

Facial Expression-Based Emotion Detection: A CNN-Based Approach for Real-Time Emotion Recognition

Miss. S N Salma¹, Dr. Manjunath N²

Student - Department of Computer Science & Engineering, R.L.Jalappa Institute of Technology,
Approved by AICTE, New Delhi Affiliated VTU, Belagavi & Accredited by NAAC - "A" grade)
Doddaballapur, Bangalore Rural District-561203

Associate Professor - Department of Computer Science & Engineering,

R L Jalappa Institute of Technology, Approved by AICTE, New Delhi Affiliated VTU, Belagavi &
Accredited by NAAC - "A" grade) Doddaballapur, Bangalore Rural District-561203

Abstract—Facial expression based emotion detection is a vital area of research in affective computing, enabling machines to interpret and respond to human emotional states. This study shows the development of the CNN based system for real-time emotion recognition for facial-expressions. This proposed systems records live video input through a webcam, preprocess facial images, and classifies expressions into eight emotional categories: Angry, disgust, fear, happy, sad, surprise, neutral and contempt. The model was trained on the FER-2013 dataset, comprising 35,887 grayscale facial images of 48×48 pixels, data augmentation techniques including rotation and brightness adjustment were applied to enhance generalization. The trained CNN model achieved a classification accuracy of approximately 66 percent on the test data. This system was integrated with OpenCV for original instance face detection and deployed using Tensor Flow and Keras. This research verifies CNN based DL models in accurately detecting facial expressions, making it useful in application like mental health monitoring, human computer interaction, virtual assistants, & user experience personalization. The result underscores the transformative role of AI in emotion recognition and its broader impact on creating emotionally aware systems.

Keywords – facial-expression recognition, emotion detection, CNN, DL, Real-Time Processing, FER-2013, OpenCV, Affective Computing, Human-Computer Interaction.

I. INTRODUCTION

Human interaction is built on communication, with emotions serving a vital component in how we connect one another for conveying feelings, intentions, and social cues. Facial expressions are a universal and compelling channel through which humans express emotions like happy, anger, sadness, surprise, and fear. However, the capability to accurately perceive and interpret the expressions are not always reliable,

especially in digital or automated environments where human emotional intelligence is absent. To fill the gap, real-time facial expression predict emotion detection utilizing AI and computer vision is emerging as a promising solution. By enabling machines to understand and respond to human emotions, the systems can enhance human-computer interaction across various domains, like healthcare, education, customer service, and mental state monitoring.

Improvement in ML and DL, particularly in Convolutional Neural Networks, have cleared the method for intelligent systems capable of recognizing facial-expression in real-time. These systems analyze facial feature & subtle muscle movements to detect and classify emotional states, making digital interactions more empathetic and responsive.

Emotion detection via facial-expressions recognition relies heavily on accurate analysis of the visual data. Traditional image processing methods are regularly struggle with challenges such as varying lighting conditions, head poses, and differences in facial structure across individuals. This is where the CNN prove invaluable. CNN represents specific DL models designed for image recognition and classification tasks, proving them to be highly effective in recognizing and distinguishing emotional expressions in real-time.

By adding powerful AI algorithms with high resolution video input, emotion detection systems can operate with high accuracy, enabling real time feedback and emotional awareness. These systems have the capability to modernise human computer interaction by introducing a layer of emotional intelligence, thus fostering more natural and effective communication between humans and machines.

The aim of research is to design, implement, & analyse a CNN-based system for facial expression-based emotion detection that operates in real-time using standard webcam input. This study provides a thorough evaluation of models performance using standard metrics and discusses the challenges and future directions in this

domain.

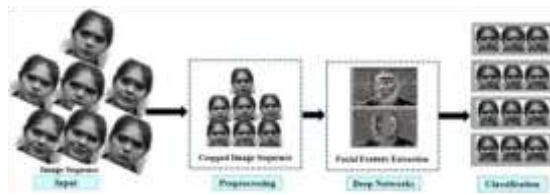


Fig. 1 Steps for facial expression detection

II. LITERATURE REVIEW

Facial expression-based emotion detection owns become a significant research domain in computer vision and emotional computing. This section reviews key studies, methodologies, and technological advancements in the field.

