**FACIAL EMOTION RECOGNITION USING CNN**

**A MAJOR PROJECT REPORT**

**SUBMITTED FOR THE FULFILLMENT OF AWARD OF THE DEGREE OF**

**Bachelor of Technology in Computer Science and Engineering**

***Submitted By***

**Satyaprakash Sethy 19BTCSE41**

**Arun Sahoo 19BTCSE08**

**Siddhant Sekhar Patra 19BTCSE43**

**Under the Supervision**

**Of**

**Dr. Prabira Kumar Sethy**

***Assistant Professor***



**Department of Computer Science Engineering and Application**

**SAMBALPUR UNIVERSITY INSTITUTE OF INFORMATION TECHNOLOGY**

**Jyoti Vihar, Burla, Odisha- 768019**

**April, 2023**



**Sambalpur University Institute of Information Technology**

**Jyoti Vihar, Burla, Odisha, Pin: 768019**

**CERTIFICATE**

This is to certify that the thesis entitled **‘Facial Emotion Recognition Using CNN’** submitted by **Satyaprakash Sethy**(**19btcse41), Arun Sahoo**(**19btcse08)** and **Siddhant Sekhar Patra**(**19btcse43)** in fulfilment of the requirements for the award of Bachelor of Technology degree in Computer Science & Engineering at Sambalpur University Institute of Information Technology is an authentic work carried out by them under my supervision. To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/Institute for the award of any Degree or Diploma.

Name and Signature of Student (s)

1. Satyaprakash Sethy

2. Arun Sahoo

3.Siddhant Sekhar Patra

**This is to certify that the above statement made by the students is correct to the best of my knowledge.**

**Dr. Prabira Kumar Sethy**

**Supervisor**

Assistant Professor

Department of Electronics and Communication Engineering

Sambalpur University Institute of Information Technology, Burla

**Counter-signed by**

**The project evaluation of the above candidates has been held on \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

|  |  |
| --- | --- |
| Dr. Sudarson Jena  HOD, Department of CSE&A | External Examiner |

**2**

**ACKNOWLEDGEMENTS**

We feel honored to avail ourselves for this opportunity to express our deep sense of gratitude to our guide **Dr. Prabira Kumar sethy, Department of Electronics and Communication Engineering**, Sambalpur University Institute of Information Technology, Burla, Odisha, India for his valuable inspiration, guidance, and warm encouragement throughout our research work. His profound knowledge and timely advice were very much helpful in our research work without which our thesis could never be able to see this day. Proud to be work under him and he has given us every freedom for working in this Project.

|  |
| --- |
| Satyaprakash Sethy 19btcse41  Arun Sahoo 19btcse08  Siddhant Sekhar Patra 19btcse43 |

**3**

**ABSTRACT**

Facial emotion recognition (FER) is an emerging and significant research area in the pattern recognition domain. In daily life, the role of non-verbal communication is significant, and in overall communication, its involvement is around 55% to 93%. Facial emotion analysis is efficiently used in surveillance videos, expression analysis, gesture recognition, smart homes, computer games, depression treatment, patient monitoring, anxiety, detecting lies, psychoanalysis, paralinguistic communication, detecting operator fatigue and robotics. In this paper, we present a detailed review on FER. The literature is collected from different reputable research published during the current decade. This review is based on conventional machine learning (ML) and various deep learning (DL) approaches. Further, different FER datasets for evaluation metrics that are publicly available are discussed and compared with benchmark results. This paper provides a holistic review of FER using traditional ML and DL methods to highlight the future gap in this domain for new researchers. Finally, this review work is a guidebook and very helpful for young researchers in the FER area, providing a general understating and basic knowledge of the current state-of-the-art methods, and to experienced researchers looking for productive directions for future work.

KEYWORDS: Facial Expression Recognition, Convolutional Neural Network, Deep Learning.

