International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I10P212

Volume 70 Issue 10, 98-110, October 2022
© 2022 Seventh Sense Research Group®

Original Article

Emotion Understanding from Facial Expressions using Stacked Generative Adversarial Network (GAN) and Deep Convolution Neural Network (DCNN)

T. Kujani[1], V. Dhilip Kumar[2]

[1,2]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Tamilnadu, India.

[1]Corresponding Author : kujani@veltech.edu.in

Received: 30 July 2022 Revised: 29 September 2022 Accepted: 06 October 2022 Published: 19 October 2022

Abstract *- Facial expressions play a vital role in nonverbal communication in understanding the behavior of human beings. Face recognition applications are applied in several fields, including authentication through biometrics, security-enhancing applications, controlling automobiles, finding mental health, in a personal interview to understand the personality, and detecting driver drowsiness. Despite several advancements, face recognition remains challenging due to variances in the effects of the images, such as illumination variation, poses, occlusions, and facial expressions. Although many techniques, such as iris and fingerprint scanning, yield prominent accuracy, face recognition is an admirable technique that is applied for many real-time recognitions and is a human-centric identification method. In this paper, facial emotion recognition with Deep Convolution Neural Network (DCNN) is used for evaluating the face expressions considering the advantage of extending the training dataset using Generative Adversarial Networks (GAN) and traditional augmentation methods. This work has explored face behavior detection using emotion identification in real-time video surveillance using combined GAN and Deep CNN. The one significant part of this work is the custom-developed deep convolution layers with suitable optimizers. Using the proposed system, the basic human expressions which play a major role in behavior understanding can be classified effectively, irrespective of gender, facial orientation, race, and age, using GAN and DCNN. The FER2013, CK+, and Custom datasets were used in experiments, and the obtained performance was compared to that of cutting-edge techniques.*

Keywords *- Behavior**,** Classification, Convolution neural network, Expression, Generative adversarial network.*

1. Introduction

One of the essential methods of expressing emotions through nonverbal communication is facial expressions. Nonverbal behavior plays a vital role in communicating states, including emotions and impressions. Human emotions can be effectively sensed using facial behavior. With the advancement of computers, facial expression recognition plays a significant role in various applications, such as the modern medical field, including health care [8] and computer-based security. A sufficient number of researches have been carried out for recognizing expressions that help to recognize and classify the basic emotional expressions from the facial pictures [11], such as fear, anger, sadness, surprise and disgust.

The basic concepts behind facial expression recognition involve image processing, extraction of features, and classification of facial expressions. Recognizing the human face plays a significant part in behavior analysis since most of the features can be extracted from facial features. Early existing facial expression recognition methods involve facial feature extraction algorithms based on feature point location, such as Local Binary Pattern [17], Gabor Wavelet [16], and multi-feature fusion. These extraction methods may lose some important information and fail under lighting conditions. The advancement of deep neural networks helps to solve the existing issues and learns the features automatically with high accuracy and recognition rate. Yet the problem that arises in these networks is the tendency of overfitting due to an increase in the parameters and increasing layers. Another prominent issue is the lack of a dataset for facial recognition. Data augmentation provides a solution to solve this issue by increasing the samples of a dataset to provide better training accuracy. Existing image augmentation techniques such as image rotation, shifting, flipping, image noising, blurring, and histogram equalization techniques are used to train deep learning models. Most of the augmented images are comparable to the sample images, which is a common issue with these techniques. Thus, sample similarity remains a prominent challenge in the existing augmentation methods. Data Augmentation techniques are required to make the deep neural network perform better. Existing research involving Convolution Neural networks and deep learning has proved with good accuracy in the area of emotion recognition by considering public datasets, but exploring the real-time dataset is still lagging in the field of emotions.

The custom dataset is developed for seven basic emotions to address this issue. The difficulty in creating the dataset is finding the right expression for each label and pre-processing the gathered images. It is easy to obtain the samples for happy, surprise, neutral, and anger but on the other hand, trying to capture the emotion of sadness, disgust correctly, and fear was a real challenge. So, more variations for these three emotions were included in the dataset. Thus, in this work, augmentation using GAN [10] is implemented, which overcomes the similarity issues. The reasons behind using GAN are that they provide super-resolution images that can be used as training data. This work investigates a nonverbal action, particularly the automatic identification of facial emotions, even in low-resolution environments. The major contributions of the proposed work are summarized as follows:

1. Custom dataset creation and pre-processing followed by face augmentation using GAN and training using DCNN for emotions prediction.
2. Experiment with the Custom dataset with and without GAN augmentation for training.
3. Comparing the performance of the CK+ dataset, FER2013 dataset, and custom dataset with the lightweight CNN Model and Deep CNN Model.
4. Accuracy of the proposed model is compared with transfer learning models.

2. Related Works

Radford et al. [1] showed that GAN could represent unsupervised learning and named the architecture DCGAN. A trained discriminator was used for classification tasks. Generators were developed to create the semantic quality of generated images. No pre-processing was done to the trained images, and the optimizer used was mini-batch Stochastic Gradient Descent (SGD). Model instability was observed when longer training collapsed to a subset of filters into a single oscillating mode.

Kosti et al. [2] proposed a CNN-based approach to predict a person's emotions. They utilized an emotic dataset which includes two different types of emotional representations. (i) 26 discrete categories of emotions, (ii) dimensions of arousal, valance, and dominance. Explained how the scene context provides prominent information to understand the human emotional states. They considered only visual information in this work. Understanding the context of the scene helps to categorize the emotions.

Dharanya et al. [3] have introduced Auxiliary Classifier GAN for regenerating the ten facial expressions, including contempt, embarrassment, and pride, in addition to seven basic emotions. The U-net architecture has been used in the generator model, and Capsule Network has been applied in the discriminator model. U-net architecture follows an encoder-decoder structure for image segmentation. The image is compressed to a small dimension at the final stage of encoding, and when it comes to decoding, it is brought back to the original dimension. The reason for using a capsule network is that it can work well even with small datasets, which is an issue in present CNN; the ADFES-BIV dataset was used for implementation and achieved 93.4% accuracy.

The six basic emotions, such as anger, disgust, fear, happiness, sadness, and surprise, were recognized using body poses. Schindler et al. [4] experimented under controlled conditions on a dataset with arbitrary poses.

Wang et al. [5] proposed using GAN to use the facial expression recognition method by combining residual networks with image processing techniques. They achieved an accuracy of 72.8% using the FER2013 dataset. They used continuous training between the generator and discriminator networks to extract the accurate dataset features by combining the residual network with image processing methods.

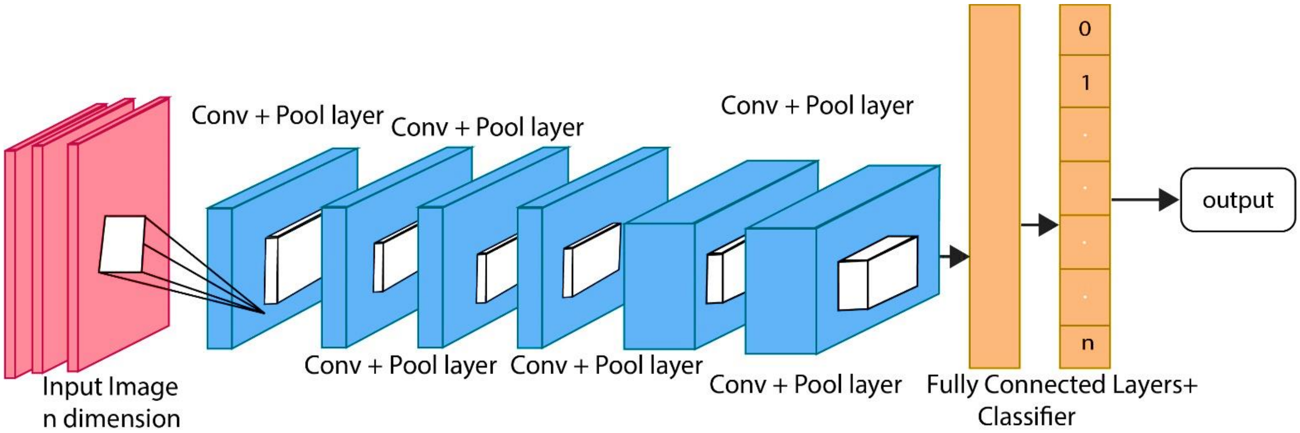


Fig. 1 Basic Architecture of CNN

Caramihale et al. [19] proposed a system for emotion classification. Images from different emotional classes were used to generate new realistic images. Facial key points were connected with the face center to determine the class of interest. Zhu et al. [12] proposed a method based on GAN to classify hyperspectral images, which overcomes the problem of overfitting. Kahou et al. [20] proposed a combined model based on ConvNet, Deep Belief Network(DBN), and a shallow network for predicting the emotion from the video clips. They combined the two different datasets to overcome the problem of overfitting.

