Deep Learning-Based Facial Emotion Recognition for Neuro-Psychological Screening

Jeeto George [1], S. Aswathy Abirami.,[2]

[1]*PG Scholar Bio-Medical Department, Udaya School of engineering, Kanyakumari, Tamil Nadu, India.*
[2]*Assistant Professor Bio-Medical Department, Udaya School of engineering, Kanyakumari, Tamil Nadu, India.*

Abstract-*In this project, an intelligent system has been developed to automatically detect and classify human emotions by analyzing facial expressions using advanced image processing and deep learning techniques. Unlike traditional methods that relied on Principal Component Analysis (PCA), this system employs a Convolutional Neural Network (CNN) to automatically extract and learn facial features. This allows the system to understand complex patterns of facial muscle movements directly from images, significantly improving accuracy and reliability. The process starts with capturing facial images and focusing on key expression areas such as the eyes, mouth, and eyebrows. The CNN model then extracts high-level features from these regions, which are used to classify emotions into categories like happiness, sadness, anger, surprise, and neutral expressions. The model is implemented using popular deep learning frameworks like MATLAB and Python, enabling it to perform real-time and dependable emotion recognition. Beyond simple emotion detection, this system analyzes subtle variations in facial expressions to support advanced neuro-screening applications. It can assist in the early detection of neurological or psychological conditions linked to impaired emotional responses, including autism spectrum disorders, Parkinson's disease, and depression. Overall, this deep learning–based approach offers a more accurate, adaptive, and automated solution for emotion-driven neuro screening, holding great promise for medical diagnostics, mental health evaluation, and the future of human-computer interaction.*

Key Words: Facial Emotion Recognition, Convolutional Neural Network (CNN), Deep Learning, Image Processing, Emotion Classification, Neuro-Screening

1. INTRODUCTION

Facial emotion recognition is a fascinating technology that helps us understand human feelings by analyzing facial expressions. By detecting emotions like happiness, sadness, anger, and surprise, it allows computers to interact with us more naturally and effectively [1]. This technology is widely used in areas such as security, healthcare, and smart devices, where understanding how people feel is crucial [2]. In the past, facial expression recognition relied on handcrafted features and classical techniques like Principal Component Analysis (PCA). However, these traditional methods struggled to capture the complex and subtle variations in real-life facial expressions [3]. They were also sensitive to changing lighting, occlusions, and head movements, which often reduced their accuracy [4]. The rise of deep learning, especially Convolutional Neural Networks (CNNs), has transformed this field. CNNs automatically learn important facial features from images without the need for manual input [5]. This has led to much better performance and greater robustness, allowing the models to detect even subtle facial muscle movements that older methods missed [6]. More recently, innovations like attention mechanisms, transfer learning, and hybrid deep learning models have further boosted the accuracy of emotion recognition, particularly in real-world and real-time scenarios [7]. Researchers are also designing lightweight CNN models that can run efficiently on edge devices with limited computing power, making this technology more accessible and practical [9]. Beyond general emotion detection, facial expression analysis is making a significant impact in mental health and neurological screening. It can help identify conditions such as depression, autism, and Parkinson's disease by spotting unusual emotional responses, providing valuable support for diagnosis and treatment [8], [10]. Despite these advances, challenges remain. Facial expressions vary widely between individuals, and obstacles like occlusion and the need for diverse, large datasets continue to pose difficulties [4]. The ongoing goal is to develop efficient, accurate models that can handle these challenges while processing data in real time. Combining cutting-edge deep learning techniques with practical application is key to creating reliable and effective emotion recognition systems for the future [5], [7], [9]. Another key part of facial emotion recognition is how good and available the data sets are for training deep learning models. Big data sets that show many different facial expressions in various lighting, angles, and backgrounds are really important for making models more accurate and able to work in different situations [2], [6]. But getting this kind of data is hard because of privacy issues and the need for careful labeling. To fix this, researchers are using more data enhancement methods and creating fake data to make training better and stop models from learning too much from one type of data [7],

