International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. IV, No. 05, pp. 8-18, May

An Innovative Multi-Stream Inception-V3 Deep Learning Strategy for Advanced Facial Expression Discrimination

Sincija C[1], D. Goldy Val Divya[2], P. Selva Rathinam[3]

[1]Assistant Professor, Department of Computer Science and Engineering, Dhanalakshmi Srinivasan College of Engineering, Coimbatore.

[2]Assistant professor, Department of Computer Science and Engineering (AIML), Marri Laxman Reddy Institute Technology and Management, Hyderabad, Telangana 500043.

[3]Assistant Professor, Computer science and Engineering department, Vel Tech Multitech Dr. Rangarajan Dr. Sagunthala Engineering College, Avadi, Chennai-600062.

***Corresponding Author E-mail**: **sincijac@dsce.ac.in**

ABSTRACT: Human Facial Expression (FE) is one of the most potent and identifying or recognising FE is a difficult task. In general, a facial expression helps people to express their feelings such as sad, anger, contempt, happy, fear, disgust and surprise. In this paper the main phases of FE techniques are pre-processing, feature extraction, and classification. The different techniques involved in FE recognition and their main contributions are explained in this paper. Initially, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to the facial expression dataset to enhance the image quality. Adaptive bilateral filtering effectively remove noise and sharpen the face image data. Scale Invariant Feature Transform (SIFT), extract features and key points for further analysis. Multi-Stream Inception-V3 is proposed to enhance the classification of facial expression images, and benefits to overcome challenges in face analysis. Using Python software the proposed Multi-Stream Inception-V3 achieved a higher accuracy of 97% which is highly efficient compared to the existing method.

Keywords: Multi-Stream Inception-V3, Scale Invariant Feature Transform (SIFT), Contrast Limited Adaptive Histogram Equalization (CLAHE), Adaptive Bilateral Filtering, Facial Expression

1. Introduction

Human FE is crucial in social communication; both verbal and non-verbal signs are typically used in conversation. Facial expressions are a kind of non-verbal communication [1]. Conversely, the intelligent signs of the broader conversation are facial expressions. Eye contact, gestures, facial expressions, body language, and paralanguage are examples of non-verbal communication between humans and animals [2]. Whereas, making eye contact is a crucial part of communication since it allows for the exchange of thoughts. Eye contact establishes a connection with others and regulates contributions and conversations [3]. Smiling, sadness, anger, dislike, surprise, and fear are examples of facial expressions. A smile on a person's face conveys their joy and showcases their curled eye. A frown and raised, twisted eyebrows are typical expressions of sadness, which is a sense of looseness. Anger on a person's face is associated with disagreeable and upsetting circumstances [4]. Whereas, FE systems widely used in many fields, including social marketing, health care systems, and computer interfaces. However, FE is extremely difficult in real-world images as the complex, noisy background environment and the foreground people's movement and fleeting motions [5-6]. The FER system's performance is significantly impacted by three primary issues: subject reliance, light variance, and head posture changes [7].

Facial expression is processed using the pre-processing technique called image normalization. It is more effective by using histogram equalization and gamma correction combined for image normalization, and a canny edge detector for edge detection in facial recognition. It takes more processing time when the edge detection is occurs [8]. Whereas, High Dynamic Range (HDR) pre-processing module is used to detect the FE. The model's capacity to adjust enhanced by applying local contrast and detail augmentation algorithms to input images. Produces a dynamic image by reducing noise and enhance the performance. Introduce artifacts particularly when kernel sizes are big and computationally intensive [9]. The proposed pre-processing technique used here is CLAHE and ABF for FE. It filtered the input image and removes the noise for further Processing.

Conversely, a hybrid descriptor using Multi-Block Local Ternary Pattern (LTP) and Gray Level Co- occurrence Matrix (GLCM) to detect the FE. Provides information about the thickness and homogeneity of texture by capturing the spatial correlations between gray levels. LTP, which uses ternary patterns for local texture, but less susceptible to changes in grayscale [10]. Moreover, a Multi-scale Block and Mean-based Local Binary Pattern (MBM-LBP) used to detect the FE. Computational effectiveness and strong texture analysis capabilities, drawbacks including noise sensitivity and the inability to save global or rotational texture information [11]. The proposed feature extraction technique used here is SIFT. It finds the key points in the FE and extracts its reaction.

To differentiate the FE or reaction, classification is used. Using a classification technique called SVM and KNN classifier to classify the facial expressions images. SVM-KNN classifier is used to increase the classification accuracy of facial expression. Yet, lower chance of over fitting, determining the relevance of a feature is simple. Time-consuming procedure, more resources are needed, more complex [12]. However, to classify facial expressions, Homogeneous Ensemble Convolutional Neural Network (HoE-CNN) is used. HoE-CNN is utilized to enhance classification performance. However, low accuracy in complex issues, particularly when functioning with high-dimensional search places [13]. Moreover, a CNN approach is used to classify the facial expression. Examining the text contained in the high-resolution face images. It improves the performance of the system and extracts more discriminating features. It takes large data to train effectively whereas takes more time to train data [14-15]. To overcome the limitations of existing classification algorithms a Multi-Stream Inception-V3 is used. Multi-Stream Inception-V3 have several convolutional layers operating in parallel, each having a different kernel size and depth to classify the FE.