Sajjad et al. [21] analyzed human behavior by considering the facial expressions from videos of television series. They employed the Viola-Jones algorithm for facial expression recognition, followed by both SVM for classification and CNN for facial recognition. Only traditional data augmentation techniques were applied for issues related to lighting conditions. Quinn et al. [22] performed facial emotion classification on CK+ and FER2013 datasets and achieved an accuracy of 66.67% on the FER2013 dataset. Data augmentation such as random rotation, flipping and cropping were used. Their focus was to improve the system for real-time facial recognition and make the dataset adapt to real-time conditions such as less lighting and noisy backgrounds.

Overfitting in facial recognition applications using neural networks is a common issue addressed in the proposed work of Chen et al. [12] using GAN, where an improved discriminator classifier was applied. Constraint Circular Consensus Cycle GAN was applied for face expression data augmentation to overcome the one-to-many mapping relationship. Liu et al. [15] proposed a boosted classifier for categorizing facial appearance using a deep belief network. A strong classifier was developed based on three training phases for facial expression, including learning and selecting features, followed by classification to obtain the effective features to classify the expression. Toisoul et al. [18] proposed an approach to represent the dimensional measure, such as valance and arousal emotions from the given input face. Their contribution was towards the combined prediction of both discrete and continuous emotions. Ozdemir et al. [23] proposed classifying emotions by combining images from public datasets. They trained a LeNet Architecture system for seven basic emotions. Haar cascade was applied for face detection and achieved 96.43% training accuracy. To improve the teaching skills, an automatic assessment of the instructor was accomplished using an emotion recognition system by incorporating Regularized Extreme Learning Machine (RELM) in which the system works based on a feed-forward approach, Bhatti et al. [24].

The following difficulties are identified in the field of facial emotion recognition based on an analysis of prior work:

- Only person-dependent databases are considered for evaluating approaches.
- The contribution to the recognition of obscured facial expressions is minimal.
- The work cited in the various publications only addresses one or two databases.
- The problem of expression recognition in low-resolution environments is still largely unresolved.

3. Materials and Methods

3.1. Convolution Neural Network

Convolution neural networks have achieved good performance in applications dealing with the processing of images [6]. The main intuition behind using CNN-based deep learning methods is that they automatically can extract features using large-scale image data and also, and they can perform well in any complex scenes. The CNN model is a deep feed-forward model that updates parameters through backpropagation. To achieve good performance, it is necessary to design the convolutional layer and pooling layer cores, which will be combined to achieve better image characteristics. Fig.1 shows the basic architecture of the Simple Convolution neural network.

Any lightweight ConvNet consists of three main layers, Convolution, Pooling and a Fully Connected Layer. When all these layers are stacked together, they form a full Convolution network. The convolution layer is the core part of the network, which does the major computation. Each convolution layer consists of learnable filters with a specific width, height and depth. These filters will convolve across the input volume in each forward move and perform the dot product between the input and filter values. These layers produce a two-dimensional activation map. For experimental purposes, the input size of [48,48,3] is considered, and if the first filter size is (5,5), then we will get a vector size of 5*5*3=75. The next important layer is the Pooling layer, which is used to reduce the dimension of feature maps, reducing the amount of computation performed inside the network. The preferred type of pooling is max pooling among the average pooling and global pooling. The activation functions used decides whether the neuron needs to be activated or not just by performing the weighted sum with added bias. The last layers in CNN are fully connected, just the feed-forward neural networks, which flatten the output coming from the previous layers. In Deep CNN(DCNN), more layers enrich the level of features. The initial layers are used for low-level feature extraction, and the last layers are for high-level extraction of features.

3.2 Data Augmentation

Data augmentation is the process of increasing the amount and diversity of data in our dataset. Since deep learning requires a huge amount of data, this is an integral part. The traditional ways to perform augmentation are either by position or by color. Some positional augmentations are scaling, rotation, zooming, cropping, padding, flipping, and changes in Brightness, Contrast, Saturation, and hue, which are examples of color augmentation. Data Augmentation can be performed using the traditional way as well as using GAN, which is described in the following section.

3.2.1 Generative Adversarial Network (GAN)

GAN was developed in 2014 by Ian Good Fellow[10] and proposed to alternate the Variational Auto Encoder (VANs) that supports generating synthetic images where the images cannot be distinguishable from the real ones. GAN can create input images by itself using the neural network instead of human input. Applications on GAN include translation of images, generating anime characters, generating new photos, creating a super-resolution image from low-resolution images, performing Pix2Pic covering grey to color conversion, aerial to map output, day picture tonight the Face ageing application.

Generator and discriminator networks are the two major networks that underpin the GAN. These networks are competitive and hostile to one another. GAN can ultimately be used to produce synthetic images from input images. The discriminator uses the generator's input to determine if the image originated from the real dataset or was altered. While the Discriminator (D) works to identify fraudulent images, the generator (G) creates brand-new images. Using the minmax game, an optimization technique used in games with two players, both the G and D enhance learning. Maximizing the discriminator's loss while minimizing the loss of the generator function is the ultimate goal of the MinMax game utilized in GAN. Fig.2 shows the basic structure of GAN.

Mini Max Formulation

The training process of GAN is described by the function given below

$min_G max_D V(D,G) = E_x \sim P_{data} log D(x) +$ $E_z \sim P_z$(z)log $(1 - D(G(x)))$ When using and training GANs, two results are possible. An image synthesis system is produced if the generator is given greater attention. Otherwise, the D component can be utilized as a classifier if the generator is simply employed to produce images for the discriminator to evaluate.

Data Augmentation using GAN

A class of generative models known as GANs allows sampling the model without explicitly building one. Deep neural networks, which take noise as random input and convert it to model distribution, are used for sampling in GAN. Let us consider the training data as T, which contains n samples of x,
$T_{data}(x) \rightarrow \{x_i\}$ where, $1 \leq i \leq n$,
GAN works to develop a new model from T
$T_{model}(x) \sim T_{data}(x)$

A discriminator aims to distinguish between the false and training data. In contrast, a generator is merely a neural network that accepts random noise as input and transforms it into a sample for model distribution.

Other Techniques implemented to assist GAN

In deep learning applications, it is required to apply normalization for scaling the features. Normalization helps to scale down the features between 0 and 1. Standardization helps to scale down our features based on Standard normal distribution (SND). In SND, mean(μ) is usually 0, and SD(σ) is 1. Normalization prefers to use MinMax Scaling. Features are scaled between 0 and 1.

$F_{norm} = \frac{F - F_{min}}{F_{max} - F_{min}}$Standardization prefers to use a standard scalar library,

$$Z = \frac{x - \mu}{\sigma}$$

3.2.2. Training process of GAN

The training process of GAN is a considerably difficult task compared with other existing methods where the loss was reduced using gradient descent for changing the weights. But in GAN, changing the weights changes the complete balance of the entire system. The main intuition of GAN is not to minimise the loss but to establish the balance between the two opposing networks.

