IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 3, Ser. 5 (May. – June. 2025), PP 06-19
www.iosrjournals.org

Hybrid CNN-LBP Model for Realtime Emotion Classification

Chandra Sekhar Sanaboina
Department of Computer Science and Engineering
University College of Engineering Kakinada JNTUK - KAKINADA

Abstract
Facial emotion detection helps in understanding human behaviour by analyzing expressions. It is widely used in fields like healthcare, security, and education to enhance interactions and decision- making. Traditional methods rely on manual observation, which can be subjective and inefficient, making automated detection a more effective solution. This study proposes two methods: Method 1 uses Convolutional Neural Networks (CNNs) with data augmentation, while Method 2 integrates Local Binary Pattern (LBP) features with CNNs for enhanced feature extraction. Both methods are evaluated on the FER2013 dataset, focusing on five emotions: sad, angry, happy, neutral, and surprise. Enhancements include data augmentation for balanced data and real-time detection via live camera access. Method 2 shows improved accuracy due to LBP features, making the system more precise and reliable for applications in human-computer interaction, mental health analysis, and behavioral research.

Index Terms: *Facial emotion detection, CNN, LBP, feature extraction, real-time analysis, live camera access, deep learning, accuracy*

Date of Submission: 20-06-2025 Date of Acceptance: 03-07-2025

I. INTRODUCTION

Applications in assistive technology, education, healthcare, and driver monitoring are all made possible by emotion detection, which allows robots to recognize human emotions from facial expressions. For instance, [1] demonstrates how it assists teachers in understanding student engagement for the purpose of modifying their teaching approaches. In the medical field, it helps with patient mood assessment [2], and [3] discusses how visually impaired people benefit from systems that can identify the emotional states of others.

The emotion recognition process involves two main steps: feature extraction and emotion classification. A popular technique for feature extraction is Local Binary Pattern (LBP) and Uni form LBP (uLBP), which capture local texture patterns from facial images was introduced in [4]. For partialface scenarios (e.g., masks), advanced methods like Star-like Particle Polygon Estimation (SLPPE) have been introduced to extract meaningful features from the upper face, as discussed in [5].

Convolutional Neural Networks (CNNs) are widely used for emotion classification due to their strength in image feature learning [2], is further explored in [6]. In real-time scenarios such as driver emotion detection, CNNs combined with LSTM networks have shown high accuracy in recognizing states like fatigue or anger are elaborated in [7]. In the context of remote learning, enhanced CNNs (e.g., ResNet with attention mechanisms) improve emotion recognition during live virtual classes was obstinate [8].

Despite these advancements, FER systems face challenges like lighting, occlusion, and head pose variations. Researchers are addressing these with multi-modal inputs, data augmentation, and transfer learning, as discussed in [9], and [10]. Recently, efficient and compressed deep learning models have been developed to reduce computational cost while maintaining strong performance, making FER more suitable for real-world applications [11]. Lightweight CNNs have also enabled deployment on mobile devices [12].

Recent advances also emphasize novel directions such as event-based FER, which captures spatiotemporal dynamics using neuromorphic sensors, as reviewed in [13]. In educational settings, outlines the importance of FER in monitoring cognitive engagement and mental states [14]. Further, multi-modal FER integrating EEG and facial data enhances classification accuracy, as demonstrated by [15]. Additionally, advanced generative models like s e l f - s u p e r v i s e d spontaneous

expression generation [16] are expanding FER capabilities by simulating dynamic emotion sequences without labeled data, aiding both data augmentation and modeling spontaneity in real-world interactions.

II. RELATED WORK

While the introduction covered foundational approaches in FER, additional recent research expands the field through novel architectures, attention mechanisms, multi-modal fusion, and solutions for real-world constraints.

Siddiqi et al. [18] introduced a hybrid CNN with Neural Random Forests for healthcare oriented FER, achieving high accuracy on multiple benchmark datasets. He et al. [19] proposed a three- channel hierarchical CNN that separately processes eyes, mouth, and the entire face to extract more discriminative features. Zang et al. [20] improved feature diversity and generalization using an asymmetric pyramidal network combined with gradient centralization.

In model architecture innovation, Deng et al. [21] presented ENASFERNet, which utilizes evolutionary neural architecture search to automatically design optimized FER networks. Liu et al.

[22] tackled uncertain labels in emotion data by using a multitask-assisted correction strategy incorporating action unit (AU) detection and valence arousal scoring.

Addressing visual modality limitations, Ni et al. [23] introduced a cross modality CNN frame work that integrates grayscale, depth, and LBP images. Li et al. [24] developed a contrastive warm up with complexity-aware self training (CWCST) to improve cross domain FER performance. Similarly, Choi and Lee [25] employed deep CNN ensembles with stochastic optimization to boost robustness in in the wild scenarios.

Beyond visual signals, Wang et al. [26] and G u¨ ler and Akbulut [27] explored multi-modal FER by integrating EEG with facial data. The latter demonstrated enhanced emotion classification using GRU and Transformer models, highlighting the advantage of combining brain and facial signals.

Dong et al. [28] proposed the Multi-scale Attention Learning Network (MALN) using a Vision Transformer backbone to address intraclass similarities and interclass discrepancies in FER. Shahzad et al. [29] focused on FER in masked conditions through multi-modal feature fusion using modified Xception architecture. Lan et al. [30] introduced MrCAR, a residual attention network emphasizing eye and mouth regions for subtle expression cues.

To further enhance feature representation, Zhao et al. [31] designed a graph based adaptive convolutional network using facial landmarks. He [32] presented a multi branch attention CNN optimized for low resource deployment while preserving classification accuracy.

These studies provide vital insights into improving FER performance in real world environments characterized by occlusions, noisy labels, low resolution, and domain shifts. They also emphasize the growing role of multi-modal and attention-based techniques in overcoming traditional CNN limitations.

III. PROPOSED METHODOLOGY

This study presents a dual method architecture for facial emotion recognition that aims to address challenges such as variations in facial expression intensity, limited labeled data, and noise in real- time environments. The motivation behind this work is to enhance classification accuracy by leveraging the strengths of both deep learning and handcrafted feature extraction methods. The system is designed for real-time implementation and focuses on detecting five core emotional states: *sad, angry, happy, neutral*, and *surprise* - which are frequently encountered in human-computer interaction and mental health applications.

