International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Evaluation of a CNN-Based Model for Predicting Human Psychological States Using Facial Features: Benefits and Limitations

Parth Shinde[1], Nikhil Satpute[2], Siddhant Nimbalker[3]

[1,2,3]Student, Department of Computer Science, PVG's College of Science & Commerce.

Abstract

This research evaluates the advantages and limitations of a CNN-based model for predicting human psychological states through facial analysis. This approach innovatively leverages CNNs to interpret facial expressions, which are essential for mental health assessment and human-computer interaction. The system provides an innovative approach to understanding human emotions by analysing facial expressions, a critical aspect of mental health assessment and human-computer interaction. However, while the method offers substantial advancements in emotion recognition, it faces challenges such as computational costs, dataset limitations, and an over-reliance on facial features. This paper explores the model's strengths and limitations, proposing improvements and suggesting directions for future research.

Keywords: Emotion Recognition, Convolutional Neural Networks, Psychological State Analysis, Facial Feature Extraction, Deep Learning, Real-time Processing.

1. Introduction

Facial expressions are an essential aspect of human communication, reflecting emotional states and influencing interpersonal interactions. Emotion recognition has advanced significantly with deep learning, especially CNNs, transforming how facial features are processed and understood[1], [2]. This research evaluates a CNN-based model for predicting human emotions, considering its benefits and limitations in real-world applications. While the model demonstrates potential for use in various domains, such as mental health diagnostics and human-computer interaction, it also encounters several constraints that need to be addressed for broader adoption.

2. Literature survey

2.1 Lightweight CNNs for Emotion Recognition

Recent studies highlight lightweight CNN architectures for real-time emotion recognition, optimizing computational efficiency with mechanisms like multitask convolutional networks (MTCNN) and global average pooling[3], [4]. These innovations allow for more portable and adaptable systems, reducing memory consumption without sacrificing performance. However, despite these improvements, real-world challenges like diverse lighting conditions and varying head poses still affect model robustness.

2.2 Context-Aware Emotion Detection

A different approach to emotion recognition includes the analysis of environmental context alongside facial expressions. Context-aware models enhance accuracy by considering not only facial cues but also the surrounding objects and interactions[5]. These models are more adept at interpreting emotions in dynamic environments but are computationally expensive and require sophisticated data collection techniques that may not be feasible in real-time applications.

2.3 Emotion Detection in Humanoid Robots

Humanoid robots benefit from integrated face and emotion recognition systems that enhance their ability to interact with humans. Real-time CNN implementations for such robots have achieved high success rates in emotion recognition [6]. However, the complexity of simultaneously performing facial recognition and emotion detection increases the computational load, making it difficult to achieve this in low-resource settings.

2.4 Implementation of CNN and Haarcascade-Based Emotion Detection

The emotion detection model uses a combination of CNNs and Haarcascade classifiers. CNNs are employed to classify facial expressions, while Haar cascade is used for face detection. The architecture consists of multiple convolutional layers for feature extraction, followed by pooling and dropout layers to reduce spatial dimensions and prevent overfitting.

3. Model architecture

3.1 Convolutional Layers

These layers capture features such as edges and textures from input images by applying filters. Each successive layer increases the number of filters to learn more complex patterns from facial features.

3.2 Pooling Layers

MaxPooling layers down-sample feature maps, reducing computational load while preserving key information.

3.3 Dropout Layers

Dropout is applied at different stages to prevent the model from overfitting to the training data. This helps generalize the model's performance across different datasets.

3.4 Fully Connected Layers

After feature extraction, the data is passed through fully connected layers to perform emotion classification. A softmax activation function in the final layer outputs probabilities for each emotion class.

4. Methodology

The model was trained on the FER-2013 dataset, which consists of labelled images across seven emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral [7]. To ensure the robustness of the model, several preprocessing techniques were applied, including resizing, normalization, and data augmentation (e.g., flipping and rotation).

4.1 Data Preprocessing

Images were resized to 48x48 pixels, with pixel values normalized between 0 and 1. Data augmentation techniques were used to increase the diversity of the dataset, preventing the model from overfitting on specific facial features or expressions[8].