2. Related work

Zhao *et al* (2021) [16] have proposed a global Multi-scale and Local Attention Network (MA-Net) for FE recognition. To improve the performance of system a feature pre-extractor, a multi-scale module, and a local attention module is used as an advanced method. The advantage of MA-Net algorithm is easy to understand because they represent a solution to a problem step-by-step. However, with a larger number of parameters, MA-Net typically needs more training data to function.

Li *et al* (2021) [17] have proposed a Deep Residual Network ResNet-50 for FE recognition. The ResNet-50 is reliable enough to accurately and efficiently identify the FE. When identifying face reaction, several issues are increased; poor identification speed, high computation costs, and inadequate precision.

Nan *et al* (2022) [18] have proposed a Attention MobileNet (A-MobileNet) network model, to enhance the model's ability to extract fine-grained features of facial expressions, The A-MobileNet model is reliable adequate to accurately and efficiently identify and classify face reaction. Whereas, slow convergence and local minima susceptibility are two drawbacks of A-MobileNet.

Gong *et al* (2025) [19] have proposed a Deep Neural Network (DNN) for facial emotion identification and another that automatically tracks and segments faces. Deep Automatic Facial Expression Recognition Model (DAFERM) for interactive Virtual Reality (VR) applications improves the classification accuracy and performance for FE. However, DAFERM takes more time to process when multiple images are used.

Alzahrani *et al* (2024) [20] have proposed a Deep Learning (DL) systems specifically Recurrent Neural Network (RNN) as well as CNN used for the task of automatically categorizing and detecting emotions from human FE. For the effective recognition and categorization of facial emotions, the Multi-head Attention Bi-directional Long Short-Term Memory (MA-BiLSTM) model is utilized. Although, it takes more training time and computational complexity in comparison to the conventional LSTMs.

The contribution of this work is follows

- CLAHE based image preprocessing technique is introduced to the facial expression for enhancing contrast of an image.
- Unwanted noises are removed using the ABF to get high quality image.
- Subsequently, feature extraction by Scale-Invariant Feature transforms to find the key points for further classification.
- Finally, a Multi-Stream Inception-V3 is proposed to enhance the classification of face recognition.

3. Proposed work

The proposed block diagram in figure 1 for FE using DL technique with facial recognition datasets. Initially using CLAHE is applied to the facial expression dataset to enhance contrast of the image quality. After that using ABF the image reduces noise. The processed face image is given as input to feature extraction here the SIFT is used to find the key points.

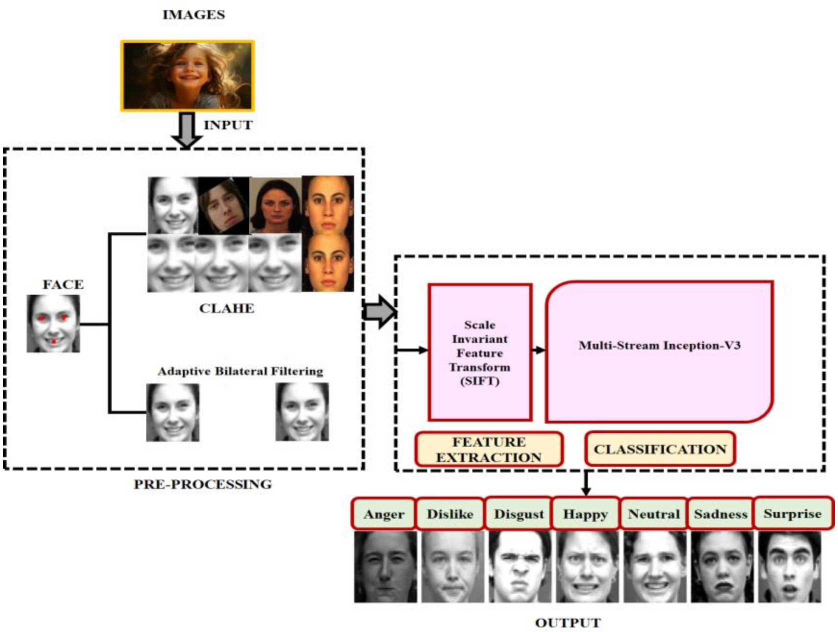

Figure 1: *Block diagram of proposed work*

Finally using classification technique the Multi-Stream Inception-V3 the process uses auxiliary classifiers, factorized convolutions, and label smoothing. A proposed Multi-Stream Inception-V3 is used to increase the accuracy and efficiency for facial recognition.

