ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1662

# Emotional Recognition Based on Faces through Deep Learning Algorithms

Saikat Goswami[1]
Department of Computer Science & Engineering
American International University-Bangladesh
Dhaka, Bangladesh

Md. Tanvir Ahmed Siddiqee[2]
Department of Software Engineering
American International University-Bangladesh
Dhaka, Bangladesh

Khurshedul Barid[3]
Department of Computer Science
American International University-Bangladesh
Dhaka, Bangladesh

Shuvendu Mozumder Pranta[4]
Department of Computer Science & Engineering
World University of Bangladesh
Dhaka, Bangladesh

**Abstract:- Facial expressions have long been a straightforward way for humans to determine emotions, but computer systems find it significantly more difficult to do the same. Emotion recognition from facial expressions, a subfield of social signal processing, is employed in many different circumstances, but is especially useful for human-computer interaction. Many studies have been conducted on automatic emotion recognition, with the majority utilizing machine learning techniques. However, the identification of basic emotions such as fear, sadness, surprise, anger, happiness, and contempt remains a challenging subject in computer vision. Recently, deep learning has gained more attention as potential solutions for a range of real-world problems, such as emotion recognition. In this work, we refined the convolutional neural network method to discern seven basic emotions and assessed several preprocessing approaches to illustrate their impact on CNN performance. The goal of this research is to enhance facial emotions and features by using emotional recognition. Computers may be able to forecast mental states more accurately and respond with more customised answers if they can identify or recognise the facial expressions that elicit human responses. Consequently, we investigate how a convolutional neural network-based deep learning technique may enhance the recognition of emotions from facial features (CNN). Consequently, we investigate how a convolutional neural network-based deep learning technique may enhance the recognition of emotions from facial features (CNN). Our dataset, which comprises of roughly 32,298 pictures for testing and training, includes multiple face expressions. After noise removal from the input image, the pretraining phase helps reveal face detection, including feature extraction. The preprocessing system helps with this.**

***Keywords**:- Facial Expression, Recognition, Classifications, Deep Learning Algorithm.*

## I. INTRODUCTION

These days, emotional awareness is one of the most important and difficult methods. There are several uses for emotion recognition, such as assessing stress levels, blood pressure, and more. The application of joyful, sad, calm, and neutral facial features are among the emotional strategies that can be used to enhance face features. Numerous methods and algorithms are useful in identifying the internal mechanisms of the human body. Emotional recognition is able to instantly identify the thoughts of others. Due to the early diagnosis of diseases through emotional awareness, it shields people against serious infections or illnesses. The primary benefit of emotional awareness is its ability to discern human mindsets without the need for direct inquiry. The video-based detection of the many sorts of emotions without their understanding involves a lot of facial recognition. This research offers two methods for calculating the facial recognition algorithm based on videos: one that measures angles and distances accurately. The other one, which lowers the set of important frames, is the organization of the all-video clips [1]. The suggested paper uses an enhanced deep learning technique that makes use of ECNN to disclose the same seven emotions as the facial acting coding system (FACS). Identifying all possible visible anatomically based face motions was the primary objective of the development of the FACS system. Numerous applications are made possible by the defined nomenclature that FACS provides for the study of face movement. It generates various facial reactions' classification without using the best method, like the seven emotions of the Facial Acting Coding System (FACS). As a result, the accuracy of the suggested system is higher than that of the current method.

## II. OVERVIEW OF PROPOSED ARCHITECTURE AND METHOD

Convolutional neural network analysis is the most popular technique for analyzing images (CNN). CNN contains hidden layers known as convolutional layers, in contrast to a multilayer perceptron (MLP). The two-level CNN architecture serves as the foundation for the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1662

