



PERFORMANCE OF FACIAL EXPRESSION RECOGNITION USING CNN AND DRLBP

Ojesvi Kumar* & Rahul Kaushik**

* PG Scholar, Computer Science and Engineering Department, Bhiwani Institute of Technology & Sciences, Bhiwani, Haryana

** Assistant Professor, Computer Science Engineering Department, Bhiwani Institute of Technology & Sciences, Bhiwani, Haryana

Cite This Article: Ojesvi Kumar & Rahul Kaushik, "Performance of Facial Expression Recognition Using CNN and DRLBP", International Journal of Advanced Trends in Engineering and Technology, Volume 8, Issue 1, Page Number 46-53, 2023.

Abstract:

Facial expression recognition is a crucial task in computer vision with applications in fields such as emotion detection, human-computer interaction, and affective computing. This research presents a novel approach that combines a convolutional neural network (CNN) classifier with the Dynamic Regional Local Binary Patterns (DRLBP) proposed method for accurate and efficient facial expression recognition. The proposed method follows a systematic process that begins with the collection of a diverse dataset of facial images displaying different expressions, along with their corresponding labels. The DRLBP algorithm is then utilized to extract discriminative features from the facial images. Unlike traditional Local Binary Patterns (LBPs), the DRLBP algorithm dynamically selects regions of interest based on facial landmarks or key points, allowing for effective capturing of expressive facial regions. The performance of the proposed method can be evaluated using various metrics such as accuracy, precision, recall, or F1 score, by leveraging the strengths of both CNNs and the DRLBP algorithm, the combined approach achieves accurate and robust facial expression recognition. CNNs excel in learning complex hierarchical representations, while DRLBP effectively captures expressive facial regions. The integration of these two methods presents promising prospects for the development of advanced facial expression recognition systems with numerous practical applications.

Key Words: DRLBP Convolutional Neural Network (CNN), Facial Expression Recognition

Introduction:

Face recognition is a task that humans perform routinely and effortlessly in their daily lives. Robert Axelrod has also shown the ability to recognize that they have met before and distinguish them from strangers is one of the bases for humans to form cooperation [1]. The last decade has witnessed a trend towards an increasingly ubiquitous computing environment, where powerful and low-cost computing systems are being integrated into mobile phones, cars, medical instruments and almost every aspect of our lives. This has created an enormous interest in automatic processing of digital images and videos in a number of applications, including biometric authentication, surveillance, human-computer interaction, and multimedia management. Research and development in automatic face recognition follows naturally. Face recognition is a visual pattern recognition problem where a three-dimensional object is to be identified based on its two-dimensional image. In recent years, significant progress has been made in this area; owing to better face models and more powerful computers, face recognition system can achieve good results under constrained situations. However because face images are influenced by several factors: illumination, head pose, expression and so on, in general conditions, face recognition is still challenging. From a computer vision point of view, among all these "noises" facial expression maybe the toughest one in the sense that expressions actually change the three-dimensional object while other factors, such as illumination and position, only affect imaging parameters. To get rid of expression "noise", one first needs to estimate the expression of an image; this is called "Facial Expression Recognition". Another, maybe more important motivation of facial expression recognition is that expression itself is an efficient way of communication: it's natural, non-intrusive, and [2] has shown that, surprisingly, expression conveys more information than spoken words and voice tone. To build a friendlier Human Computer Interface, expression recognition is essential.



Anger Surprise Sadness Disgust Fear Happy

Figure 1: Six Prototypical Expressions

Face Expressions - The face is a unique and important feature of human beings, conveying identity and emotion. The face can send many subtle signals. A facial expression is worth a thousand words. Human face as a whole signifies a lot regarding their mood, and expressions are vibrant features which commune the speakers emotions, thoughts, intentions and so on. For example, an array of facial expressions a smile of happiness, a frown of sadness or disapproval, wide-open eyes of surprise, or lips curled in disgust - all shows a wide range of emotions. Facial expressions are the gesture executed as a result of the movement of the facial muscles and these movements express the emotional status of a person to the observer. Elman and Friesen [11] postulated six primary emotions whose combination fabricates every other composite facial expression and this typical expressions are shown in figure 1 These prototypical emotions possess a distinguishing content together with a unique facial expression and they appear to be universal across the ethnicity of human beings.

