

Facial Emotion Detection: Advancing Emotional and Social Skills in Autistic Children Using NAO Robot

RULA SHARQI, LOURDE HAJJAR
Engineering and Physical Sciences
Heriot-Watt University
Knowledge Park Dubai
UAE

Abstract: Autism Spectrum Disorder (ASD) presents significant challenges in emotional recognition and social interactions, making it difficult for affected individuals to identify and respond to emotions. This paper focuses on the role of robotics, specifically the NAO robot, in facilitating facial emotion detection to address these challenges. The NAO robot, integrated with advanced facial recognition software, is capable of detecting a wide range of human emotions and providing real-time feedback. This paper examines how robotics equipped with emotion recognition technology can be utilized to enhance the understanding of emotional signs. Through interactions with NAO, users can practice identifying emotions, thus improving emotional awareness and response. The study emphasizes the potential of humanoid robots in applying facial emotion detection technology for broader applications in therapeutic and educational settings. The findings suggest that integrating facial emotion detection into robotic systems, like NAO, offers significant promise in advancing emotional understanding and interaction.

Keywords: *Facial Emotion Detection, NAO Robot, Behavioral Robotics, Human-Robot Interaction.*

Received: March 23, 2025. Revised: July 2, 2025. Accepted: August 5, 2025. Published: October 16, 2025.

1. Introduction

Autism Spectrum Disorder (ASD) presents significant challenges for individuals, particularly in the areas of emotional and social development. Children with ASD often struggle to recognize, understand, and respond to emotions, both in themselves and in others. These challenges can lead to difficulties in forming relationships, understanding social cues, and engaging in meaningful communication. As a result, effective interventions that can help children with ASD develop emotional and social skills are crucial. One promising approach is the use of humanoid robots, such as the NAO robot, in therapeutic settings.

The NAO robot, a small, programmable humanoid, is equipped with advanced technologies such as facial emotion recognition software, allowing it to detect and respond to a wide range of human emotions. This technology enables the robot to provide real-time feedback to children, helping them practice identifying emotions and engage in social interactions in a controlled,

non-threatening environment. NAO's interactive capabilities allow it to simulate human-like responses, which can foster emotional understanding and social learning in children with ASD.

In recent years, there has been increasing interest in leveraging robots for educational and therapeutic purposes, particularly in supporting children with special needs. **The integration of emotion detection and response systems in robots** offers a novel and effective way to address emotional and social deficits in children with ASD. By engaging with NAO, children have the opportunity to practice identifying emotions, responding appropriately, and improving their emotional awareness in a dynamic, interactive setting. This paper explores the potential benefits of using the NAO robot for advancing the emotional and social skills of children with autism, and investigates the implications of this technology for improving therapeutic interventions in the field of behavioral robotics.

Through the implementation of real-time facial emotion detection, this study aims to highlight how robots like NAO can serve as valuable tools in

supporting children with autism in developing essential emotional and social competencies.

The findings from this research may offer new insights into the role of humanoid robots in therapeutic environments and their potential to revolutionize the way we approach ASD treatment.

2. Background and Related Work

Human Emotion and Classification

Human emotions are diverse, complex, and often exist on a spectrum, with expressions that may be transitional or compounded. Figure 1 illustrates the "Wheel of Emotions," which attempts to represent this complexity visually. These characteristics raise the question of how humans are inherently able to perceive and recognize emotions at such a fundamental level. In the context of Facial Emotion Recognition (FER), emotions are primarily conveyed through facial expressions. These expressions can be understood as patterns, formed by the coordination of specific facial features, which together create a cohesive representation of an emotional state.[2]

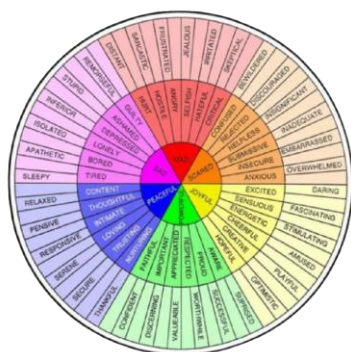


Figure 1: Wheel of Emotions [5]