suggested strategy. Background removal is the first step that is recommended because it is used to extract emotions from an image. Here, the standard CNN network module (EV) is used to extract the primary expressional vector. Identifying relevant facial points of significance results in the expressional vector (EV) being created. Variations in EV are tightly correlated with changes in expression. The EV is generated by using a basic perceptron unit on a background-free facial image. We also use a nonconvolutional perceptron layer as the last stage in the suggested FERNEC model. Each convolutional layer receives the input data (or image), alters it, and then sends the output to the layer after it. This transformation makes use of a convolutional procedure. Any of the applied convolutional layers can be utilized to find patterns. A convolutional layer has four filters in total. In addition to the face, the input image that is given to the first-part CNN (used for backdrop removal) usually contains shapes, edges, textures, and objects. The edge detector, circle detector, and corner detector filters are used at the start of convolutional layer 1. Once the face has been detected by applying other filtering techniques like the median and Gaussian noise filters, the second-part CNN filter identifies facial features like the eyes, ears, mouth, nose, and cheeks. Consequently, this research uses the kernel filter to help minimize the dimensional space borders. One of the most important methods in emotional recognition is the extraction of face features, which is accomplished in this work through the use of four different approaches: template-based, geometric, hybrid, and holistic. The next step of the preprocessing procedure is the feature extraction technique. Consequently, once feature extraction is complete, the output is sent to the convolutional neural network. The output keeps going since the enhanced CNN model of the face expression is provided by the previously discussed feature extraction technique. The suggested method's overview demonstrates that the preprocessing technique is applied to the input image. The preprocessing method aids in lowering the image's noise level. The proposed ECNN's overview is shown in Figure 1

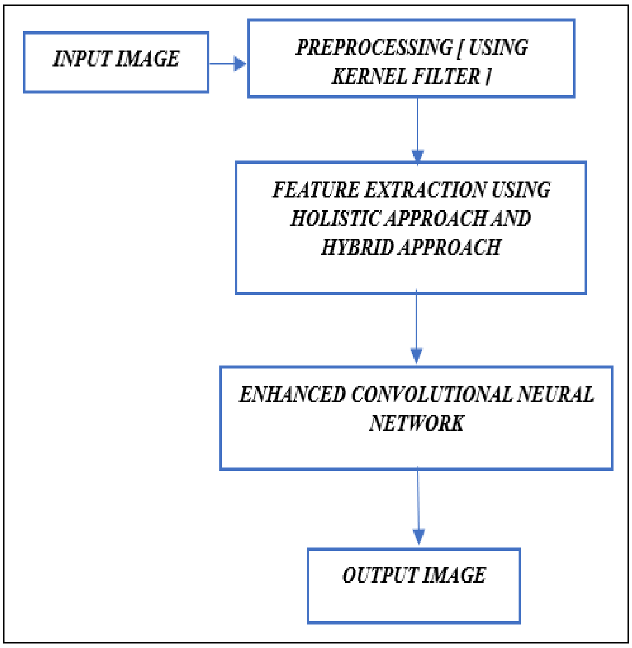


Fig 1: Overview of the Proposed ECNN Approach

### A. *Proposed Method*

The proposed architecture is summarized in Figure 2. The first dataset, which is fed through a preprocessing approach, includes the different facial expressions.