Related Work:

In the past few years, facial features detection and landmarks analysis plays a vital role in several practical application such as surveillance system, crime detector and age estimation. In this paper, we proposed a novel approach of recognizing facial expressions based on multi landmark detectors, local transform features and recognizer classifier. The proposed system is divided into four stages. (a) Face detection using skin color segmentation and ellipse fitting, (b) Plotting landmarks on facial features, (c) Feature extraction using euclidean distance, HOG and LBP. While, (d) SVM classification learner is used to classify six basic facial expressions like Neutral, Happy, Sad, Anger, Disgust, and Surprise. The proposed method is applied on two facial expression datasets i-e. MMI facial expressions dataset and Chicago Face dataset and achieved accuracy rates of 80.8% and 83.01%, respectively. The proposed system outperforms the state-of-the-art facial expression recognition system in terms of recognition accuracy. The proposed system should be applicable to different consumer application domains such as online business negotiations, consumer behavior analysis, E-learning environments, and virtual reality practices [12]

Shintaro Kondo et.al. (2022) So that dialogue agents look more human, we've been looking into a way to make a model for making facial emotions that dialogue agents can use to show how they feel. In this work, we change the model from the previous study by using video as the input data and increasing the frame rate of the results to improve the quality of the results. Research from before was used to make both of these changes. We can also make facial pictures for emotional speech videos by putting the video of the expression points made by the model into the model for making facial expression videos from real photos. [13]

Proposed Approach:

Facial expression recognition using a CNN classifier and DRLBP (Dynamic Regional Local Binary Patterns) proposed method combines the power of convolutional neural networks (CNNs) with the descriptive capabilities of the DRLBP algorithm to achieve accurate and efficient facial expression recognition. The process begins with data collection, where a diverse dataset of facial images displaying different expressions is obtained. Each image is labeled with the corresponding emotion or expression, forming the ground truth for training and evaluation. Next, the DRLBP algorithm is applied to extract discriminative features from the facial images. DRLBP calculates local binary patterns (LBPs) by comparing the intensity values of each pixel with its neighbors in a specific region. Unlike traditional LBP, DRLBP dynamically selects regions of interest based on facial landmarks or key points, allowing it to capture expressive facial regions effectively.

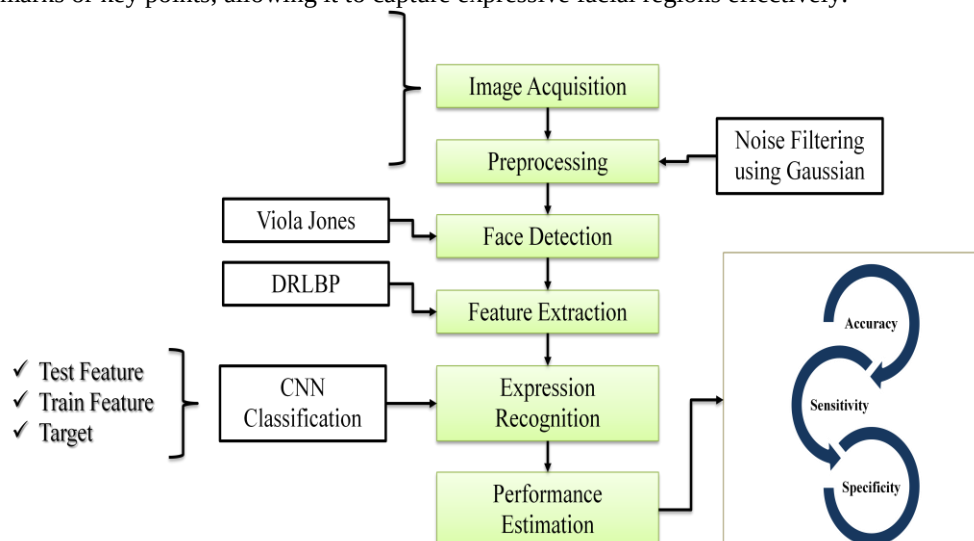


Figure 2: Proposed flow Diagram

Once the DRLBP features are extracted, they are fed into a CNN classifier for training. The CNN architecture consists of multiple layers, including convolutional layers for feature extraction and pooling layers

for spatial down sampling. These layers enable the network to learn hierarchical representations of facial expressions, automatically identifying relevant patterns and features.

