiences [1].

However, with the advent and development of deep learning, researchers using strategic models belonging to the ML decision-making type have been relatively fewer in facial emotion recognition. In a paper by Smith K. Khare et al., they conducted a survey summarizing works from 2014 to 2023 of facial emotion recognition models [2]. Among nearly 30 models, those using ML models accounted for less than 30%, with models being commonly used for facial emotion classification.

2.1 Support Vector Machine (SVM) Model

SVM, an effective supervised approach, is broadly applied for regression and classification. It operates by seeking a hyperplane that optimally separates data into various classes. The key principle is to maximize the margin among different classes, which are known as support vectors. They are samples that closest to the hyperplane and heavily influence its position and orientation. SVM can handle high-dimensional spaces and non-linear classification by using kernel functions, to represent input samples as high dimensional vectors for convenient separation. Its effectiveness lies in its ability to handle a wide range of problems, from simple to complex, and its strong mathematical foundation.

The disadvantage of the SVM is a common issue with ML models, which is the use of manually designed fixed designs to extract features, wasting big data resources.

In the study by Horisu Abdullashi Shehu et al., they chose SVM and gave reasons: even though deep learning models perform very well in many classification-based tasks, as an anti-attack model, SVM is more capable of resisting various types of attacks [3]. At the same time, SVM has a larger application in small sample datasets. In the model by Mohseni et al., features were extracted by measuring the proportions of facial graphics profiles. Haq et al. used facial landmark positions (MP) and landmark angles (MA) as features. In this work, validated on the Surrey Audio-Visual Expressed Emotion (SAVEE) dataset, the proposed method achieved a result of 88% [3].

2.2 Other Machine Learning Models

As a neural network (NN), the application of Artificial Neural Networks (ANN) in emotion classification is far less than that of Convolutional Neural Networks (CNN), limited by image size, which may lead to a dramatic increase in the number of trainable parameters, causing gradient disappearance or explosion. This method is also not very common in ML. However, in the research by A.K. Hassan, S.N. Mohammed, due to the starting point of graph mining, they proposed a new facial emotion detection method, using graph mining technology for minimizing the dimension of extracted features, making ANN capable of effectively processing image data. In Munasinghe's research, by only extracting facial features from the distance features closest to the eyes and nose [4]. In the test of this method using an emotional analysis database, the average success rate for only four emotions reached 90%.

2.3 Summary

Manually designing and extracting fixed features is an unavoidable issue for ML models, which can limit their effectiveness in handling vast datasets efficiently. The reliance on predetermined feature sets can restrict the model's capabilities to obtain the nuanced variations present in large and complex datasets, which are often required for accurate emotion classification.

Despite these limitations, ML models still offer significant value, particularly in scenarios where data is limited or the computational resources are constrained. Their simplicity and interpretability make them a preferred choice for applications where understanding the decision-making process of the model is crucial. Furthermore, ML models can be highly effective when tailored to specific tasks. For instance, in environments where certain types of attacks are anticipated, ML models can be designed with built-in defenses, leveraging their adaptability to specific threat vectors. This makes them indispensable in cybersecurity applications where robustness against adversarial attacks is paramount.

In conclusion, while deep learning's advanced capabilities are undeniable, ML models still hold their ground in specific contexts. Researchers should weigh the trade-offs between the two approaches carefully, taking into account the data distribution, the task-specific requirements at hand, and the accessible computational resources. By doing so, they can harness the full potential of ML models, leveraging their strengths in scenarios that align with their design and capabilities.

3 Deep Learning-based Facial Emotion Recognition

In 2006, a seminal breakthrough for machine learning was published by the doyen of the field, Hinton, and his team [5]. Their revolutionary article introduced the concept of deep learning (DL) for the first time. It elucidates that the difficulties within the learning process of deep neural networks could be alleviated by layer-wise initialization.