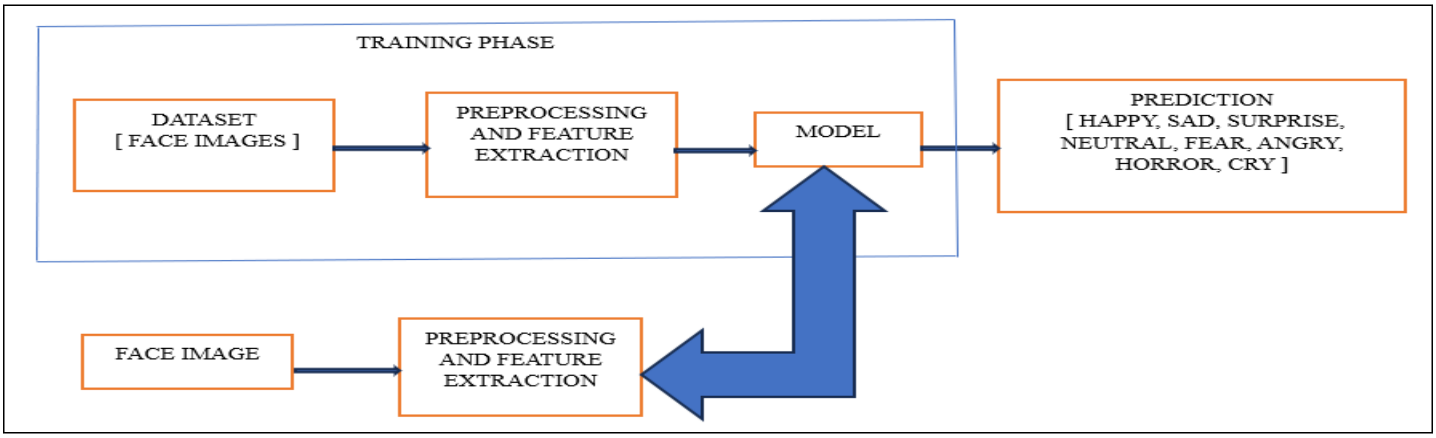


Fig 2: Proposed Architecture

### B. *Prepocessing and Feature Extraction*

One of the most important filters for facial recognition in an image is the kernel filter, which also helps to minimize dimensionality, smooth edges, and reduce unfocus in the image. The kernel filter for image preprocessing is implemented in this work. Edge detection is one of the most significant kernel filter techniques since it lowers the dimensional space and predicts the optimal edge. The box, average, Gaussian, and median filters comprise the three distinct filter evaluations that make up the smoothening kernel.

➢ *Preprocessing is Mostly Used to:*

- Add a filter to eliminate noise;
- Convert an RGB image to a greyscale image.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1662

One of the most important image processing techniques is feature extraction. As a result, face recognition involves three steps: face detection, which is thought of as the identification of the input image, face extraction, which happens subsequent to the image identification. The process of feature extraction aids in categorizing an image's accuracy. The primary goal of feature extraction is to minimize the dimensionality of the input images as well as the image's dimensional edges once feature recognition is complete. Feature recognition is regarded as the image's identification. The two methods of feature recognition [25] are implemented in this research, specifically:

- Holistic feature-based approach
- Hybrid Approach

*C. Holistic Approach*

One of the most important ways for identifying the many emotions on the face, such as sobbing, rage, sadness, happiness, etc., is the holistic approach. It does this by detecting the entire face in the input. As a result, many applications of the holistic approach such as the Eigen face and the Fisher facial are used [26].

*D. Hybrid Approach*

The hybrid feature extraction in facial recognition is defined as the fusion of the hybrid and the all-image feature; it accurately recognizes facial images, especially those with expressive mouths, noses, and eyes. This strategy uses a variety of facial expression strategies, like as