**4**

# LIST OF ABBREVIATIONS

## Abbreviation Description

CNN Convolutional Neural Network

FER Facial Expression Recognition

VGG Visual Geometry Group

**5**

**LIST OF FIGURES**

**Chapter 3 Page**

**Figure 1**: Training phase V

**Figure 2:** Testing phase V

**Figure 3:** Sample of images that are in the FER2013 dataset. VI

**Figure 4:** Architecture of our CNN model VII

**CHAPTER 4**

**Figure 4:** Confusion matrix of the model XI

**Figure 5:** Accuracy curve in training and validation phase XII

**Figure.6.** Loss curve in training and validation phase XII

**6**

# LIST OF TABLES

# Chapter 3

**Table****1:**Pseudocode of the proposed model. . VI

**Table 2:** Functions and parameters used for model during the training. IX

**7**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **CONTENTS** |  |
| **Content Details** | |  | **Page No.** |
| **Title Page** | |  | **1** |
| **Certificate by the Supervisor** | |  | **2** |
| **Acknowledgements** | |  | **3** |
| **Abstract** | |  | **4** |
| **List of Abbreviations** | |  | **5** |
| **List of Figures** | |  | **6** |
| **List of Tables** | |  | **7** |
| **Contents** | |  | **8-9** |
| **Chapter 1** | |  | **I-II** |
| **1.1** | | **Introduction** | **I** |
| **1.2** | | **Problem Statement** | **I** |
| **1.3** | | **Objective** | **II** |
| **1.4** | | **Motivation** | **II** |
| **Chapter 2** | | **Literature Survey** | **III-IV** |
| **Chapter 3** | | **Materials and method** | **V-X** |
| **3.1** | | **Methodology** |  |
| **3.2** | | **Dataset** |  |
| **3.3** | | **Architecture of CNN** | **VII-VIII** |
|  | **3.3.1** | **Input Layer** | **VII** |
| **3.3.2** | **Convolution and Pooling Layers** | **VII** |
| **3.3.3** | **Fully Connected Layer** | **VIII** |
| **3.3.4** | **Output Layer** | **VIII** |
| **3.4** | | **Data Preprocessing** | **VIII-IX** |
|  | **3.4.1** | **Data Augmentation** | **IX** |
| **3.5** | | **Functions and Parameters** | **IX** |
| **3.6** | | **Model Evaluation**  **8** | **X** |
| **Chapter 4** | | **Results and Discussion** | **XI-XIII** |
| **Chapter 5** | | **Conclusion and Future Work** | **XIV** |
|  | | **Conclusion** | **XIV** |
|  | | **Future works** | **XIV** |
|  | | **References** | **XV-XVI** |

**9**

**CHAPTER: -1**

**1.1. INTRODUCTION**

Automatic emotion recognition is a large and important research area that addresses two different subjects, which are psychological human emotion recognition and artificial intelligence (AI). The emotional state of humans can obtain from verbal and non-verbal information captured by the various sensors, for example from facial changes [1], tone of voice [2] and physiological signals [3]. In 1967, Mehrabian [4] showed that 55% of emotional information were visual, 38% vocal and 7% verbal. Face changes during a communication are the first signs that transmit the emotional state, which is why most researchers are very interested by this modality.

Extracting features from one face to another is a difficult and sensitive task in order to have a better classification. In 1978 Ekman and Freisen [5] are among the first scientific interested in facial expression which are developed FACS (Facial Action Coding System) in which facial movements are described by Action Units AUs, they are broken down the human face into 46 AUs action units each AU is coded with one or more facial muscles.

The automatic FER is the most studied by researchers compared to other modalities to statistics which made by Philipps et al. [6], but it is task that is not easy because each person presents his emotion by his way. Several obstacles and challenges are present in this area that one should not neglect like the variation of head pose, luminosity, age, gender and the background, as well as the problem of occlusion caused by Sunglasses, scarf, skin illness…etc.

Several traditional methods exist are used for the extraction facial features such as geometric and texture features for example local binary patterns LBP [7], facial action units FAC [5], local directional patterns LDA [8], Gabor wavelet [9]. In recent years, deep learning has been very successful and efficient approach thanks to the result obtained with its architectures which allow the automatic extraction of features and classification such as the convolutional neural network CNN and the recurrent neural network RNN; here what prompted researchers to start using this technique to recognize human emotions. Several efforts are made by researchers on the development of deep neural network architectures, which produce very satisfactory results in this area.

