Smart Mood Detection: Real-Time Emotion Analysis for Human-Computer Interaction Using Deep Learning

Chandrappa S
Dept. of Computer Science and Engineering
JAIN (Deemed to-be University)
Bengaluru, India
chandrucs21@gmail.com
https://orcid.org/0000-0001-6508-5103

Tapabrata Banerjee
Dept. of Computer Science and Engineering(AI and DE)
JAIN (Deemed to-be University)
Bengaluru, India

Sohan Saha Mohd Zaid
Dept. of Computer Science and Engineering (AI and DE)
JAIN (Deemed to-be University)
Bengaluru, India

***Abstract*—A convolutional neural network (CNN)-based face emotion recognition system that analyzes human emotions in real-time, including happy, sorrow, rage, and surprise, is described. Existing approaches often struggle with accurately recognizing subtle or complex emotions due to limited feature extraction capabilities and the need for large, diverse training datasets. Additionally, they face challenges in adapting to real-world conditions, such as variations in lighting, angles, and occlusions. In order to improve the model's generalizability, the proposed research used a preprocessing pipeline that included gray scale conversion, scaling, normalization, and augmentation techniques including rotation and flipping on a labeled and balanced dataset of facial photos. Dropout and regularization techniques were used to reduce over fitting, and the CNN architecture was meticulously adjusted to maximize speed. High accuracy and resilience across emotion categories are demonstrated by preliminary results, indicating the model's potential for real-world uses in security surveillance, healthcare, and human-computer interaction.**

Keywords—Object Detection, Mood Detection, Point Cloud Library, Computer Vision, Facial Expression Analysis

I. INTRODUCTION

Facial emotion recognition, or FER for short, refers to the process of attributing one or more of the universal emotions including happy, sorrow, anger, fear, surprise, and neutral to a face as its emotions and classifying them into specific categories. In the earlier days of recognizing various emotions, there used to be a human interpreter but that has all changed with the latest innovations in deep learning and machine learning where this entire process is now fully automated with much greater precision. With advancements in human-computer interaction continuing their development, FER systems breathe new life into intuitive computer interfaces by accurately analysing, interpreting and predicting human emotions with minimal interaction.

This work aims to develop a robust and accurate face emotion recognition model in Python using the principles of machine learning and deep learning frameworks. In order to enhance the understanding of facial emotions, we introduced a balanced dataset comprising diverse tagged emotions to provide the model with numerous facial expressions. The methodology included several methodological steps such as minimization of dataset pre-processing, feature extraction of relevant facial signals and the deployment of a neural network capable of detecting and clustering stimulus features into structural emotional categories.

II. LITERATURE REVIEW

Mood detection is an emerging area in the field of affective computing, applied to the assessment of mental health status along with customer sentiments. In the early stages, these studies relied on text and facial cues and were based on heuristic methods for developing a more structured framework. The introduction of deep learning approaches, especially Convolutional Neural Networks, greatly improved the accuracy and scale of mood detection systems, particularly image-based assessments.

Anand M and Babu S [1], developed Facial Expression Recognition (FER) by employing techniques such as CNNs and DBNs, but they tend to encounter overfitting, high costs, and real-world variability. Recent techniques such as DBN with QPSO and WKELM have better performance but are limited by accuracy and flexibility. The presented EGRU-KOA model overcomes the above issues through the integration of Enhanced Gated Recurrent Units (EGRU) and the Kookaburra Optimization Algorithm (KOA) for enhanced parameter tuning and error minimization. Utilizing FER-2013 and EMOTIC datasets with preprocessing methods such as LBP and CBP, the model performs greatly better than traditional approaches in accuracy and efficiency

Text based mood detection, Sentiment analysis in the early days used NLP for text polarity detection. Such as positive or negative tone, whether, neutral or not. For example, it used resources like Word Net and emotion ontologies to gauge mood [2-3].

Early research on mood detection concentrated on feature extraction that starts from facial expression, speech prosody, and text analyses, assuming the existence of six standard facial_expressions elicited in humans for emotion categorization, as proposed by Ekman's work [4].

Convolutional Neural Networks (CNNs) have become the standard for mood detection, particularly in image-based applications. These networks are highly effective because of their ability to automatically generate hierarchical representations from raw pixel data, thereby reducing the dependency on manually crafted features that were once essential in traditional image processing approaches. For example, the LeCun et al [5], in 1998 CNNs invention for image recognition opened the avenue for application in facial emotion detection. In fact, CNNs work perfectly on perceiving complex and subtle expressions through learning spatial hierarchies within facial features, which makes them good at distinguishing slight mood indicators. In fact, the CNN architectures VGGNet and ResNet have been applied to emotion detection that achieves the best performance on the FER-2013 dataset by Goodfellow et al.[6].