1. Generate sample images by providing randomly generated noisy images.
2. Pick some sample images from the original dataset and combine them with the generated images.
3. The combined set of images shall be given to the discriminator network, which trains the network and updates the weights dynamically.
4. The feedback from the Discriminator network is sent back to the generator to generate a better synthetic image.

Training the Discriminator Network

The discriminator performs optimally when the cost function is maximized. During initial training, there is more chance that the discriminator may miss classifying the data.

$$max_D[E_{x \sim P_{data}(x)} log\left(D(x)\right) + E_{z \sim P_z(z)} log\left(1 - D\left(G(z)\right)\right)]$$

The output of the discriminator is 1 when the input is obtained from training data, and the output is 0 when the input is from generated data.

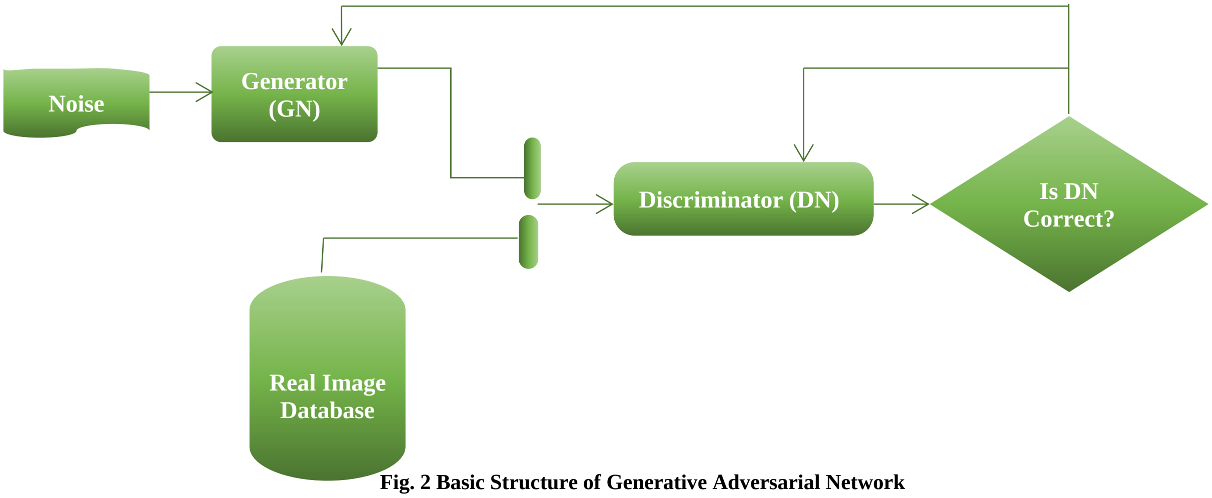


Fig. 2 Basic Structure of Generative Adversarial Network

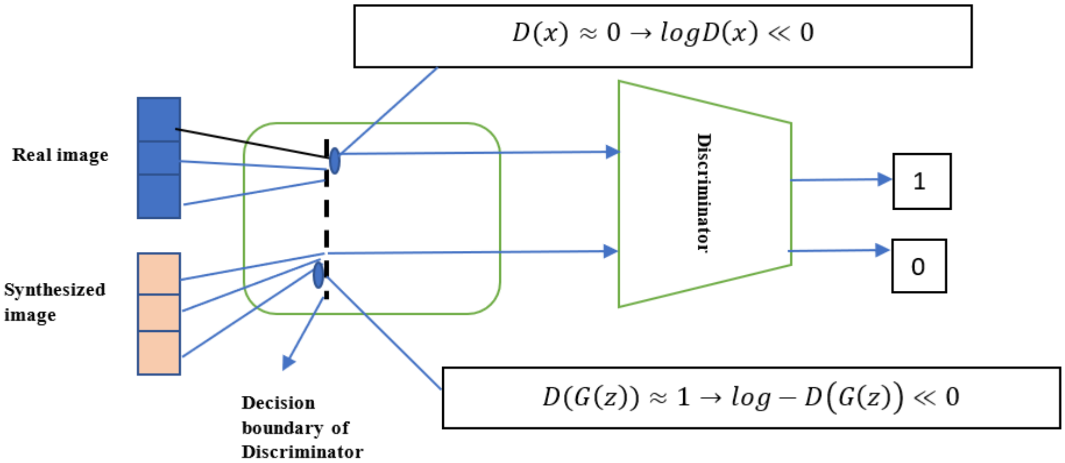

$D(x) \approx 0 \rightarrow logD(x) \ll 0$
Real image
Synthesized image
Discriminator
1
0
Decision boundary of Discriminator
$D(G(z)) \approx 1 \rightarrow log - D(G(z)) \ll 0$

Fig. 3 Training process of the discriminator

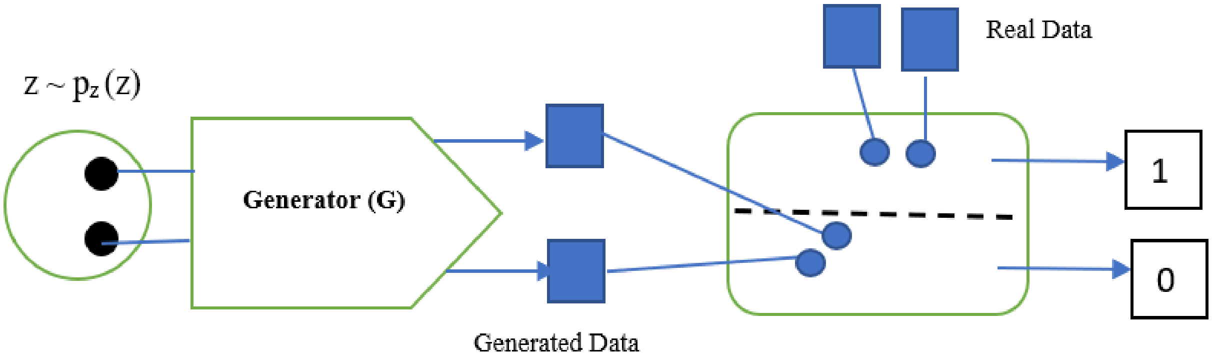


Fig. 4 Training process of the generator

According to Fig.3, one image from the real data is to the right of the decision boundary, which is miss classified, so the value of D(x) is 0. Similarly, one data from generated data is to the left of the decision boundary, so it is misclassified. Hence, the value of D(G(z)) is 1, so these cases show when the discriminator is not performing correctly. The training process of the discriminator is shown in Fig. 3.