During the training phase, the CNN classifier learns to map the extracted DRLBP features to the corresponding emotion or expression labels using a labeled training dataset. This process involves optimizing the network's weights and biases through backpropagation and gradient descent algorithms, minimizing the classification error. After training, the CNN classifier can be used for inference on new, unseen facial images. The DRLBP algorithm is applied to extract features from the test image, and these features are then fed into the trained CNN for prediction. The output of the classifier represents the predicted emotion or expression label for the input image. The performance of the proposed method can be evaluated using various metrics, such as accuracy, precision, recall, or F1 score, by comparing the predicted labels with the ground truth labels of a separate test dataset. Fine-tuning techniques, such as regularization, dropout, or learning rate scheduling, can be employed to enhance the model's performance and prevent overfitting. The combined approach of utilizing a CNN classifier and the DRLBP proposed method leverages the strengths of both techniques. CNNs excel at learning complex hierarchical representations, while DRLBP effectively captures expressive facial regions. By integrating these two methods, accurate and robust facial expression recognition systems can be developed, with potential applications in emotion detection, human-computer interaction, and affective computing..

The proposed system has architecture. It consists of different modules depending on the connection of the input image, and the facial database shows the method of storing the recognition results. The image obtained from the dataset library, the first module deals with "resize and normalize", and then pre-processes, the pre-processing module converts the color image to its grayscale representation and then performs detection of faces. In the next step, the detected face is further removed from the image. This module can also detect the face of the object in the detected face area and transform and resize the face. The feature extraction module is used to create facial representations.

Binary Image:

A binary image is a type of digital image that consists of only two possible pixel values, typically black and white or 0 and 1. It can be thought of as a black-and-white image where each pixel is either fully black (represented by 0) or fully white (represented by 1).

In a binary image, each pixel represents a single element of information and the image as a whole conveys a specific pattern or structure.

Grayscale Images:

Grayscale images, also known as black-and-white or monochrome images, are digital images where each pixel is represented by a single intensity value, typically ranging from 0 (black) to 255 (white). Unlike color images, grayscale images do not contain color information and are composed solely of shades of gray.

RGB:

RGB (Red, Green, Blue) images are digital images that use the RGB color model to represent and display colors. In an RGB image, each pixel is composed of three color channels: red, green, and blue. The combination of these three primary colors in varying intensities forms the full spectrum of colors that can be perceived by the human eye.

In the RGB color model, each color channel is represented by an 8-bit value, ranging from 0 to 255. A value of 0 indicates the absence of that particular color, resulting in a minimum intensity (black), while a value of 255 represents the maximum intensity (full brightness) of that color channel.

To display a specific color, the RGB values of the corresponding color channels are combined. For example, pure red would have maximum intensity in the red channel (255) and zero intensity in the green and blue channels (0, 0). Pure green would have maximum intensity in the green channel (255) and zero intensity in the red and blue channels (0, 0). Pure blue would have maximum intensity in the blue channel (255) and zero intensity in the red and green channels (0, 0). By varying the intensities of the red, green, and blue channels, all other colors can be formed.

System Implementation:

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

The facial expression recognition pipeline typically consists of several stages, including pre-processing, face detection, feature extraction using DRLBP, expression recognition using CNN, and performance estimation. Here's an overview of each step:

Input Image:

The input to the system is a facial image, which can be captured using a camera or obtained from a dataset. The image may contain one or multiple faces.

Pre-processing:

In this stage, the input image undergoes pre-processing operations to enhance its quality and make it suitable for subsequent analysis. Pre-processing may include operations such as resizing, normalization, denoising, and illumination normalization to improve the overall quality and consistency of the image.

Face Detection:

Face detection is performed to locate and extract facial regions within the pre-processed image. Various face detection algorithms, such as Viola-Jones, Haar cascades, or deep learning-based approaches (e.g., SSD, Faster R-CNN), can be employed to identify and localize faces accurately. Once the faces are detected, they are isolated for further analysis.

Feature Extraction DRLBP:

Dynamic Regional Local Binary Patterns (DRLBP) is a feature extraction technique used to capture expressive facial regions and extract discriminative features. DRLBP calculates local binary patterns by comparing the intensity values of pixels in specific regions around facial landmarks or key points. These patterns encode information about local texture and shape, which are useful for capturing facial expression characteristics.