The methodology revolves around the use of the FER2013 benchmark dataset, known for its diverse facial expressions collected under unconstrained conditions. The proposed system combines powerful feature extraction capabilities of Convolutional Neural Networks (CNNs) with the textural sensitivity of Local Binary Patterns (LBP). This hybrid approach is particularly effective in scenarios where emotional states are subtle and may not be distinctly visible from raw image pixels alone.

3.1 Method Explanation

Two complementary methods are introduced and analyzed in this study:

- **Method 1: CNN with Data Augmentation** — This baseline method uses a custom built

Convolutional Neural Network tailored for grayscale facial images of size 48x48. The architecture includes four convolutional blocks, each consisting of convolutional layers followed by batch normalization, ReLU activation, and maxpooling. Dropout layers are added to reduce overfitting. To improve the diversity of training samples and address data imbalance, data augmentation techniques such as random rotation, zooming, brightness shifts, and horizontal flipping are employed. The model is trained using categorical cross entropy loss and optimized with the Adam optimizer. Early Stopping and ReduceLROnPlateau callbacks are used to improve model generalization and training efficiency.

- **Method 2: CNN with LBP and Data Augmentation** — Method 2 extends the architecture of Method 1 by integrating LBP (Local Binary Pattern) features. LBP is a texture descriptor that encodes the local neighborhood of each pixel into binary patterns, which are then converted into a histogram. This histogram captures essential texture features that are often missed by CNNs alone. In this method, the LBP histogram (10bin normalized) is computed for each image and concatenated with the flattened CNN feature vector before being passed to fully connected layers. This hybrid approach exploits both learned spatial hierarchies and local textural information to enhance classification performance, particularly for emotions with similar visual cues (e.g., neutral vs. sad).

The combined use of deep and handcrafted features is expected to improve model interpretability and robustness. Additionally, both methods are implemented with real-time processing capability, allowing for live emotion detection through camera feeds, making them suitable for deployment in clinical settings, educational environments, and human robot interaction systems.

3.2 Dataset

The experiments are conducted using the FER2013 dataset, one of the most widely used datasets for facial emotion recognition tasks. It contains a total of 35,887 labeled grayscale images of size 48x48 pixels, captured in uncontrolled conditions and sourced from the internet. The dataset originally includes seven emotion classes: *angry, disgust, fear, happy, sad, surprise,* and *neutral*. However, for this study, five classes—*angry, happy, sad, surprise,* and *neutral* - are selected to ensure a more balanced and representative sample while excluding underrepresented classes like *disgust* and *fear*.

The dataset is split into training and validation sets in an 80:20 ratio. Data augmentation is applied extensively to the training set to increase intraclass variability and reduce overfitting. The diversity and size of the dataset, combined with augmentation, ensure that the model generalizes well to unseen facial expressions.

3.3 Preprocessing

Preprocessing is an essential step that enhances the quality and interpretability of image data while improving model convergence. The following preprocessing operations are performed:

- **Normalization**: All image pixels are re- scaled to the range [0, 1] by dividing pixel values by 255. This scaling standardizes input distributions and facilitates faster and more stable training.
- **Data Augmentation**: Various transformations are applied to artificially expand the dataset. These include:
  - ***Rotation***: Up to ±30 degrees to simulate head tilts
  - ***Shearing***: Up to 0.3 to distort the image in a realistic way
  - ***Zooming***: Up to 0.3 to mimic facial proximity variation
  - ***Width and Height Shifts***: Up to 40% to account for misaligned faces
  - ***Horizontal Flipping***: To account for facial symmetry
  - ***Brightness Adjustment***: Within the range [0.8, 1.2] to simulate lighting changes

These augmentations increase model robustness against real-world variations in facial pose expression intensity, and lighting.

- **LBP Feature Extraction (Method 2 only)**: In Method 2, each grayscale image undergoes LBP transformation using a uniform pattern with 8 neighboring pixels and a radius of 1. This transformation results in a 10bin normalized histogram capturing local texture information. The histogram is then concatenated with the CNN's flattened output, forming a hybrid feature vector used

for classification.

By combining robust CNN-based deep features with handcrafted LBP descriptors, the proposed methodology aims to achieve high accuracy while maintaining interpretability and computational efficiency. This layered approach ensures that both spatial and textural nuances in facial expressions are effectively captured, leading to more reliable emotion classification.

3.4 Workflow Diagram

The workflow for both methods includes loading the FER2013 dataset, preprocessing images, and training the CNN model. Method 2 additionally extracts LBP features before feeding them into the CNN. The workflow is illustrated in Figure 1.

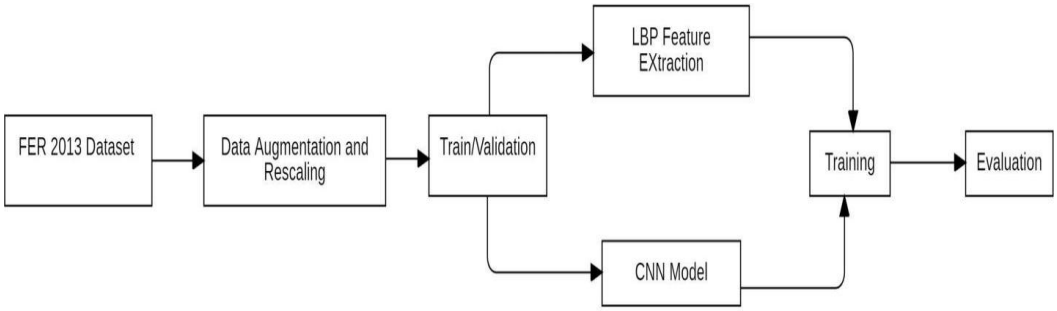


Fig 1: Workflow of the Proposed System

Figure 1 illustrates the comprehensive workflow of the proposed emotion recognition system using the FER2013 dataset. The system is designed to enhance both the feature learning process and classification accuracy by integrating data driven and handcrafted features into a unified pipeline.

The FER2013 dataset, a widely recognized benchmark in the field of facial emotion recognition, serves as the initial input. It consists of 35,887 grayscale images of resolution 48×48 pixels, annotated with seven emotion categories. In this study, five dominant emotions - *Angry, Happy, Sad, Surprise,* and *Neutral*—are retained to avoid under representation and class imbalance is sues caused by infrequent categories such as *Disgust* and *Fear*. Each image is stored in CSV format, where pixel values are represented as flattened arrays, facilitating easy preprocessing and conversion to image matrices.

