International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Performance Evaluation of Facial Expression Recognition Using CNN and DRLBP

Dr. Divya Shree
Assistant Professor, Department of computer Science, Tika Ram Girls College Sonipat

ABSTRACT- Facial expression recognition is a fundamental problem in computer vision with wide-ranging applications in emotion analysis, human–computer interaction and affective computing. This study proposes an effective facial expression recognition framework that integrates a Convolutional Neural Network (CNN) classifier with the Dynamic Regional Local Binary Pattern (DRLBP) feature extraction technique to enhance recognition accuracy and robustness. The proposed approach follows a structured methodology, beginning with the acquisition of a diverse dataset comprising facial images representing multiple emotional expressions along with their corresponding class labels. Subsequently, discriminative facial features are extracted using the DRLBP algorithm. Unlike conventional Local Binary Pattern (LBP) methods, DRLBP dynamically identifies and analyzes expressive regions of the face based on facial landmarks or key facial points, thereby improving the representation of emotion-related facial variations. The extracted features are then supplied to the CNN model, which efficiently learns hierarchical and high-level representations for classification. The performance of the proposed system is evaluated using standard metrics such as accuracy, precision, recall and F1-score. By effectively combining the powerful feature-learning capability of CNNs with the region-adaptive feature extraction strength of DRLBP, the proposed method achieves reliable and robust facial expression recognition, demonstrating strong potential for real-world applications.

KEYWORDS -DRLBP convolutional neural network (CNN), facial expression recognition

I INTRODUCTION

Face recognition is an ability that humans perform naturally and effortlessly in everyday life. The capability to recognize familiar individuals and distinguish them from strangers plays a crucial role in social interaction and cooperation, as highlighted by Robert Axelrod, who emphasized its importance in the formation of cooperative behavior [1]. Over the past decade, rapid advancements in computing technology have led to the emergence of a pervasive computing environment, where powerful yet affordable computational systems are embedded in mobile devices, automobiles, medical equipment and numerous other aspects of daily life. This technological evolution has generated significant interest in the automatic analysis of digital images and videos for a wide range of applications, including biometric authentication, surveillance systems, human–computer interaction and multimedia content management. Consequently, automatic face recognition has become a prominent research area within computer vision.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Face recognition is fundamentally a visual pattern recognition problem in which a three-dimensional human face must be identified using its two-dimensional image representation. Although substantial progress has been achieved in recent years due to improved facial models and increased computational power, most face recognition systems still perform reliably only under controlled conditions. In real-world scenarios, facial images are affected by several challenging factors such as illumination variations, head pose changes and facial expressions. Among these factors, facial expression poses the greatest challenge because it alters the actual three-dimensional structure of the face, whereas other factors primarily affect imaging conditions. To mitigate the influence of expression-related variations, it becomes necessary to first identify or estimate the facial expression present in an image, a process referred to as Facial Expression Recognition (FER).

Beyond its role in enhancing face recognition systems, facial expression recognition is vital because facial expressions themselves serve as a powerful and natural medium of communication. Expressions are non-intrusive and intuitive and studies have shown that they often convey more information than spoken words or vocal tone [2]. As a result, accurate recognition of facial expressions is essential for developing more intuitive and human-friendly human–computer interaction systems.

The human face is a distinctive and expressive feature that conveys both identity and emotional state. Facial expressions communicate subtle yet meaningful signals related to an individual's emotions, intentions and psychological condition. Expressions such as a smile indicating happiness, a frown reflecting sadness or disapproval, widened eyes expressing surprise, or a curled lip denoting disgust exemplify the wide spectrum of human emotions. These expressions arise from coordinated movements of facial muscles, which collectively transmit emotional information to observers. Ekman and Friesen [11] proposed six fundamental emotions happiness, sadness, anger, fear, surprise and disgust from which all other complex expressions can be derived. These prototypical expressions possess distinct facial patterns and are considered universal across different cultures and ethnic groups.

