e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:06/June-2025 Impact Factor- 8.187 www.irjmets.com

MOOD DETECTION USING FACIAL EXPRESSION AND DEEP LEARNING

Miss. Manisha Raghoba Agarwadekar*1

*1Student, Department Of Information Technology, Sindhudurg Sub Campus, University Of Mumbai, Sawantwadi, Maharashtra, India.

DOI: https://www.doi.org/10.56726/IRJMETS80333

ABSTRACT

This research project, titled "Mood Detection Using Facial Expression and Deep Learning", explores the design and implementation of an intelligent system capable of recognizing persons through facial expressions in real time. Emotional analysis plays a major role in enhancing human-computer interaction, especially in fields such as mental health, education, and customer service. The proposed system employs a Convolutional Neural Network (CNN) trained on a labelled dataset of grayscale facial images to classify emotions into five categories: Happy, Sad, Angry, Surprise, and Neutral. The model utilizes methods like batch normalization, dropout regularization, and data augmentation to enhance performance and minimize overfitting. A well-organized training pipeline and early stopping strategy help in building a reliable model with a final test accuracy of approximately 75%. The application is developed using Streamlit, offering users two input options—capturing an image via webcam or uploading a file. It includes real-time face detection, emotion prediction, and a text-to-speech feedback mechanism for better accessibility, especially for users with visual impairments. Additionally, detected results, along with user details (name, age, gender), are securely stored in a MySQL database for record-keeping and further analysis. This system provides an intuitive, cost-effective, and user-friendly solution for mood detection using deep learning techniques.

Keywords: Mood Detection, CNN, Deep Learning, Streamlit, Pyttsx3, Opencv, FER 2013, Real-Time Mood Detection.

I. INTRODUCTION

In today's era of artificial intelligence, enabling machines to understand and respond to human emotions has become a key research focus. Facial expressions are among the most natural and universal indicators of human mood, playing a vital role in communication and decision-making. This research project aims to develop a deep learning-based mood detection system that identifies emotions such as Happy, Sad, Angry, Surprise, and Neutral from facial images. Using Convolutional Neural Networks (CNNs), the system is trained on grayscale facial images to accurately classify emotions. The model is supported by preprocessing steps like grayscale conversion, normalization, and data augmentation to enhance accuracy. A user-friendly web interface, built with Streamlit, allows users to upload or capture images in real time. The system also includes voice-based feedback for accessibility and stores results in a MySQL database for future analysis. This project contributes to the growing field of affective computing by building intelligent systems that recognize and respond to human emotions. It has wide applicability in healthcare, education, customer service, accessibility support, and workplace well-being.

II. METHODOLOGY

Overview of FER

Facial Emotion Recognition (FER) is a subfield of Instinctive computing that focuses on detecting human emotions by facial expressions. Paul Ekman detected six basic emotions—happiness, sadness, anger, surprise, and neutral—that are universally noticeable. Recognizing these emotions automatically using computer vision and artificial intelligence has numerous applications in health monitoring, education, customer service, and security. FER systems Analyze facial movements and patterns to infer emotions. They leverage psychological theories of emotion expression and perception, translating these visual cues into computational models. This theoretical foundation verifies the reliability of facial expressions as indicators of emotional states.

Overview of CNN

Convolutional Neural Networks (CNNs) are a specialized part of neural networks, most effective for analyzing visual data. They are composed of some layers.

Convolution Layers: These apply filters (kernels) to input images to extract features like edges or corners.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:06/June-2025 Impact Factor- 8.187 www.irjmets.com

Pooling Layers: These decrease the spatial size of the feature maps to make the computation capable and to reduce overfitting.

Fully Connected Layers: These merge all features and produce the final output, like a classification of the emotion. Activation Functions: ReLU (Rectified Linear Unit) is commonly used to introduce non-linearity into the model.

Flow of the Mood Detection System

The methodology followed a structured approach starting from dataset selection and preprocessing to model training and deployment. The FER-2013 dataset was selected due to its large, labelled collection of 48x48 grayscale facial images captured in real-world scenarios. This dataset provides a diverse range of expressions, making it suitable for training robust models. Only five emotion classes—Happy, Sad, Angry, Surprise, and Neutral—were considered for simplicity and accuracy. Preprocessing steps included converting all images to normalized pixel values between 0 and 1, resizing to fit the input layer dimensions, and encoding labels into numerical form. These actions guarantee uniformity and the best performance of the model. The application was developed in Python, using key libraries such as TensorFlow for model building, OpenCV for face detection, and Streamlit for GUI creation. Extra tools such as pyttsx3 and MySQL were utilized for providing text-to-speech feedback and for data storage, respectively.