[10]. Also, combining facial emotion recognition with other types of information like speech, text, and body signals has created new ways to study emotions using multiple sources. Using several types of data can make emotion detection systems more reliable and better at handling real-life challenges [8]. Future work is centered on making systems that can handle all these different data types efficiently while still working quickly. These improvements will help in areas like virtual assistants, education, health monitoring, and smarter ways people interact with computers [5], [9]. An emerging trend in facial emotion recognition involves the application of explainable artificial intelligence (XAI) methods to enhance the transparency and interpretability of model decisions. It is crucial to understand how and why a model assigns a specific emotion, particularly in sensitive areas such as healthcare and psychological evaluation [8]. Explainable models assist in uncovering biases, increasing trust, and ensuring fairness in emotion detection systems. Additionally, researchers are concentrating on enhancing the ethical dimensions of these systems, including data privacy, informed consent, and the responsible handling of facial data [2], [7]. The integration of explainability and ethical considerations will be vital for the broad implementation of facial emotion recognition technologies in practical applications [5], [10].

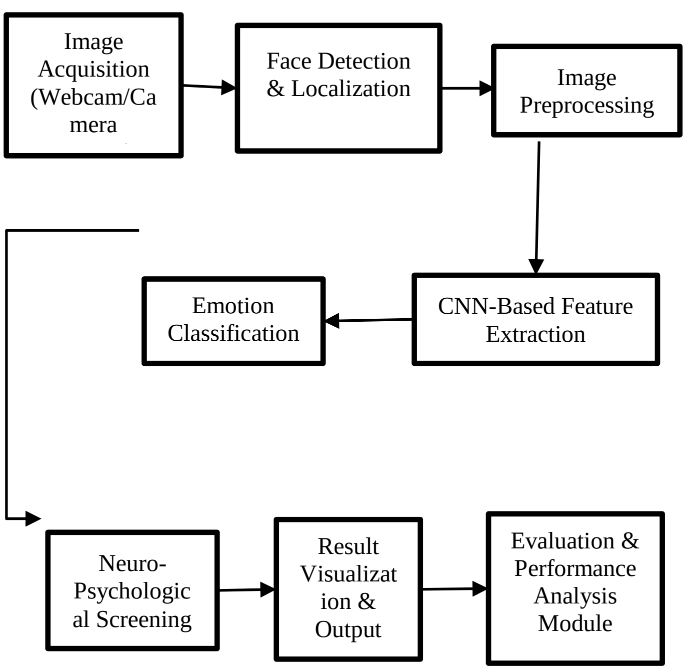

Fig 1: Block Diagram

2. METHODOLOGIES

The proposed system uses a new method for recognizing emotions from facial expressions, based on deep learning. It uses a type of neural network called a Convolutional Neural Network (CNN) to improve upon older methods. Traditional systems needed people to manually pick out important features from faces, like using PCA. This new system does it automatically by learning from the images directly. The system starts with collecting facial images, either from live video or from stored image sets. These images go through some steps to make them ready for processing. These steps include finding the face, aligning it, resizing it, adjusting brightness and contrast, and removing any noise to make the images clearer and more consistent. Once processed, the images are sent to the CNN model, which is the main part of the system. The CNN learns to recognize features in a layered way. At the beginning, it looks for basic details like lines, textures, and shapes. As it goes deeper, it starts recognizing more complex shapes like facial structures and how muscles move, which are important for telling emotions apart. This automatic learning process means the system doesn't need people to design features manually. It becomes more flexible and better at recognizing emotions. The system is especially good at picking up small changes in facial muscles, like around the eyes, eyebrows, and mouth, which are key for showing emotions. This helps the system tell similar or subtle expressions apart with high accuracy. The model is trained using images that have labels showing the emotions they represent, like happy, sad, angry, surprised, or neutral. To make the model work well in different situations, it uses techniques like adjusting the data, randomly dropping parts of the images during training, and normalizing the data. These steps help the model work well in different lighting, angles, and backgrounds. The system can also detect and classify emotions in real time. By making the CNN efficient and using good computing techniques, the system can process live video and give instant results. This real-time ability makes it useful for things like smart security systems, virtual assistants, and online learning tools. Another important use for this system is in checking for neuro-psychological conditions. By looking at patterns in how people express emotions, the system can help detect conditions like autism, Parkinson's disease, and depression. Since it uses data automatically, it's less biased and provides more consistent results. Overall, this CNN-based system is strong, accurate, and can handle a lot of data. It replaces the old manual methods with smart learning, leading to better emotion recognition and real-time use. It works well for healthcare, human-computer interaction, and systems that understand emotions.