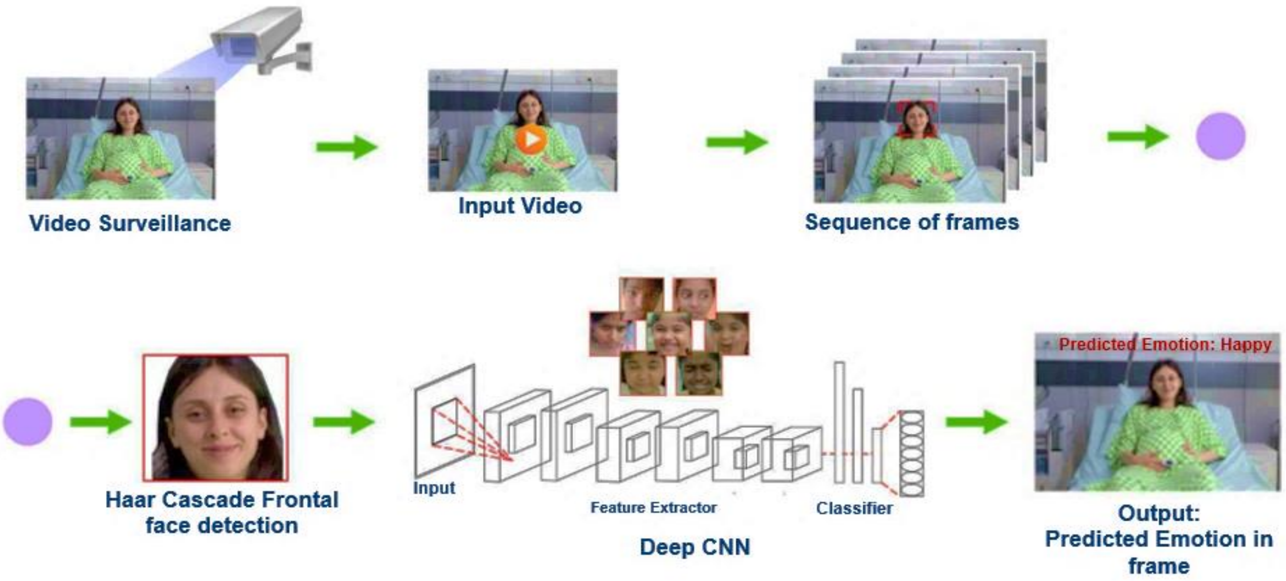


Fig. 5 Graphical Abstract flow of the proposed system

Training the Generator Network

The generator performs optimally when the cost function is minimized. The idea is to minimize the likelihood of discriminators classifying the fake data as fake. The cost function of a generator function is mentioned below:

$$min_G[E_{x\sim P(x)}log\big(D(x)\big) + E_{z\sim P_z(z)}log\left(1 - D\big(G(z)\big)\right)]$$

In the above expression, the first term is not considered since there is no G parameter; hence only the second parameter is considered.

So, it tries to minimize the cost of classifying G as 0. The discriminator will initially identify the real and fake data when the real and fake data are given to the discriminator. The training process of the generator is described in Fig. 4.

When the discriminator correctly classifies the data,

$$\mathrm{D}\big(\mathrm{G}(\mathrm{z})\big) \approx 0 \rightarrow \log\big(D\big(G(z)\big)\big) \ll 0$$

When the generator is trained well, the discriminator cannot classify the data correctly; that is, it will be able to identify the fake data correctly.

$$\mathrm{D}\big(\mathrm{G}(\mathrm{z})\big) \approx 1 \rightarrow \log\big(D\big(G(z)\big)\big) \approx 0$$

3.3. Graphical Representation of Proposed work

Fig. 5 visually shows the proposed model of the facial expression system. The first stage is receiving video input and converting it into a sequence of frames. The face in the frames is now recognized using the frontal face Haar Cascade. The detected face is fed as input to the proposed DCNN for classifying the expression.

4. Results and Discussion

This section outlines the tests carried out, the findings, and a performance evaluation of deep CNN models. The hardware and software settings used for the experiment are shown in Table 1.

Table 1. System Development Environment

Hardware Environment	Software Environment
CPU: Intel Core i5 Processor	OS: Windows 10
Graphics card: NVIDIA CUDA 11.0	Development Framework: Anaconda and Tensorflow
Memory card: 16 GB	Language: Python

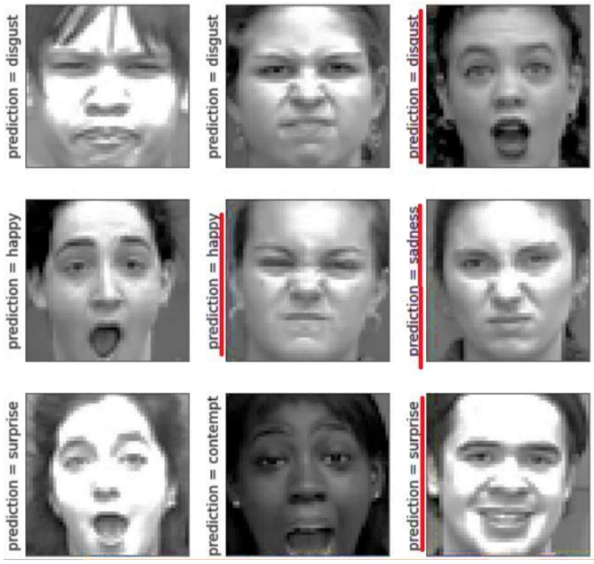


Fig. 6 Predictions from the test dataset where some misclassification samples are underlined.[13]

4.1. Experimental results using CNN for CK+ dataset

Extended Cohn-Kanade, CK+ Dataset [13] is one of the most common datasets used for emotion detection and consists of seven different expressions for Anger, Contempt, Disgust, Fear, Happy, Sadness and Surprise. For implementation purposes, CK+ Dataset is used, which has 918 overall samples, including all categories of seven emotions [25].

A Sequential CNN model has been developed for training with a training size of 85%, and a test size of 15% was considered. The input image_size used is 48x48 was used. Three convolution and max pool layers were used sequentially, followed by a flattened layer and a dense layer with 128 neurons. The final layer used is the dense layer with seven classes, and the Softmax classifier is applied for classification. An accuracy of 97.46% was obtained after 50 epochs. The details of hyperparameters used for training the model using the CK+ dataset are given in Table 2. Predictions from the test dataset with some misclassification samples are shown in Fig.6. The Loss and accuracy achieved in 50 epochs using CK+ Dataset are depicted in Fig. 7. Confusion matrix of test data in the CK+ Dataset is depicted in Fig. 8.

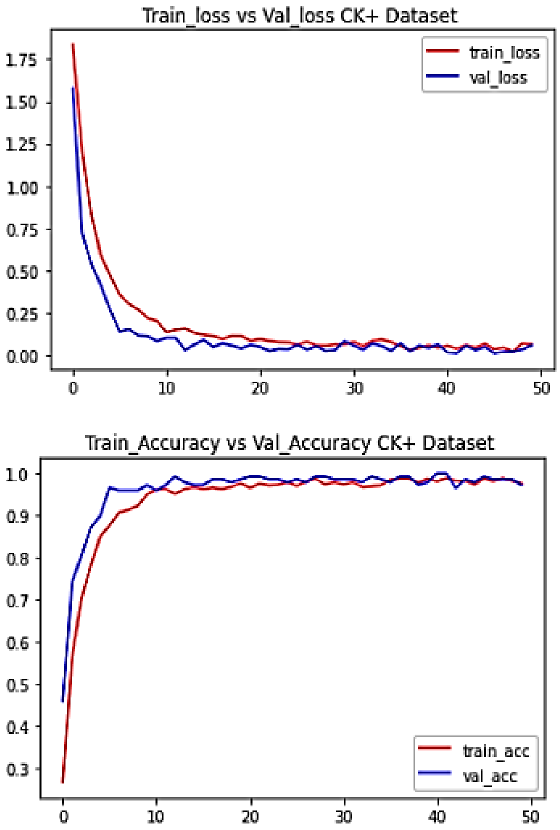


Fig. 7 Train Loss, Validation Loss, Train Accuracy and Validation accuracy using CK+ Dataset for 50 iterations

Table 2. Parameters used in training CNN model in CK+ dataset

Image_shape	48x48
Optimizer	Adam
Loss function	categorical_crossentropy
Classifier in the dense layer	Softmax
Normalization	Batch Normalization
Metrics	Accuracy
No_of_epochs	50