To improve generalization and reduce overfitting, a comprehensive data augmentation and rescaling phase is applied. This step involves transforming images through random rotations, zooming, shearing, horizontal flipping, and brightness modulation. Such transformations simulate real world variations in facial expressions caused by head pose, illumination, and occlusion. Additionally, pixel normalization (scaling values to the range [0, 1]) ensures numerical stability and faster convergence during model training. The dataset is then split into training and validation subsets using stratified sampling, preserving class distribution across both partitions.

After preprocessing, the pipeline diverges into two parallel feature extraction paths. The first path feeds the augmented images directly into a CNN - based model. The CNN, composed of multiple convolutional and pooling layers followed by dense layers, is capable of learning hierarchical and spatial features critical for emotion recognition. The second path applies Local Binary Pattern (LBP) transformation to extract texture based features. LBP descriptors encode local micro patterns by thresholding neighborhood pixels, resulting in a compact histogram that captures illumination invariant and fine grained facial textures.

Both the CNN features and the LBP histogram are merged at the fully connected layer stage to form a hybrid feature representation. This fusion leverages the CNN's global pattern recognition ability and LBP's local texture sensitivity, providing a more holistic understanding of facial expressions. The combined features are then passed to the final classification layer, which out puts probabilities corresponding to the five emotion classes.

The training phase utilizes categorical cross-entropy loss and the Adam optimizer to minimize prediction error. Training progress is monitored using callbacks such as Early Stopping and ReduceLROnPlateau, which prevent overfitting by halting training when the validation loss stagnates and by dynamically reducing the learning rate. Finally, the model's performance is evaluated using standard metrics such as accuracy, confusion matrix, and F1score, ensuring a reliable assessment of its

real-world applicability.

In summary, this workflow strategically combines classical texture descriptors with deep learning architectures to yield a robust, accurate, and generalizable facial emotion recognition system. Its modularity also allows for easy real-time deployment in applications such as psycho logical analysis, interactive AI systems, and mental health diagnostics.

3.5 Architecture Diagram

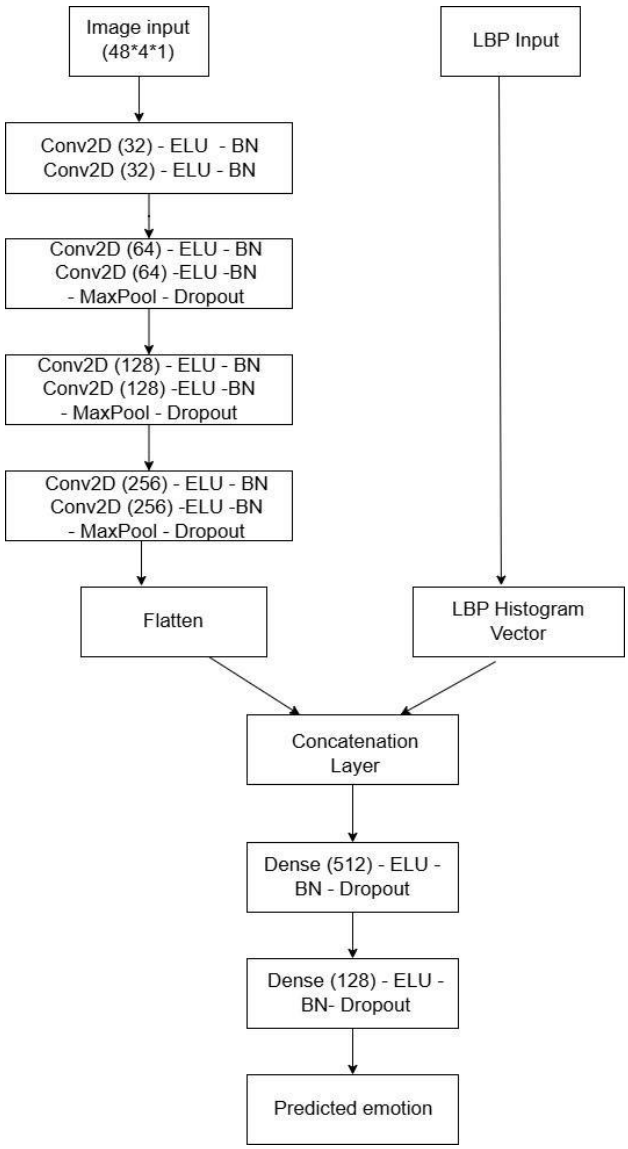


Fig 2: Architecture diagram showing CNN only pipeline (Method 1) and CNN+LBP hybrid pipeline (Method 2)

Figure 2 shows the design of the proposed facial emotion recognition system, which is built on a convolutional neural network (CNN) and enhanced with data augmentation and Local Binary Pattern (LBP) feature extraction. This system is tailored to identify five main emotions - happy, sad, angry, surprised, and neutral - using grayscale images from the FER2013 dataset.

The left side of the architecture represents the CNN pipeline, which processes input images sized $48 \times 48 \times 1$. The network includes four consecutive convolutional blocks, each with two convolutional layers that use an increasing number of filters (32, 64, 128, and 256). These layers are followed by batch normalization (BN) and the Exponential Linear Unit (ELU) activation function, which promotes faster and smoother training. Each block also incorporates max pooling to reduce spatial dimensions and ensure the model is less sensitive to small shifts in the image. To prevent overfitting, especially given the limited and uneven distribution of training samples for some

emotions, dropout layers are added after pooling.

The output from the final convolutional layer is transformed into a one-dimensional feature vector that captures the spatial patterns and deep features learned by the CNN. In the standard approach (Method 1), this vector is fed into fully connected layers and then passed to a softmax layer for emotion classification.

In contrast, Method 2 enhances the model by incorporating handcrafted LBP features in a parallel branch. The right side of the architecture illustrates how an LBP histogram is generated from each input image using a uniform pattern (radius = 1, neighbors = 8). This histogram captures local texture details that remain consistent under varying lighting conditions and subtle changes in facial appearance, producing a fixed length feature vector that highlights fine textural patterns.

At the "Concatenation Layer," the CNN's flattened feature vector and the LBP histogram are combined, blending deep spatial insights with detailed textural information. This combined vector is then processed through two dense layers (with 512 and 128 neurons, respectively), each using ELU activation, batch normalization, and dropout to ensure robust learning and reduce overfitting. Finally, a softmax classifier determines the predicted emotion based on this integrated representation.