4.2 Training and Evaluation

For the analysis, the dataset was organized into training, validation, and testing subsets. The model training process utilized the Adam optimizer and employed categorical cross-entropy as the loss function. Validation helped monitor performance and tune hyperparameters. The final model was evaluated on the test set using accuracy, precision, recall, and F1-score metrics[9], [10].

4.3 Comparison of Algorithms

The CNN model, while achieving high accuracy, does not outperform more lightweight architectures such as MobilenetV2. This highlights the trade-off between model complexity and performance, with simpler models being more efficient while maintaining competitive accuracy[11].

5. Benefits

5.1 Effective Feature Extraction

The use of multiple convolutional layers (32, 64, 128 filters) allows the model to capture detailed and intricate facial features, leading to high accuracy in emotion detection.

5.2 Real-Time Capability

The architecture supports near real-time emotion detection, making it suitable for interactive systems, such as human-computer interactions or mental health monitoring.

5.3 Regularization and Generalization

Dropout layers (rates 0.25 and 0.5) minimize overfitting, allowing the model to generalize well across diverse datasets.

5.4 Adaptability

The model's structure is flexible and can be easily modified to include new emotion categories or additional refined layers, making it versatile for different applications.

5.5 Efficient Face Detection

The integration of the Haar Cascade algorithm for face detection is a reliable and well-established method that ensures accurate facial region detection before emotion classification.

5.6 Fast and Efficient Predictions

The model delivers rapid predictions, which is crucial for real-time applications like interactive systems and sentiment analysis.

6. Limitations

6.1 Overfitting Risk

Despite dropout layers, the model’s complex architecture may overfit, particularly on limited datasets like FER-2013.[12]

6.2 Context Ignorance

The model solely relies on facial features for emotion detection, ignoring other critical contextual information like body language, voice tone, or environmental cues, which may lead to misinterpretations[13].

6.3 Limited Dataset Generalization

The reliance on the FER-2013 dataset, which has known biases, may hinder the model’s ability to generalize well across diverse populations or cultural differences in emotional expressions[14].

6.4 Individual Variability in Emotion

Emotions can vary significantly between individuals, which can challenge the model's ability to accurately identify emotions across different users and facial structures[15].

6.5 Dataset Bias and Lower Accuracy

The FER-2013 dataset is imbalanced, which may skew results. Additionally, the model’s performance is inferior compared to more advanced alternatives like MTCNN, which offer better accuracy[16].

6.6 Instability in Frame-by-Frame Analysis

The model's tendency to analyse emotions on a per-frame basis can result in inconsistent and unstable results, making emotion tracking less reliable over time[17].

7. Expected outcome

In this research, we aim to evaluate the effectiveness of CNN-based models in predicting individuals' psychological states through facial feature analysis. By comparing the benefits and limitations of these models, we seek to clarify their capabilities, potential accuracy, and practical applications. Training CNNs with a diverse dataset of facial expressions, we intend to develop insights into the model's reliability and scope in interpreting a range of human emotions. This evaluation could lead to advancements in emotional intelligence technology, potentially enhancing fields such as mental health assessment, human-computer interaction, and automated emotional analysis. We anticipate providing a balanced view of the model's strengths in real-time emotion detection and its limitations, offering a foundation for future improvements and research directions in understanding psychological states through AI-driven facial analysis.

8. Conclusion

The CNN-based model for predicting human psychological states through facial features offers significant advancements in the field of emotion recognition. It demonstrates high accuracy and real-time capabilities, making it suitable for practical applications. However, challenges like dataset biases, computational demands, and limited contextual awareness remain. Addressing these limitations will be crucial for the future development and deployment of emotion recognition systems.