Expression Recognition CNN:

The extracted DRLBP features are then fed into a Convolutional Neural Network (CNN) for expression recognition. The CNN is trained on a labeled dataset containing facial images and their corresponding expression labels. During training, the CNN learns to associate the extracted DRLBP features with specific expressions, enabling it to predict the expression class for new, unseen images. The CNN architecture consists of convolutional layers, pooling layers, and fully connected layers, which enable it to learn hierarchical representations of the input features.

Result Discussion:

To evaluate the performance of the system, the predicted expression labels are compared against the ground truth labels of a separate test dataset. Performance metrics such as accuracy, precision, recall, or F1 score can be used to measure the effectiveness of the facial expression recognition system. These metrics provide insights into the system's ability to correctly identify and classify facial expressions.

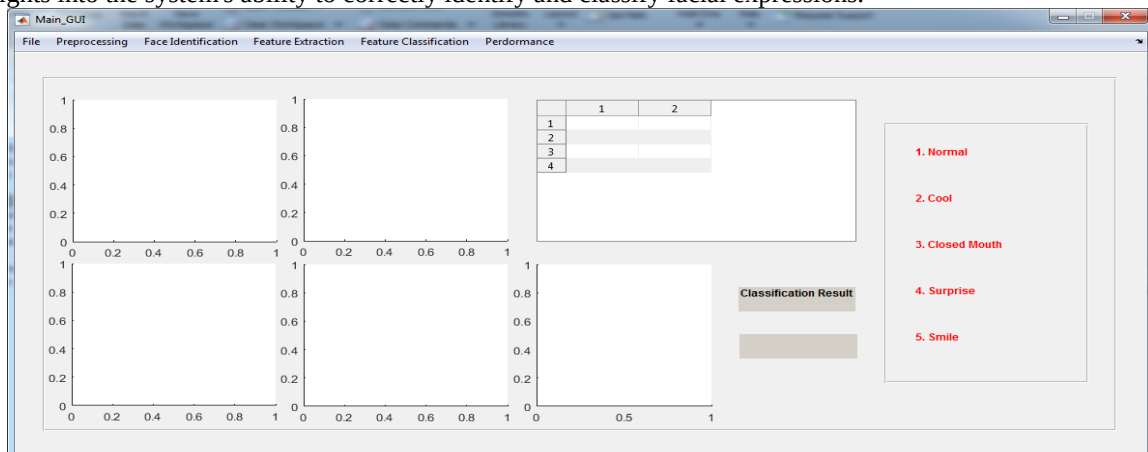


Figure 3: GUI Window

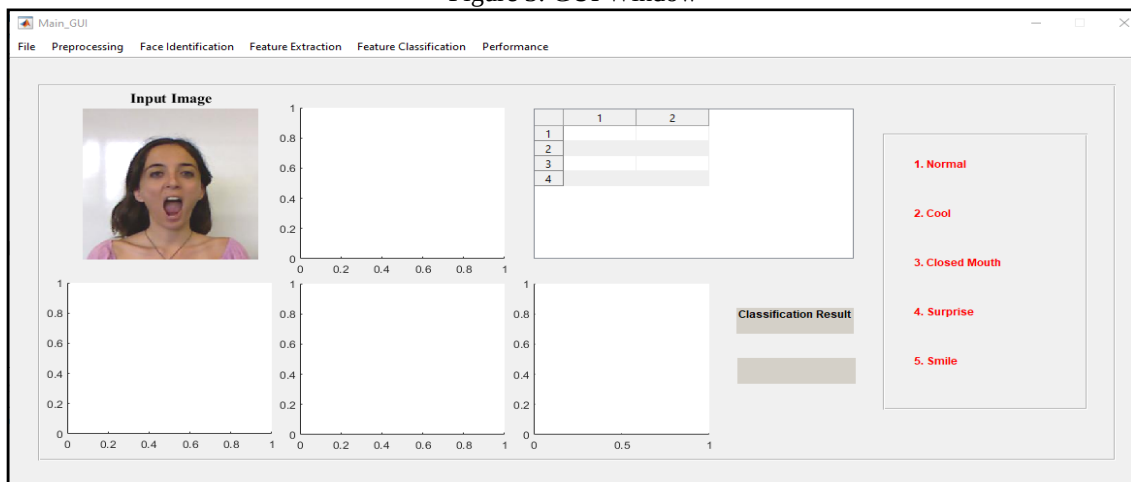


Figure 4: GUI Window input image

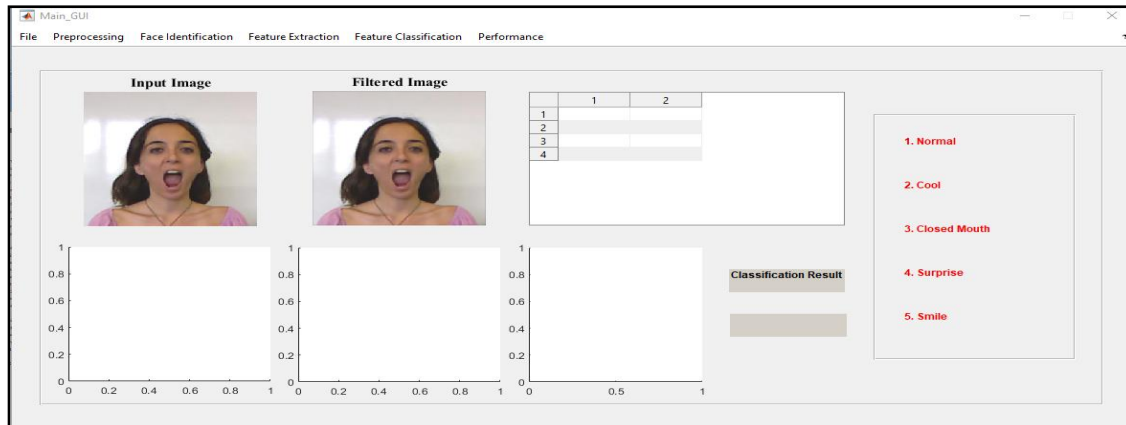


Figure 5: GUI Detected Image

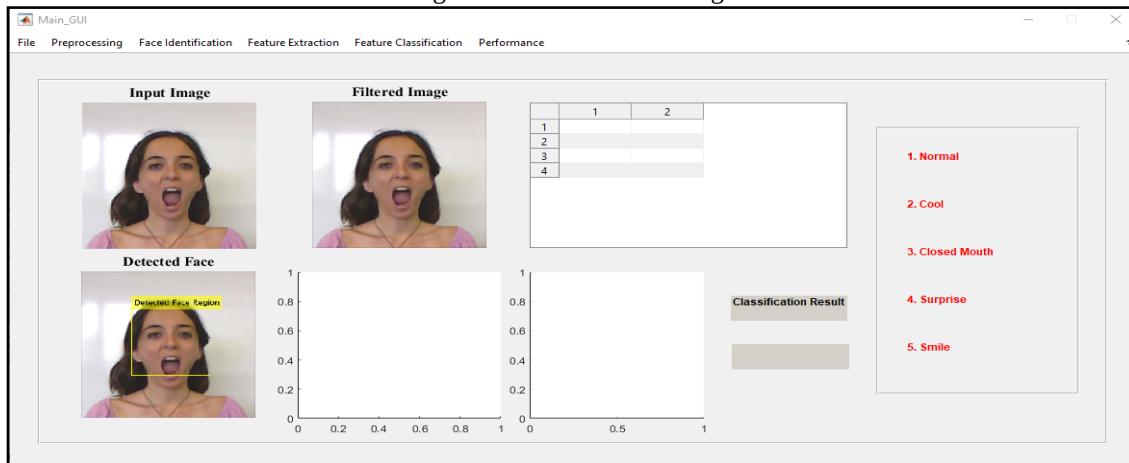


Figure 6: DRLBP Window

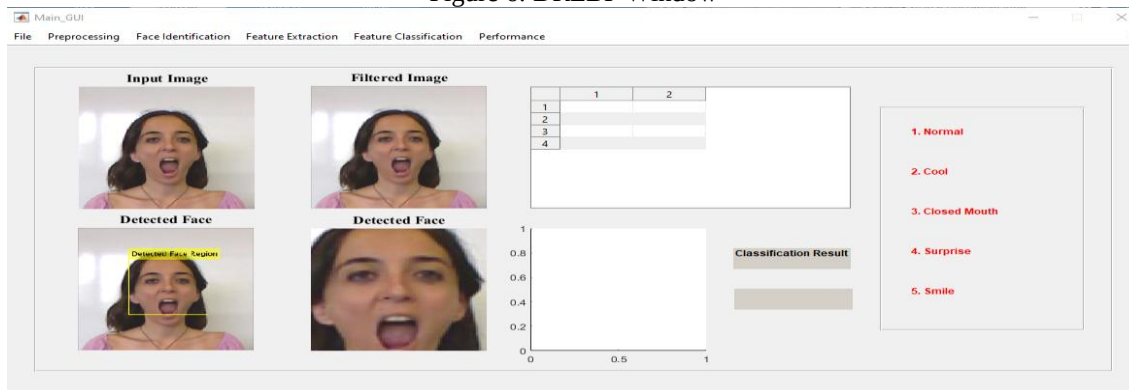