The most significant characteristic of deep learning models is their ability to enhance classification accuracy through autonomous learning, which is also quite convenient when performance improvements are needed. Due to these advantages, DL models have garnered favor among researchers in recent years, with as many as 20 out of 28 articles prioritizing DL models for classification. Among these, 17 models related to neural networks, indicating that DL has become the preferred choice for scholars in facial emotion classification in recent years.

3.1 Convolutional Neural Network (CNN) Model

CNNs operate on the principle of filtering input images with a set of convolutional filters to extract features. These filters automatically learn to recognize patterns like edges, textures, and shapes through a deep learning process. The superiority of CNNs lies in their efficiency and ability to handle spatial hierarchies. They reduce computational load through parameter sharing, where the same filter is applied across the entire input image. This feature learning is hierarchical, with deeper layers capturing more complex patterns. CNNs are translation invariant, allowing them to detect features regardless of their position in the image, and are highly effective for tasks such as image classification and object detection.

In the research of Irfan Haider and others, neural networks can better learn from the most challenging data triads. This work presents a robust approach for human face emotion classification by leveraging CNNs with triplet-loss-based features. It combines the power of CNNs for feature extraction with the classification capabilities of SVM, resulting in improved accuracy and reliability in recognizing human emotions from facial expressions [6].

In the research of Asad Khattak, et al, a convolutional neural network model was applied for categorizing emotions reading from human face and meanwhile recognizing age and gender of the specific people. The emotional recognition accuracy of all models reached an astonishing 95.65%, age recognition accuracy was 98.5%, and gender recognition accuracy was 99.14% [7].

It can be seen that the improvement of CNNs in accuracy is immediate. In addition, composite CNN models are also noteworthy. In the research of Ninad Mehendale, they proposed a new facial emotion recognition technology (FERC) using CNNs. FERC contains two sub-models: the first one removes the background from the image, and the second one conducts feature extraction depending on facial input [8]. This study, by integrating CNNs to form a composite convolutional neural network, achieved significant results, even though they did admit some limitations of the method, such as the high computational power during CNN tuning and the issues caused by facial hair. However, apart from these issues, their algorithm's accuracy was very high (i.e., 96%).

3.2 Other Deep Learning Models

In 2022, Professor Zhang used a facial expression recognition method based on learning spatial attention networks (SAN), aiming to make the model focus on the importance of different areas of the face through deep learning. He improved recognition accuracy by focusing on capturing information from key areas in expressions [9].

Another paper proposed a facial expression recognition method using an adaptive graph attention network (AGAN), which can automatically learn the relationships and importance of facial images and dynamically adjust the structure of the graph during the recognition process [10]. The characteristic of the adaptive graph attention network is its ability to dynamically adjust the network structure according to the relationships found between facial expressions (adaptive), thereby improving recognition performance and adapting to different data distributions and scenarios.

A paper tackled facial expression recognition exploiting hierarchical graph convolutional networks (HGNN), which models the hierarchical structure of facial images, extracts feature at different levels, and performs joint learning and inference [11]. This is similar to the previously mentioned FERC. Hierarchical graph convolutional networks can effectively capture multi-scale information and hierarchical structures in facial expressions, thereby improving recognition performance and achieving better results in tasks of different complexities and granularities of facial expression recognition.

3.3 Summary

In summary, non-CNN classification models are still mainly neural networks, relying on the accurate identification of data with spatial structures by neural networks. In addition, neural networks can adapt to different tasks via increasing layers or adjusting the network structure. This is also what researchers are keen to see.

Even though CNNs can automatically learn and extract features and are suitable for processing data with spatial structures, this does not mean they are perfect. The demand for data volume makes them a model with considerable costs. At the same time, because humans cannot interpret the decision-making process inside AI, this will also affect people's trust in the model.

4 Discussion and Prospect

This article aims to assist researchers in selecting the appropriate decision-making models. Clearly, for current research, CNN and most DL models are the ones with higher accuracy. The choice of ML models needs to consider the actual situation; for instance, their characteristic of being less susceptible to attacks can be applied in terms of security. Before more powerful decision-making models emerge, choosing CNN as a classification model may become the norm, or perhaps modifying and stacking neural networks to achieve higher accuracy.