References

1. Ayush Bhalerao, Suraj Chamate, Shivprasad Bodke, Shantanu Desai, Somesh Kalaskar, "A Comprehensive Survey on Prediction of Human Psychological State by Analysing the Facial Features," IJRASET, vol. 12, no. 3, March 2024.
2. N. Zhou, R. Liang, W. Shi, "A Lightweight Convolutional Neural Network for Real Time Facial Expression Detection," IEEE Access, vol. 9, pp. 55735-5584, 2021, doi:10.1109/ACCESS.2020.3046715.
3. Heredia et al., "Adaptive Multimodal Emotion Detection Architecture for Social Robots," IEEE Access, vol. 10, pp. 20727-20744, 2022, doi:10.1109/ACCESS.2022.3149214.
4. M.-H. Hoang, S.-H. Kim, H.-J. Yang, G.-S. Lee, "Context-Aware Emotion Recognition Based on Visual Relationship Detection," IEEE Access, vol. 9, pp. 90465-90474, 2021, doi:10.1109/ACCESS.2021.3091169.
5. S. Dwijayanti, M. Iqbal, B. Y. Suprapto, "Real-Time Implementation of Face Recognition and Emotion Recognition in a Humanoid Robot Using a Convolutional Neural Network," IEEE Access, vol. 10, pp. 89876-89886, 2022, doi:10.1109/ACCESS.2022.3200762.
6. N. Samadiani, G. Huang, Y. Hu, X. Li, "Happy Emotion Recognition from Unconstrained Videos Using 3D Hybrid Deep Features," IEEE Access, vol. 9, pp. 35524-35538, 2021, doi:10.1109/ACCESS.2021.3061744.
7. H. Zhang, A. Jolfaei, M. Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing," IEEE Access, vol. 7, pp. 159081-159089, 2019, doi:10.1109/ACCESS.2019.2949741.

8. S. K. Jarraya, M. Masmoudi, M. Hammami, "Compound Emotion Recognition of Autistic Children During Meltdown Crisis Based on Deep Spatio-Temporal Analysis of Facial Geometric Features," IEEE Access, vol. 8, pp. 69311-69326, 2020, doi: 10.1109/ACCESS.2020.2986654.
9. S. Kumar, R. Rekha, "Improving Smart Home Safety with Face Recognition using Machine Learning," 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 478-482, doi:10.1109/ICISCoIS56541.2023.10100592.
10. T. Dar, A. Javed, S. Bourouis, H. S. Hussein, H. Alshazly, "Efficient-SwishNet Based System for Facial Emotion Recognition," IEEE Access, vol. 10, pp. 71311-71328, 2022, doi:10.1109/ACCESS.2022.3188730.
11. V. Savchenko, L. V. Savchenko, I. Makarov, "Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network," IEEE Transactions on Affective Computing, vol. 13, no. 4, pp. 2132-2143, 2022, doi:10.1109/TAFFC.2022.3188390.
12. S. K. Vishnumolakala, V. S. Vallamkonda, S. C. C., N. P. Subheesh, J. Ali, "In-class Student Emotion and Engagement Detection System (iSEEDS): An AI-based Approach for Responsive Teaching," 2023 IEEE Global Engineering Education Conference (EDUCON), Kuwait, 2023, doi:10.1109/EDUCON54358.2023.10125254.
13. S. Haddad, O. Daassi, S. Belghith, "Emotion Recognition from Audio-Visual Information Based on Convolutional Neural Network," 2023 International Conference on Control, Automation and Diagnosis (ICCAD), Rome, 2023, doi: 10.1109/ICCAD57653.2023.10152451.
14. A. Sharma, V. Bajaj, J. Arora, "Machine Learning Techniques for Real-Time Emotion Detection from Facial Expressions," 2023 IEEE Delhi Section Flagship Conference (DELCON), Rajpura, India, 2023, pp. 1-6, doi:10.1109/DELCON57910.2023.10127369.
15. Y. S. Deshmukh, N. S. Patankar, R. Chintamani, N. Shelke, "Analysis of Emotion Detection of Images Using Sentiment Analysis and Machine Learning Algorithm," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 1071-1076, doi: 10.1109/ICSSIT55814.
16. S. Pande, S. Shinde, "A Survey on: Emotion Recognition with Respect to Database and Various Recognition Techniques," Int. J. Computer. Appl., vol. 58, no. 3, pp. 9-12, Nov. 2012.
17. H. Xiao, W. Li, G. Zeng, Y. Wu, J. Xue, J. Zhang, C. Li, G. Guo, "On-Road Driver Emotion Recognition Using Facial Expression," Applied Sciences, Jan. 2022, doi: 10.3390/app12020807.
18. Cirneanu, D. Popescu, D. D. Iordache, "New Trends in Emotion Recognition Using Image Analysis by Neural Networks," Italian National Conference, Aug. 2023, doi: 10.3390/s23167092.
19. B. Zhang, D. Wei, Q. Zhang, W. Si, X. Li, Q. Zhu, "Classroom Monitoring System Based on Facial Expression Recognition," 2021, doi: 10.1109/DCABES52998.2021.00034.
20. Y. Liu, L. Wei, K. Liu, Y. Zhan, "Smile Upon the Face but Sadness in the Eyes: Emotion Recognition Based on Facial Expressions and Eye Behaviours," Nov. 2024, doi: 10.48550/arXiv.2411.05879.
21. T. Teixeira, E. Granger, A. L. Koerich, "Continuous Emotion Recognition with Spatiotemporal Convolutional Neural Networks," arXiv preprint arXiv:2011.09280, 2020.
22. D. Qu, S. Dhakal, D. Carrillo, "Facial Emotion Recognition Using CNN in PyTorch," arXiv preprint arXiv:2312.10818, 2023.
23. F. Ghaffar, "Facial Emotions Recognition Using Convolutional Neural Net," arXiv preprint arXiv:2001.01456, 2020.