A. Emotion Recognition Approaches

Some techniques are analyzed for the emotion recognition, ranging from traditional ML to progressed deep learning architectures.

Handcrafted feature based approaches: Initial approaches involved extracting facial features using methods such as Local Binary Patterns, Histogram of Oriented Gradients, and Gabor filters, followed by classification using K-Nearest Neighbors (KNN). Although, these methods faced limitations in real time performance and generalization across diverse populations.

Deep Learning-Based Approaches: The evolution of DL, CNN have become frequently used for the feature extraction & classification in emotion detection tasks. Researchers have proposed CNN based models trained on datasets such as FER-2013, managing notable accuracy in detecting basic emotions. additional recent works are merged recurrent neural networks (RNNs) networks to capture temporal changes in expressions

B. Dataset Challenges and Solutions

A critical challenge in emotion recognition is the availability of diverse, annotated datasets with balanced emotion classes. Several publicly available datasets have been utilized in past research:

- *FER-2013:* A widely used dataset for emotion recognition, containing the 35,888 grayscale images with 7 emotion classes: angry, disgust, fear, happy, sad, surprise, & neutral. The dataset is divided for a training set with 28,709 example and a public test set alongside 3,589 examples.
- *AffectNet:* A large scale dataset containing approximately 0.4 million facial images manually labeled for the presence of eight facial expressions. It is the one of the largest emotion recognition databases, comprising over one million facial images

gathered from the internet.

To improve model stability and reduce overfitting, researchers have incorporated data augmentation techniques including horizontal flipping, brightness adjustments, and random cropping throughout training.

C. Real-Time Implementation and Applications

Real-time emotion recognition has been enabled through facial landmark detection and efficient frame processing. Frameworks such as OpenCV and dlib have been instrumental in capturing facial features from live webcam feeds. Studies demonstrated integration of trained CNN models with OpenCV for real time video capture and face detection, enabling on the fly emotion prediction, emotion detection system has been utilized in the fields like health monitoring, smart classrooms, customer feedback analysis, and interactive virtual assistants. Future improvements could involve expanding emotion categories.

III. METHODS AND MATERIALS

A. Dataset Description

This study utilizes the FER-2013 (Facial Expression Recognition 2013) dataset, namely a comprehensive collection held by facial images labeled with various emotions. The dataset contains 35,888 grayscale facial images, with a size of 48×48 pixels. The images were labeled with 7 emotion categories: angry, disgust, fear, happy, sad, and neutral. For this project, an additional class, Contempt, was incorporated to expand the emotion recognition capability to eight classes.

The datasets are divided into preparation and testing subsets to ensure model evaluation to prevent overfitting.

B. Data Preprocessing

Proper data preprocessing is basic for improving model performance and ensuring meaningful predictions. The following steps were undertaken:

Gray-scale conversion: The input image is converted to gray scale to reduce computational complexity and focus for facial structure rather than color variations.

Image resizing: All images are resized to 48×48 pixels to maintain consistency with the CNN model's expected input dimensions.

Normalization : pixel intensity value is scaled to the range[0, 1] to improve convergence during training and ensure consistent feature distribution.

Face Detection: in OpenCV Haar cascade classifier is employed to detect and crop facial regions from input frames, ensuring that only relevant facial features were analyzed.

Data Augmentation: To enhance model robustness and moderate overfitting, data augmentation techniques are applied, including rotation, flipping, brightness adjustment, and random cropping.

C. ML Models

In this, CNN was developed and implemented for facial expression recognition. CNN is specialized DL designed for image recognition and classification task, creating them highly effective in identifying and distinguishing emotional expressions.

The CNN architecture consists of the following layers:

- **Input Layer:** Accepts 48×48 pixel grayscale images.
- **Convolutional Layers:** Extract key facial features using learnable filters. Three convolutional layers were implemented along with multiple 32, 64, and 128 filters continuously, each with a 3×3 kernel size and ReLU activation function.
- **Pooling Layers:** Max-Pooling layers (2×2) implemented after each convolutional layer

for dimensionality reduction and to retain dominant features.