2.1 Image Acquisition

The Image Acquisition Module is the starting point of any facial emotion recognition system. Its job is to capture clear and accurate images of faces, either from live camera feeds or from stored datasets, which then serve as the foundation for analyzing emotions. The quality of these images is crucial because it directly impacts how well the system can understand and interpret facial expressions. When working in real time, this module grabs video through webcams or built-in cameras and continuously extracts frames to detect emotions as they happen. This real-time capture is essential for applications like human–computer interaction and behavioral studies, where immediate feedback is needed. At the same time, the module can process offline datasets filled with labeled facial expressions. These datasets are vital for training and testing the system's deep learning models, helping improve accuracy and reliability. To ensure great results, the module captures images with enough resolution and clarity for detailed feature extraction. It also adapts to different facial poses and angles, so it can recognize faces even if people aren't looking straight at the camera. Additionally, the images are standardized into common formats like RGB or gray scale, making them compatible with the system's preprocessing and analysis steps. In short, this module keeps a steady and reliable flow of facial images coming in, whether in real-time or from saved data. This continuous stream allows the system to effectively detect and analyze emotions, making the technology practical and powerful in a variety of settings.

2.2 Face Detection & Localization

Imagine a system designed to spot faces effortlessly in images or videos, focusing only on what matters most: the face itself. This is the role of the Face Detection & Localization Module. It scans input visuals and accurately identifies where a face is, trimming away all unnecessary background to streamline what gets analyzed next. To accomplish this, the module uses advanced techniques like Haar Cascade, MTCNN, and deep learning models. Among these, deep learning stands out for its ability to handle challenging situations whether the lighting is tricky, faces are turned at odd angles, or parts are hidden behind glasses or masks. Once a face is found, the system draws a bounding box around it, zeroing in on the exact area to keep. But it doesn't stop there. The module also pinpoints important facial landmarks key features such as the eyes, nose, mouth, and eyebrows. These details are essential for understanding expressions and making sure the face is aligned properly for the next steps, such as recognizing emotions or preparing data for further processing. Finally, the detected face is neatly cropped and resized according to the bounding box. This ensures that every face fed into the system is consistent and focused, making the subsequent analysis more accurate and efficient. In this way, the Face Detection & Localization Module lays the crucial groundwork for powerful facial recognition and emotion detection technologies.

2.3 Image Preprocessing

When raw facial images come in, they often differ in size, lighting, and quality making it tough for a CNN model to analyze them accurately. That's where the Image Preprocessing Module steps in, transforming these varied images into a consistent, clean format that the model can work with more efficiently. First, the images are converted into either grayscale or normalized RGB. Grayscale helps reduce the amount of data the model has to process while keeping the key facial details intact. If color is important, the RGB format is used instead. Next, all images are resized to a standard dimension, like 48×48 or 224×224 pixels, so they fit perfectly into the CNN's expected input size. But size and color aren't the only challenges. To make the images clearer, the module applies noise reduction and contrast enhancement techniques. For instance, histogram equalization boosts the contrast, making facial features like eyes and mouth stand out more. Then, pixel normalization adjusts the brightness values to a common scale, which helps the model learn more steadily and quickly. One crucial step is face alignment, which ensures that key facial features eyes, nose, and mouth are positioned consistently across all images. This alignment means the model can focus better on the right details every time. In the end, this preprocessing work cleans up the data and organizes it neatly, setting the stage for the CNN model to extract features more effectively and boost its emotion recognition performance. It's all about turning messy input into reliable, ready-to-analyze images, making the whole system smarter and more accurate.