Fig. 1. Six prototypical expressions

II RELATED WORK

In recent years, facial feature detection and landmark-based analysis have become increasingly important in a wide range of practical applications, including surveillance systems, crime analysis and age estimation. This paper proposes an efficient facial expression recognition framework that integrates multiple landmark detectors, local feature transformation techniques and a supervised classification model. The proposed system follows a four-stage architecture. First, face detection is performed using skin color segmentation combined with ellipse fitting to accurately localize facial regions. Second, facial landmarks are identified and plotted on key facial components. Third, discriminative features are extracted using Euclidean distance measurements, Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) to capture both geometric and textural information. Finally, a Support Vector Machine (SVM) classifier is employed to categorize facial expressions into six fundamental emotion classes: Neutral, Happy, Sad, Anger, Disgust and Surprise. The effectiveness of the proposed method is evaluated on two benchmark datasets, namely the MMI Facial Expression Dataset and the Chicago Face Dataset, achieving recognition accuracies of 80.8% and 83.01%, respectively. Experimental results demonstrate that the proposed approach outperforms existing state-of-the-art facial expression recognition techniques in terms of accuracy. Owing to its robustness and efficiency, the proposed system is well-suited for diverse real-world applications, including online business negotiations, consumer behavior analysis, e-learning environments and virtual reality systems [12].

Shintaro Kondo et.al. (2022) To enhance the human-like behavior of dialogue agents, this study explores an approach for generating facial emotional expressions that enable agents to visually convey their emotional states. Building upon a previously proposed model, the present work introduces two key improvements. First, video sequences are used as input data instead of static images, allowing the model to capture temporal dynamics of facial expressions more effectively. Second, the frame rate of the generated output is increased to improve the smoothness and visual quality of the emotional expressions. Both enhancements are informed by findings from prior research. Furthermore, the proposed framework enables the generation of realistic facial images for emotionally expressive speech videos by integrating the expression-point video produced by the model with an existing facial expression synthesis system based on real facial photographs. This approach contributes to more natural and expressive visual communication in dialogue agents. [13]

III PROPOSED APPROACH

Facial expression recognition using a Convolutional Neural Network (CNN) classifier combined with the Dynamic Regional Local Binary Pattern (DRLBP) method leverages the strengths of deep learning and robust feature extraction to achieve accurate and efficient emotion recognition. The methodology begins with the collection of a diverse facial image dataset containing multiple expressions, where each image is annotated with its corresponding emotion label to serve as ground truth for training and evaluation. Subsequently, discriminative features are extracted from the facial images using the DRLBP algorithm. DRLBP computes local binary patterns by

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

comparing the intensity of each pixel with that of its neighboring pixels within selected regions. Unlike conventional LBP techniques, DRLBP dynamically identifies regions of interest based on facial landmarks or key facial points, enabling more effective capture of expression-sensitive facial areas. This adaptive regional analysis enhances the representation of facial variations associated with different emotional states.