- **Flatten Layer:** Translates the subsequent 2D feature maps into a 1D feature vector.
- **Fully connected layers:** A dense layer with 128 units and ReLU activation, followed by a dropout-layer (0.5) to prevent overfitting.
- **Output Layer:** A dense layer with 8 units and Softmax activation for multi class emotion classification.

This is compiled using Adam optimizer, categorical cross-entropy-loss function, and accuracy as the evaluation metric.

D. Model Training and Evaluation

To ensure generalizability, these datasets were divided into training, validation, and testing sets. And it is trained for a max of 40 epochs with a size of 32. Data augmentations are implemented during training to enhance generalization.

Its performance is evaluated using the following metrics:

- **Accuracy:** Measures percentage of correct predictions.
- **Precision:** Evaluates how many predicted emotions were correctly classified.
- **Recall (Sensitivity):** measure the model identifies each specific emotion.
- **F1-Score:** harmonic mean precision and recall, ensuring a balanced evaluation.
- **Confusion Matrix:** Provides a breakdown of true hits, true negatives, false hits and false negatives for each emotion class.

E. Real-Time Deployment

This trained model is integrated into a real-time webcam-based emotion detection system. The system architecture follows a layered approach:

1. **Input (Webcam):** Detects the user's facial expression using a webcam.
2. **Preprocessing Layer:** Apply grey scale conversion, face detection using OpenCV haar cascade classifier and image resizing to 48×48 pixels.
CNN Model Layer: Processes the preprocessed facial image through the trained CNN model to predict one of the eight emotional states.
3. **Output (Monitor):** Displays the detected emotion on the screen with the prediction confidence.

OpenCV is used for real-time video capture and face detection, and the trained CNN model is deployed to predict emotions on-the-fly.

IV. RESULT AND ANALYSIS

I. Model Performance Evaluation

The working of CNN model is assessed relying on accuracy, precision, recall, F1-score along with confusion matrix analysis. It achieved the classification accuracy of approximately 66% on test dataset.

The confusion matrix analysis provided deeper insights into model performance. Misclassifications were observed primarily among emotions with subtle differences, such as Neutral versus Contempt, and Sad versus Fear. Common issues affecting accuracy included:

- Varying lighting conditions
- Differences in expression intensity across individuals
- Cultural and demographic variations in facial expressions



Fig.2 CK+dataset sequences

B. Analysis of Results

The results were analyzed to identify advantages and limitations of it. The CNN effectively captured spatial hierarchies and fine-grained visual features, making it suitable for detecting subtle changes in facial muscles that represent emotions. Although, the models performance was influenced by following factors:

Dataset Limitations: Meanwhile FER-2013 is a widely

used dataset, its grayscale nature and limited diversity in terms of the age, ethnicity, and real world conditions may impact generalization.

Class Imbalance: The dataset exhibit class imbalance, with some emotions having significantly more samples than others. This can lead to biased predictions favoring the majority classes.

Environmental Factors: Real-world deployment introduces challenges such as varying brightening conditions, head poses, and occlusions that were not fully represented in the train data.

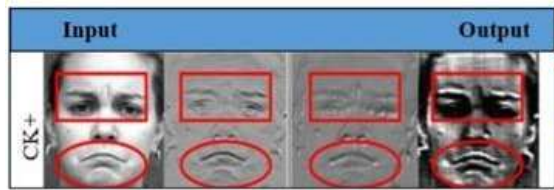


Fig. 3 Input and output sequence using DenseNet-201

C. Real-Time Performance

The integrated webcam-based system demonstrated real-time emotion recognition capability with acceptable latency on standard hardware. The system effectively detected faces and predicted emotions on-the-fly, making it suitable for practical applications.