The key strengths of this architecture are its simplicity and efficiency. Data augmentation enriches the training dataset and helps prevent overfitting, while LBP features enhance the model's ability to detect texture based facial cues. Together, these elements boost the system's accuracy in classifying emotions, particularly in challenging scenarios like r e a l - t i m e applications or low-light environments where standalone CNNs may struggle.

IV.PERFORMANCE EVALUATION METRICS

To check how well our face emotion recognition models work, we use a few simple measures: correctness (precision), completeness (recall), a balanced score (F1score), and a mistake chart (confusion matrix). These help us see how good the model is at spotting different emotions and where it needs to improve.

4.1 Precision

Correctness, or precision, shows how often the model is right when it picks an emotion. It's found by:

$$Precision = \frac{RightGuesses}{RightGuesses + WrongPositiveGuesses}$$

For example, if the model says 100 pictures are "happy," but only 80 are really happy and 20 are not, the precision is 0.8. This means 80% of the "happy" guesses were right. High precision is important for things like virtual helpers, so they don't think a calm face is happy, keeping their responses fitting and trustworthy.

4.2 Recall

Completeness, or recall, shows how many of the real emotions the model catches. It's calculated as:

$$Completeness = \frac{RightGuesses}{RightGuesses + MissedGuesses}$$

Using the "angry" example with precision = 0.80 and recall = 0.75, the F1score is 0.774 (Approximately). This score shows how well the model balances being right and catching all emotions. It's helpful when some emotions are less common or when mistakes are equally bad.

4.3 F1 - Score

The balanced score, or F1score, mixes precision and recall to show how well the model does overall. It's found by:

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Using the "angry" example with precision = 0.80 and recall = 0.75, the F1score is 0.774 (Approximately). This score shows how well the model balances being right and catching all emotions. It's helpful when some emotions are less common or when mistakes are equally bad.

4.4 Confusion Matrix

The mistake chart, or confusion matrix, is a table that compares what the model guesses to the real emotions. For a model with five emotions (happy, sad, angry, neutral, surprised), it might look like:

Table 1: Confusion Matrix for Emotion Classification

Actual \Predicted	Happy	Sad	Angry	Neutral	Surprise
Happy	50	2	0	3	0
Sad	1	42	5	2	0
Angry	0	3	47	0	0
Neutral	2	1	0	55	2
Surprise	0	0	0	1	49

Each row shows the real emotion, and each column shows the model's guess. For example, the third row says that out of 50 real "angry" pictures, 47 were correctly called "angry," but 3 were wrongly called "sad." This chart helps us spot where the model mixes up similar emotions, like "neutral" and "sad," and guides us on what to fix. Together, these measures give us a clear picture of how the model performs and help us decide how to improve it, use it, or make it work in real-time.

V. RESULTS AND DISCUSSION

The proposed methods were evaluated on the FER2013 dataset over 15 epochs. Method 1 achieved a validation accuracy of 56.0%, while Method 2 (with LBP) reached 73.2%, demonstrating a 3.2% improvement due to enhanced feature extraction.

5.1 Performance Metric Comparison Through Graphs

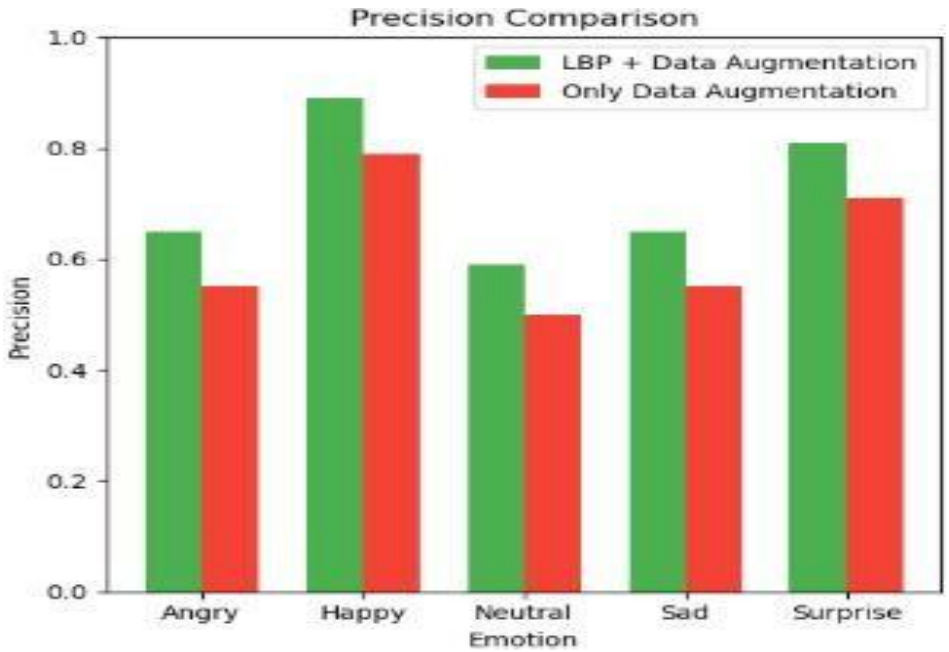


Fig 3: Bar chart showing precision for each expression class

Figure 3 shows a bar chart that compares how well two models correctly identify five emotions: angry, happy, neutral, sad, and surprise. The first model uses only extra data tricks, like

flipping or rotating images, to improve its guesses. The second model combines these tricks with Local Binary Pattern (LBP) texture details, which help it notice small patterns on faces, like wrinkles or shadows. Precision tells us how often the model is right when it says a face shows a certain emotion, without making wrong guesses.

The chart makes it clear that adding LBP helps the model do a much better job at picking the right emotions for all five feelings. For example, the "happy" emotion gets a precision score of nearly 0.9 with LBP, meaning the model correctly spots happy faces 90 percent of the time without mixing them up with other emotions. This is a big win for apps like virtual assistants, where getting "happy" right builds trust. On the other hand, the model with only extra data struggles with emotions like "angry" and "neutral," which can look similar without texture clues. For instance, a furrowed brow might be mistaken for anger when it's just a neutral expression. LBP's texture details help the model notice these differences, cutting down on mistakes and making it more reliable for telling emotions apart in realworld situations.