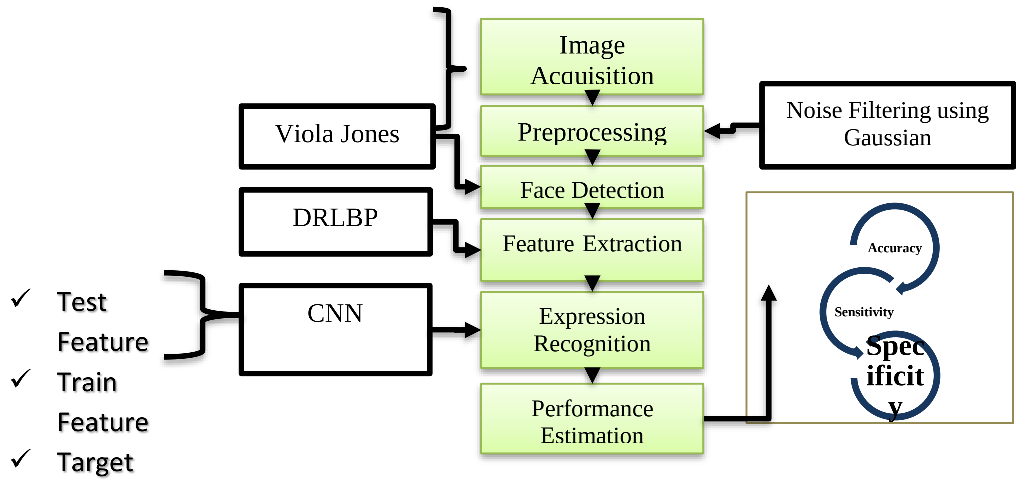

Fig.2 Proposed flow Diagram

After extracting the DRLBP features, they are provided as input to a Convolutional Neural Network (CNN) classifier for training. The CNN architecture comprises multiple layers, including convolutional layers for deep feature learning and pooling layers for spatial down-sampling. These layers enable the network to learn hierarchical representations of facial expressions by automatically identifying salient patterns and discriminative features. During the training phase, the CNN model learns to associate the extracted DRLBP features with their corresponding emotion or expression labels using a labeled training dataset. This learning process involves optimizing the network's weights and biases through backpropagation and gradient descent, with the objective of minimizing classification error. Once trained, the CNN classifier can be employed for inference on previously unseen facial images. For testing, DRLBP features are extracted from the input image and passed through the trained CNN to predict the corresponding emotion or facial expression class. The performance of the proposed approach is evaluated using standard metrics such as accuracy, precision, recall and F1-score by comparing the predicted labels with ground-truth annotations from an independent test dataset. To further improve generalization and reduce overfitting, optimization strategies such as regularization, dropout and learning-rate scheduling can be incorporated.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

The proposed system follows a modular architecture designed to efficiently process facial images and manage recognition outputs. The process begins with the acquisition of facial images from a dataset repository. In the initial stage, the input images are resized and normalized to ensure uniformity. Subsequently, a preprocessing module converts color images into grayscale format and performs face detection. The detected facial region is then isolated from the background, transformed and resized for further analysis. Finally, the feature extraction module generates robust facial representations that are used for classification. This structured architecture ensures effective feature learning and reliable facial expression recognition.

Binary Image: A binary image is a digital image composed of only two possible pixel values, commonly represented as black and white or as 0 and 1. In such images, each pixel exists in one of two states: completely black (0) or completely white (1). Binary images simplify visual information by representing it in a discrete form, making them particularly useful for image analysis tasks. Each pixel in a binary image carries a single unit of information and collectively, the pixels form patterns or structures that represent meaningful visual features within the image.

Grayscale images: Grayscale images, commonly referred to as monochrome or black-and-white images, are digital images in which each pixel is represented by a single intensity value. This intensity typically ranges from 0, corresponding to black, to 255, representing white, with intermediate values indicating varying shades of gray. Unlike color images, grayscale images contain no color information and rely solely on brightness variations to represent visual content.

RGB - RGB (Red, Green, Blue) images are digital images that represent color information using the RGB color model. In this model, each pixel is defined by three separate color channels red, green and blue whose combined intensities produce a wide range of colors visible to the human eye. By varying the contribution of these primary colors, the full color spectrum can be generated. Typically, each color channel is encoded using an 8-bit value ranging from 0 to 255, where 0 denotes the absence of that color and 255 represents its maximum intensity. Different combinations of these intensity values determine the final color and brightness of each pixel in the image.

A specific color in an RGB image is produced by combining the intensity values of the red, green and blue channels. For instance, pure red is represented by a maximum intensity value in the red channel (255) and zero intensity values in both the green and blue channels (0, 0). Similarly, pure green is obtained by assigning a maximum intensity value to the green channel (255) while setting the red and blue channels to zero and pure blue is generated by assigning the maximum intensity value to the blue channel (255) with no contribution from the red and green channels. By adjusting the intensity levels of these three color channels in different combinations, a wide variety of colors can be created.

IV SYSTEM IMPLEMENTATION

- Input Image
- Pre-processing
- Face Detection
- Feature Extraction DRLBP
- Expression Recognition CNN
- Performance Estimation

The facial expression recognition framework generally follows a multi-stage pipeline that includes image preprocessing, face detection, feature extraction using the Dynamic Regional Local Binary Pattern (DRLBP) method, expression classification using a Convolutional Neural Network (CNN) and performance evaluation. The following section provides an overview of each stage.

Input Image: The system takes a facial image as input, which may be acquired directly through a camera or retrieved from an existing dataset. The input image may contain a single face or multiple faces. During the preprocessing stage, the image is subjected to various enhancement operations to improve its quality and ensure consistency for subsequent analysis. These operations typically include resizing to a standard dimension, normalization to adjust pixel intensity values, noise reduction to eliminate unwanted artifacts and illumination normalization to reduce the effects of varying lighting conditions. Preprocessing plays a crucial role in improving the reliability and accuracy of the facial expression recognition process.

