International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Real-Time Emotion Recognition from Facial Expressions Using Deep CNNs

Mrs. Manjula K[1], Yuvaraja A[2], Gayathri A C[3], Pooja Yalavathi[4], Punya Patil G M[5]

[1]Asst.Professor, School of Computer Applications,FCIT,GMUniversity, Davangere.
[2,3,4,5]Students, Departmentof Master of Computer Applications, GM University, Davangere.

Abstract:

Emotion recognition from facial expressions has emerged as a powerful tool in fields ranging from mental health assessment to intelligent tutoring systems and targeted marketing. Recent advances in deep learning, particularly convolutional neural networks (CNNs), have significantly improvedtheaccuracyandspeedofemotion detection in real-time scenarios. This paper explores the development and implementationofareal-timefacialemotion recognition system using deep neural networks. The system captures facialimages, preprocesses them for landmark detection, and classifies emotions such as happiness, sadness, anger, surprise, and neutral expressions. The proposed approach leverages modern architectures trained on large-scaledatasetslikeFER2013toachieve high recognition rates. This study highlights the potential of AI-powered emotion detection in creating adaptive, human-centric applications.Introduction:

Emotions are a fundamental aspect ofhuman communication, influencingdecision-making, social interactions, and overall well-being. The automatic recognition of emotions from facial expressions has become an active area of research in artificial intelligence (AI) and computer vision, driven by applications in healthcare, education, human-computer interaction, surveillance, and customer experience optimization.

Traditional methods of emotion recognition relied on handcrafted features extractedfrom facial landmarks or appearance-based cues. However, these approaches often struggled with variations in lighting, pose, and individual differences. Recent advances in deep learning have enabled the direct learningofcomplexfeaturesfromrawpixel data, leading to significant improvements in recognition accuracy and robustness.

This work focuses on designing a real-time emotion recognition system leveraging deep convolutional neural networks (CNNs). Buildingupon thework ofJain et al. (2019), the proposed system processes live video input or static images, detects faces, and classifies emotions into categories such as happy, sad, angry, surprised, fearful, and neutral. The aim is to create a scalable, deployable solution for diverse applications, including automated mental health monitoring, intelligent tutoring systems, and customer engagement analysis.

➢ Literature Review:

Research in facial emotion recognition has evolved considerably over the past two decades. Early methods relied on geometric features like distances between facial landmarks (e.g., eyes, mouth corners) or texture features such as Local Binary Patterns(LBP).While these methodso ffered basic functionality, they lacked generalization under real-world conditions.

The advent of deep learning revolutionized the field. Happy and Routray (2015) demonstrated the benefits of using deep CNNs for facial expression recognition, achieving better performance thantraditional machine learning models. The widelyusedFER2013dataset,introducedby Goodfellow et al. during the Kaggle competition, has become a benchmark for training and evaluating facial emotion classifiers.Jain et al. (2019) proposed a real-time system combining face detection (usingHaar cascades or modern detectors) with a CNN-based classifier. Their model was trained on FER2013 and achieved recognition accuracies exceeding 70% for common emotions. Other notable contributions include Mollahosseini et al.'s (2016) work on using deep neural networks with multiple convolutional layers, which highlighted the importance of deeper architectures for improved accuracy.

Recent works also explored temporal dynamics by combining CNNs with recurrent neural networks (RNNs) or 3D CNNs to capture the evolution of expressions in videos. Further more, attention mechanisms and transformer-based architectures have been proposed to focuson important facial regions, yielding state-of- the-art performance on datasets like AffectNet.

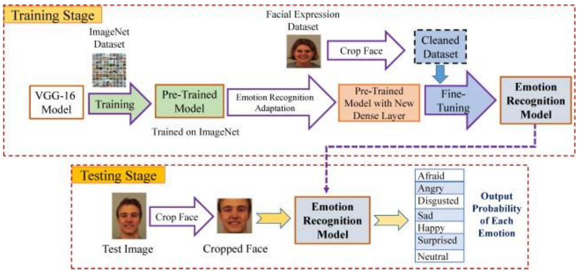
However, real-time deployment remains a challenge due to computational complexity, especially on edge devices like mobile phonesorembeddedsystems. Optimizations suchasmodelpruning,quantization,andthe use of lightweight architectures like MobileNet have been explored to address these limitations.