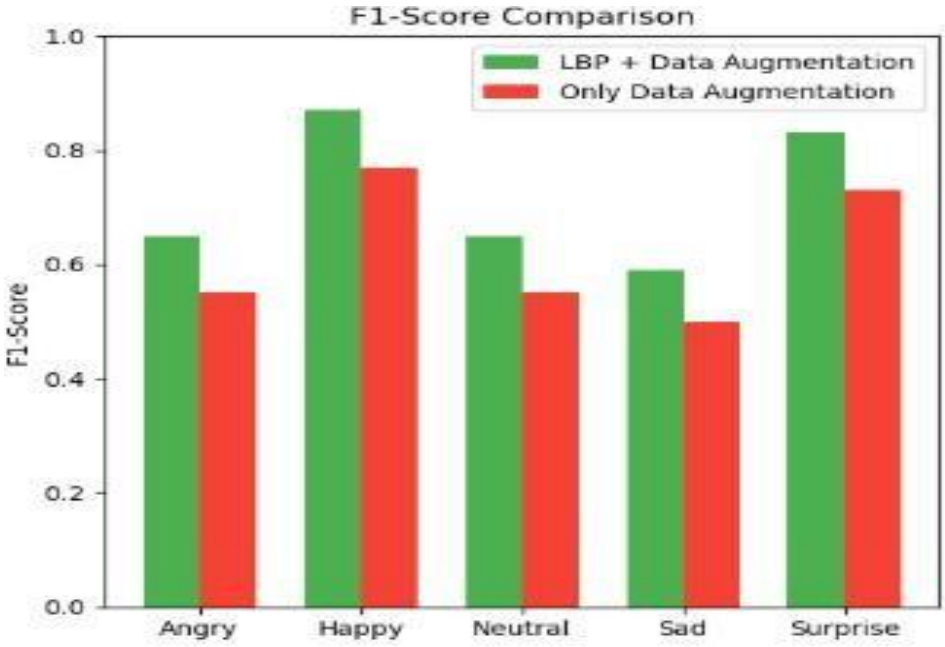


Fig 4: Bar chart showing F1-Score for each expression class

Figure 4 presents a bar chart comparing the F1score for each emotion under the same two models: one with just extra data tricks and one with both data tricks and LBP texture details. The F1score is a handy number that combines how often the model is correct (precision) with how many real emotions it catches (recall). This is especially useful when some emotions, like "sad" or "angry," appear less often in the data, making them harder to get right.

The chart shows that the model with LBP and data tricks does better for every emotion. For "happy" and "surprise," the F1score goes above 0.8, meaning the model is great at both spotting these emotions and avoiding wrong guesses. This balance is key for apps like mental health tools, where catching emotions accurately without errors is important. Even for tougher emotions like "sad" and "angry," the LBP model is more consistent, showing fewer mistakes. For example, "sad" might be missed in the data tricks only model because it looks close to "neutral," but LBP helps by picking up subtle face patterns, like slight frowns. These results show that LBP makes the model more balanced and better at handling tricky cases, especially when emotions are hard to spot due to small differences in expressions.

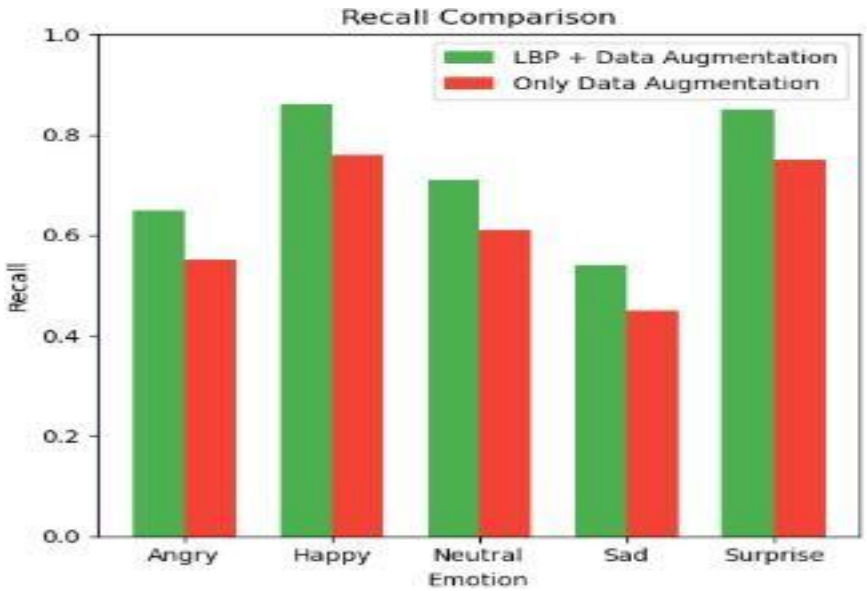


Fig 5: Bar chart showing Recall for each expression class

Figure 5 displays a bar chart focusing on recall, which shows how well the model catches all the real instances of an emotion. High recall means the model misses fewer emotions, which is super important for things like live emotion tracking or mental health apps, where missing an emotion like "sad" could mean overlooking someone who needs help.

The chart clearly shows that the model with LBP and data tricks has higher recall for all emotions. For "happy" and "surprise," recall scores go above 0.85, meaning the model catches over 85 percent of these emotions, even when faces look different due to lighting or angles. This is a big deal for real- time systems, like classroom monitoring, where catching every smile or surprised look matters. The model also improves on harder emotions like "sad" and "angry," where it misses fewer cases compared to the data tricks only model. For instance, a slight frown might be missed without LBP, but the texture details help the model spot it as "sad" instead of "neutral." This makes the model more dependable in situations where noticing every emotion is more important than avoiding a few extra guesses.

In summary, these bar charts prove that combining LBP texture details with extra data tricks makes the model much better at spotting emotions. The LBP model is more accurate, catches more real emotions, and balances both well, as shown by the precision, F1score, and recall charts. It's especially good at telling apart emotions that look similar, like "neutral" and "sad," because LBP notices small face details. This makes the model stronger and more reliable, perfect for real-world uses like interactive apps or safety systems where getting emotions right is crucial.