Similarly, it is possible to replicate this ability to detect and combine various facial features in computers, enabling them to autonomously recognize the emotions expressed through a person's facial expression. By analyzing individual facial components, we can identify the underlying emotion, a methodology that can be adapted for use in AI systems for emotion detection, including in applications such as robotic interactions with children with autism. [3]

Deep Learning – Convolutional Neural Networks

Convolutional Neural Networks (CNNs) play a crucial role in the proposed face emotion detection system by providing an effective means for classifying facial expressions. CNNs are designed to automatically detect and learn hierarchical features from images, making them particularly well-suited for image recognition tasks such as emotion detection. In the context of this system, CNNs are employed to process facial images captured by the NAO robot's camera. The images are pre-processed (resized, normalized, and converted to grayscale) and then passed through a series of convolutional layers to extract relevant features. These layers utilize ReLU activation functions, batch normalization, and max-pooling for dimensionality reduction, followed by dropout layers to mitigate overfitting. The CNN model is trained using the FER2013 dataset, which consists of 48x48 grayscale facial images categorized into seven emotion classes. The model architecture, consisting of multiple convolutional blocks, culminates in a softmax output layer that classifies each image into one of the emotion categories. This CNN-based approach not only improves the model's ability to recognize emotions accurately but also demonstrates robustness and generalization capabilities, as evidenced by the test accuracy of 68.61% achieved on the unseen data. The use of data augmentation, regularization techniques, and optimization strategies such as the Adam optimizer further enhances the performance and stability of the model.

3. System Overview

This section will describe the design and implementation process of the face emotion detection system.

The proposed system is an integrated pipeline for real-time emotion recognition, leveraging the FER-2013 dataset for model training and the humanoid robot (NAO) for real-world interaction. The system combines camera-based input capture and deep learning-based emotion inference as shown in figure 2.

The flow of the system is as follows:

1. *Camera Input:* A camera mounted on the NAO robot captures facial images of a human subject.

2. *Preprocessing*: Captured images are resized, normalized, and converted to grayscale for efficient processing.
3. *Deep Learning Inference*: The pre-processed image is passed to a Convolutional Neural Network (CNN) for emotion recognition. The CNN maps the image to one of the seven emotional categories.

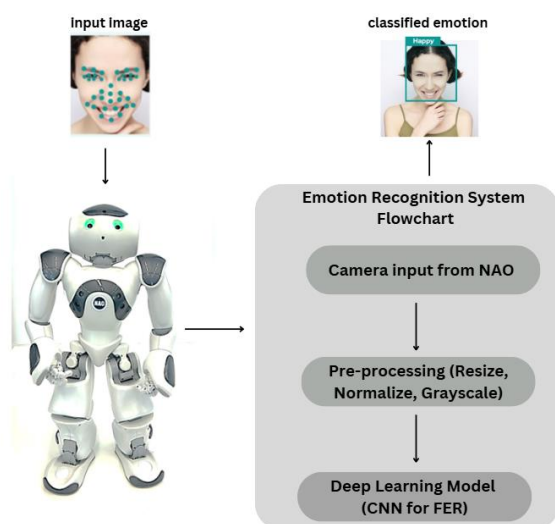


Figure 2: System Overview

4. Methodology

This work employs a convolutional neural network (CNN)-based approach to perform facial emotion recognition on the FER2013 dataset, following methodologies that have demonstrated state-of-the-art performance as discussed by Khairuddin and Chen [6]. The FER2013 dataset comprises 48×48 grayscale images categorized into seven emotional classes. Preprocessing steps included normalization to scale pixel values to the range $[0, 1]$ and reshaping the images into tensors compatible with the model. The dataset was split into training (80%), validation (10%), and testing (10%) subsets to facilitate training and evaluation. To enhance generalization and increase robustness, data augmentation techniques such as random rotations, zooming, horizontal flipping, and shifting were applied to the training dataset.

The proposed CNN architecture consists of three convolutional blocks, each comprising two convolutional layers with ReLU activation functions, batch normalization, and max-pooling

layers for dimensionality reduction. Dropout layers were incorporated after each block with rates of 0.25, 0.3, and 0.4, respectively, to reduce the risk of overfitting. At the final stage, the output layer utilizes a softmax activation function to classify images into the seven emotion categories.