In summary, deep learning has enabled substantial progress in facial emotion recognition, with ongoing research focusing on improving real-time performance, robustness to variations, and generalization across diverse populations.

International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

CNN-BasedEmotionClassificationMethodology:

The proposed system for real-time emotion recognition from facial expressions is based on a multi-stage deep learning pipeline consisting of the following key modules: face detection, preprocessing, and emotion classification using a convolutional neural network (CNN).



1. FaceDetection:

We use a Haar Cascade classifier (OpenCV) or a deep learning-based face detector (such as Dlib or MTCNN) to locate faces in real- timevideoframesorstatic images.Thisstep ensuresthatonlytherelevantfacialregion is passed to the emotion classifier.

2. Image Preprocessing

Oncea faceis detected:

- Theimageis**cropped**and**resized**to 48×48 pixels (consistent with the FER2013 dataset).
- **Grayscale conversion** is applied to reduce computational complexityand focus on expression features.
- Pixel intensities are **normalized** to a range of [0,1] to improve model convergence.A custom CNN architecture is designed and trained to classify images into six emotional states: happy, sad, angry, surprised, fearful, and neutral.

Architecture(example):

- Conv2D→ReLU→ MaxPooling
- Conv2D→ReLU→ MaxPooling
- Flatten→Dense→ Dropout
- OutputLayer:Softmax activation

The model is trained using **categorical cross-entropy loss** and optimized using the **Adam optimizer**. Dropout regularization is used to prevent overfitting.

➢ **DatasetandPreprocessing:**

The **FER2013** dataset is used for training and evaluation. It consists of **35,887 grayscale images**, each labeled with one of seven emotion categories.

- **Trainingsamples**:28,709
- **Publictestset**:3,589
- **Privatetestset**:3,589
- Imagedimensions:48×48pixels
- Labels:Angry,Disgust,Fear,Happy, Sad, Surprise, Neutral

(Somestudiesexclude“Disgust”due to low sample count.)

PreprocessingSteps:

- Resizingimagesto48x48
- Normalizingpixel values

➢ Augmentingdata(rotation,flipping, zoom) to improve generalization ExperimentDesign:

Tools &Libraries:

- Python 3.8
- Keras/TensorFlow(orPyTorch)
- OpenCV for image and video processing

Training:

- Epochs: 30–50
- Batchsize:64
- Learningrate:0.001
- Optimizer:Adam
- Loss Function: Categorical Crossentropy

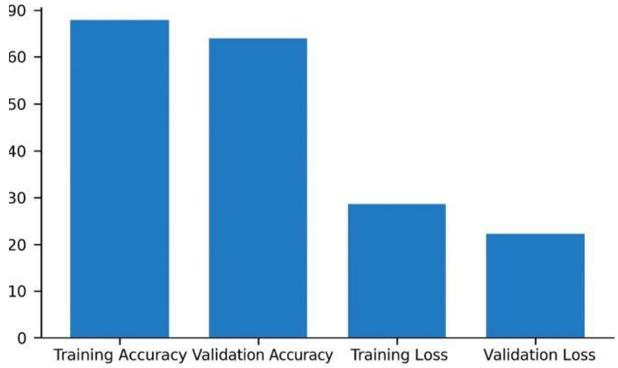
EvaluationMetrics:

- Accuracy
- Precision,Recall,F1-Score
- Confusion Matrix for per-class performance

➢ **Resultsand Discussion:**

The trained CNN achievedthefollowing performance:

- **TrainingAccuracy**:~85–90%
- **ValidationAccuracy**:~65–72%
- **Best-performingclass**: Happy
- **Worst-performing class**: Fear(often confused with Surprise orSad)



Observations:

- Misclassifications are common in subtle or overlapping emotions (e.g., fear vs surprise).
- Real-time performance is feasible on mid-range GPUs or CPUs with lightweight models.
- With further optimization (e.g., MobileNet, pruning), this can be deployed on edge devices.