D. Key Findings and Insights

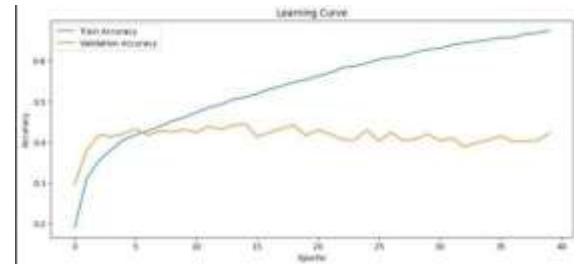
- CNN-based deep learning models were effective for facial expression recognition, achieving reasonable accuracy in classifying eight emotion categories.
- Data augmentation technique significantly improve model stability and reduce overfitting.
- Real-time deployment using OpenCV enables practical applications in human-computer interaction.
- Challenges remain in handling environmental variations and demographic diversity, highlighting the need for larger and more different datasets.
- The model achieved approximately 66 percent accuracy, demonstrating the potential for emotion aware technologies while also suggesting room for improvement through more advanced architectures and massive datasets.

IV. CONCLUSION

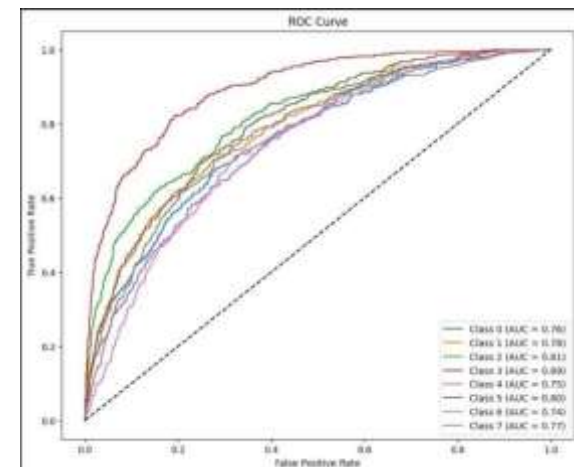
This specific study presented the design, implementation, and evaluation of CNN based on facial expression based emotion detection system. The intended system captures real time video input through

webcam, preprocesses facial images, and classifies expressions into 8 emotional categories:angry, disgust,fear,happy,surprise,neutral and contempt.

The model trained on FER-2013 dataset, comprising 35,887 grayscale facial images of 48×48 pixels, data augmentation technique including rotation, flipping, and brightness adjustment were used to enhance generalization. The trained CNN model achieved a classification accuracy of approximately 66% on test data.



The system was integrated with OpenCV for real-time face detection and deployed using TensorFlow and Keras. The result demonstrate the potential of CNN-based deep learning model in accurately detecting facial expressions, making it used in applications like mental health monitoring, human-computer interaction, virtual assistants, and user experience personalization.



The findings underscore the transformative role of AI in emotion recognition and its broader impact on creating emotionally aware systems. However, challenges remain in handling environmental variations, cultural differences, and demographic diversity, highlighting the need for larger and more diverse datasets.

V DISCUSSION AND FUTURE WORK

A. Discussion

The result for this study proves that CNN based DL

models are effective for facial expression recognition, achieving reasonable accuracy in classifying eight emotion categories. The systems real time performance demonstrates its potential for practical applications in human-computer interaction.

However, several limitations were identified:

Dataset Constraints: The FER-2013 dataset, while widely used, as limitation in terms of diversity and real world representation. The grayscale nature of images and the controlled conditions under collected may not fully capture the complexity of real-world facial expressions.

Class Imbalance: The dataset exhibit class imbalance, with some emotions having significantly more samples than others. This can lead to biased predictions favoring the majority classes.

Environmental Factors: Real-world deployment introduces challenges such as differential lighting conditions, head poses, and occlusions that were not fully represented in the training data.

Model Architecture: While the CNN architecture used in this research is effective, more advanced architectures like ResNet, EfficientNet, or hybrid CNN-RNN models could potentially achieve higher accuracy

B. Future Work

Although the current facial expression-based emotion detection system yields promising results, several enhancements and research directions can furthermore improve its performance, usability, and real-world applicability. Key areas for additional development include:

1. *Dataset Expansion and Diversity:* This model is currently trained on a limited dataset; incorporating major and more diverse datasets such as AffectNet, which contains approximately 0.4 million manually labeled images, would improve robustness and generalization. Including real-world images and video data from varied environments (e.g., natural lighting, occlusions, different head poses) would make the model more adaptive and resilient.
2. *Improving Model Accuracy with Advanced Architectures:* Exploring more advanced CNN architectures (e.g., ResNet, EfficientNet) to enhance feature mining and segregation performance. Investigating hybrid models that combine CNN with Transformers for improved temporal analysis in video-based emotion detection. Using attention mechanism to focus on key facial regions contributing to emotion recognition.
3. *Real -Time Performance Optimization:* Reducing model latency and computational complexity for smoother real-time execution on low-power devices. Applying techniques such as model pruning, quantization, and edge computing to enable deployment on mobile and embedded

systems.