Most of the CNN-based mood detection models are based on well-established facial emotion datasets, and FER-2013, CK+, and JAFFE are, in fact, some of the largest. Those are good sources for the annotated facial expressions needed for training and testing the emotion recognition models. For example, the CK+ dataset is famous for its well-balanced class distribution that enables efficient robust detection across a broad range of different emotional expressions, which further enhances the credibility of mood detection systems in diversified contexts. The preprocessing process generally involves normalizing images [8-10] from these datasets into grayscale followed by resizing and cropping them to the size specifications of the input layer of the CNN. This pre-processing would standardize the data, but it would also reduce the computational costs while allowing smother training and faster inference.

Real-time mood detection systems have commonly been used in applications such as virtual assistants, customer service automation, and mental health support tools. They use video feeds taken from webcams to analyze facial expressions in real time and provide immediate feed-back on the emotional state of a person. This is very often achieved by combining Haar cascades for facial detection into CNN models intended for emotion detection so that the system becomes responsive and effective even when light or environmental conditions change[7].This hybridization allows for video capturing and processing to be performed at the frame level in which the emotional indicators of every frame are incessantly detected while an ongoing assessment of the user's mood is conducted.

A. Limitations and Future Directions

The Current CNN-based mood-detection systems are indeed highly potent, yet they also have some disadvantages and are indeed far from the complexity of human emotions. Future developments may, therefore, focus on multimodal approaches where visual data will be merged with text, voice, and physiological signals to paint a better and more holistic picture of mood and emotion. Additionally, techniques like transfer learning—where models pre-trained on large datasets transfer learned features to new tasks—can enhance the performance of mood detection systems, especially in low-data scenarios. Generative models also hold promise for generating synthetic expression data to improve the diversity of training datasets, potentially addressing challenges in detecting subtle or rare expressions. These advancements could significantly expand the capabilities of mood detection, making it more accurate, adaptable, and widely applicable across various domains.

III. Proposed System

In this work, an attempt is made to develop a model for face emotion recognition by using Python based on both the machine and deep learning frameworks. Ensuring that the data was as diversified as possible and reasonably representative of most facial expressions has led to curating a balanced dataset comprising a wide variety of faces, with each tagged with distinct emotional labels. This allows it to be generalized with a variety of expressions, and thus is good model for subtle changes made between two different emotions. The main stages involved in this work include rigorous data preprocessing and noise and inconsistency removal [8-10] in the dataset in order to ensure the quality of data. Advanced feature extraction would be required to extract critical facial signals capturing subtle changes in expressions. Finally, we opt for a more elaborate architecture of our neural network that learns and classifies facial signals in real-time with regard to responding to the complexity of varied emotions. Our model seeks not only to classify well but characterizes the features underlying these classifications to support its application in sentiment analysis, human-computer interaction, or even behavioral studies.

The webcam-based architecture for instant mood classification will capture, process, and analyze video frames to classify a user's mood. Initialization of the camera feed initiates the real-time frame capture in which the system activates the webcam. While initialized, the system captures individual frames from the video feed, then analyzes these captured frames for mood detection. Each frame has to undergo preprocessing so that it can be further analyzed; the preprocessing includes grayscale conversion, resizing of the frame, and normalization of pixel values, which help in accurate detection.

After that, the face detection algorithm is applied to detect the face and isolate it from the video frame, where the face region of interest that will be crucial for mood classification is emphasized. The isolated face region is then input to a CNN model for feature extraction.

The CNN structure utilizes convolutional layers to extract patterns and features from facial images, applies pooling layers for dimensionality reduction, and includes a flattening step to feed the input to the classifier. After having extracted the features, the following layers that consist entirely of fully connected layers process the information on detailed facial features and prepare it for classification use. The final stage of the classification process is performed by a softmax layer that produces a probability distribution across predefined mood categories, such as Neutral, Happy. The mood with the highest probability is selected as the identified mood for that frame. The identified mood is then displayed on the live video feed, thus providing the user with instant feedback. It then allows the system to cycle back, capture the next frame, and continue this process for continued mood detection. This architecture enables effective real-time mood classification to be responsive due to the use of CNN-based feature extraction and classification in a streamlined cyclic workflow. The proposed architecture of the system is mentioned in Fig. 1 below.