- Geometric-based technique
- Appearance-based technique

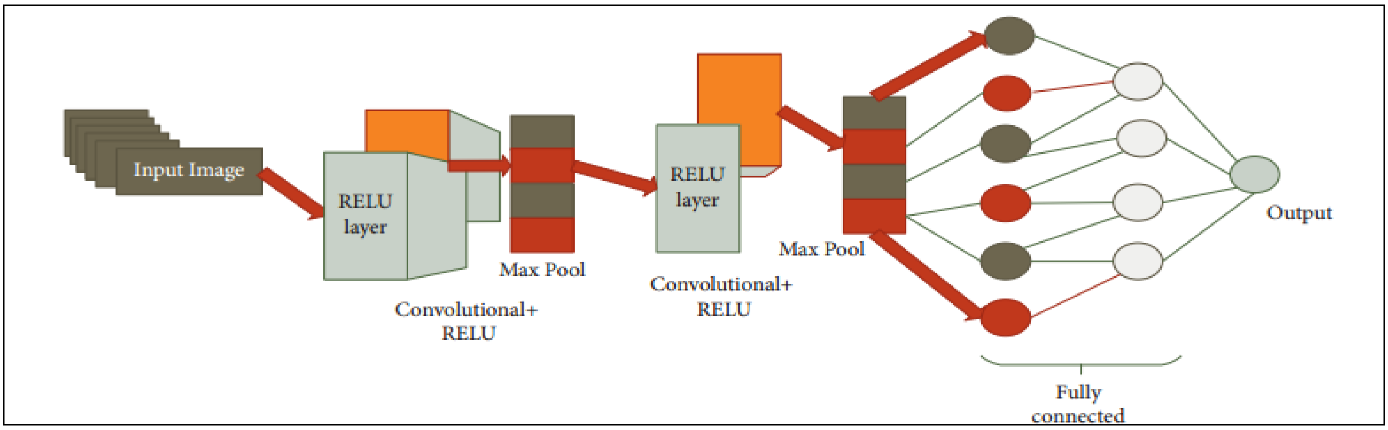


Fig 3: Geometric-based Approach

According to Figure 3, the geometric feature technique in the geometric-based approach uses the gradient magnitude and the canny filter's output to handle size and relative position. Feature extraction is used in the appearance-based approach, which is based on principle compound analysis (PCA). PCA's primary purpose is to reduce a picture's enormous dimensionality and transform it into a compact, dimensionality-independent image. It produces an image of flawless quality.

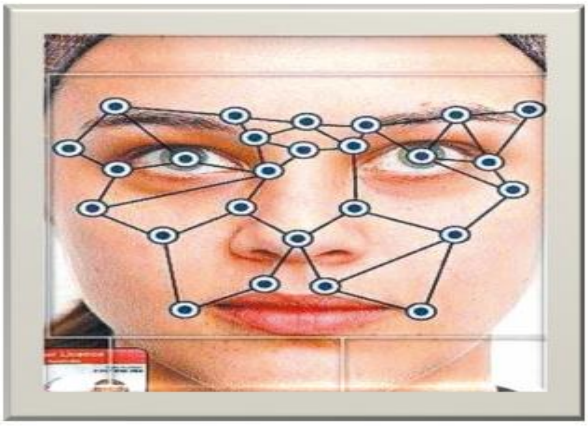

Fig 4: Different Skin Maps in the Original Image

Using the picture's threshold function, Figure 4 constructs the different kinds of skin mappings from the source image. Using the deep learning approach, the feature-extracted image provides the input for the convolutional neural network.

*E. Convolutional Neural Network*

One kind of deep learning algorithm is the convolutional neural network. Convolutional neural networks operate as integral layer networks, nothing more.

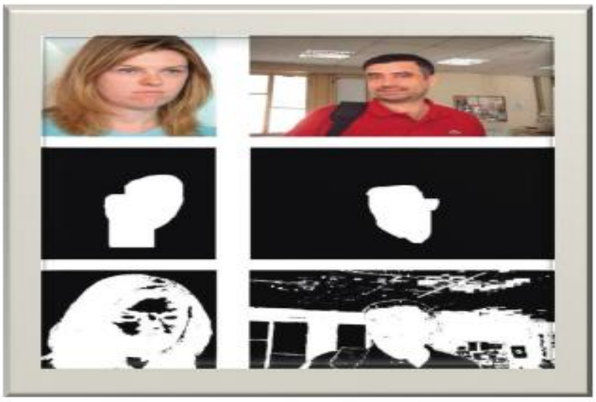

Fig 5: Convolutional Neural Network

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1662

Figure 5 suggests that the multilayer particular task is carried out by the convolutional neural network as a classification approach. The convolutional neural network's role is to pass the input picture through the convolutional RELU layer before using the maxpooling algorithm. The output image is passed via the completely linked layer since this process is repeated once. It facilitates the correct output's separation. The image's classification is additionally aided by the completely connected layer.

$$\text{Clarity} = \frac{\text{Real Positive}}{\text{Real Positive} + \text{Fake Positive}}$$

$$\text{Accuracy} = \frac{\text{RP+RN}}{\text{RP} + \text{RN} + \text{FP} + \text{FN}}$$

## III. EXPERIMENTAL RESULT AND ANALYSIS

### *A. Dataset*

About 32,298 distinct tagged photos make up the dataset. Since the photos had 48 ∗ 48 size pixels, the FER2013 dataset was utilized for additional analysis. There are two columns in the dataset: one is a pixel and the other is emotional [27]. The emotional acts are examined in the code value ranging from 0 to 6, and the pixel value is shown in the pixel column. The dataset for classifying emotions using different feature extraction techniques is displayed in Table 1.