2.3 Facial Feature Extraction (CNN-Based)

Imagine a system that can understand human emotions just by looking at a face. The Facial Feature Extraction Module makes this possible using a powerful technology called Convolutional Neural Networks (CNNs). Unlike older methods that required experts to manually select important facial features, CNNs learn directly from the data, making the process smarter and more accurate. First, the system takes preprocessed facial images and passes them through several layers of the CNN. Early layers focus on capturing simple details like edges and textures. As the data moves deeper, the network recognizes more complex patterns such as facial structures and expressions. This step-by-step learning allows the system to grasp both basic and subtle emotions. What makes CNNs especially effective is their ability to detect tiny muscle movements around the eyes, eyebrows, and mouth areas crucial for expressing feelings. Through convolution and pooling operations, the network extracts important spatial and texture information while keeping the process efficient by reducing unnecessary data. Another strength of this module is its resilience. It performs well even when faces are seen under different lighting, angles, or partially obscured. By training on a wide variety of images, the CNN adapts to these changes seamlessly. In the end, it produces detailed feature maps that feed into the next stage, where the emotion behind the expression is identified. In essence, this technology combines deep learning with thoughtful design to decode human emotions from facial images, opening doors to more intuitive and responsive applications in fields like security, healthcare, and human-computer interaction.

2.4 Emotion Classification

The Emotion Classification Module marks the final and essential step in a facial emotion recognition system. Its role is to make sense of the features extracted by a convolutional neural network (CNN) and assign a specific emotion label to a facial image. After the CNN identifies detailed feature maps that highlight important aspects of the face, these are sent to classification layers designed to accurately recognize emotions. This module uses fully connected layers to combine all the learned features from earlier convolutional stages. By doing so, it merges spatial and texture information into a comprehensive representation of the facial expression. The system is capable of distinguishing multiple emotions such as happiness, sadness, anger, surprise, and neutral. At the output stage, the module applies a Soft max activation function. This converts the network's raw outputs into probabilities, showing how confident the model is about each emotion category. This approach allows the system to evaluate all emotions simultaneously and select the most likely one. During training, the system improves its accuracy by minimizing classification errors, using techniques like cross-entropy loss combined with optimization methods such as Adam or Stochastic Gradient Descent (SGD). Once trained, the model supports real-time emotion detection, processing live video or images instantly. The final output includes the predicted emotion label and a confidence score, reflecting the model's certainty. This feature helps applications make informed decisions based on how sure the system is about the detected emotion, making the module both powerful and practical for real-world use.

2.5 Neuro-Psychological Screening

In the world of clinical and psychological analysis, understanding emotions plays a vital role. The Neuro-Psychological Screening Module takes emotion recognition a step further by studying how emotional responses change over time. This helps experts spot unusual behaviors that might signal neurological or mental health issues. By examining sequences of emotions detected across moments, the module can find patterns like delayed reactions, inconsistencies, or sudden emotional swings. It's designed to pick up signs such as reduced emotional expression or overly intense responses, which are often linked to conditions like depression, autism, or Parkinson's disease.

Beyond just identifying emotions, the module measures how strong and consistent these feelings are expressed, offering deeper insights into a person's emotional health. These detailed observations can support early diagnosis and provide healthcare professionals with valuable information during assessments. Ultimately, this module transforms emotion recognition technology into a powerful decision-support tool, delivering objective and automated analysis that enhances neuro-psychological screening and mental health monitoring. It's a step forward in helping clinicians better understand and care for their patients.

2.6 System Implementation & Visualization

Imagine a system that brings together the latest in technology to understand human emotions in real time. This is exactly what the System Implementation & Visualization Module achieves. Using powerful tools like MATLAB or Python frameworks such as Tensor Flow, Keras, and PyTorch, it builds and runs advanced convolutional neural network (CNN) models that can detect emotions as they happen. What makes this system truly user-friendly is its clear and interactive display. Users see emotion recognition results instantly, with visual overlays, labels, and graphs that make emotional patterns easy to grasp. Beyond just showing results, it also saves the data for future review, whether for clinical studies or assessing how well the system performs over time. Designed with growth in mind, this module can handle multiple users simultaneously and easily connect with other software, including healthcare platforms. In the end, it combines efficient processing, intuitive visualization, and practical usability to make real-time emotion recognition accessible and effective.