4. *Multimodal Emotion Recognition:* Combining facial expressions with other modalities such as voice tone, body language, and physiological signals to more holistic emotion detection system.
5. *User - Centric Application Development:* Developing real-time applications such as emotion-aware chatbots, virtual therapists, and smart classroom tools to personalize user interactions.
6. *Ethical, Social, and Inclusive Deployment:* Ensuring fairness and mitigating bias in emotion recognition across diverse user populations. Collaborating with psychologists and human behavior researchers to validate the model's effectiveness and interpretability.
7. *Continuous Learning and Adaptability:* Implementing online learning techniques to allow the model to continuously adapt to new data and user-specific emotional patterns.

V. ACKNOWLEDGMENT

The author expresses sincere gratitude to the academic mentors and The department of computer science and engineering at R.L. Jalappa Institute of Technology for their invaluable guidance and support. The open-source communities and the creators of the Fake Job Postings dataset are also acknowledged for providing the essential resources that made this research possible

VI. REFERENCES

1. Krishna Kant, Dipti B. Shah Lightweight Attention Based Deep CNN Framework for Human Facial Emotion Detection from Video Sequences, Received: 23 July 2024 / Accepted: 14 November 2024 © The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2024.
2. S. Woo, J. Park, J. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), Jan. 2018, pp. 3–19.
3. D. Yang, Z. Chen, Y. Wang, S. Wang, M. Li, S. Liu, X. Zhao, S. Huang, Z. Dong, P. Zhai, and L. Zhang, "Context de-confounded emotion recognition," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 19005–19015.
4. Z. Yang and S. S. Narayanan, "Modeling dynamics of expressive body gesture in dyadic interactions,"

- IEEE Trans Affect. Comput., vol.8, no.3, pp. 369–381, Jul. 2017.
5. P. Barros, D. Jirak, C. Weber, and S. Wermter, “Multimodal emotional state recognition using sequence-dependent deep hierarchical features,” *Neural Netw.*, vol.72, pp. 140–151, Dec. 2015..
 6. L. F. Barrett, B. Mesquita, and M. Gendron, “Context in emotion perception,” *Curr. Dir. Psychol. Sci.*, vol. 20, no.5, pp. 286–290, Oct. 2011.
 7. S. Zhang, Y. Pan, and J. Z. Wang, “Learning emotion representations from verbal and nonverbal communication,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18993–19004.
 8. F. Chollet, “Xception: DL along with depthwise separable convolutions,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.
 9. S. Woo, J. Park, J. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 3–19.
 10. K. Xu et al., “Show, attend and tell: Neural image caption generation with visual attention,” in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.
 11. W. Li, X. Dong, and Y. Wang, “human emotion recognition with relational region-level analysis,” *IEEE Trans. Affect. Comput.*, vol.14, no.1, pp. 650–663, Jan. 2023.
 12. R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, “Context based emotion-recognition using EMOTIC dataset,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2755–2766, Nov. 2020.
 13. D. Ghimire and J. Lee, “Geometric feature-based facial-expression recognition within image sequences using multi-class AdaBoost and support vector machines,” *Sensors*, vol. 13, no.6, pp. 7714–7734, Jun. 2013.
 14. B. Fasel and J. Luetten, “Automatic facial expression analysis: A survey,” *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, Jan. 2003.
 15. S.N.M. Sayed Ismail, N.A. Ab. Aziz, S.Z. Ibrahim, and M.S. Mohamad, “A systematic review of emotion recognition using cardio-based signals,” *ICT Exp.*, vol.10, no. 1, pp.156–183, Feb. 2024..