Face Detection: Face detection is carried out to identify and extract facial regions from the preprocessed image. Various face detection techniques, including traditional methods such as Viola–Jones and Haar cascade classifiers, as well as modern deep learning–based approaches like Single Shot Detector (SSD) and Faster R-CNN, can be utilized for accurate face localization. After detection, the facial regions are segmented and isolated from the background to enable focused and efficient analysis in subsequent stages.

Feature Extraction DRLBP: Dynamic Regional Local Binary Pattern (DRLBP) is a feature extraction technique designed to capture expressive facial regions and generate discriminative features for facial expression analysis. The DRLBP method computes local binary patterns by comparing the intensity values of pixels within dynamically selected regions centered around facial landmarks or key facial points. These binary patterns effectively encode local texture and structural information, making them particularly useful for representing subtle facial variations associated with different expressions.

Expression Recognition CNN: The extracted DRLBP features are subsequently provided as input to a Convolutional Neural Network (CNN) for facial expression classification. The CNN is trained using a labeled dataset comprising facial images along with their corresponding expression

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

categories. During the training process, the network learns to map the DRLBP feature representations to specific facial expressions, enabling accurate prediction for previously unseen images. The CNN architecture is composed of convolutional layers for feature learning, pooling layers for spatial reduction and fully connected layers for classification, allowing the model to effectively learn hierarchical and discriminative representations of facial expressions.

V RESULT DISCUSSION

The performance of the proposed system is assessed by comparing the predicted facial expression labels with the ground truth annotations of an independent test dataset. Standard evaluation metrics, including accuracy, precision, recall and F1-score, are employed to quantify the effectiveness of the facial expression recognition model. These metrics offer a comprehensive understanding of the system's capability to accurately detect and classify different facial expressions.