Table 1: Emotions Tabulations Dataset

| Emotions | Code |
|---|---|
| Happy | 0 |
| Sad | 1 |
| Surprise | 2 |
| Neutral | 3 |
| Fear | 4 |
| Horror | 5 |
| Cry | 6 |

Table 2: Accuracy Rate

| Epochs | Hrs | Accuracy rate |
|---|---|---|
| 20 | 5 | 80 |
| 200 | 15 | 95 |
| 600 | 100 | 97 |

### *B. Testing and Training Model*

Since the preprocessing method reduces noise in the filter, 80% of the data are used as input for the training stages. Reducing the high dimensionality of the space in the filter is made easier by feature extraction. In order to minimize the dimensional space, our paper applies different kinds of feature extraction, together with edge detection. Consequently, it generates output images with exact dimensional quality and clarity [28]. In the data, the categorization and feature extraction processes operate simultaneously. The testing and training phase features are transmitted after the categorization [29]. As a result, we get 20, 200, and 600 models from the training iterations. The training time is 5, 15 and 100.

The accuracy percentage for the data that was gathered is shown in Table 2. It demonstrates that, in comparison to the current method, our suggested solution utilizing the CNN technique yields more accurate findings [30].

### *C. Results Comparison*

The comparative findings demonstrate that the collection of templates is extracted using the current CNN and SVM methodology, a support vector machine-based method utilizing the Gabor filter, and the Monto Carlo algorithm [26–30]. This current technique's feature extraction yields less than ideal outcomes.

The comparison results for face feature identification are displayed in Table 3. Our study achieves higher accuracy when compared to existing methodologies, especially several feature extraction kinds that aid in accurately extracting features.

Table 3: Comparison Table

| Results | Accuracy rate | Emotions |
|---|---|---|
| CNN [30] | 79.98 | 7 |
| SVM [31] | 85–90 | 6 |
| ECNN (proposed method) | 97 | 7 |

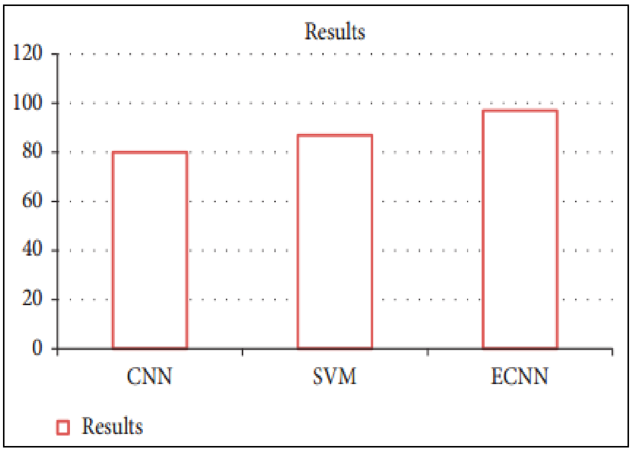

Fig 6: Results

The accuracy results for both the suggested and the current strategy are displayed in Figure 6. Comparing our suggested method to the current methodology, the results are superior.

However, the vehicle body is very reflective and there is a large amount of inter object reflection in the photograph which is interpreted to damage. So, proposed a method for classifying reflection in photographs.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1662

In this chapter, there is some limitation work to detect the reflection image edge in close photographs in vehicle panels. At last, it's possible to implement vehicle damage.[9]

*D. Experimental Setup*

Using a sequence of photos of any subject ranging from neutral to the specified expression, the experiment uses the Cohn-Kanade dataset to extract the extreme image displaying the desired expression. These photos are then passed through the appropriate filters, which in turn turn the image into a CSV file containing the image's data. This file is then fed into the classifier, which makes predictions about the outcome. Smartphones are used to capture real-time photos, which are subsequently processed by the system as needed.

## IV. LITERATURE REVIEW

In this paper, Yadan Lv et al. [5] used a deep learning technique to accomplish facial recognition. We do not need to add any extra features for noise reduction or adjustment because the parsed components aid in identifying the different kinds of feature identification techniques. One of the most crucial and distinctive strategies is the parsed technique.