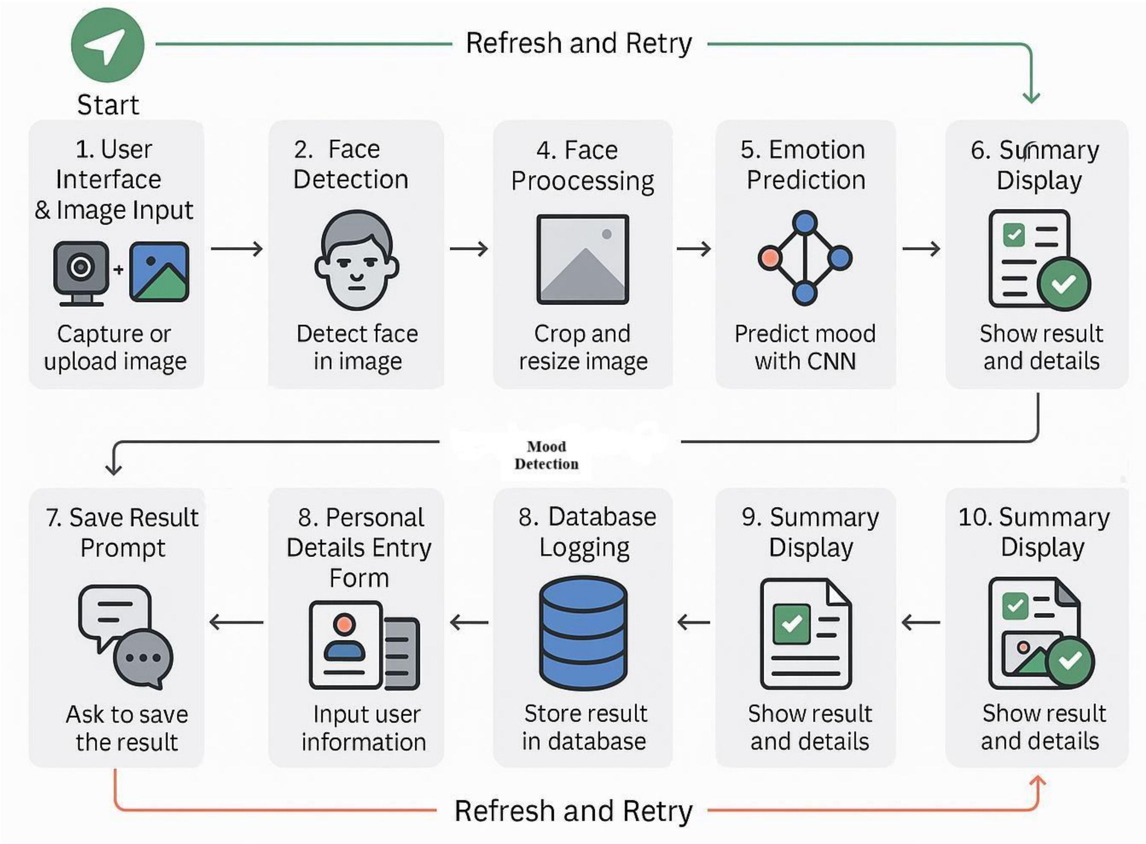

Figure 1: Mood detection application flow Chart

III. MODELING AND ANALYSIS

The CNN architecture consisted of an input layer for 48x48 grayscale images, followed by multiple convolutional layers that extracted visual features such as eyes, mouth, and eyebrows. Pooling layers reduced spatial dimensions while preserving key features. Dropout layers prevented overfitting, and batch normalization accelerated training by stabilizing activations. Finally, dense layers processed the extracted features to produce a probability distribution over emotion classes via SoftMax activation. Training was carried out using the Adam optimizer, known for its adaptive learning rate capabilities, along with the categorical cross-entropy loss function suited for multi-class classification tasks. The model was assessed using metrics including accuracy and precision. Techniques like early stopping and model checkpoints were employed to monitor validation performance and avoid overfitting. Overall, the model achieved a validation accuracy within the range of 70%–

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:06/June-2025 Impact Factor- 8.187 www.irjmets.com

80%, demonstrating effective emotion classification under standard conditions.