2.7 Performance Analysis

Facial emotion recognition systems are becoming increasingly important, especially in sensitive fields like neuro-psychological screening and healthcare monitoring. To ensure these systems work effectively, a dedicated Performance Evaluation Module measures how accurately and reliably they detect and classify emotions from facial expressions in various real-world conditions. This evaluation relies on several key metrics based on four outcomes: True Positives (correctly detected emotions), True Negatives (correctly identified non-targets), False Positives (incorrectly predicted emotions), and False Negatives (missed emotional expressions). These form the foundation for calculating performance measures such as accuracy, precision, recall, and the F1-score.

Accuracy measures the overall correctness of the system by calculating the proportion of correctly classified emotions out of all predictions. It is expressed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy provides a general idea of the system's performance but may be insufficient when dealing with imbalanced emotion classes.

Precision quantifies how many of the emotions predicted as positive are actually correct. High precision reduces false detections and ensures that identified emotions are reliable:

$$\text{Precision} = \frac{TP}{TP + FP}$$

This metric is particularly important in applications where false detection could lead to misinterpretation, such as clinical assessments or mental health screening.

Recall (Sensitivity) measures the system's ability to correctly detect actual emotions, indicating how many true emotional expressions are identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN}$$

High recall is crucial for medical and neuro-psychological applications, where missing subtle emotional cues can affect the accuracy of diagnosis or monitoring.

F1-Score provides a balanced evaluation by combining precision and recall into a single metric, making it especially useful when emotion classes are imbalanced:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-score reflects the system's overall robustness in correctly classifying emotions while minimizing both false positives and false negatives.

Overall, the Performance Evaluation Module ensures that the facial emotion recognition system is reliable, accurate, and suitable for both real-time applications and clinical assessments, providing quantitative evidence of its effectiveness and areas for improvement.