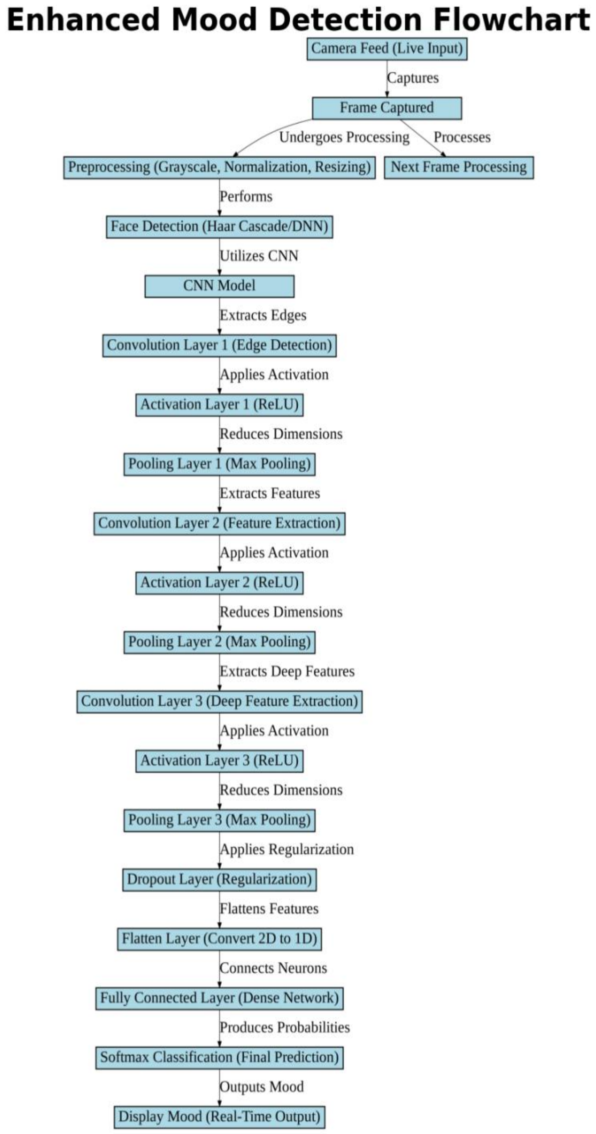


Fig. 1. System Flowchart

CNNs extract hierarchical spatial patterns in three stages. Early layers detect basic features like edges, gradients, and textures, helping identify shapes and contrast differences. Mid-level layers recognize parts of objects such as eyes, noses, and mouths in facial images, along with local textures and geometric patterns. Deeper layers combine these features to understand complete objects, such as recognizing a full face and classifying emotions like happiness or sadness. This step-by-step learning enables CNNs to automatically analyze images, making them highly effective for tasks like facial recognition and emotion detection

IV. Implementation and Testing

A mood detection system was developed by programming with Python, integrating computer vision and deep learning for analyzing real-time mood expressions. A vigorous testing procedure was adopted to make the system reliable and accurate. Testing is, after all, confirmation of functionality and identification of bugs for rectification, which would eventually promise the quality and user requirements for the system. An elaborate test plan incorporating facial expression dataset has been conducted for validation across the range of moods taken in the dataset.

The very motivation for moving to Python is that it offers a rich and vast ecosystem of libraries and tools. Libraries like TensorFlow and Keras have been instrumental in easing the CNN model development process with pre-built layers and the possibility of acceleration on a GPU as well as easy APIs. OpenCV also happens to be quite useful for capturing and processing images in real-time, as is necessary in video-based applications. One can utilize OpenCV to take webcams into real-time captures and preprocess facial data. For example, a mood-detecting model might use OpenCV in handling both face detection and preprocessing while Keras constructs the CNN architecture, and it controls the training process to make it smooth and seamless for developers.

Table 1 illustrates the test cases, the Testing included various aspects, each designed to confirm quality and strength of the software, which benefits include the following:

- Cost Reduction: Early diagnosis decreases total development and maintenance costs.
- Enhanced User Experience: Clear, Predictable interface really recognizes moods; in-line with users expectations improves satisfaction.