Figure 7: Face Detection Image

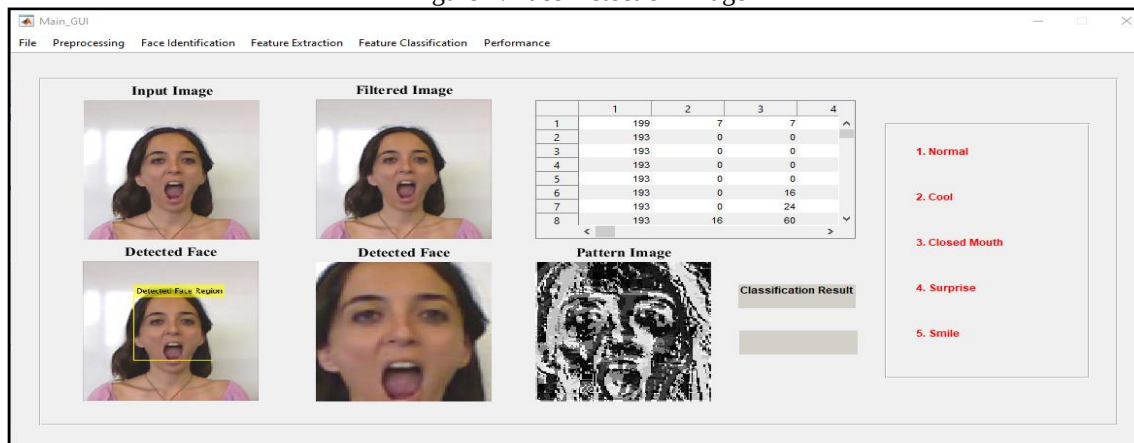


Figure 8: Face Classification

The final part of the confidence measure system, which is followed by the classifier, is in charge of figuring out how accurate the recognition result is.

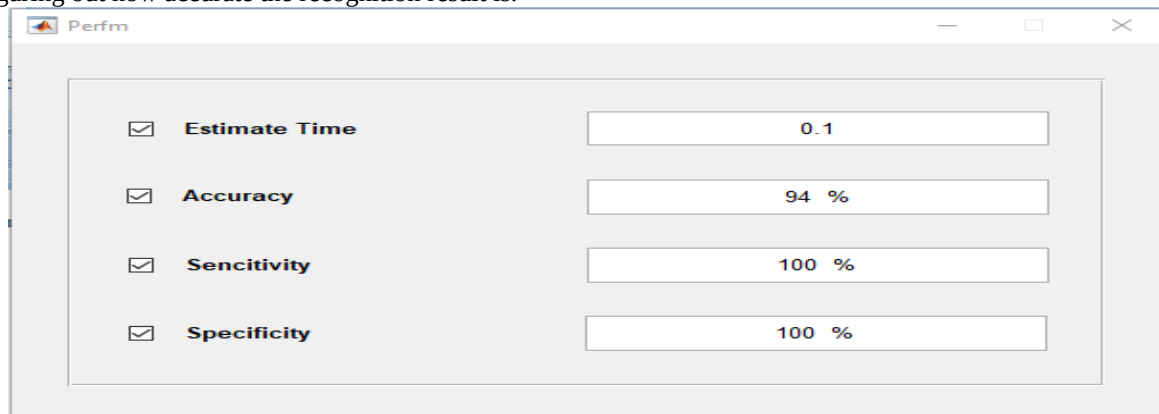


Figure 9: Performance Result

Performance Evolution:

When evaluating the performance of a facial expression recognition system, accuracy, sensitivity, and specificity are commonly used metrics. Here's a brief explanation of each metric:

True Positive (TP):

TP represents the number of samples that are correctly classified as positive (correctly recognized expressions).