RESULT & DISCUSSION

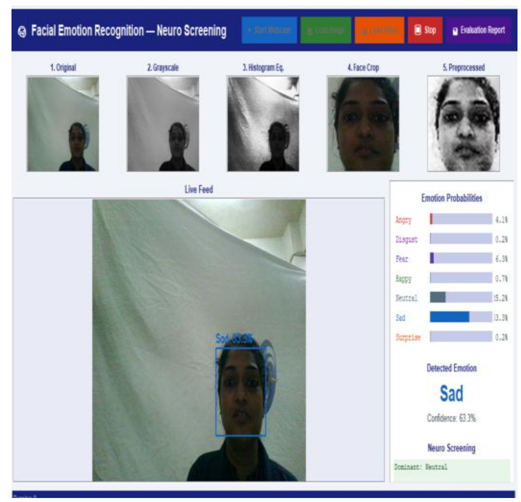

Fig.2. Face Emotion Recognition

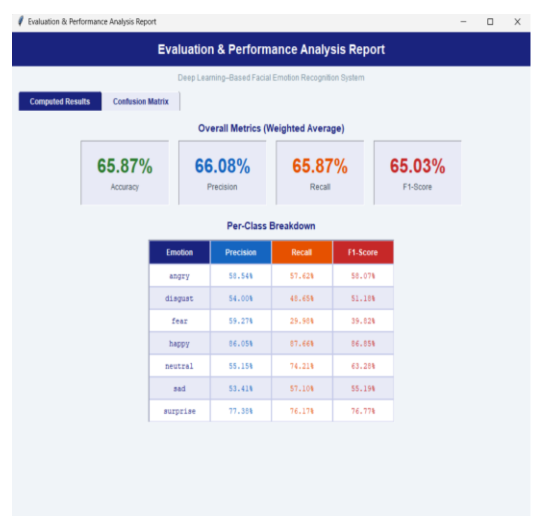

Fig 3: Evaluation & Performance Analysis Report

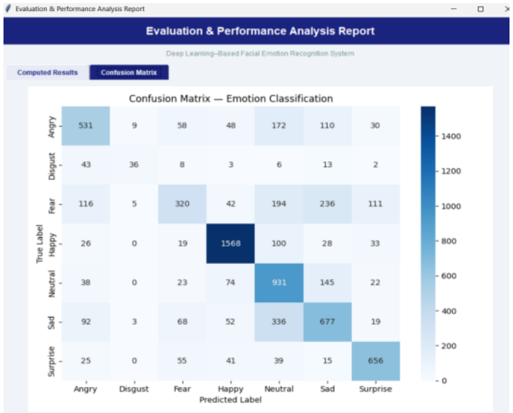

Fig 4: Confusion Matrix

The experimental results show that the facial emotion recognition system works well in both real-time and testing situations. As shown in Fig.2, the system can handle all the steps needed to process images, including taking the image, preparing it, finding the face, and using a CNN to extract features, all in real time using a webcam. The images showing each step help confirm that each part of the system works correctly and helps make the system more reliable. The results from the performance evaluation in Fig.3 show that the CNN model has an overall accuracy of 65.87%, with good balance between precision, recall, and F1-score. These numbers suggest the model works reasonably well across different emotion categories, but there is still room for improvement. The moderate performance can be because of difficulties like changes in facial expressions, different lighting conditions, and the challenge of telling apart subtle emotions. The confusion matrix in Fig. 4 gives more detailed information about how the model makes its predictions. Emotions like Happy, Neutral, and Surprise are predicted more accurately, which means the model has learned to recognize these clear expressions well. However, emotions like Fear and Sad are often predicted incorrectly, especially being mistaken for Neutral. This is expected because these emotions have similar facial features and involve slight muscle movements that are hard to tell apart, even for people. Overall, the results show that the system can detect emotions in real time with acceptable performance. However, improvements could be made by using bigger and more varied data, better model designs, or adding other types of input. Despite these challenges, the system has strong potential for use in areas like human-computer interaction, analyzing behavior, and screening for neurological or psychological conditions, where understanding emotions in real time is important.

CONCLUSION

A new facial emotion recognition system has been developed using convolutional neural networks (CNN) that automatically detects and classifies human emotions from facial images. Unlike traditional methods that relied on PCA-based feature extraction, this system uses deep learning to automatically capture detailed facial features—such as subtle muscle movements around the eyes, eyebrows, and mouth—leading to much better accuracy and reliability. Implemented with Python deep learning tools, the system can classify basic emotions like happiness, sadness, anger, surprise, and neutral expressions in real time. It integrates key components including image acquisition, preprocessing, CNN-based feature learning, classification, and neuro-psychological screening. This makes it a complete, end-to-end solution suitable for both research purposes and practical use. Beyond simple emotion detection, the system shows promise for early neuro-psychological screening by identifying subtle emotional patterns linked to conditions like depression, autism spectrum disorders, and Parkinson's disease. Thanks to its adaptability, real-time processing speed, and automated feature extraction, it stands out as a scalable and dependable tool for applications in healthcare, education, and human-computer interaction.

REFERENCES

[1] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," IEEE Trans. Affect. Comput., vol. 12, no. 3, pp. 1–10, 2023.

[2] S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Trans. Affect. Comput., vol. 13, no. 3, pp. 1195–1215, 2023.

[3] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion-aware facial expression recognition using CNN," IEEE Trans. Image Process., vol. 33, pp. 456–468, 2024.

[4] H. Yang, U. Ciftci, and L. Yin, "Facial expression recognition by de-expression residue learning," IEEE Trans. Image Process., vol. 33, pp. 890–902, 2024.

[5] Z. Zhao, Q. Liu, and S. Wang, "Facial expression recognition from image sequences using deep learning," Pattern Recognit. Lett., vol. 179, pp. 45–52, 2024.

[6] R. K. Gupta, S. Verma, and A. Sharma, "Real-time facial emotion recognition using CNN and transfer learning," Comput. Vis. Image Underst., vol. 235, 2025.

[7] M. H. Khan, J. Zhang, and L. Chen, "Attention-based CNN for facial emotion recognition in real-world scenarios," Expert Syst. Appl., vol. 244, 2025.

[8] P. Kumar and R. Singh, "Deep learning-based emotion detection for mental health monitoring," IEEE Access, vol. 13, pp. 10234–10245, 2025.

[9] T. Nguyen, H. Nguyen, and K. Nguyen, "Lightweight CNN model for real-time facial expression recognition on edge devices," Sensors, vol. 25, no. 4, 2025.

[10] S. Das, A. Dutta, and P. Banerjee, "Hybrid deep learning framework for facial emotion recognition using CNN and LSTM," Neurocomputing, vol. 580, pp. 112–123, 2025.