TABLE I. Test Cases

Test Case ID	Description	Input	Expected Result	Actual Output
TC01	Initializing Face Detection	Starting Camera Feed	Face detected by the system in the frame	Face detected correctly
TC02	Neutral Mood Classifying	User mood is neutral	The system has given an output as: "Neutral"	Mood: Neutral detected
TC03	Happy Mood Classifying	User mood is happy	The system has given an output as: "Happy"	Mood: Happy detected
TC04	Frame Capture when light is too low	Camera light is very low	System captures mood efficiently	Accurate mood in low lighting
TC05	Real-Time Classification	Mood change really fast	System updates with the mood label within seconds	Real-time mood detection
TC06	Invalid Image Format	Non-face object in frame	System does not attempt mood classification	Error handled properly

V. Results and Discussion

The FER-2013 dataset is used in this research work, downloaded from the 2013 Kaggle Challenge, contains 35,887 grayscale images (48x48 pixels) across seven emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Dataset Split into training, validation, and test sets; it supports deep learning models like CNNs with its size and diversity. Despite low resolution and class imbalance, it remains a key benchmark for emotion recognition.

The deployed mood detection system did quite well in real-time mood classification and detection using a live webcam feed. Summary of test results below:

1. High Accuracy: The proposed classification system picked the moods accurately, with very few errors between neutral and happy moods. This is because the pre-trained neural network can distinguish facial expressions regardless of the conditions.

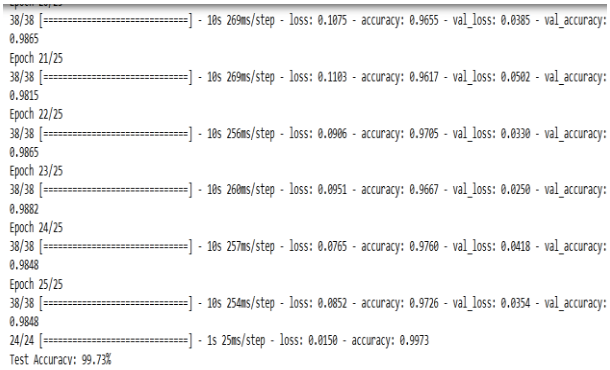
2. Real-Time Performance: The mood detection system worked with minimum latency and updated classifications in near real-time, hence the system was able to act on changes of mood almost in real-time, making the application interactive.

3. Strong Error Handling: A large number of error scenarios were tested in advance, including camera disconnection and other non-face objects appearing in the frame. the system was able to handle this

4. Case correctly, either by proper error messages or by attempting non-classification.

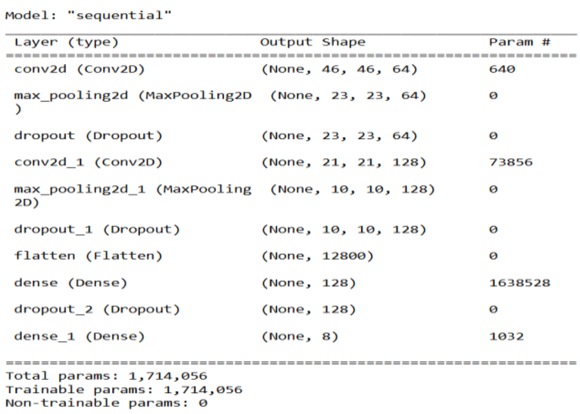
5. Adaptability in Varied Conditions: The model was constant in all conditions of lighting. Extremely minor alterations in natural and artificial illumination were inconsequential to the system's accuracy; which it serves towards the variety of real-life settings.

6. System Usability and User Feedback: Following the initial testing sessions, the software was intuitive enough in being easy to set up, which makes it applicable in the usage even of less-technically-equipped users. The graphical interface clearly presented the classification results for users.



```
38/38 [==============================] - 10s 269ms/step - loss: 0.1075 - accuracy: 0.9655 - val_loss: 0.0385 - val_accuracy:
0.9865
Epoch 21/25
38/38 [==============================] - 10s 269ms/step - loss: 0.1103 - accuracy: 0.9617 - val_loss: 0.0502 - val_accuracy:
0.9815
Epoch 22/25
38/38 [==============================] - 10s 256ms/step - loss: 0.0906 - accuracy: 0.9705 - val_loss: 0.0330 - val_accuracy:
0.9865
Epoch 23/25
38/38 [==============================] - 10s 260ms/step - loss: 0.0951 - accuracy: 0.9667 - val_loss: 0.0250 - val_accuracy:
0.9882
Epoch 24/25
38/38 [==============================] - 10s 257ms/step - loss: 0.0765 - accuracy: 0.9760 - val_loss: 0.0418 - val_accuracy:
0.9848
Epoch 25/25
38/38 [==============================] - 10s 254ms/step - loss: 0.0852 - accuracy: 0.9726 - val_loss: 0.0354 - val_accuracy:
0.9848
24/24 [==============================] - 1s 25ms/step - loss: 0.0150 - accuracy: 0.9973
Test Accuracy: 99.73%
```