In this work, Mehmood et al. [6] applied deep learning ensemble techniques and optimal feature selection for human brain EEG sensor-based emotional identification. This work presents the implementation of feature selection and EEG feature extraction techniques based on facial recognition methodology optimization. There are four different categories of emotions at play here: happy, calm, sad, and fearful. Better accuracy in the emotional classification is provided by the feature extraction method, which is based on the optimal picked feature, such as the balanced one-way ANOVA technique. More methods, such as the arousal-valence space, improve EEG recognition. Li et al. [14] implemented the deep facial expression recognition survey. Recognition of facial expressions is regarded as one of the main problems in the network system. Facial expression recognition (FER) faces two main challenges: insufficient training sets and non-relatable expression variations. The dataset is organized using the neural pipeline technique in the first instance.

This will lessen the difficult problems with the FER approach. In this research, Wang et al. [23] applied the most recent deep learning technique. In this paper, the four-category deep learning model is implemented. Convolutional neural networks and deep architectures make up the first group. The deep learning model largely persuades the deep neural networks. It is among the machine learning algorithm's most crucial operations. It is essential to the correctness of the data, and the classification includes both linear and nonlinear special functions. The convolutional neural layer, the pooling layer, and the fully connected layer are the three most important layers of a convolutional neural network. The first layer to apply filters to lower the noise and dimensional space in the filter is called the convolution layer. The CNN's pooling layer aids in lessening the over-fitting issue. The pooling layer and convolutional layer are put first, followed by the fully connected layer. As a result, it eliminates the function's erroneous data. Gnana et al. [24] used the literature review for feature selection of high-dimensional data in this research. The most straightforward feature selection method for data is to send all of the data to a statistical measurement approach; this aids in the selection of the feature selection approach.

## V. CONCLUSION AND FUTURE WORK

We used a convolutional neural network with a deep learning algorithm to propose a facial feature for emotional recognition in our study. We did this by using various facial features with appropriate dimensional space reduction, and we sharpened the edges using a kernel filter as a preprocessing technique. The outcomes demonstrate enhanced accuracy and the convolutional neural network's classification methodology as compared to the present approach. The results are superior when contrasting the present method with our recommended strategy. The capacity of the model seems to meet the difficulty of complex facial expression recognition at those resolutions. We can enhance CNN's performance by employing data augmentation strategies like adding noise and combining data from stages (b), (cropping), and (f). The highlighted study looks into picture synthesis approaches that may be seen as a deep learning augmentation data solution. It aims to prevent data hunger and overfitting for tiny amounts of data. Our study's future depends on including additional information, classifying facial emotions into ten categories, and investigating automatic facial emotion detection.

➢ *Data Availability*

The corresponding author can provide the datasets used and/or analyzed in the current work upon request.

➢ *Conflicts of Interest*

There are no conflicts of interest, according to the authors.

## REFERENCES

[1]. F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera, and G. Anbarjafari, "Audio-visual emotion recognition in video clips," IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 60–75, 2019.

[2]. C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of big data, vol. 6, no. 1, p. 60, 2019

[3]. B. J. Park, C. Yoon, E. H. Jang, and D. H. Kim, "Physiological signals and recognition of negative emotions," in Proceedings of the 2017 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1074–1076, IEEE, Jeju, Korea, October 2017.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1662

[4]. L. Wiskott and C. Von Der Malsburg, "Recognizing faces by dynamic link matching," NeuroImage, vol. 4, no. 3, pp. S14–S18, 1996

[5]. Y. Lv, Z. Feng, and C. Xu, "Facial expression recognition via deep learning," in Proceedings of the 2014 International Conference on Smart Computing, pp. 303–308, IEEE, Hong Kong, China, November 2014.

[6]. R. Majid Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," IEEE Access, vol. 5, pp. 14797–14806, 2017

[7]. R. Majid Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," IEEE Access, vol. 5, pp. 14797–14806, 2017

[8]. S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2852–2861, Honolulu, HI, USA, July 2017.