3.1 Pre-processing for Face Recognition

In the face recognition identification pre-processing have two methods CLAHE and Adaptive Bilateral filtering.

3.1.1 Contrast Limited Adaptive Histogram Equalization (CLAHE)

To improve image quality for facial expression each frame delivered have been subjected to an image enhancement process known as CLAHE. After that, the frame is transformed into a grayscale image in order to identify a face. The face image is aligned to maintain both eyes in the same horizontal position if it is discovered to be tilted. The face that detected is shrunk and cropped. Histogram Equalization is an image processing technique that is used to alter the image's intensity. This enhances the contrast of a face image and it could be explained by a histogram. The histogram is said to be equalized when every gray level in the image is utilized equally. The intensities in the histogram are thus uniformly distributed. CLAHE have replaced AHE for process in figure 2.

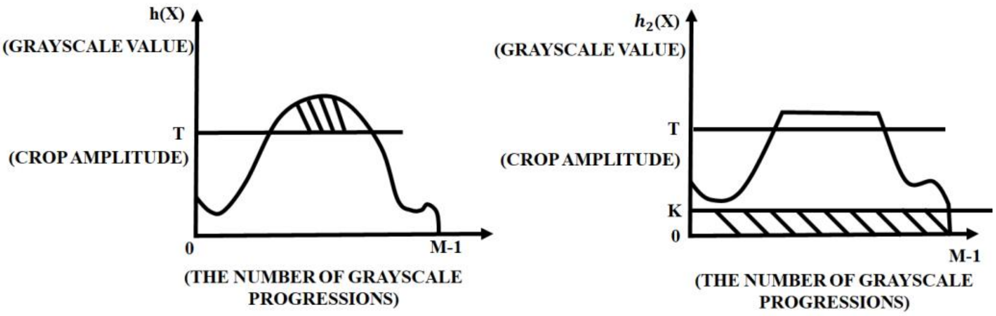

Figure 2: *Basic CLAHE*

The problem with AHE is that it amplified noise too much. CLAHE cuts the histogram at a predetermined value to lessen the amplification. To improve noise in almost constant regions, adaptive histogram equalization is employed. To reduce amplified noise, CLAHE restricts contrast amplification. By impartially allocating the portion of the histogram that crosses the clip boundary across all histograms. The CLAHE enhance and contrast the quality of the face image.

3.1.2 Adaptive Bilateral Filtering

In order to remove noise in the face image an ABF is used. For facial recognition an adaptive version of the bilateral filter in which the Gaussian range kernel's width and center are permitted to vary from pixel to pixel in figure 3. It is used to reduce noise and improve image precision. Sharpening is not possible with the conventional bilateral filter, so the range kernel had to be modified.

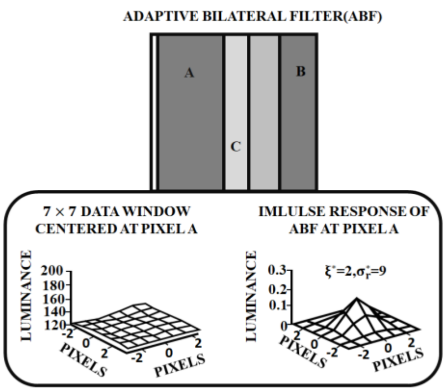

Figure 3: *Adaptive bilateral filtering*

The kernel's width and center is adjusted to determine the amount of noise reduction and sharpening at a specific pixel. Assume that $f : I \rightarrow R$ remain the input face image, wherever I $\subset$ Z2$I \subset Z^2$ is the image domain. The output image $g : I \rightarrow R$ is known as

$$g(i) = \eta(i)^{-1} \sum_{j\in\Omega} \omega(j)\varphi_i(f(i-j) - \theta(i)f(i-j) \quad (1)$$

Where

$$\eta(i) = \sum_{j\in\Omega} \omega(j)\varphi_i(f(i-j) - \theta(i)) \quad (2)$$

Now, $\varphi_i : R \rightarrow R$ remains the local Gaussian range kernel, and Ω is a window with the origin at its center:

$$\varphi_i(l) = \exp(-\frac{t^2}{2\sigma(i)^2}) \quad (3)$$

The widths $\sigma(i)$ now (3) as well as center $\theta(i)$ in (1) are both spatially changing functions, which is significant. In (1), the spatial kernel $\omega \sim \Omega \rightarrow R$ stands Gaussian.

$$\omega(j) = \exp(-\frac{(\|j\|2}{2\rho^2}) \quad (4)$$

Usually, the window is set to be $\Omega = [-3\rho, 3\rho]^2$ and (1) be referred to as ABF. The processed image is filtered and noises are removed using CLAHE and ABF. Next, the pre-processed image is given as an input to feature extraction.