Fig. 2. Proposed Model Testing Accuracy



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 46, 46, 64)        640

 max_pooling2d (MaxPooling2D  (None, 23, 23, 64)       0
 )

 dropout (Dropout)           (None, 23, 23, 64)        0

 conv2d_1 (Conv2D)           (None, 21, 21, 128)       73856

 max_pooling2d_1 (MaxPooling  (None, 10, 10, 128)      0
 2D)

 dropout_1 (Dropout)         (None, 10, 10, 128)       0

 flatten (Flatten)           (None, 12800)             0

 dense (Dense)               (None, 128)               1638528

 dropout_2 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 8)                 1032

=================================================================
Total params: 1,714,056
Trainable params: 1,714,056
Non-trainable params: 0
```

Fig. 3. Proposed Model layers, output shape and parameters

7. Future Enhancements: The system was effective but definitely suggests further improvements in terms of mood classification that would include detecting mood such as sad or surprised expressions, amongst others. It would also be helpful to optimize its processing aspect especially suited for mobile or embedded platforms

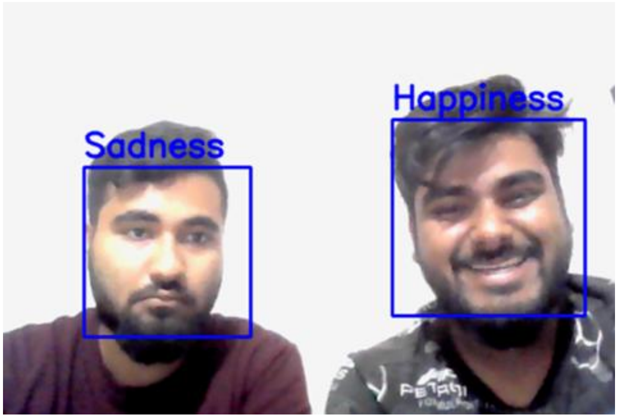


Fig. 4. Sadness and Happiness Mood Detected

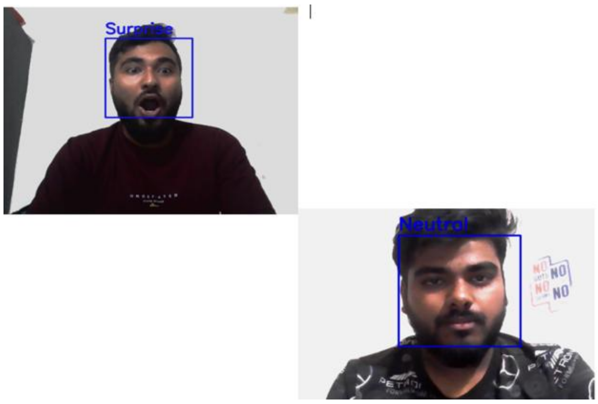


Fig. 5. Surprise and Neutral Mood Detected

The innovative factor of proposed mood detection system is fundamentally based on its ability to combine facial detection with emotion recognition, based on Convolutional Neural Networks, commonly referred to as CNNs, in real time. This remarkable system has an accuracy rate of more than 99%. Moreover, it successfully combines the efficiency of Haar cascades in face detection with sophisticated preprocessing techniques and CNNs, which allow for highly accurate mood classification. This not only ensures effective performance but also guarantees robustness across a wide range of diverse conditions.

TABLE II. COMPARATIVE ANALYSIS WITH EXISTING MODELS

Model/ Approach	Methodology	Dataset	Accuracy
Heuristic Methods	Manual feature extraction from facial expressions & text	Custom datasets	~70%-80%
VGGNet-CNN	Deep CNN for spatial feature learning	FER-2013, CK+, JAFFE	~90%-92%
ResNet-CNN	Deep CNN with residual connections for better learning	FER-2013, CK+	~93%-95%
Proposed Approach	CNN with preprocessing (grayscale, resizing, dropout)	FER-2013, Custom dataset	**99.73%**

In proposed experimental configuration, the CNN model was trained with appropriately optimized parameters for precise emotion classification. In preprocessing, images were resized to 48x48, converted into grayscale, and normalized to range pixel values from 0 to 1. Data augmentation methods such as flipping, rotation, and cropping were used to enhance the model's capability to handle variations in real data. To avoid overfitting, a dropout rate of 0.5 was employed in fully connected layers, randomly disabling some neurons during training. Besides, L2 regularization was used to regulate large weights in the convolutional layers, making the model attend to salient features. The above techniques combined assisted the model to achieve high accuracy (99%) while being robust and generalized.

With 99% proposed method accuracy, the work is well-suited for applications needing accurate and dependable emotion detection. For example, in mental health monitoring, accuracy ensures proper identification of subtle emotions such as anxiety or sadness with minimal false detections that may mislead treatment. In driver monitoring systems, the high accuracy ensures timely detection of critical states such as drowsiness or distraction, improving road safety. In security systems as well, it minimizes false alarms when detecting aggressive or suspicious behaviors, providing reliable surveillance. The precision and robustness of our method make it versatile for real-world applications such as customer sentiment analysis, remote learning, and virtual reality, where high accuracy is crucial to enhance user experience and decision-making.