[9]. L. Chen, M. Zhou, W. Su, M. Wu, J. She, and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction,"Information Sciences, vol. 428, pp. 49–61, 2018

[10]. P. Babajee, G. Suddul, S. Armoogum, and R. Foogooa, "Identifying human emotions from facial expressions with deep learning," in Proceedings of the 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), pp. 36–39, IEEE, Novi Sad, Serbia, May 2020

[11]. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," Informatics in Medicine Unlocked, vol. 20, Article ID 100372, 2020.

[12]. Tan, M. Sarlija, and N. Kasabov, "NeuroSense: short-termˇ emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns,"Neurocomputing, vol. 434, pp. 137–148, 2021

[13]. P. Satyanarayana, D. P. Vardhan, R. Tejaswi, and S. V. P. Kumar, "Emotion recognition by deep learning and cloud access," in Proceedings of the 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 360–365, IEEE, Greater Noida, India, December 2021.

[14]. K. Jayanthi and S. Mohan, "An integrated framework for emotion recognition using speech and static images with deep classifier fusion approach," *International Journal of Information Technology*, pp. 1–11, 2022

[15]. S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, vol. 7, no. 3, pp. 1195–1215, 2020

[16]. S. S. Yadahalli, S. Rege, and S. Kulkarni, "Facial micro expression detection using deep learning architecture," in Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC), pp. 167–171, IEEE, Trichy, India, September 2020

[17]. Yang, X. Han, and J. Tang, "&ree class emotions recognition based on deep learning using staked autoencoder," in Proceedings of the 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1–5, IEEE, Shanghai, China, October 2017

[18]. Asaju and H. Vadapalli, "A temporal approach to facial emotion expression recognition," in Proceedings of the Southern African Conference for Artificial Intelligence Research, pp. 274–286, Springer, Cham, January 2021

[19]. Sati, S. M. Sanchez, N. Shoeibi, A. Arora, and´ J. M. Corchado, "Face detection and recognition, face emotion recognition through NVIDIA Jetson Nano," in Proceedings of the International Symposium on Ambient Intelligence, pp. 177–185, Springer, Cham, September 2020

[20]. S. Rajan, P. Chenniappan, S. Devaraj, and N. Madian, "Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM," IET Image Processing, vol. 14, no. 7, pp. 1373–1381, 2020

[21]. S. Mirsamadi, E. Barsoum, and C. Zhang, "Automatic speech emotion recognition using recurrent neural networks with local attention," in Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2227–2231, IEEE, New Orleans, LA, USA, March 2017

[22]. Domnich and G. Anbarjafari, "Responsible AI: Gender Bias Assessment in Emotion Recognition," 2021, https://arxiv.org/pdf/2103.11436.pdf.

[23]. O. Ekundayo and S. Viriri, "Multilabel convolution neural network for facial expression recognition and ordinal intensity estimation," *PeerJ Computer Science*, vol. 7, p. e736, 2021.

[24]. X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 4, pp. 747–750, 2020.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1662

[25]. D. Asir, S. Appavu, and E. Jebamalar, "Literature review on feature selection methods for high-dimensional data," *International Journal of Computer Application*, vol. 136, no.1, pp. 9–17, 2016.
[26]. E. Owusu, J. A. Kumi, and J. K. Appati, "On Facial Expression Recognition Benchmarks," Applied Computational Intelligence and Soft Computing, vol. 2021, Article ID 9917246, 2021
[27]. S. Dhawan, "A review of face recogntion," IJREAS, vol. 2, no. 2, 2012
[28]. J. Goodfellow, D. Erhan, P. L. Carrier et al., "Challenges in representation learning: a report on three machine learning contests," in Proceedings of the International Conference on Neural Information Processing, pp. 117–124, Springer, Berlin, Heidelberg, November 2013
[29]. P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.
[30]. N. Bhoi and M. N. Mohanty, "Template matching based eye detection in facial image," International Journal of Computer Application, vol. 12, no. 5, pp. 15–18, 2010
[31]. U. Bakshi and R. Singhal, "A survey on face detection methods and feature extraction techniques of face recognition," International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol. 3, no. 3, pp. 233–237, 2014.