Final Model

Custom CNN with Preprocessing (Final Model)

From all the testing and some other testing, I just got 70% accuracy, but I want to increase my accuracy. So, for this time, I enhanced some operations of my basic custom CNN. This model uses a custom-built CNN architecture with grayscale input pictures resized to 48×48 pixels and includes layers like Conv2D, Batch Normalization, MaxPooling2D, and Dropout. It employs the FER-2013 dataset and adds manual preprocessing using Haar cascades to arrange faces before training. The model benefits from image augmentation (rotation, zoom, flip, shifts), and is trained for 70 epochs with callbacks such as Early Stopping and Reduce LR On Plateau. Finally, after that model training, I got 75% accuracy. I used this model developed from this model for my further execution in that research project work.

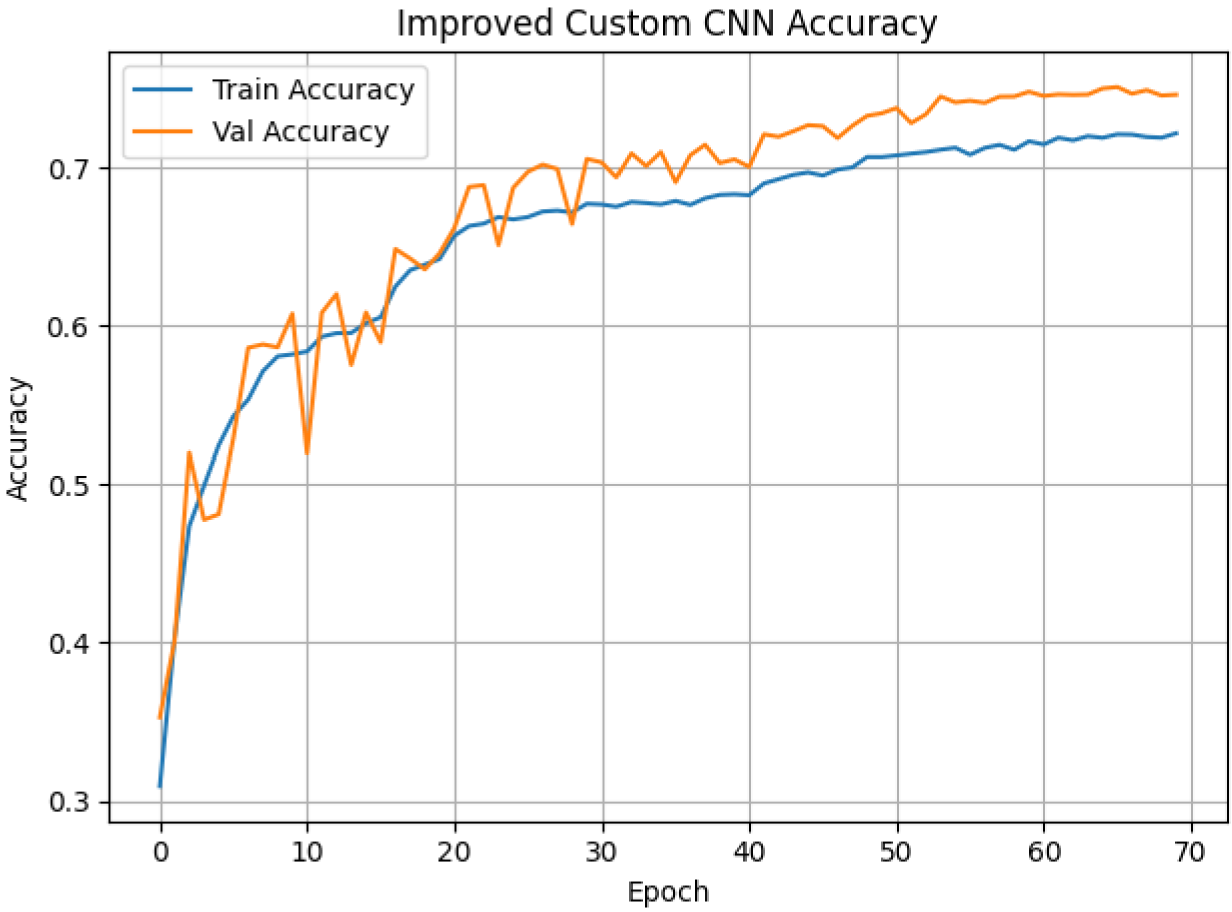

Figure 4.6 Accuracy of My Final Model

IV. RESULT AND DISCUSSION

The mood detection system achieved reliable real-time emotion recognition using static facial images. When tested under well-lit environments with frontal face positioning, the CNN consistently identified emotions with high confidence. The integration of text-to-speech provided immediate voice feedback, improving accessibility, especially for visually impaired users. The interface built with Streamlit offered smooth interactions, allowing users to capture or upload an image, receive predictions, and optionally store results in a MySQL database. The recorded information contained the individual's name, age, gender, emotional state, confidence level, and timestamp, which can be valuable for future analysis. The application performed efficiently on systems without a GPU, proving its suitability for educational institutions and small-scale healthcare environments. However, certain limitations were observed. The accuracy of emotion prediction diminished in dimly lit or chaotic backgrounds. Additionally, nuanced expressions occasionally led to erroneous classifications. These challenges suggest the need for advanced data augmentation, illumination normalization, and inclusion of temporal data from video inputs for future improvements.

Comparative Analysis

Table: Comparison between the existing model and my model

Model	Dataset	Technique	Real-time	Audio Feedback	Complexity	Accuracy
My model	FER-2013 (5 images)	CNN	Yes	Yes	Low	75%
Emotion Net	Affect Net	VGG16 + Transfer Learning	No	No	High	85%
Deep Mood	JAFFE	ResNet + Data Aug.	Partial	No	Medium	72%
Mobile FER	Custom	MobileNet	Yes	No	Medium	67%
Real- Time FER (2022)	CK+	Lightweight CNN	No	No	High	72%

V. CONCLUSION