True Negative (TN):

TN represents the number of samples that are correctly classified as negative (correctly rejected expressions).

False Positive (FP):

FP represents the number of samples that are incorrectly classified as positive (expressions mistakenly recognized when they are negative).

False Negative (FN):

FN represents the number of samples that are incorrectly classified as negative (expressions mistakenly rejected when they are positive).

Accuracy:

Accuracy measures the overall correctness of the system's predictions by comparing them to the ground truth labels. It is calculated as the ratio of the number of correct predictions to the total number of predictions, expressed as a percentage. Accuracy provides an overall assessment of how well the system performs across all expression classes.

$$\text{Accuracy} = (\text{Number of Correct Predictions} / \text{Total Number of Predictions}) * 100$$

Sensitivity (Recall):

Sensitivity, also known as recall or true positive rate, measures the system's ability to correctly identify positive instances (expressions) out of the total positive instances in the dataset. It is calculated as the ratio of true positive predictions to the sum of true positives and false negatives.

$$\text{Sensitivity} = (\text{True Positives} / (\text{True Positives} + \text{False Negatives})) * 100$$

Sensitivity is particularly relevant in scenarios where correctly detecting positive instances is critical, such as identifying specific expressions associated with particular emotional states.

Specificity:

Specificity measures the system's ability to correctly identify negative instances (non-expressions) out of the total negative instances in the dataset. It is calculated as the ratio of true negative predictions to the sum of true negatives and false positives.

$$\text{Specificity} = (\text{True Negatives} / (\text{True Negatives} + \text{False Positives})) * 100$$

Specificity is important when accurately identifying non-expressions is crucial, such as distinguishing neutral or non-emotional states from specific expressions.

Table 1: Accuracy comparison with existing work

	Classification Technique	Accuracy (%)
Existing System [14]	SVM	83.01
Proposed System	CNN	94.00

Based on the table 1, the existing system utilizes a Support Vector Machine (SVM) classifier and achieves an accuracy of 83.01%. On the other hand, the proposed system incorporates a Convolutional Neural Network (CNN) classifier and achieves a higher accuracy of 94.00%. This comparison suggests that the proposed system with the CNN classifier outperforms the existing system with the SVM classifier in terms of accuracy. The CNN model is known for its ability to effectively extract and learn complex features from images,

which can contribute to improved performance in facial expression recognition tasks. The higher accuracy achieved by the proposed system indicates its capability to better classify and recognize facial expressions compared to the existing system.

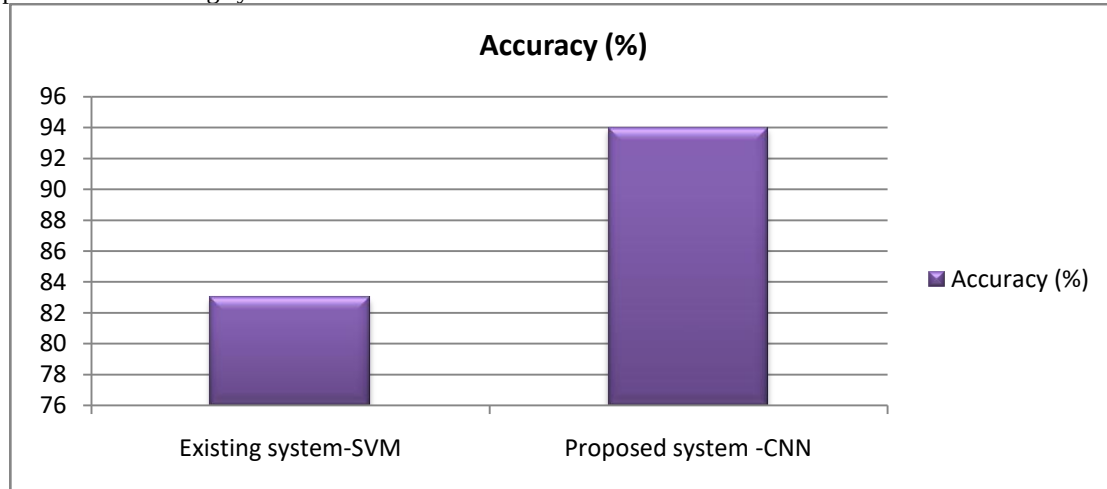


Figure 10: Accuracy comparison with existing work

Conclusions:

In conclusion, facial expression recognition using a combination of CNN classifier and DRLBP (Dynamic Regional Local Binary Patterns) as a feature extraction method offers promising results in accurately identifying and classifying facial expressions. The CNN classifier leverages its ability to learn complex patterns and hierarchical representations from image data, while DRLBP captures local texture and shape information to extract discriminative features.