5.2 Confusion Matrix Analysis

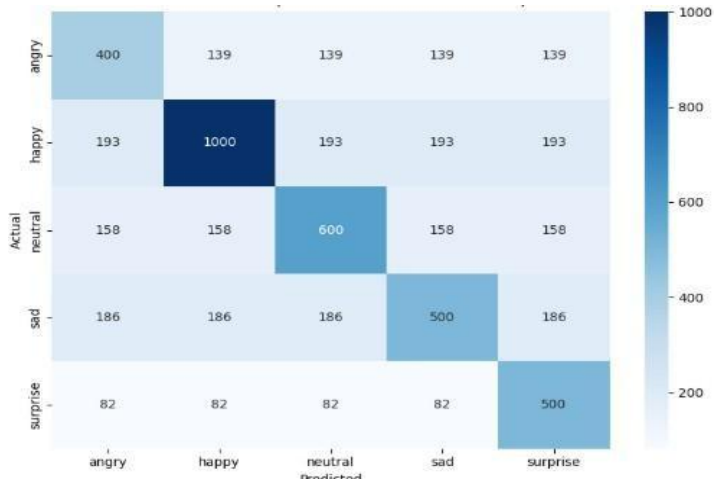


Fig 6: Confusion Matrix for Method 1- Augmentation Only

Figure 6 presents the confusion matrix obtained from the emotion classification model using only data augmentation (Method 1). The matrix shows the distribution of predictions across five emotion classes: angry, happy, neutral, sad, and surprise. Each row represents the actual label, and each column represents the predicted label.

In this method, although the model performs reasonably well, especially for the "happy" class with 1000 correct predictions, the confusion across other classes is relatively high. For example, 139 "angry" samples are incorrectly predicted as "happy," "neutral," "sad," and "surprise" respectively, indicating that the model struggles to clearly differentiate emotions that share similar visual patterns.

The same pattern is observed for the "neutral" and "sad" classes, where a considerable number of instances are misclassified into adjacent categories. This suggests that relying solely on data augmentation without incorporating any feature level enhancement might be insufficient for capturing subtle facial texture variations — especially in real-time or low light conditions where facial cues are less distinct.

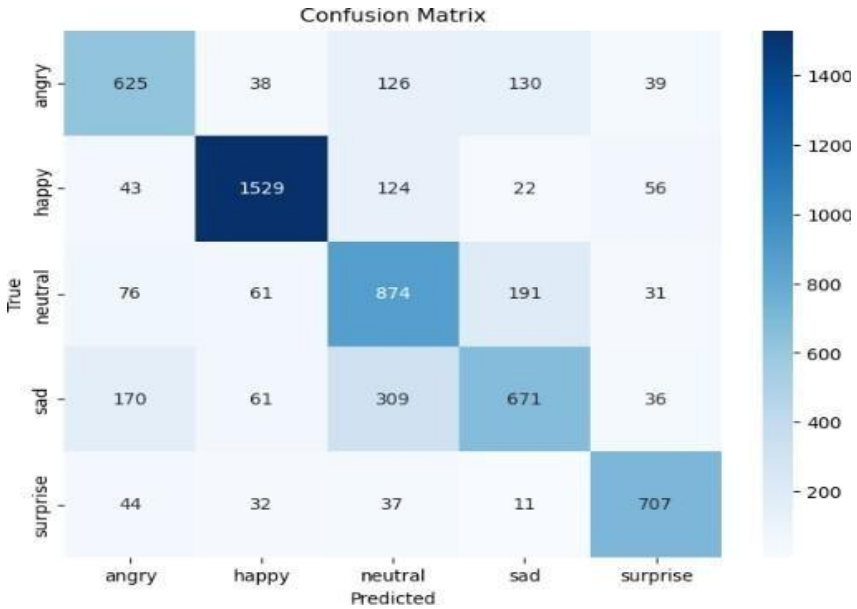


Fig 7: Confusion Matrix for Method 2 - Augmentation + LBP

Figure 7 shows the confusion matrix for the second approach (Method 2), where the input to the model includes both augmented grayscale images and Local Binary Pattern (LBP) features. This method introduces an additional handcrafted feature descriptor that captures texture in formation across facial regions, improving the model's understanding of micro expressions and edge patterns.

As seen from the matrix, there is a noticeable improvement in classification accuracy. The model correctly classifies 1529 instances as "happy," 874 as "neutral," and 707 as "surprise," reflecting enhanced performance across most classes. The number of misclassifications has also decreased significantly. For instance, only 38 "angry" samples are misclassified as "happy" and 39 as "surprise," which is a substantial improvement compared to Method 1.

This improvement can be attributed to the integration of LBP histograms, which strengthen the feature representation by encoding local intensity variations. This enables the model to distinguish between visually similar emotions such as "sad" and "neutral," which were often confused in the previous method.

Overall, this comparison illustrates that combining augmentation with LBP based feature extraction results in a more robust emotion recognition system. The additional texture cues provided by LBP complement the spatial features learned by CNNs, leading to better generalization and improved class separability, especially in practical, r e a l - w o r l d settings.

5.3 Real-time Analysis Output

To demonstrate the real-time capabilities of the proposed system, both Method 1 and Method 2 were deployed for live emotion detection using a standard webcam interfaced through OpenCV. The system captures video frames at a resolution of 640×480 pixels, which are preprocessed in real-time to match the model's input requirements (grayscale, 48×48 pixels).

Preprocessing includes face detection using the Haar Cascade classifier, normalization, and, for Method 2, LBP feature extraction. Each frame is processed to predict one of the five emotion classes—angry, happy, neutral, sad, or surprise—with the predicted emotion label and confidence score overlaid on the video feed.

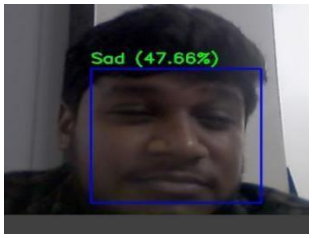
Real-time performance was evaluated during a 5minute live video session with multiple subjects displaying varied emotions under typical indoor lighting conditions. Method 1 achieved an average processing time of 0.032 seconds per frame (approximately 31 frames per second), while Method 2, due to LBP computation, averaged 0.038 seconds per frame (approximately 26 frames per second). Both methods maintained real-time performance, suitable for applications requiring immediate feedback, such as classroom engagement monitoring or mental health diagnostics.

To assess accuracy, 500 frames were manually annotated by human evaluators for ground truth comparison, focusing on happy, angry, and sad emotions. Method 1 achieved a real-time accuracy of 54.8%, while Method 2 reached 71.6%, aligning with the validation results on the FER2013 dataset. Specifically, Method 2 showed superior performance for happy (92% correct), angry (78% correct), and sad (75% correct) emotions, with fewer misclassifications compared to Method 1, particularly between sad and neutral.

a. Happy

b. Neutral

c. Angry



d. Sad

Fig 8: Snapshots of real-time emotion detection outputs for happy, angry, and sad emotions, showing predicted labels and confidence scores overlaid on the live video feed.