3.2 Feature extraction by Scale-Invariant Feature Transform

The processed image is extracted using SIFT methodology for local feature detection, which have subsequently grown to be a face expression technology in the imaging sector. Finding robustly identifiable image characteristics is its aim in order to provide object detection in various viewing scenarios and matching across many images and image sequences. This not only expands the amount of characteristics that are available, but it also makes the method extremely forgiving to scale changes. Scale space calculation and feature recognition have been streamlined to provide accelerated versions of the SIFT technique. With sub-pixel positioning accuracy and a rotation-invariant feature descriptor linked to every candidate point, SIFT functions in theory similarly to a multi-scale corner detector. This (usually 128-dimensional) feature descriptor be thought of as a "face recognition" since it summarizes the distribution of the gradient directions in a spatial neighborhood surrounding the relevant feature point.

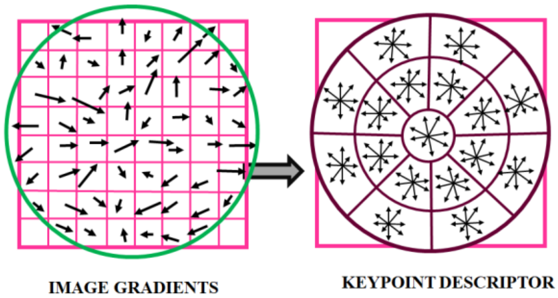

Figure 4: *SIFT-key point descriptor*

The following are the primary steps in calculating SIFT features:

- To find possible locations of interest, extremes are detected in a Laplacian-of-Gaussian (LoG) scale space.
- Refinement of key points by fitting a continuous model to ascertain exact scale and placement in figure 4.
- Orientation assignment based on the feature point's dominant orientation derived from the gradient directions of the surrounding image.
- The local gradient histogram is normalized to create the feature descriptor

Next, the extracted features are given to the proposed multi-stream Inception-V3.

3.3 Classification using Multi-Stream Inception-V3

To classify the facial expression a proposed Multi-Stream Inception-V3 is used. The final step in the face recognition system is classification, where the classifier groups expressions like "smile," "sadness," "surprise," "angry," "fear," "disgust," and "neutral." Expanding the network model's breadth and depth is the most straightforward method of enhancing network performance. More parameters will be produced as the network's depth and width increase, which will result in over fitting; and a large increase in computation will be required. Although the complete connection layer is converted to sparse connections and sparse features be introduced to address the aforementioned issues, the current models are unable to efficiently handle the rapid processing of non-uniform sparse data. Figure 6 architecture of multi-stream Inception-V3

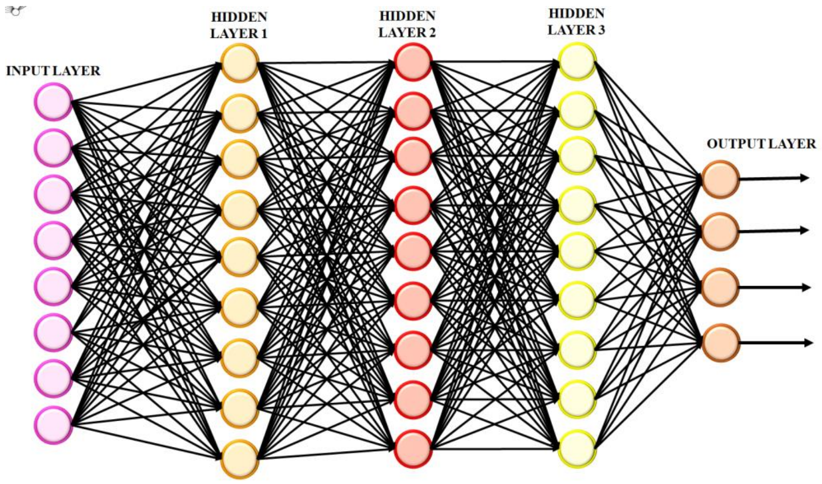

Figure 5: *Layers for multi-stream Inception-V3*

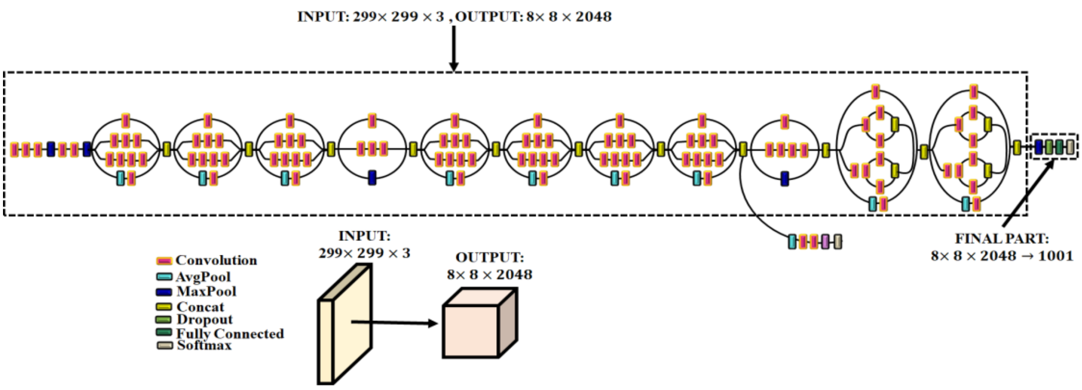

Figure 6: *Architecture of multi-stream Inception-V3*

Based on the multi-branch module, the Inception network family software is an ideal choice for supporting both the type of high-performance computing that fully employs dense matrices and the sparse feature at the filter level.