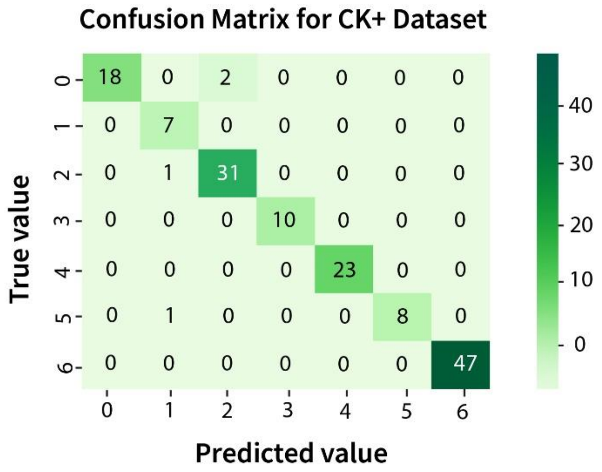


Fig. 8 Confusion matrix for emotion detection using test data of CK+ Dataset

4.2 Experimental results using lightweight CNN for the FER2013 dataset

FER2013 [26] is another facial emotion recognition dataset commonly used for research. This dataset contains labeled images for basic emotions: anger, disgust, fear, happiness, neutrality, sadness, and surprise. After using CNN, it is again proved that FER2013 has many misclassifications since the number of individuals was more. It is observed that the FER2013 dataset with lightweight CNN has low bias and high variance in accuracy and loss, showing more overfitting. The same dataset is trained with deep CNN with more Convolution layers and batch normalization to overcome the overfitting issue. The observed overfitting during training is shown in Fig. 9. Confusion matrix of test data in FER2013 is depicted in Fig. 10.

4.3. Experimental results using DCNN for the FER2013 dataset

The custom Deep CNN model is applied to FER 2013 dataset. The accuracy of the experiments has improved from 68.8% to 78.32%. The layers of Deep CNN, parameters computation, and the total number of parameters are mentioned in Table 3. Fig.11 depicts the loss and accuracy achieved using DCNN on the FER2013 dataset. The plot demonstrates minimal bias and low variance, indicating that neither overfitting nor underfitting occurred. The proposed DCNN layers have trained the model efficiently. The accuracy for 25 epochs is given in Fig. 12.

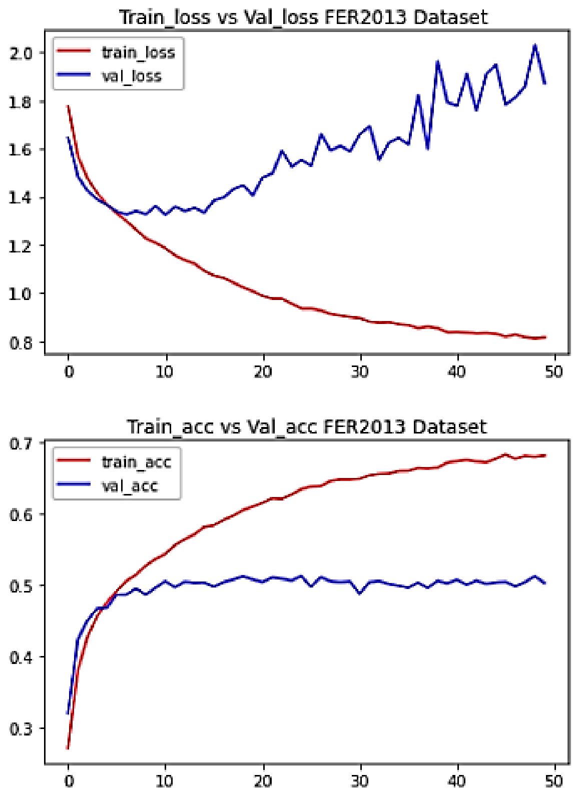


Fig. 9 Los and Accuracy using FER2013 dataset using lightweight CNN

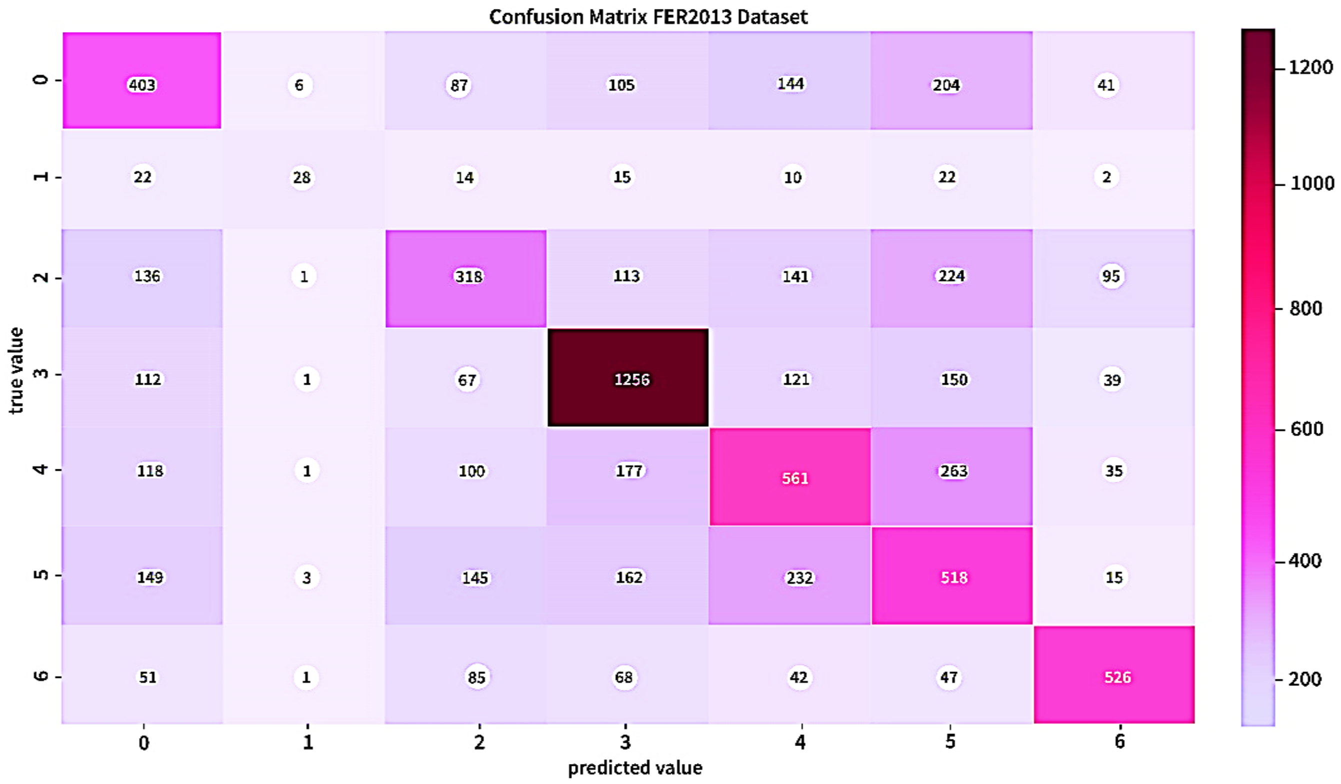


Fig. 10 Confusion matrix for using test data of FER2013 Dataset

Table 3. Details of parameters applied in training DCNN model with FER2013 dataset