The model was optimized using the Adam optimizer with a learning rate of 0.01, chosen for its adaptive learning rate capabilities and efficient convergence. Early stopping was used to terminate training if validation loss did not improve for 15 consecutive epochs, preventing overfitting. Additionally, the learning rate was reduced by a factor of 0.75 when validation performance plateaued for 5 epochs, enabling fine-tuning during later stages of training. Training metrics, including accuracy and loss for both training and validation sets, were logged to monitor performance and identify convergence trends. The final model weights were saved based on the best validation accuracy achieved during training.

5. Result

The CNN model for facial emotion recognition achieved a test accuracy of 68.61% on the FER2013 dataset. The training process demonstrated consistent improvement in both accuracy and loss, with the model effectively learning to classify emotions from the augmented dataset.

Validation accuracy and loss trends indicated stable generalization capabilities, as validation accuracy improved progressively while validation loss initially decreased before stabilizing, prompting the use of early stopping at the 65th out of 100 epochs to finalize the model weights.

The final test accuracy on unseen data underscores the model's robustness and its ability to generalize to new inputs. These results highlight the potential of the proposed architecture and training methodology for reliable emotion recognition tasks.

To verify the system, an interview was conducted with a psychiatrist based in Dubai, and it shows the following: The psychiatrist highlighted that integrating facial emotion detection technology, like NAO robots, can significantly improve emotional recognition and social skills in children

with autism. As shown in Fig. 3 below, the concept is that children with autism can use NAO to help translate and understand emotions by interacting with the robot, which provides a controlled, non-judgmental environment for practice. These children often struggle with interpreting facial expressions, and NAO's consistent feedback can enhance their ability to recognize and respond to emotions. The psychiatrist emphasized that such technology complements traditional therapies, especially when access to human therapists is limited. However, challenges such as the need for context recognition and individual emotional variability were noted. Collaboration between psychiatric professionals and technologists is crucial to ensure the system aligns with therapeutic goals and ethical standards.



Figure 3: NAO Robot System Overview for Emotion Detection

6. Discussion

The proposed CNN emotion recognition model achieved a test accuracy of 68.61%, demonstrating effective generalization to unseen data. Key findings include the significant role of data augmentation and regularization techniques, such as dropout and early stopping, in stabilizing learning and mitigating overfitting. The model effectively handled the challenges of the FER2013 dataset, including variability in facial expressions and

noise, though performance could be improved for real-time therapeutic applications. Integration with the NAO robot highlights the system's practical potential in ASD therapy, enabling real-time emotional feedback and fostering engaging interactions.

7. CONCLUSION

This research demonstrates the potential of using CNN-based emotion recognition systems integrated with humanoid robots like NAO to support therapeutic interventions for children with ASD. The proposed methodology achieved a robust test accuracy of 68.61% on the FER2013 dataset, showcasing the model's capability in emotion detection.

While the results are promising, they also highlight areas for further research. Future work could explore incorporating more diverse datasets, refining model architectures, and assessing the long-term impact of such systems on emotional and social learning in children with ASD. The insights gained from this research pave the way for innovative approaches to behavioral therapy, emphasizing the role of technology in improving quality of life for individuals with special needs.

References

- [1] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. on Neural Networks*, vol. 4, pp. 570-578, July 1993.
- [2] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124-129.
- [3] Mundt, M., Yekta, A., & Hossain, S. (2018). Emotional recognition in children with autism using facial expression and machine learning algorithms. *International Journal of Social Robotics*, 10(3), 355-368.
- [4] Plutchik, R. (1980). *Emotion: A Psychoevolutionary Synthesis*. Harper & Row.
- [5] Modeling Emotions in a Computational System - Emotional Modeling in the Independent Core Observer Model Cognitive Architecture - Scientific Figure on ResearchGate. Available

from:https://www.researchgate.net/figure/C-D-r-Gloria-Willcoxs-Feelings-Wheel-9_fig1_303445396

- [6] Yousif Khaireddin and Z. Chen, “Facial Emotion Recognition: State of the Art Performance on FER2013,” *ArXiv* (Cornell University), Jan. 2021, doi: <https://doi.org/10.48550/arxiv.2105.03588>.