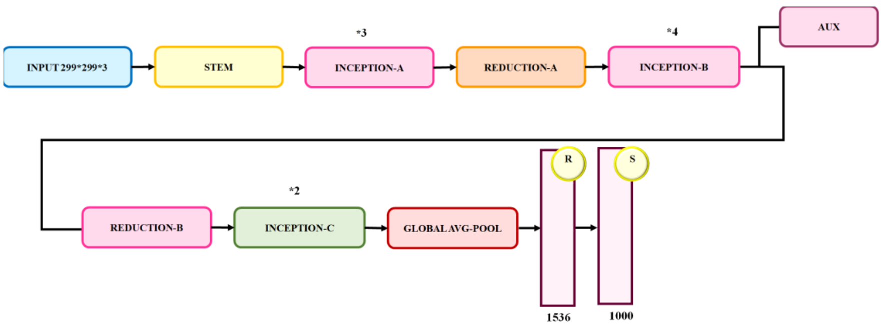

Figure 7: *Modules in multi-stream Inception-V3*

By lowering the scale of the network parameters needed for each training iteration, dropout reduces over fitting and, based on a specific probability, briefly "drops" the neural network unit from the network throughout each training phase in figure 7. Here is the precise procedure:

(1) Stop a subset of neurons at random without altering the i/p or o/p;

(2) The loss conclusion still makes use of the network's back propagation even after adding the input for forward propagation. After the forward propagation of training samples and the back propagation of loss are finished, the constraints of the functioning neurons are adjusted, while the parameters of the paused neurons remain unchanged;

(3) Repeat steps (1) as well as (2) until the loss function becomes stable.

The dropout algorithm's calculating formula is as follows:

$$r_j^{(l)} \sim Bernoulli(p) \tag{5}$$

$$y_i^{(l+1)} = f(w_i^{l+1)}.I(x).r_j^{(l)} + b_i^{(l+1)}) \tag{6}$$

Where $r_j^{(l)}$ is a probability vector that follows a Bernoulli distribution and only has 0 and 1 entries. $w_i^{l+1)}$ and $b_i^{(l+1)}$ denote the ith node's weight and bias in layer l's subsequent layer, correspondingly, $I(x)$ denotes using the Inception V3 network to extract the feature vector, as well as $y_i^{(l+1)}$denotes the result following the activation function's passing ().With Inception V3, the original bottom auxiliary classifier is dropped, four new Inception structures were created, and just one middle-level auxiliary classifier was kept. It improved processing efficiency and created a new Inception V3 model that is deeper, wider, and has superior performance and expressive capabilities. Two 3 × 3 convolution kernels are used in place of the original 5 × 5 convolution kernel in Inception Module A. When using the proposed multi-stream Inception-V3 the classification is done in depth for better accuracy for FE.

4. Result and Discussion

In this section the seven classes of face images are predicted using the proposed Multi-stream Inception-V3. For identifying the face reaction the dataset is taken from the kaggle.com. Face recognition is implemented using Python software. The train images have the count of 783 and test images have the count of 198.

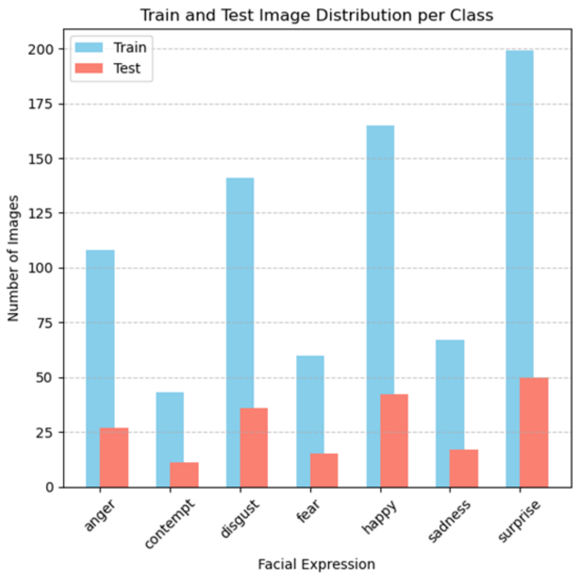

Figure 8: *Train and test image distribution per class*

Figure 8 shows the train and test image distribution per class. Here number of images taken is 200 and seven classes (anger, contempt, disgust, fear, happy, sadness and surprise) reactions are trained using the facial recognition dataset. For anger (test data is 27 and train data is 108), for contempt (test data is 36 and train data is 43), for disgust (test data is 36 and train data is 141), for fear (test data is 15 and train data is 60), for happy (test data is 42 and train data is 165), for sadness (test data is 17 and train data is 67) and for surprise (test data is 50 and train data is 199).