**1.2. PROBLEM STATEMENT**

Human emotions and intentions are expressed through facial expressions and deriving an efficient and effective feature is the fundamental component of facial expression system. Facial expressions convey non-verbal cues, which play an important role in interpersonal relations. Automatic recognition of facial expressions can be an important component of natural human-machine interfaces; it may also be used in behavioral science and in clinical practice.

**I**

An automatic Facial Expression Recognition system needs to solve the following problems: detection and location of faces in a cluttered scene, facial feature extraction, and facial expression classification.

**1.3. OBJECTIVE**

The objective of this project is:

1. By using Convolutional Neural Network to recognize/identify the facial emotions.

**1.4. MOTIVATION**

In this project facial emotion recognition system is implemented using convolution neural network. Facial images are classified into seven facial expression categories namely Anger, Disgust, Fear, Happy, Sad, Surprise and 'Neutral. Kaggle dataset FER2013 is used to train and test the classifier.

**II**

**CHAPTER: -2**

**LITERATURE SURVEY**

Two different approaches are used for facial expression recognition, both of which include two different methodologies, exist [10]. Dividing the face into separate action units or keeping it as a whole for further processing appears to be the first and the primary distinction between the main approaches. In both of these approaches, two different methodologies, namely the ‘Geometric based’ and the ‘Appearance-based’ parameterizations, can be used.

Making use of the whole frontal face image and processing it in order to end up with the classifications of 6 universal facial expression prototypes: disgust, fear, joy, surprise, sad ness and anger; outlines the first approach. Here, it is assumed that each of the above-mentioned emotions have characteristic expressions on face and that’s why recognition of them is necessary and sufficient. Instead of using the face images as a whole, dividing them into some sub-sections for further processing forms up the main idea of the second approach for facial expression analysis. As expression is more related with subtle changes of some discrete features such as eyes, eyebrows and lip corners; these fine-grained changes are used for analyzing automated recognition.

There are two main methods that are used in both of the above explained approaches. Geometric Based Parameterization is an old way which consists of tracking and processing the motions of some spots on image sequences, firstly presented by Suwa et al to recognize facial expressions [11]. Cohn and Kanade later on tried geometrical modeling and tracking of facial features by claiming that each AU is presented with a specific set of facial muscles [12]. The disadvantages of this method are the contours of these features and components have to be adjusted manually in this frame, the problems of robustness and difficulties come out in cases of pose and illumination changes while the tracking is applied on images, as actions & expressions tend to change both in morphological and in dynamical senses, it becomes hard to estimate general parameters for movement and displacement. Therefore, ending up with robust decisions for facial actions under these varying conditions becomes to be difficult.

Rather than tracking spatial points and using positioning and movement parameters that vary within time, color (pixel) information of related regions of face are processed in Appearance Based Parameterizations; in order to obtain the parameters that are going to form the feature vectors. Different features such as Gabor, Haar wavelet coefficients, together with feature extraction and selection methods such as PCA, LDA, and Adaboost are used within this framework.

For classification problem, algorithms like Machine learning, Neural Network, Support Vector Machine, Deep learning, Naive Bayes are used.

**III**

Raghuvanshi A. et al have built a Facial expression recognition system upon recent research to classify images of human faces into discrete emotion categories using convolutional neural networks [13]. Alizadeh, Shima, and Azar Fazel have developed Facial Expression Recognition system using Convolutional Neural Networks based on Torch model [14].