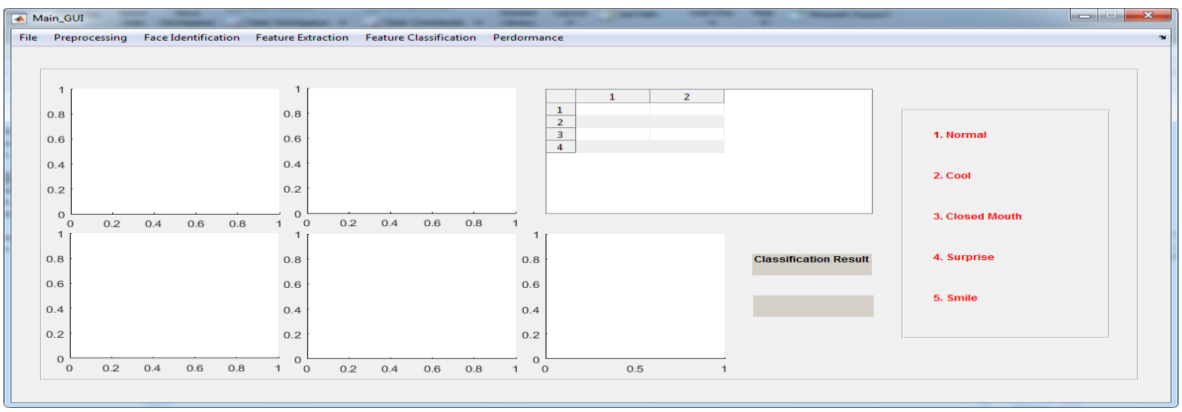

Fig.3 GUI Window

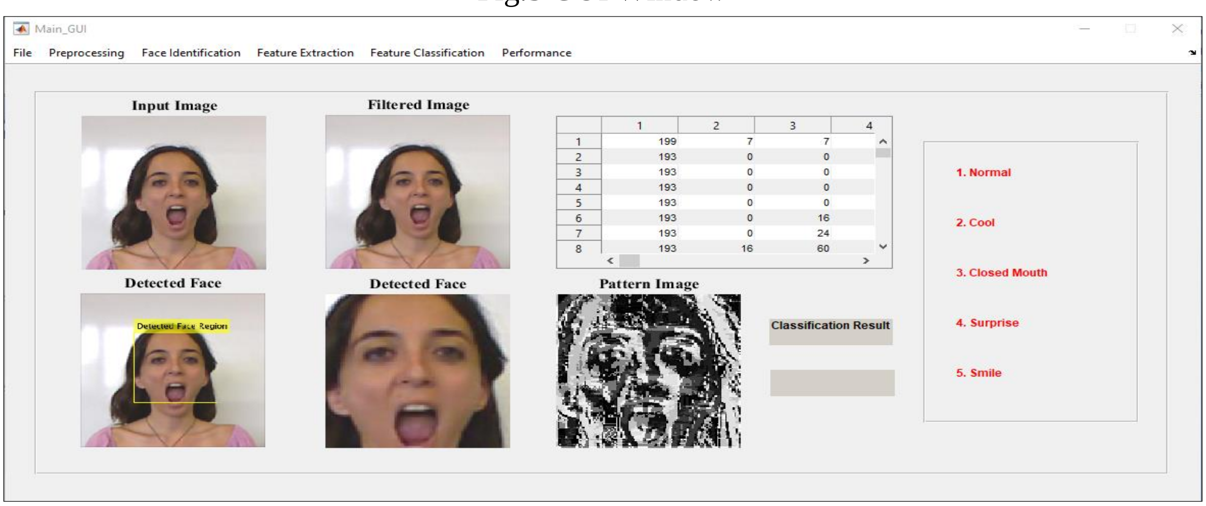

Fig.4 face classification

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

The final component of the confidence measurement system, which follows the classification stage, is responsible for determining the reliability and accuracy of the recognition outcome.

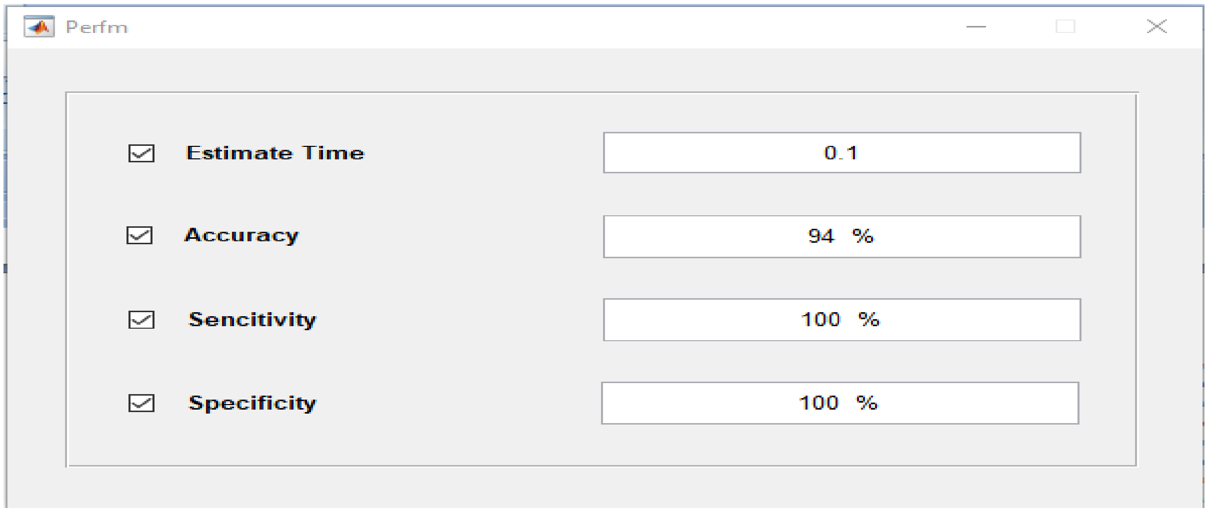

Fig. 5 performance result

VI PERFORMANCE EVOLUTION

To assess the performance of a facial expression recognition system, commonly used evaluation metrics include accuracy, sensitivity and specificity. The following provides a brief description of each metric.

True Positive (TP):

TP denotes the number of samples that are correctly identified as positive, meaning the facial expressions are accurately recognized.

True Negative (TN):

TN refers to the number of samples that are correctly identified as negative, indicating that non-target expressions are accurately rejected.

False Positive (FP):

FP represents the number of samples that are incorrectly identified as positive, where expressions are mistakenly recognized despite being negative.

False Negative (FN):

FN indicates the number of samples that are incorrectly identified as negative, where expressions are mistakenly rejected even though they are positive.

Accuracy: Accuracy quantifies the overall correctness of a system's predictions by comparing the predicted labels with the corresponding ground truth labels. It is computed as the ratio of correctly classified samples to the total number of samples and is typically expressed as a percentage. Accuracy provides a general measure of the system's performance across all facial expression classes.