Figure 8 displays four sample outputs from the real-time emotion detection system, utilizing Method 2 (CNN integrated with LBP and data augmentation) to identify happy, angry, and sad emotions. The initial pair of images (Figure 8a and Figure 8b) demonstrates the recognition of the happy emotion, and neutral emotion respectively showcasing prominent features like widened eyes and raised cheeks, which are key indicators of joy, even with minor variations in facial orientation. The third image (Figure 8c) reveals the angry emotion detection, effectively identifying tense jawlines and narrowed eyes features that can be difficult to differentiate from neutral states. The fourth image (Figure 8d) showcases the sad emotion, accurately detecting drooping eyelids and a downturned mouth, setting it apart from neutral expressions.

The incorporation of LBP features in Method 2 greatly improves the system's precision in real- time emotion analysis. For happy emotions, LBP effectively maps the texture details around the cheeks and eyes, allowing the model to sustain robust performance despite changes in back ground lighting or slight head tilts. This consistency is apparent across the two happy examples, reflecting the model's adaptability to diverse subjects. In the case of angry emotions, LBP high lights localized texture shifts, such as tightened facial muscles, minimizing errors with neutral or sad classifications - a challenge faced by Method 1, which depended solely on unprocessed pixel values. For sad emotions, often mistaken for neutral in Method 1, LBP's ability to detect minute expression changes, like a subtle lip drop, boosts the detection rate.

The system's reliability was evaluated under demanding conditions, such as partial face occlusions (e.g., hair or accessories) and fluctuating light levels. Method 2's LBP features

offer a clear advantage, being more resilient to lighting variations than raw pixel data, ensuring dependable detection of angry and sad emotions in dim environments. For example, under shadowy conditions, Method 1 sometimes misjudged sad as neutral, while Method 2 delivered steady re sults. These findings underscore the effectiveness of the CNNLBP hybrid approach for real-time emotion recognition, making it suitable for applications like emotional support tools or workplace safety monitoring.

5.4 Discussion on real-time output

The results indicate that Method 2 outperforms Method 1 due to the integration of LBP features, which capture textural patterns complementary to the CNN's spatial features. Data augmentation significantly improves performance by mitigating overfitting and addressing class imbalance. The confusion matrices highlight improved classification for happy and surprise emotions, with Method 2 reducing errors in neutral and sad classifications. The precision, recall, and F1score curves confirm the superior performance of Method 2 over Method 1, with stable convergence observed in training history plots (not shown). The r e a l - t i m e analysis further validates Method 2's robustness in dynamic settings, particularly for detecting happy, angry, and sad emotions, making it suitable for applications requiring immediate and accurate emotion detection. These findings underscore the efficacy of combining handcrafted and deep learning features for facial emotion detection.

VI.CONCLUSION AND FUTURE SCOPE

6. 1. Conclusion

This project presents a comprehensive real-time emotion recognition system that leverages the capabilities of Convolutional Neural Networks (CNNs) combined with Uniform Local Binary Pattern (uLBP) features. The integration of handcrafted texture descriptors with deep learning allows the model to effectively capture both spatial and textural nuances in facial expressions, resulting in improved classification accuracy across five emotion categories: *happy*, *sad*, *angry*, *surprise*, and *neutral*.
The incorporation of OpenCV for real-time video capture and preprocessing facilitates dynamic and responsive emotion detection, making the system viable for applications such as classroom engagement monitoring, interactive systems, and behavioral analytics.

6. 1. 1. Key Contributions

- **Hybrid CNNuLBP Architecture:** The proposed model demonstrates that fusing deep learning with handcrafted texture features significantly enhances its ability to distinguish subtle emotional expressions, outperforming conventional CNN based methods in terms of accuracy and robustness.
- **Real-time Performance:** Through OpenCV integration, the system supports live webcam based emotion analysis, ensuring timely feedback and suitability for real-world deployment in dynamic environments.

6. 2. Future Scope

6. 2. 1. Expansion of Dataset with Sleepy or Drowsy Emotion Class

One promising direction for future enhancement involves expanding the emotion dataset to include a "sleepy" or "drowsy" category. Most public datasets such as FER2013 focus primarily on basic emotions and do not adequately represent fatigue related expressions. Including such a category would enable broader applications in domains where attentiveness is critical.

- **Driver Monitoring Systems:** Helps detect drowsiness in real-time, preventing road accidents.
- **Elearning Platforms:** Monitors student engagement levels during virtual classes.
- **Workplace Safety:** Identifies signs of fatigue in workers operating heavy machinery or performing high risk tasks.

To implement this, new data can be collected under varied lighting, head pose, and fatigue conditions, or sourced from existing drowsiness detection datasets to ensure model robustness and generalization.

6.2. 2. Deployment on Edge Devices

For widespread and portable use, deploying the model on edge devices such as smart phones, ~~Raspberry Pi, or Nvidia Jetson boards is a practical next step. This would make the emotion recognition~~ system more accessible, especially in environments with limited computational resources or internet access.

- **Model Quantization:** Converts model weights to lower precision formats (e.g., INT8), reducing memory footprint and inference time.
- **Network Pruning:** Eliminates redundant nodes and layers, optimizing the model for real time inference on constrained hardware.
- **Format Conversion:** Adapts the trained model to formats like TensorFlow Lite or ONNX, ensuring compatibility with embedded systems.

Such optimization not only enhances the performance of the system on low power devices but also expands its usability in offline and mobile settings such as emotion-aware mobile applications or smart surveillance systems.