Figure 9: *Random images from each facial expression*

Figure 9 shows random images in training dataset of facial expression image. Here, the random images are taken from the facial expression dataset. Whereas, seven classes of images such as sad, anger, contempt, happy, fear, disgust and surprise are expressed.

Figure 10: *Original in to resize image*

Figure 10 shows the original image to resize image. The original image is taken from the facial expression dataset. Here the input face image is resized using image resizing technique. The noises present in the original image are removed.

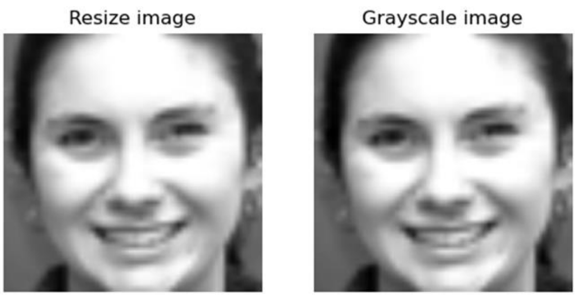

Figure 11: *Resize image into grayscale image*

Figure 11 shows the resized image into gray scale image for facial expression approach. Here resized image is transformed into grayscale image for enhancing the FE recognition. Face image in to a gray scale or black and white representation is gray scale image.

Figure 12: *CLAHE into adaptive bilateral filtering*

Figure 12 shows the CLAHE image into adaptive bilateral filtering image. Here the image is enhanced and contrast of the pixel value is improved. It smooths the face image and reduces the noise using adaptive bilateral filtering.

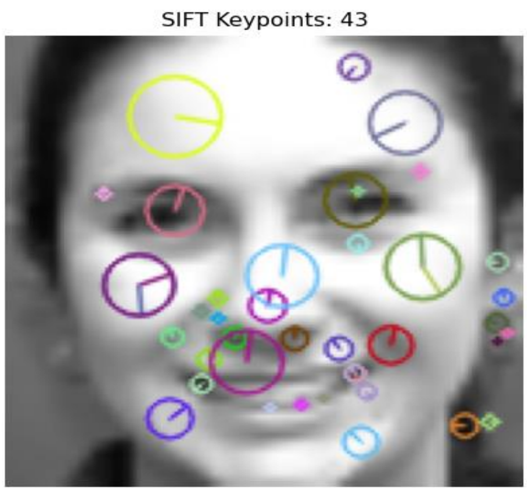

Figure 13: *SIFT key points: 43*

Figure 13 shows the SIFT Key points for face image. The SIFT finds the key points from the face image. Each key point values are pointed with the variables i.e. 0 to 43.

4.1 Performance Metrics

Standard medical criteria like accuracy, sensitivity and specificity which indicate the likelihood that a face reaction be correctly detected and appropriately diagnosed as the probabilities calculated by the multi-stream Inception-V3. The true positive, true negative, false positive and false negative results are denoted by TP, TN, FP, and FN, respectively. Sensitivity$[TP/(TP + FN)]$, specificity $[TN/(TN + FP)]$, F-value $[2 \times (recall \times precision)/(recall + precision)]$, and accuracy $[(TP + TN)/(TP + FP + FN + TN)]$.

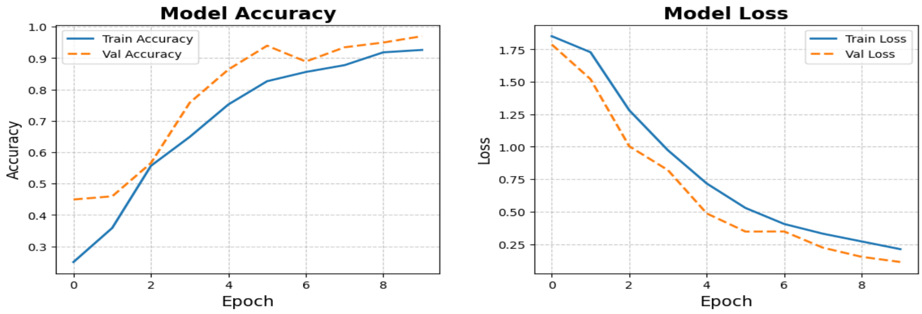

Figure 14: *Model accuracy and model loss*

Figure 14 shows the model loss and model accuracy. The X-axis denotes the epochs and Y-axis is accuracy for model accuracy simultaneously for model loss X-axis for epochs and Y-axis for loss. The Model Accuracy graph illustrates increasing accuracy of 97%, with the model achieving high performance and minimal over fitting. Model loss is reliably declining in training and validation loss, representing improved learning and reduced error.