By following the proposed methodology, which includes pre-processing, face detection, DRLBP-based feature extraction, expression recognition using CNN, and performance estimation, a robust facial expression recognition system can be developed. This system has the potential to enhance human-computer interaction, emotion detection, healthcare, security, and other domains where understanding and responding to human emotions is crucial.

However, it is important to address challenges such as variability in facial expressions, robustness to environmental conditions, limited training data, generalization to unseen individuals, real-time performance, ethical considerations, and privacy concerns. Overcoming these challenges will ensure the development of an accurate, robust, and ethically sound facial expression recognition system that meets the requirements of various applications. As research and advancements in the field of facial expression recognition continue, there is potential for further improvements in accuracy, real-time performance, and applicability to diverse scenarios. The integration of deep learning techniques, such as CNNs, with innovative feature extraction methods like DRLBP opens up new opportunities for capturing and understanding human emotions, leading to enhanced human-machine interaction and a wide range of practical applications.

Future Enhancement:

The future scope of facial expression recognition using CNN classifier and DRLBP is promising. Here are some potential areas of development and advancement:

Improved Accuracy and Robustness:

There is ongoing research to improve the accuracy and robustness of facial expression recognition systems. This includes advancements in CNN architectures, feature extraction techniques like DRLBP, and the integration of other deep learning approaches. The aim is to enhance the system's ability to handle variations in facial expressions, lighting conditions, occlusions, and other environmental factors.

References:

1. Majumder, L. Behera, and V. K. Subramanian, "Emotion recognition from geometric facial features using self-organizing map," *Pattern Recognition*, vol. 47, no. 3, pp. 1282-1293, 2014
2. Y. Tong, R. Chen, J. Yang, and M. Wu, "Robust facial expression recognition based on local tri-directional coding pattern," in *Complex, Intelligent, and Software Intensive Systems - Proceedings of the 12th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS-2018*, 2018, pp. 606-614.
3. H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2983-2991
4. H. Yang, U. A. Ciftci, and L. Yin, "Facial expression recognition by de-expression residue learning," in *Conference on Computer Vision and Pattern Recognition, CVPR*, 2018, pp. 2168-2177.

5. S. Xie, H. Hu, and Y. Wu, "Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition," *Pattern Recognition*, vol. 92, pp. 177-191, 2019
6. A. Majumder, L. Behera, and V. K. Subramanian, "Automatic facial expression recognition system using deep network-based data fusion," *IEEE transactions on cybernetics*, vol. 48, no. 1, pp. 103-114, 2018.
7. S. Xie and H. Hu, "Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks," *IEEE Transactions on Multimedia*, vol. 21, no. 1, pp. 211-220, 2018.
8. Z.-H. Jiang, Q. Wu, K. Chen, and J. Zhang, "Disentangled representation learning for 3d face shape," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 957-11 966.
9. T. Zhang, H. Wang, and Q. Dong, "Deep disentangling siamese network for frontal face synthesis under neutral illumination," *IEEE Signal Processing Letters*, vol. 25, no. 9, pp. 1344-1348, 2018.
10. T. Hinz and S. Wermter, "Image generation and translation with disentangled representations," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1-8.
11. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *Conference on Neural Information Processing Systems*, 2014, pp. 2672-2680.
12. A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint*, vol. arXiv: 1511.06434, 2015.
13. Syeda Amna Rizwan, "An Accurate Facial Expression Detector using Multi -Landmarks Selection and Local Transform 978-1-7281-4235-7/20/\$31.00 ©2020 IEEE
14. Shintaro Kondo; Seiichi Harata; Takuto Sakuma; Shohei Kato acial Expressions Generating Model Reflecting Agent's Emotion Response Using Facial Landmark Residual Networks2022 IEEE 11th Global Conference on Consumer Electronics (GCCE) Year: 2022