Accuracy = (Number of Correct Predictions / Total Number of Predictions) * 100

Sensitivity (Recall): Sensitivity, also referred to as recall or the true positive rate, evaluates the system's ability to correctly identify positive instances, such as specific facial expressions, from all actual positive samples in the dataset. It is calculated as the ratio of true positive predictions to the combined total of true positives and false negatives.

Sensitivity = (True Positives / (True Positives + False Negatives)) * 100

Sensitivity is especially important in applications where accurate detection of positive instances is crucial, such as recognizing specific facial expressions associated with distinct emotional states. In contrast, specificity assesses the system's ability to correctly identify negative instances, or non-target expressions, from all actual negative samples in the dataset. It is computed as the ratio of true negative predictions to the combined total of true negatives and false positives.

Specificity = (True Negatives / (True Negatives + False Positives)) * 10

Specificity is particularly important in situations where the correct identification of non-expressive states is essential, such as accurately distinguishing neutral or non-emotional facial states from specific emotional expressions.

VII CONCLUSIONS

In conclusion, facial expression recognition based on the integration of a Convolutional Neural Network (CNN) classifier with the Dynamic Regional Local Binary Pattern (DRLBP) feature extraction technique demonstrates strong potential for accurately identifying and classifying facial expressions. The CNN model effectively learns complex patterns and hierarchical representations from facial data, while DRLBP contributes by capturing discriminative local texture and structural information from expressive facial regions. The complementary strengths of these two approaches enhance the overall robustness and reliability of the recognition system. By adopting a systematic methodology comprising image preprocessing, face detection, DRLBP-based feature extraction, CNN-based expression classification and performance evaluation, an efficient and robust facial expression recognition framework can be developed. Such a system holds significant potential for improving applications in human–computer interaction, emotion analysis, healthcare monitoring, security systems and other domains where accurate interpretation of human emotions is essential. Despite these advantages, several challenges remain, including variations in facial expressions, sensitivity to environmental conditions, limited availability of labeled training data, generalization to unseen individuals, real-time processing constraints and ethical and privacy considerations. Addressing these issues is crucial for developing facial expression recognition systems that are not only accurate and robust but also socially responsible and ethically compliant. With ongoing research and technological advancements, further improvements in recognition accuracy, real-time performance and adaptability to diverse real-world scenarios are expected. The continued integration of deep learning models such as CNNs with advanced feature extraction methods like

DRLBP offers promising opportunities for deeper understanding of human emotions, ultimately leading to more natural, intelligent and effective human–machine interaction across a wide range of practical applications.

5.2 Future Enhancement

The future scope of facial expression recognition based on the integration of a CNN classifier and the Dynamic Regional Local Binary Pattern (DRLBP) method is highly promising. Several potential directions for further development and advancement are outlined below.

Improved Accuracy and Robustness: Continuous research efforts are focused on enhancing the accuracy and robustness of facial expression recognition systems. These efforts include improvements in CNN architectures, advanced feature extraction techniques such as DRLBP and the incorporation of other deep learning methods. The primary objective is to strengthen the system's ability to effectively handle variations in facial expressions, changing illumination conditions, occlusions and diverse environmental factors.

REFERENCES

1. P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.
2. P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.
3. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. 511–518.
4. T. Ojala, M. Pietikäinen and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
5. C. Shan, S. Gong and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.
6. Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
7. A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
8. I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

9. S. L. Happy and A. Routray, “Automatic facial expression recognition using features of salient facial patches,” *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 1–12, 2015.
10. Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu, “Comparison between geometry-based and Gabor-wavelets-based facial expression recognition,” in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 1998, pp. 454–459.
11. M. Pantic and L. J. M. Rothkrantz, “Automatic analysis of facial expressions: The state of the art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.
12. S. A. Rizwan, “An accurate facial expression detector using multi-landmarks selection and local transform,” *IEEE Access*, vol. 8, pp. 9781–9790, 2020.
13. S. Kondo, S. Harata, T. Sakuma and S. Kato, “Facial expressions generating model reflecting agent’s emotion response using facial landmark residual networks,” in *Proc. IEEE 11th Global Conf. on Consumer Electronics*, 2022, pp. 1–6.
14. I. J. Goodfellow et al., “Generative adversarial nets,” in *Proc. Conf. on Neural Information Processing Systems*, 2014, pp. 2672–2680.
15. Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, “A survey of affect recognition methods: Audio, visual and spontaneous expressions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.