This research project successfully demonstrated the feasibility of designing a real-time facial emotion recognition mechanism using Convolutional Neural Networks (CNNs) and open-source implementations like Python, TensorFlow/Keras, OpenCV, and Streamlit. Despite the limitations of a small dataset and simple hardware, the system achieved approximately 65 75% validation accuracy and performed consistently across five core emotion classes: Happy, Sad, Angry, Surprise, and Neutral. A key strength of the system lies in its inclusive design, particularly the integration of a text-to-speech feedback module that enhances Reachability for visually impaired users. Additionally, the support for both webcam input and image upload offers practical flexibility in different real-world settings. The Flexible architecture, lightweight model, and user-friendly interface showcase how machine learning can be applied successfully even under resource constraints. This research not only meets its core objectives but also focuses on the real-world potential of emotion-aware applications in fields such as mental health monitoring, educational tools, customer Engagement systems, and assistive technologies.

VI. REFERENCE

[1] GeeksforGeeks. (2023). Emotion Detection using Convolutional Neural Networks (CNNs). https://www.geeksforgeeks.org/deep-learning/emotion-detection-using-convolutionalneural-networks- cnns/ 2.

[2] Sambare, M. (2022). FER-2013: Facial Expression Recognition Dataset. Kaggle. https://www.kaggle.com/datasets/msambare/fer2013

[3] Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order. Pattern Recognition, 61, 610–628. https://www.sciencedirect.com/science/article/pii/S1746809416301010

[4] Mollahosseini, A., Chan, D., & Mahoor, M. H. (2016). Going deeper into facial expression recognition using deep neural networks. IEEE Winter Conference on Applications of Computer Vision (WACV). https://ieeexplore.ieee.org/document/7358934

[5] Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). Real-time Convolutional Neural Networks for Emotion and Gender Classification. arXiv preprint.https://arxiv.org/abs/1704.04861

[6] Levi, G., & Hassner, T. (2015). Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. Proceedings of the ACM ICMI. https://ieeexplore.ieee.org/document/7553523

[7] GeeksforGeeks. (2021). Introduction to Convolutional Neural Network (CNN). https://www.geeksforgeeks.org/machine-learning/introduction-convolution-neural network/

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:06/June-2025 Impact Factor- 8.187 www.irjmets.com

[8] IJNRD. (2023). Real-time Emotion Detection using CNN and Streamlit. International Journal of Novel Research and Development. https://www.ijnrd.org/papers/IJNRD2305049.pdf

[9] GeeksforGeeks. (2021). Convert Text to Speech in Python. https://www.geeksforgeeks.org/python/convert-text-speech-python/

[10] Viso.ai. (2022). Understanding ResNet – Residual Neural Network. https://viso.ai/deep-learning/resnet-residual-neural-network/

[11] ScienceDirect. (2022). MobileNetV2: Efficient Deep Learning Model. https://www.sciencedirect.com/topics/computer-science/mobilenetv2

[12] GeeksforGeeks. (2022). A Beginner's Guide to Streamlit. https://www.geeksforgeeks.org/a-beginners-guide-to-streamlit/

[13] Streamlit Documentation. (2023). Connect Streamlit with MySQL. https://docs.streamlit.io/develop/tutorials/databases/mysql