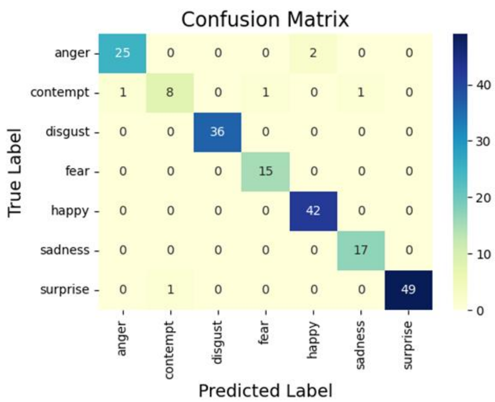

Figure 15: *Confusion matrix for proposed method*

The confusion matrix is shown in figure 15. Proposed method evaluates the model's performance in predicting different levels of face expression. It shows high accuracy for all classes, with 25 faces of anger, 8 faces of contempt, 36 faces of disgust, 15 faces for fear, 42 faces for happy, 17 faces for sadness and 49 faces of surprise is correctly classified.

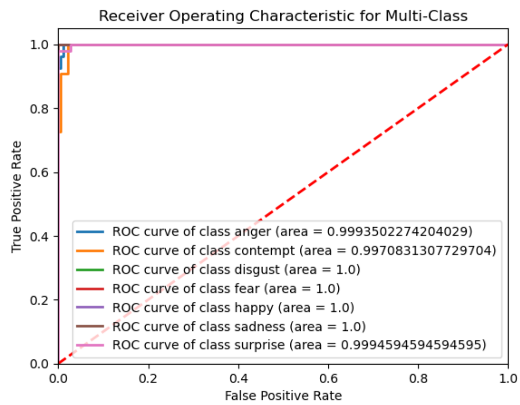

Figure 16: *ROC curve for Multi-class classification*

Figure 16 shows the ROC curve for multi-class classification. The suggested multi-stream Inception–V3 continues to have the highest AUC value of 1 for happy, disgust, sadness and fear, and for anger, contempt, and surprise the AUC value of 0.99 demonstrating the approach.

4.2 Comparison for the Proposed Method

The comparison of proposed method with existing method is presented in table 1. The MA-

Net have the classification accuracy of 88.42% [16] and ResNet-50 have the accuracy of 96% [17]. The proposed Multi-stream Inception-V3 have the higher accuracy of 97% compared to the existing method in table 1. The proposed Multi-stream Inception-V3 have better performance for facial expression.

Table 1: *Comparison for the proposed method*

Study	Method	Accuracy
Zhao *et al* [16]	MA-Net	88.42%
Li *et al.*[17]	ResNet-50	96%
Proposed approach	Multi-stream Inception-V3	97%

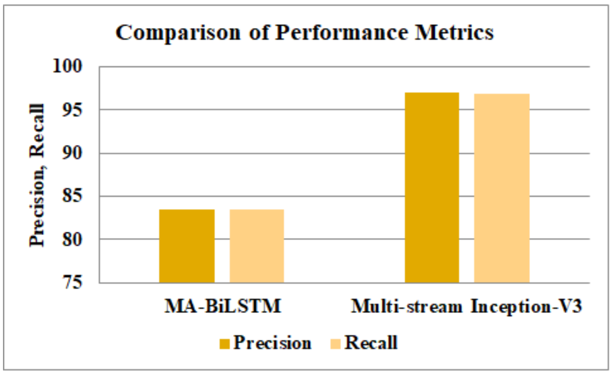

Figure 17: *Comparison of precision and recall*

Figure 17 shows the comparison for Precision and recall value proposed work. The MA-BiLSTM have the precision value of 83.42% and recall value of 83.48% [20] and proposed Multi-stream Inception-V3 have the precision value of 96.94% and recall of 96.97% for recognising FE.

5. Conclusion

In this study, Multi-Stream Inception-V3 is proposed for the classification of facial expression image. The preprocessing stage effectively enhanced the contrasts of the face image using CLAHE and noises are removed using Adaptive bilateral filtering. Furthermore, SIFT-based feature extraction finds the key points from the facial expression images. Finally, the proposed Multi-Stream Inception-V3 classifies the facial expression with better accuracy. The performance evaluation, conducted on the facial expression dataset, demonstrates the superior capabilities of the Multi-Stream Inception-V3 related to existing methods. The improved accuracy of 97% is achieved as a greater accuracy in contrast to the existing techniques.

References

1. Hangyu Li; Nannan Wang; Xinpeng Ding; Xi Yang; Xinbo Gao, Year: 2021, "Adaptively learning facial expression representation via cf labels and distillation", IEEE Transactions on Image Processing, Vol: 30, pp. 2016 – 2028.

2. Yuanling Lv; Yan Yan; Jing-Hao Xue; Si Chen; Hanzi Wang, Year: 2024, "Relationship-guided knowledge transfer for class-incremental facial expression recognition", IEEE Transactions on Image Processing, Vol: 33, pp. 2293 – 2304.