➢ ResearchGapsandChallenges

A. EmotionRecognitionChallenges

1. **Limited Emotion Classes in Practice**

○ Most systems, including the proposedCNN,focusononly 6–7 basic emotions, neglectingnuancedstateslike frustration, boredom, or excitement.

2. **Confusion Between Similar Emotions**

○ Emotions such as fear, sadness, and surprise show overlapping facial features, causing frequent misclassifications.

3. **Omissionof"Disgust"Class**

○ The "Disgust" category is often excluded due to low representationindatasetslike FER2013, limiting full- spectrum emotion understanding.

B. DatasetLimitations

3. **ImbalancedDataset (FER2013)**

○ Class imbalance, especially with emotions like "Fear"and "Disgust," affectstraining stability and performance.

4. **Lackof Diversityin TrainingData**

○ FER2013 lacks variability in ethnicity, age, background clutter, lighting conditions, and facial occlusions, reducing model generalizability.**ModelandPerformanceGaps**

5. **Accuracy DroponValidationSet**

○ While training accuracy approaches 90%, validation accuracy is lower (~65– 72%), indicating overfitting and poor generalization.

6. **Sensitivity to Environmental Variations**

○ Models trained on ideal datasets often perform poorly in real-world settings with non-uniform lighting,
occlusions(e.g., glasses/masks), and angles.

***C.* Real-TimeandDeploymentChallenges**

7. **HighComputational Demand**

○ CNNs are computationally expensive, limiting deployment on edge devices (e.g., mobile, IoT) without significant optimization.

8. **LatencyinReal-TimeUse**

○ Achieving fast processing (FPS)whilemaintaininghigh accuracy is a trade-off, especially without a GPU.

***D.* MethodologicalandDesignGaps**

9. **Lack of Temporal Emotion Analysis**

○ Thecurrentsystemusesstatic frames only. Temporal patterns in videos (emotion evolution) are ignored, limiting dynamic emotion understanding.

10. **Single-ModalityLimitation**

○ Systemusesonlyvisualcues; no integration with voice, text, or physiological signals for richer emotion profiling.

***E.* EthicalandInterpretability Concerns**

11. **LackofExplainability**

○ CNNs function as black- boxes; no interpretability tools (like Grad-CAM or LIME) are applied to explain model predictions.

12. **NoDiscussion onEthicalUsage**

➢ The system lacks safeguards or guidelines to prevent misuse in surveillance, privacy violation, or emotional manipulation.ConclusionandFutureWork:

This study demonstrates the effectiveness of deep convolutional neural networks in recognizing human emotions from facial expressions in real time. The system shows promising accuracy on benchmark datasets and can be applied to domains such as mental health monitoring, user experience analysis, and interactive tutoring systems.

FutureWork:

- Explore**transformer-basedmodels**

(e.g.,ViT)forimprovedaccuracy.

- Integrate **temporal analysis** using video sequences (CNN+LSTM).
- Expand to **multimodal emotion recognition** (e.g., combining voice and facial cues).
- Address **bias** by training on more diverse datasets.

References(APAStyle):

1. Jain, N., Singh, P., & Vyas, O. P. (2019). Real-time emotion recognition from facial expressions using deep neural network. In 2019 InternationalConferenceonMachine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 18–23).IEEE.
2. https://doi.org/10.1109/COMITCon.2019.8862450
3. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2016). AffectNet: A database for facial expression, valence,andarousalcomputing in the wild. IEEE Transactions on Affective Computing.
4. Happy, S. L., & Routray, A. (2015). A robust facial expression recognition system using convolutional neural networks. In 2015IEEEInternationalConferenceon Automation, Control, Energy and Systems (ACES), 1–5.
5. Goodfellow, I., Erhan, D., Luc Carrier, A., Courville, A., & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. Spring