Layers	Output_Shape	Parameters Computation	Total Number of Parameters
Input shape	(48,48,1)	-	-
Conv2D	(48,48,64) Filters=64, Kernel_size=(5x5)	((5 x 5 x 1) + 1) x 64 =1664	1664
Batch Normalization	(48,48,64)	256	256
Conv2D	(48,48,64) Filters=64, Kernel_size=(5x5)	((5x5x64) + 1) x 64 =102464	102464
Batch Normalization	(48,48,64)	64 x 4 = 256	256
MaxPooling2D	(24,24,64)	No parameters	0
Dropout	(24,24,64)	No parameters	0
Conv_2D	(24,24,128) Filters=128, Kernel_size=(3x3)	((3x3x64) +1) x 128=73856	73856
Batch Normalization	(24,24,128)	128 x 4 = 512	512
Conv_2D	(24,24,128) Filters=128, Kernel_size=(3x3)	((3x3x128) + 1) x 128=147584	147584
Batch Normalization	(24,24,128)	128 x 4 = 512	512
MaxPooling2D	(12,12,128)	No parameters	0
Dropout	(12,12,128)	No parameters	0
Conv_2D	(12,12,256) Filters=256, Kernel_size=(3x3)	((3x3x128) + 1) x 256=295168	295168
Batch Normalization	(12,12,256)	256 x 4 = 1024	1024
Conv_2D	(12,12,256) Filters=256, Kernel_size=(3x3)	((3x3x256) + 1) x 256 = 590080	590080
Batch Normalization	(12,12,256)	256 x 4 = 1024	1024
MaxPooling2D	(6,6,256)	No parameters	0
Dropout	(6,6,256)	No parameters	0
Flatten	9216	No parameters	0
Dense	128	((128 x 9216) +128) = 1179776	1179776
Batch Normalization	(128)	(128x4) = 512	512
Dropout	128	No parameters	0
Dense	7	((7x128) +7) = 903	903

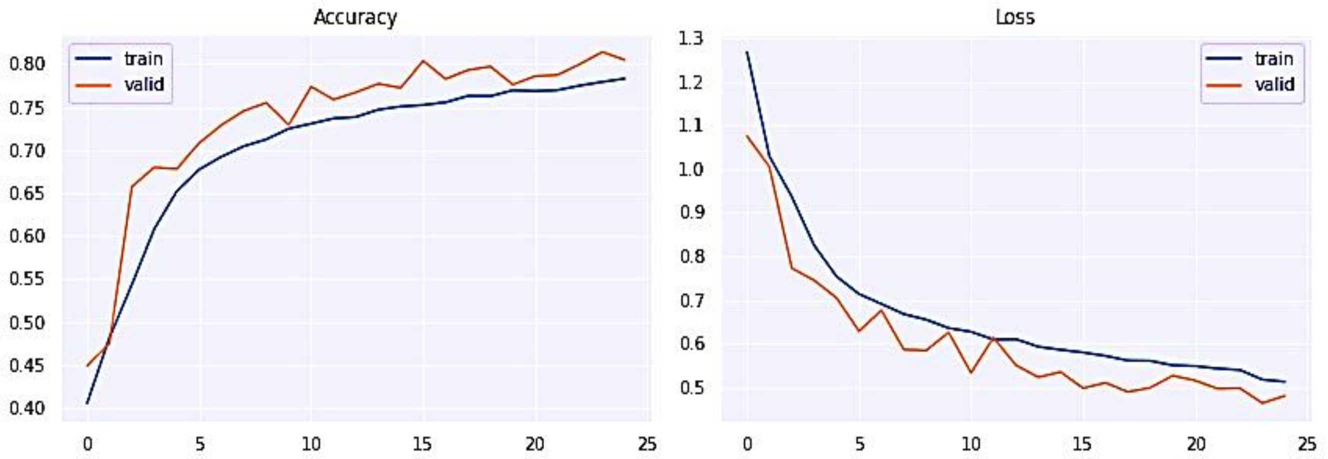


Fig. 11 Loss and Accuracy achieved using DCNN of FER2013 Dataset

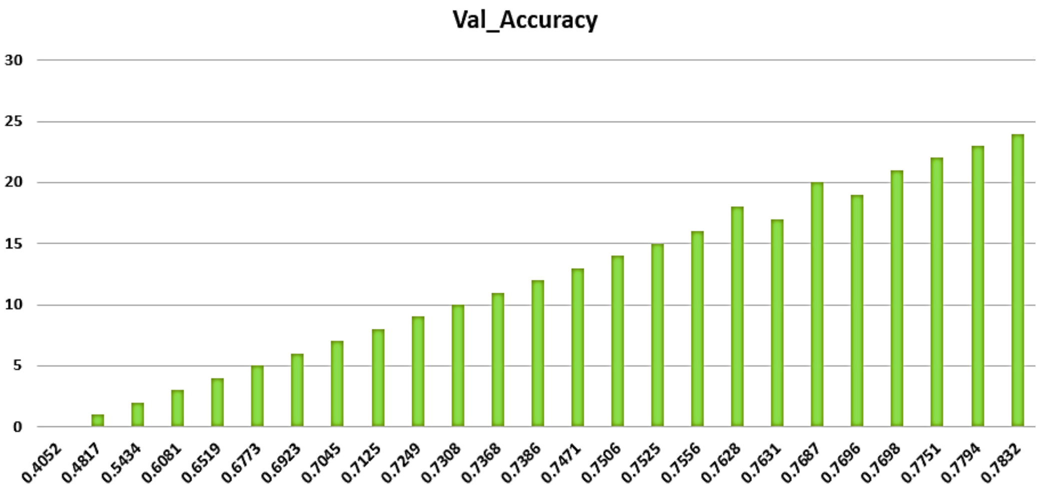


Fig. 12 Accuracy plot using DCNN

Table 4. Total Number of images in the dataset

Dataset	Train	Test	Total
Custom Dataset without augmentation	3064	767	3831
Custom Dataset with augmentation	22220	5556	27776

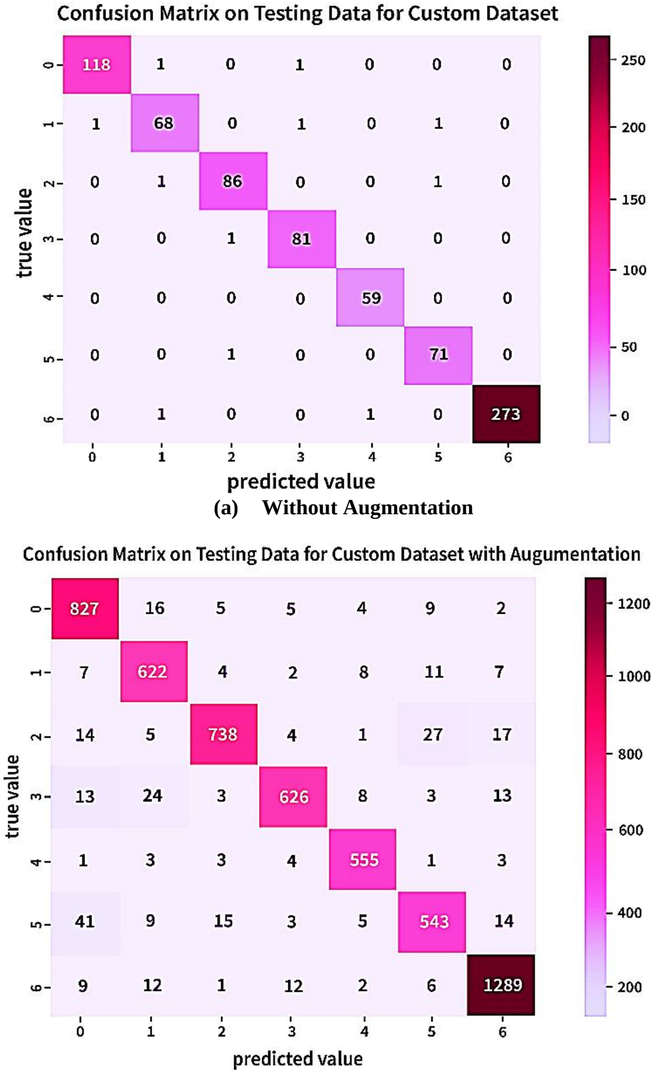
Table 5. Number of images in Custom Dataset for each expression

Category of Expression	Total Number of Images before augmentation	Total Number of Images after augmentation
Anger	563	4229
Disgust	368	3332
Fear	442	3830
Happy	382	3393
Neutral	320	2954
Sad	358	3206
Surprise	1398	6832
Total	3831	27776

4.3.1. Custom Dataset Generation / Image Acquisition

Another contribution is to create a custom dataset for facial expressions. The created dataset consists of seven basic expressions. The videos of expressions were collected with mixed variations, such as age group between 6 years to 40 years, gender including both male and female, and different race. A standard digital camera of 13 megapixels was used to capture the video. The number of images used in training with and without augmentation is given in Table 4. and Table 5.