VI. Conclusion and Future Scope

This paper proposes an effective mood detection system based on Python and deep learning techniques. Images were used to analyze facial expression data primarily through the utilization of convolutional neural networks. Our approach achieved quite promising results in the classification of emotions from facial expressions after systematically preprocessing image data, performing feature extraction, and training a CNN-based model. This paper shows that deep learning architectures, trained and evaluated appropriately, are capable of effectively classifying emotional states, with important implications for surveillance of mental health status, human-computer interaction, and studies of social robotics. Analysis of the system describes its overall robustness, excellent performance metrics, and suggests that facial image data are a suitable resource for mood classification.

A. Future Perspectives

Although the results of our mood detection system are encouraging, numerous opportunities exist for additional investigation and improvement:

Improve Model Robustness Future work should continue to explore larger, and more diverse datasets to improve model generalizability to different demographics, lighting conditions, and facial expressions.

Multi-modal Emotion Recognition: Audio cues with the analysis of text and physiological signals combined with visual data can further improve more accuracy and more holistic understanding of mood.

Privacy-preserving methodologies: Models that work on anonymized or less granular facial data may manage to mitigate privacy-related fears and make mood detection technologies more user-friendly and ethical. Extended Application Fields: This covers basic applications in areas as elementary as education, social robotics, or health monitoring, including individual therapy and mood monitoring.

References

[1] Anand M and Babu S, Predicting the facial expression recognition using novel enhanced gated recurrent unit-based Kookaburra Optimization algorithm. (2024). International Journal of Intelligent Engineering and Systems, 17(3), 159–170. https://doi.org/10.22266/ijies2024.0630.14

[2] Li, S. (2025). Application of entertainment e-learning mode based on genetic algorithm and facial emotion recognition in environmental art and design courses. Entertainment Computing, 52, 100798

[3] Zhang, F., Liu, Y., Yu, X., Wang, Z., Zhang, Q., Wang, J., & Zhang, Q. (2025). Towards facial micro-expression detection and classification using modified multimodal ensemble learning approach. Information Fusion, 115, 102735

[4] Ekman, P. (2016). What Scientists Who Study Emotion Agree About. *Perspectives on Psychological Science*, 11(1), 31-34

[5] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-Based Learning Applied to Document Recognition." In Proceedings of the Ieee, 2278–2324

[6] Goodfellow, D. Erhan, P. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D. Lee, et al. Challenges in representation learning: A report on three machine learning contests. In International conference on neural information processing. Springer, 2013

[7] Khan, Jawad & Jeong, Byeong-Soo. (2016). Summarizing customer review based on product feature and opinion. 10.1109/ICMLC.2016.7860894.

[8] Chandrappa S, Dharmanna L, Basavaraj Anami, " A Novel Approach for Early Detection of Neovascular Glaucoma Using Fractal Geometry", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.14, No.1, pp. 26-39, 2022. DOI: 10.5815/ijigsp.2022.01.03

[9] Akrout, B. (2025). Deep facial emotion recognition model using optimal feature extraction and dual - attention residual U - Net classifier. Expert Systems, 42(1), e13314.

[10] S. Chandrappa, L. Dharmanna and K. I. R. Neetha, "Automatic Elimination of Noises and Enhancement of Medical Eye Images through Image Processing Techniques for better glaucoma diagnosis," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 2019, pp. 551-557, doi: 10.1109/ICAIT47043.2019.8987312.