Compared to choosing the right decision-making model, the author believes that the greater challenge is the level of trust humans have in AI. Researchers who have not yet left the era of weak AI cannot expect all experts and psychologists to fully trust AI algorithms. The worst-case scenario is that doctors completely distrust the results given by AI, while a better scenario is that doctors take AI's suggestions into consideration. This is also the greatest contribution that AI emotion recognition can currently make in psychotherapy.

At the same time, exploring more application scenarios is something that researchers should pay more attention to. Although AI emotion recognition has a broad range of application prospects, in reality, researchers have explored less in areas outside of psychological research. For example, in the author's previous research, an application model was sought in education, but only the EmotionCues system, which focuses on students' classroom attention, was found (2). Further research, such as students' satisfaction with classroom education, is also a field that can be studied in depth. In addition, the fields of education and security mentioned at the beginning are also worth delving into.

5 Conclusion

This article delves into the intricacies of classification decision models within the realm of facial emotion recognition, drawing upon an extensive review of academic papers published in recent years. It examines the various models. By undertaking a comparative analysis that highlights the merits and demerits of each model, the article serves as a valuable guide for researchers who are on the cusp of selecting a decision model tailored to their specific needs and objectives.

In addition to offering a critical evaluation of existing models, the article discusses the prevalent trends, the challenges faced, and the milestones achieved. Furthermore, the article extends beyond theoretical discussions and ventures into the practical implications of these models. It explores the potential applications of facial AI emotion recognition technology across different sectors and contemplates how these models can be effectively utilized in real-world scenarios. The prospects for model application are examined, with an emphasis on the adaptability and scalability of these systems to meet diverse requirements. By providing a thorough analysis and forward-looking insights, the article equips readers with the knowledge to navigate the complex domain of facial AI emotion recognition. It underscores the importance of making informed decisions when choosing a classification model and encourages the exploration of innovative applications that can harness the full potential of this technology.

References

1. Nikolić, D. Strong machine learning: A way towards human-level intelligence. ResearchGate. 1-21 (2024)
2. Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R.: Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. Information Fusion, **102**, 102019 (2023).
3. Shehu, H. A., Sharif, M. H., & Uyaver, S.: Facial expression recognition using deep learning. In International Conference of Mathematical Sciences, vol. 2334, pp. 17-21. AIP Publishing, Istanbul (2021).
4. Hassan, A. K., & Mohammed, S. N.: A novel facial emotion recognition scheme based on graph mining. Defence Technology, **16**(5), 1062-1072 (2020).
5. Hinton, G. E., Osindero, S., & Teh, Y. W. A fast learning algorithm for deep belief nets. Neural computation, **18**(7), 1527-1554 (2006).
6. Haider, I., Yang, H. J., Lee, G. S., & Kim, S. H.: Robust human face emotion classification using triplet-loss-based deep CNN features and SVM. Sensors, **23**(10), 4770 (2023).
7. Khattak, A., Asghar, M. Z., Ali, M., & Batool, U.: An efficient deep learning technique for facial emotion recognition. Multimedia Tools and Applications, **81**(2), 1649-1683 (2022).
8. Mehendale, N.: Facial emotion recognition using convolutional neural networks (FERC). SN Applied Sciences, **2**(3), 446 (2020).
9. Zhang, J., Wang, W., Li, X., & Han, Y.: Recognizing facial expressions based on pyramid multi-head grid and spatial attention network. Computer Vision and Image Understanding, **244**, 104010 (2024).
10. Li, Y., Xi, M., & Jiang, D.: Cross-view adaptive graph attention network for dynamic facial expression recognition. Multimedia Systems, **29**(5), 2715-2728 (2023).
11. Liu, D., Zhang, H., & Zhou, P.: Video-based facial expression recognition using graph convolutional networks. In International Conference on Pattern Recognition, pp. 607-614. IEEE, Taichung (2021).