The comparison of correctly classified testing data using Custom Dataset with and without augmentation is shown in Fig. 13. Performance metrics considered in this work are F1-Score, Precision, and Recall, which is represented in Fig. 14. The graphical representation of the performance metrics and the ROC curve which helps in identifying the correctness of classification is given in Fig. 15.



(a) Without Augmentation

(b) With Augmentation

Fig. 13. Classification matrix of Testing Data in Custom Dataset with and without augmentation

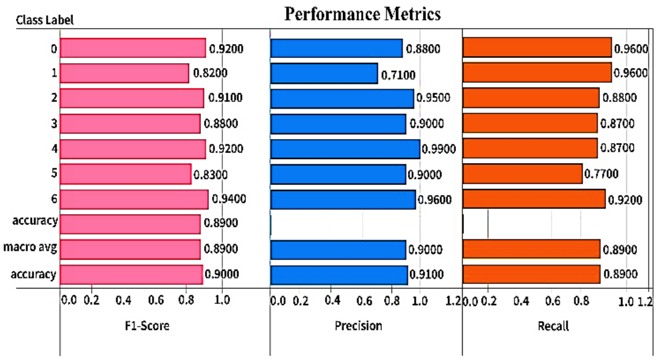
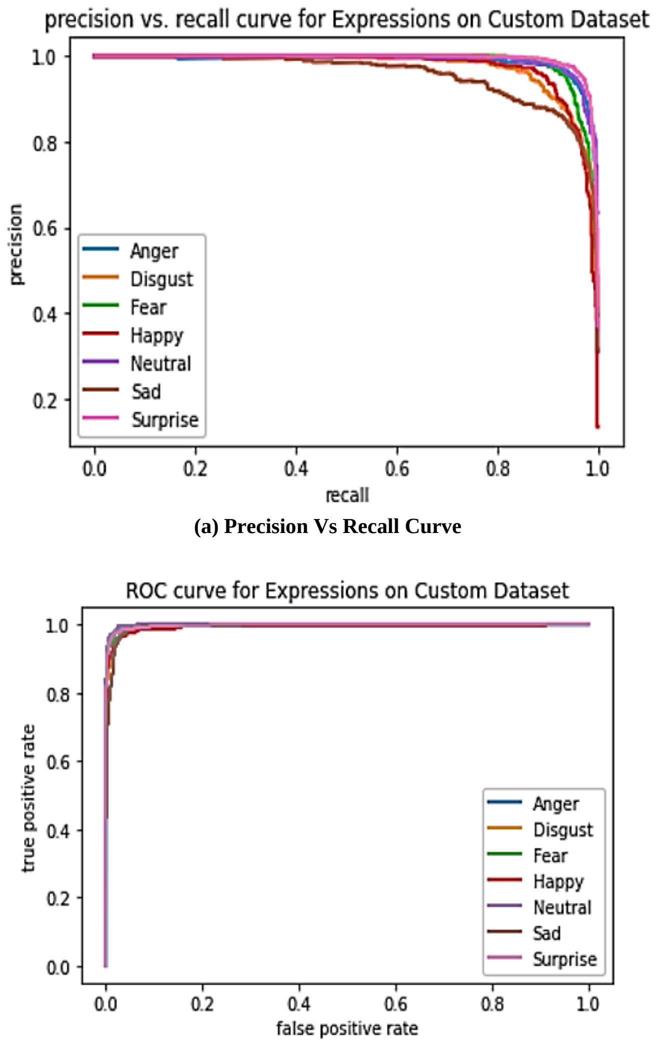


Fig. 14 Performance metrics of F1-Score, Precision and Recall

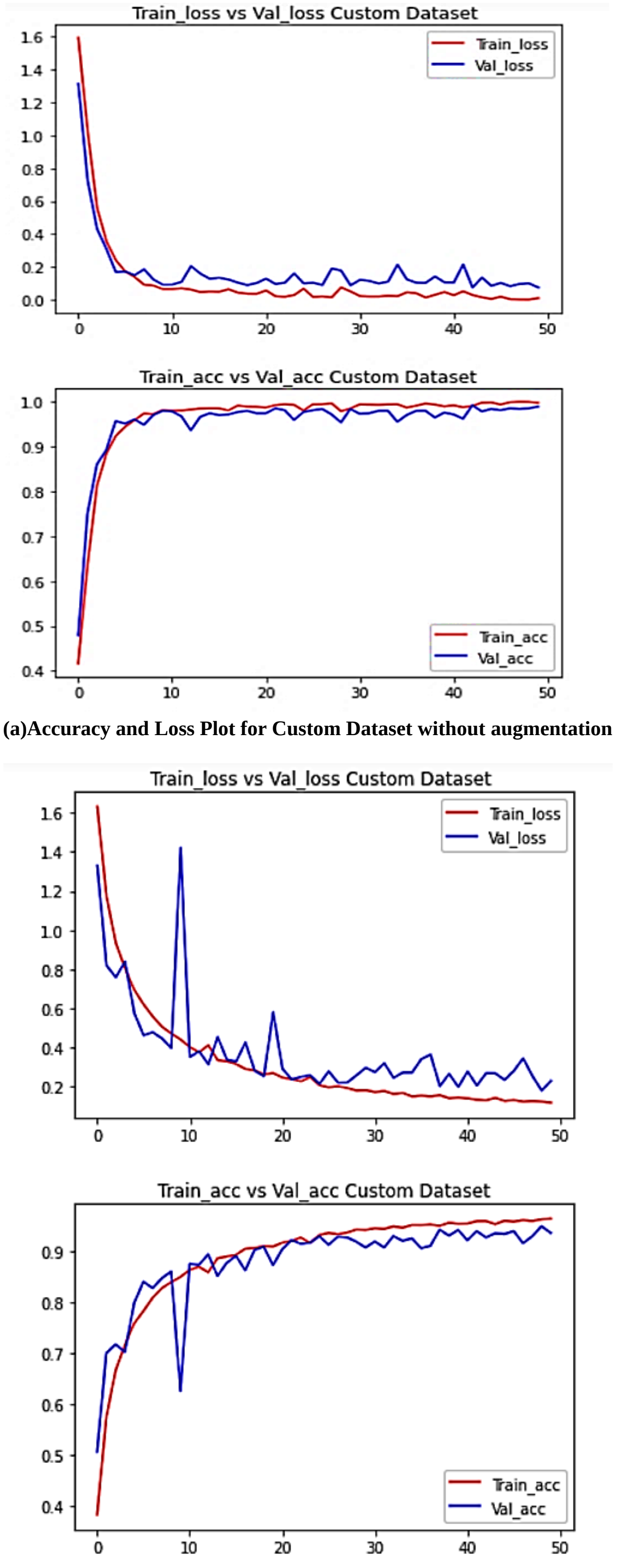


(a) Precision Vs Recall Curve

(b) ROC curve

Fig. 15 (a) Precision and Recall curve (b) ROC curve

Fig. 16, without augmentation, in 50 epochs, the training accuracy of the model was about 99%. When the model is trained using the augmented dataset, the training accuracy reaches 97% in 50 epochs with low bias.