3. Bin Jiang; Nanxing Li; Xiaomei Cui; Qiuwen Zhang; Huanlong Zhang; Zuhe Li; Weihua Liu, Year: 2024, "Research on facial expression recognition algorithm based on improved MobileNetV3", EURASIP Journal on Image and Video Processing, No: 1, p. 22.

4. Fan Zhang; Gongguan Chen; Hua Wang; Caiming Zhang, Year: 2024, "CF-DAN: Facial-expression recognition based on cross-fusion dual-attention network", Computational Visual Media, Vol: 10, No: 3, pp. 593 – 608.

5. Weicheng Xie; Zhibin Peng; Linlin Shen; Wenya Lu; Yang Zhang; Siyang Song, Year: 2024, "Cross-Layer Contrastive Learning of Latent Semantics for Facial Expression Recognition", IEEE Transactions on Image Processing.

6. Rongkang Dong; Kin-Man Lam, Year: 2024, "Bi-center loss for compound facial expression recognition", IEEE Signal Processing Letters, Vol: 31, pp. 641– 645.

7. Jiawei Mao; Rui Xu; Xuesong Yin; Yuanqi Chang; Binling Nie; Aibin Huang; Yigang Wang, Year: 2024, "Poster++: A simpler and stronger facial expression recognition network", Pattern Recognition, p. 110951.

8. Afolabi I. Awodeyi; Omolegho A. Ibok; Idama Omokaro; Jones U. Ekwemuka; Michael O. Ighofiomoni, Year: 2025, "Effective Preprocessing Techniques for Improved Facial Recognition under Variable Conditions", Franklin Open, p. 100225.

9. Yuntao Zhou; Thiyaporn Kantathanawat; Somkiat Tuntiwongwanich; Chunmao Liu, Year: 2025, "High dynamic range preprocessing, ParallelAttention Transformer and CoExpression analysis for facial expression recognition", Computers and Electrical Engineering, Vol: 123, p. 110110.

10. Sachinkumar Veerashetty; Virupakshappa; Ambika, Year: 2024, "Face recognition with illumination, scale and rotation invariance using multiblock LTP-GLCM descriptor and adaptive ANN", International Journal of System Assurance Engineering and Management, Vol: 15, No: 1, pp. 174 – 187.

11. Nitin Arora; Ishan Budhiraja; Deepak Garg; Sahil Garg; Bong Jun Choi; M. Shamim Hossain, Year: 2025, "Revolutionizing facial image retrieval: Multi-block and mean based local binary patterns with sign and magnitude analysis", Alexandria Engineering Journal, Vol: 116, pp. 601– 608.

12. Soumya Ranjan Mohanta; Karan Veer, Year: 2024, "Soft computing based comparative model for the classification of facial expression recognition", Wireless Personal Communications, Vol: 136, No: 4, pp. 2573 – 2594.

13. Rit Lawpanom; Wararat Songpan; Jakkrit Kaewyotha, Year: 2024, "Advancing facial expression recognition in online learning education using a homogeneous ensemble convolutional neural network approach", Applied Sciences, Vol: 14, No: 3, p. 1156.

14. Huilin Ge; Zhiyu Zhu; Yuewei Dai; Biao Wang; Xuedong Wu, Year: 2022, "Facial expression recognition based on deep learning", Computer Methods and Programs in Biomedicine, Vol: 215, p. 106621.

15. Carmen Bisogni; Aniello Castiglione; Sanoar Hossain; Fabio Narducci; Saiyed Umer, Year: 2022, "Impact of deep learning approaches on facial expression recognition in healthcare industries", IEEE Transactions on Industrial Informatics, Vol: 18, No: 8, pp. 5619 – 5627.

16. Zengqun Zhao; Qingshan Liu; Shanmin Wang, Year: 2021, "Learning deep global multi-scale and local attention features for facial expression recognition in the wild", IEEE Transactions on Image Processing, Vol: 30, pp. 6544 – 6556.

17. Bin Li; Dimas Lima, Year: 2021, "Facial expression recognition via ResNet-50", International Journal of Cognitive Computing in Engineering, Vol: 2, pp. 57– 64.

18. Yahui Nan; Jianguo Ju; Qingyi Hua; Haoming Zhang; Bo Wang, Year: 2022, "A-MobileNet: An approach of facial expression recognition," Alexandria Engineering Journal, Vol: 61, No: 6, pp. 4435 – 4444.

19. Qingzhen Gong; Xuefang Liu; Yongqiang Ma, Year: 2025, "Real-Time Facial Expression Recognition Based on Image Processing in Virtual Reality", International Journal of Computational Intelligence Systems, Vol: 18, No: 1, p. 8.

20. Ahmad A. Alzahrani, Year: 2024, "Bioinspired image processing enabled facial emotion recognition using equilibrium optimizer with a hybrid deep learning model", IEEE Access, Vol: 12, pp. 22219 – 22229.