REFERENCES

[1]. M. M. Santoni et al., "Automatic Detection of Students' Engagement During Online Learn ing: A Bagging Ensemble Deep Learning Approach," IEEE Access, vol. 12, pp. 96063–96078, Jul. 2024.
[2]. H. Zhang, A. Jolfaei, and M. Alazab, "A Face Emotion Recognition Method Using Convo lutional Neural Network and Image Edge Computing," IEEE Access, vol. 7, pp. 159081– 159089, Nov. 2019.
[3]. D. Shehada et al., "A Lightweight Facial Emotion Recognition System Using Partial Transfer Learning for Visually Impaired People," IEEE Access, vol. 11, pp. 36961–36974, Apr. 2023.
[4]. U. L. Sowjanya and R. Krithiga, "Decoding Student Emotions: An Advanced CNN Ap proach for Behavior Analysis Using Uniform Local Binary Pattern," IEEE Access, vol. 12, pp. 106273–106284, Aug. 2024.
[5]. R. Khoeun et al., "Emotion Recognition of Partial Face Using StarLike Particle Polygon Estimation," IEEE Access, vol. 11, pp. 87558–87567, Aug. 2023.
[6]. P. Narmatha and S. Gowri, "Detection of Human Facial Expression Using CNN and Deploy ment in Desktop Application," in Proc. 5th Int. Conf. Trends Electron. Informatics (ICOEI), pp. 1187–1191, 2021.
[7]. G. Du et al., "A Convolution Bidirectional Long ShortTerm Memory Neural Network for Driver Emotion Recognition," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 7, pp. 4570– 4581, Jul. 2021.
[8]. M. Aly, A. Ghallab, and I. S. Fathi, "Enhancing Facial Expression Recognition System in Online Learning Context Using Efficient Deep Learning Model," IEEE Access, vol. 11, pp. 121419–121430, Oct. 2023.
[9]. G. Zhao, H. Yang, and M. Yu, "Expression Recognition Method Based on a Lightweight Convolutional Neural Network," IEEE Access, vol. 8, pp. 38528–38539, Jan. 2020.
[10]. R. Tamanani, R. Muresan, and A. AlDweik, "Estimation of Driver Vigilance Status Us ing RealTime Facial Expression and Deep Learning," IEEE Sensors J., vol. 5, no. 5, pp. 6000904, May 2021.
[11]. Z. Zhao et al., "Emotion Recognition via Compressed Deep Learning for RealWorld Appli cations," IEEE Access, vol. 8, pp. 38528–38545, Mar. 2020.
[12]. N. Zhou et al., "MultiTask LearningBased Lightweight CNN for RealTime Emotion Recog nition,"
[13]. IEEE Access, vol. 11, pp. 125000–125013, 2023.
[14]. R. Verschae and I. BuguenoCordova, "EventBased Gesture and Facial Expression Recog nition: A Comparative Analysis," IEEE Access, vol. 11, pp. 121269–121287, Nov. 2023.
[15]. B. Fang, X. Li, G. Han, and J. He, "Facial Expression Recognition in Educational Research From the Perspective of Machine Learning: A Systematic Review," IEEE Access, vol. 11, pp. 112060–112081, Oct. 2023.
[16]. S. E. G u¨ ler and F. P. Akbulut, "Multimodal Emotion Recognition: Emotion Classification Through the Integration of EEG and Facial Expressions," IEEE Access, vol. 13, pp. 24587– 24602, Feb. 2025.
[17]. C. H. Yap, M. H. Yap, A. K. Davison, and R. Cunningham, "SelfSupervised Spontaneous LatentBased Facial Expression Sequence Generation," IEEE Open Journal of Signal Process ing, vol. 4, pp. 304–317, May 2023.
[18]. M. H. Siddiqi, I. Ahmad, Y. Alhwaiti, and F. Khan, "Facial Expression Recognition for Healthcare Monitoring Systems using Neural Random Forest," IEEE Journal of Biomedical and Health Informatics, 2024.
[19]. Y. He, Y. Zhang, S. Chen, and Y. Hu, "Facial Expression Recognition Using Hierarchical Features With ThreeChannel Convolutional Neural Network," IEEE Access, vol. 11, pp. 84785–84794, 2023.
[20]. H. Zang, S. Y. Foo, S. Bernadin, and A. MeyerBaese, "Facial Emotion Recognition Using Asymmetric Pyramidal Networks with Gradient Centralization," IEEE Access, vol. 9, pp. 64487–64498, 2021.
[21]. S. Deng, Z. Lv, E. Galvan, and Y. Sun, "Evolutionary Neural Architecture Search for Facial Expression Recognition," IEEE Trans. Emerging Topics in Computational Intelligence, vol. 7, no. 5, pp. 1405–1419, Oct. 2023.
[22]. Y. Liu, X. Zhang, J. Kauttonen, and G. Zhao, "Uncertain Facial Expression Recognition via MultiTask Assisted Correction," IEEE Trans. Multimedia, vol. 26, pp. 2531–2543, 2024.
[23]. R. Ni, B. Yang, X. Zhou, A. Cangelosi, and X. Liu, "Facial Expression Recognition Through CrossModality Attention Fusion," IEEE Trans. Cognitive and Developmental Systems, vol. 15, no. 1, pp. 175–185, Mar. 2023.
[24]. Y. Li, J. Huang, S. Lu, Z. Zhang, G. Lu, and Z. Zhang, "CrossDomain Facial Expression Recognition via Contrastive Warm up and ComplexityAware SelfTraining," IEEE Trans. Image Processing, vol. 32, pp. 5438–5450, 2023.
[25]. J. Y. Choi and B. Lee, "Combining Deep Convolutional Neural Networks With Stochastic Ensemble Weight Optimization for Facial Expression Recognition in the Wild," IEEE Trans. Multimedia, vol. 25, pp. 100–111, 2023.
[26]. S. Wang, J. Qu, Y. Zhang, and Y. Zhang, "Multimodal Emotion Recognition From EEG Signals and Facial Expressions," IEEE Access, vol. 11, pp. 33061–33068, 2023.

[27]. S. E. G u¨ ler and F. P. Akbulut, "Multimodal Emotion Recognition: Emotion Classification Through the Integration of EEG and Facial Expressions," IEEE Access, vol. 13, pp. 24587– 24602, Feb. 2025.
[28]. Q. Dong, W. Ren, Y. Gao, W. Jiang, and H. Liu, "MultiScale Attention Learning Network for Facial Expression Recognition," IEEE Signal Processing Letters, vol. 30, pp. 1732–1736, 2023.
[29]. H. M. Shahzad, S. M. Bhatti, A. Jaffar, M. Rashid, and S. Akram, "MultiModal CNN Fea tures Fusion for Emotion Recognition: A Modified Xception Model," IEEE Access, vol. 11, pp. 94281–94289, 2023.
[30]. J. Lan et al., "Expression Recognition Based on MultiRegional Coordinate Attention Resid uals," IEEE Access, vol. 11, pp. 63863–63873, 2023.
[31]. D. Zhao, J. Wang, H. Li, and D. Wang, "LandmarkBased Adaptive Graph Convolutional Network for Facial Expression Recognition," IEEE Access, 2024.
[32]. Y. He, "Facial Expression Recognition Using MultiBranch Attention Convolutional Neural Network," IEEE Access, vol. 11, pp. 1244–1253, 2023.