(a)Accuracy and Loss Plot for Custom Dataset without augmentation

(b)Accuracy and Loss Plot for Custom Dataset with augmentation

Fig. 16. Achieved Accuracy and Loss plot for Custom DCNN Model on Custom Dataset

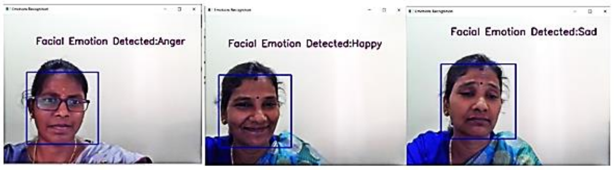

Facial Emotion Detected:Anger
Facial Emotion Detected:Happy
Facial Emotion Detected:Sad

Fig. 17 Facial Expression Results from Video Input

Table 6. Performance Accuracy Achieved

Models	**Accuracy Achieved**
Proposed Model	97%
Resnet50 Model	83%
Densenet121 Model	89%

The model developed is evaluated using a video sequence, and the sample results of detected expression are shown in Fig. 17. The proposed model for expression understanding is compared with Transfer learning pre-trained models Resnet50 and DenseNet121. The experiments show that the proposed model's accuracy is higher than the pre-trained models. The performance comparison is shown in Table 6.

5. Conclusion

Human behavior is a complex relationship of three components: actions, cognition, and emotions. Actions are observed, cognitions are thoughts and the knowledge possessed, and emotions can be understood by observing facial expressions and tracking the arousal rate. In this work, the system is developed to understand the behavior using one important type of non-verbal communication: facial expressions. Facial expression examination is one of the most prominent types of evidence to determine an individual's behaviour, which assists in identifying people's intentions. A deep convolution neural network is explored in this work to classify the basic emotions. Experiments were conducted using CK+, FER, and custom dataset as well as an augmentation to extend the dataset was performed using traditional augmentation methods and GAN. The classification accuracy of the method developed is compared for the three datasets with GAN-based image augmentation and measured performance. Custom Dataset is built with 3831 images for seven basic emotions, and then the dataset is trained with augmented images of a total size of 27776. The proposed method shows better accuracy for the custom dataset and correctly classifies the frames in a video sequence with good operating efficiency. In the future, the plan is to use the same dataset to identify the valence and arousal level, which can assist in classifying the positive and negative emotions that can be used in applications, including health care, psychology, surveillance, and other security-related applications.

Acknowledgement

The authors sincerely thank the members who contributed their expressions for dataset collection and those who helped us to complete this research work successfully.

References

[1] Radford, Alec, Luke Metz, and Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," arXiv preprint arXiv:1511.06434, 2015.

[2] R. Kosti, J.M. Álvarez, A. Recasens and A. Lapedriza, "Context based Emotion Recognition using Emotic Dataset", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI),* 2019.

[3] Dharanya V, Raj A.N.J. and Gopi V.P, “Facial Expression Recognition through Person-Wise Regeneration of Expressions using Auxiliary Classifier Generative Adversarial Network (AC-GAN) Based Model,” *Journal of Visual Communication and Image Representation*, vol. 77, pp. 103110, 2021.

[4] K. Schindler, L. Van Gool, and B. de Gelder, “Recognizing Emotions Expressed by Body Pose: A Biologically Inspired Neural Model,” *Neural Networks,* vol. 21, no. 9, pp. 1238–1246, 2008.

[5] Wang, Junhuan. "Improved Facial Expression Recognition Method Based on GAN," *Scientific Programming*, vol. 2021, 2021.

[6] T. Nishime, S. Endo, K. Yamada, N. Toma, and Y. Akamine, “Feature Acquisition from Facial Expression Image using Convolutional Neural Networks,” *Journal of Robotics, Networking and Artificial Life,* vol. 3, no. 1, pp. 9-12, 2016

[7] Chen, An, Hang Xing, and Feiyu Wang, "A Facial Expression Recognition Method using Deep Convolutional Neural Networks Based On Edge Computing," *IEEE Access,* vol. 8, pp. 49741-49751, 2020.

[8] M. Alhussein, “Automatic Facial Emotion Recognition using Weber Local Descriptor for E-Healthcare System,” *Cluster Computing,* vol. 19, no. 1, pp. 99–108, 2016. http://dx. doi.org/10.1007/s10586-016-0535-3

[9] H. Yang, Z. Zhang, L. Yin, “Identity-Adaptive Facial Expression Recognition through Expression Regeneration using Conditional Generative Adversarial Networks,” in: 2018 *13th IEEE International Conference on Automatic Face & Gesture Recognition FG 2018,* Xi’an, pp. 294–301, 2018. http://dx.doi.org/10.1109/FG.2018. 00050.

[10] Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” in *Advances in Neural Information Processing Systems,* pp. 2672–2680, 2014.

[11] G. Ali, A. Ali, F. Ali et al., “Artificial Neural Network-Based Ensemble Approach for Multicultural Facial Expressions Analysis,” *IEEE Access*, vol. 8, no. 1, pp. 134950–134963, 2020.

[12] Zhu L, Chen Y, Ghamisi P, & Benediktsson J. A, “Generative Adversarial Networks for Hyperspectral Image Classification,” *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, no. 9, pp. 5046-5063, 2018.

[13] Lucey, Patrick, et al., “The Extended Cohn-Kanade Dataset (Ck+): A Complete Dataset for Action Unit and Emotion-Specified Expression,” *Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society Conference on IEEE,* 2010.

[14] Nikhil Kumar Singh, Gokul Rajan V, "Facial Emotion Recognition in Python," *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 6, pp. 20-23, 2020. *Crossref,* https://doi.org/10.14445/23488387/IJCSE-V7I6P106

[15] Liu, Ping, Shizhong Han, Zibo Meng, and Yan Tong. ”Facial Expression Recognition via a Boosted Deep Belief Network,” *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1805-1812, 2014.

[16] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. “Comparison between Geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition using Multi-Layer Perceptron,” *In FG,* pp. 454–459, 1998.

[17] G. Zhao and M. Pietiainen, “Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 6, pp. 915–928, 2007.

[18] Toisoul A, Kossaifi J, Bulat A, Tzimiropoulos G, & Pantic M, “Estimation of Continuous Valence and Arousal Levels from Faces in Naturalistic Conditions,” *Nature Machine Intelligence,* vol. 3, no. 1, pp. 42-50, 2021.

[19] Caramihale, Traian, Dan Popescu, and Loretta Ichim, "Emotion Classification using a Tensorflow Generative Adversarial Network Implementation," *Symmetry,* vol. 10, no. 9, pp. 414, 2018.

[20] Kahou SE, Bouthillier X, Lamblin P, et al., “EmoNets: Multimodal Deep Learning Approaches for Emotion Recognition in Video,” *Journal on Multimodal User Interfaces*, vol. 10, pp. 99-111, 2016.

[21] Sajjad, Muhammad, et al., "Human Behavior Understanding in Big Multimedia Data using CNN-Based Facial Expression Recognition," *Mobile Networks and Applications*, vol. 25, no. 4, pp. 1611-1621, 2020.

[22] Quinn, Minh-An, Grant Sivesind, and Guilherme Reis. "*Real-Time Emotion Recognition from Facial Expressions*," Standford University, 2017.

[23] Ozdemir, Mehmet Akif, et al., "Real Time Emotion Recognition from Facial Expressions using CNN Architecture," *2019 Medical Technologies Congress tiptekno, IEEE,* 2019.

[24] Bhatti Y. K, Jamil A, Nida N, Yousaf M. H, Viriri S, & Velastin S. A, “Facial Expression Recognition of Instructor using Deep Features and Extreme Learning Machine,” *Computational Intelligence and Neuroscience*, 2021.

[25] P. Deivendran, P. Suresh Babu, G. Malathi, K. Anbazhagan, R. Senthil Kumar, "Emotion Recognition for Challenged People Facial Appearance in Social using Neural Network," *International Journal of Engineering Trends and Technology,* vol. 70, no. 6, pp. 272-278, 2022. Crossref, https://doi.org/10.14445/22315381/IJETT-V70I6P228

[26] Carrier P. L, Courville A, Goodfellow I. J, Mirza M, & Bengio Y, “*FER-2013 Face Database*,” University of Montreal, 2013.