24. S. D. Lalitha, K. K. Thyagharajan, "Micro-Facial Expression Recognition in Video Based on Optimal Convolutional Neural Network (MFEOCNN) Algorithm," arXiv preprint arXiv:2009.13792, 2020.
25. T.-D. Pham, M.-T. Duong, Q.-T. Ho, S. Lee, M.-C. Hong, "CNN-Based Facial Expression Recognition with Simultaneous Consideration of Inter-Class and Intra-Class Variations," Sensors, vol. 23, no. 24, p. 9658, 2023.
26. I. Ali, F. Ghaffar, "Robust CNN for Facial Emotion Recognition and Real-Time GUI," AIMS Electronics and Electrical Engineering, vol. 8, no. 2, pp. 227-246, 2024.
27. S. Sen, D. Sahana, K. S. Varsha, R. Priyanka, "A CNN-Based Approach for Facial Emotion Detection," ResearchGate, 2023.
28. D. Sahana, K. S. Varsha, S. Sen, R. Priyanka, "Enhancing Facial Emotion Recognition Using Image Processing Techniques and CNN," San Jose State University Scholar Works, 2023.
29. N. Gogolla, "Facial Expressions of Emotion States and Their Neuronal Correlates in Mice," Science, 2020.
30. X. Liu, "Recognition and Analysis of Human Facial Images Using Convolutional Neural Networks," Wikipedia, 2023.
31. S. Sun, C. Luo, J. Chen, "A Review of Natural Language Processing Techniques for Opinion Mining Systems," Information Fusion, 2017.
32. C.-H. Wu, W.-B. Liang, "Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels," IEEE Transactions on Affective Computing, 2011.
33. L.-P. Morency, R. Mihalcea, P. Doshi, "Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web," ACM, 2011.
34. S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, "Context-Dependent Sentiment Analysis in User-Generated Videos," Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 1, 2017.
35. F. Eyben, M. Wöllmer, B. Schuller, "Open EAR - Introducing the Munich Open-Source Emotion and Affect Recognition Toolkit," IEEE Conference Publication, 2009.
36. S. Poria, E. Cambria, A. Hussain, G.-B. Huang, "Towards an Intelligent Framework for Multimodal Affective Data Analysis," Neural Networks, 2015.
37. V. Perez Rosas, R. Mihalcea, L.-P. Morency, "Multimodal Sentiment Analysis of Spanish Online Videos," IEEE Intelligent Systems, 2013.
38. A. Yadollahi, A. G. Shahraki, O. R. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining," ACM Computing Surveys, 2017.
39. C.-H. Wu, W.-B. Liang, "Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels," IEEE Transactions on Affective Computing, 2011.
40. S. Sun, C. Luo, J. Chen, "A Review of Natural Language Processing Techniques for Opinion Mining Systems," Information Fusion, 2017.