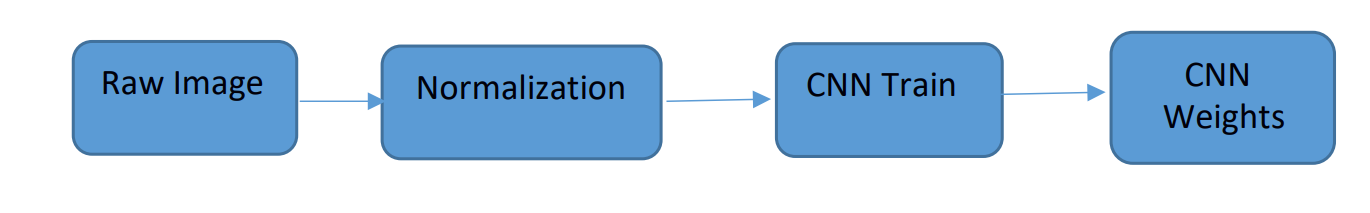
**IV**

**CHAPTER: -3**

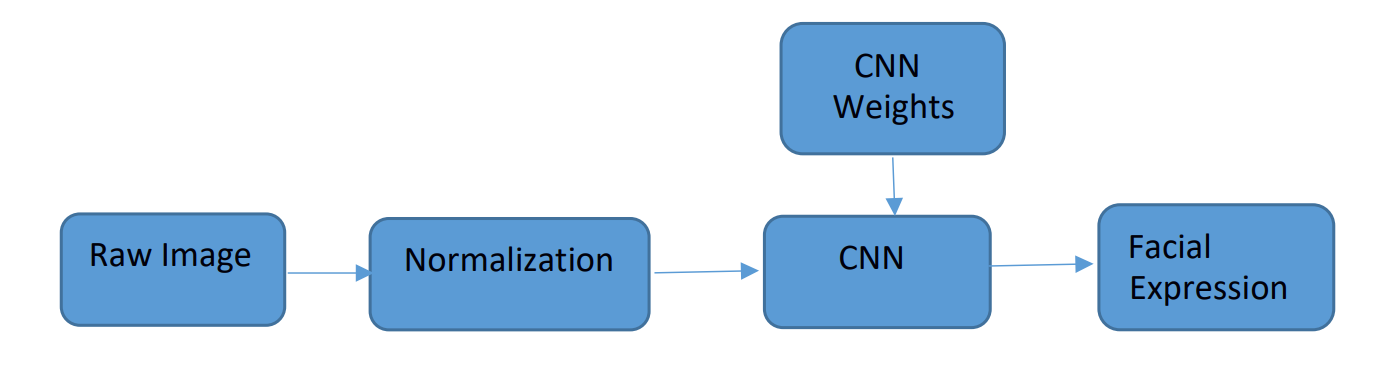
**MATERIALS AND METHODS**

**3.1. METHODOLOGY**

The facial expression recognition system is implemented using convolutional neural network. The block diagram of the system is shown in following figures.



**Fig.1.** Training phase

****

**Fig.2.** Testing phase

During training, the system received a training data comprising grayscale images of faces with their respective expression label and learns a set of weights for the network. The training step took as input an image with a face. Thereafter, an intensity normalization is applied to the image. The normalized images are used to train the Convolutional Network. To ensure that the training performance is not affected by the order of presentation of the examples, validation dataset is used to choose the final best set of weights out of a set of trainings performed with samples presented in different orders. The output of the training step is a set of weights that achieve the best result with the training data. During test, the system received a grayscale image of a face from test dataset, and output the predicted expression by using the final network weights learned during training. Its output is a single number that represents one of the seven basic expressions.

The Pseudocode of the model is in the following table.

**V**

|  |  |
| --- | --- |
| STEPS | FUNCTION |
| 1 | Start |
| 2 | Input FER2013 Dataset from Local Storage |
| 3 | Perform Pre-processing on FER2013 data |
| 4 | Normalize the dataset |
| 5 | Load the Model |
| 6 | Define the loss and optimization function |
| 7 | Train the model on the training data |
| 8  9    10 | Test the model on new/testing data  Compute the Performance and Accuracy of the proposed model by using the Performance Matrix  Finish |

**Table 1.** Pseudocode of the proposed model.

**3.2. DATASET**

The dataset from a Kaggle Facial Expression Recognition Challenge (FER2013) is used for the training and testing. It comprises pre-cropped, 48-by-48-pixel grayscale images of faces each labeled with one of the 7 emotion classes: (anger, disgust, fear, happiness, sadness, surprise, and neutral). Dataset has training set of 35887 facial images with facial expression labels. The dataset has class imbalance issue, since some classes have large number of examples while some has few. The dataset is balanced using oversampling, by increasing numbers in minority classes. The balanced dataset contains 40263 images, from which 28709 images are used for training, 3589 images are used for testing.

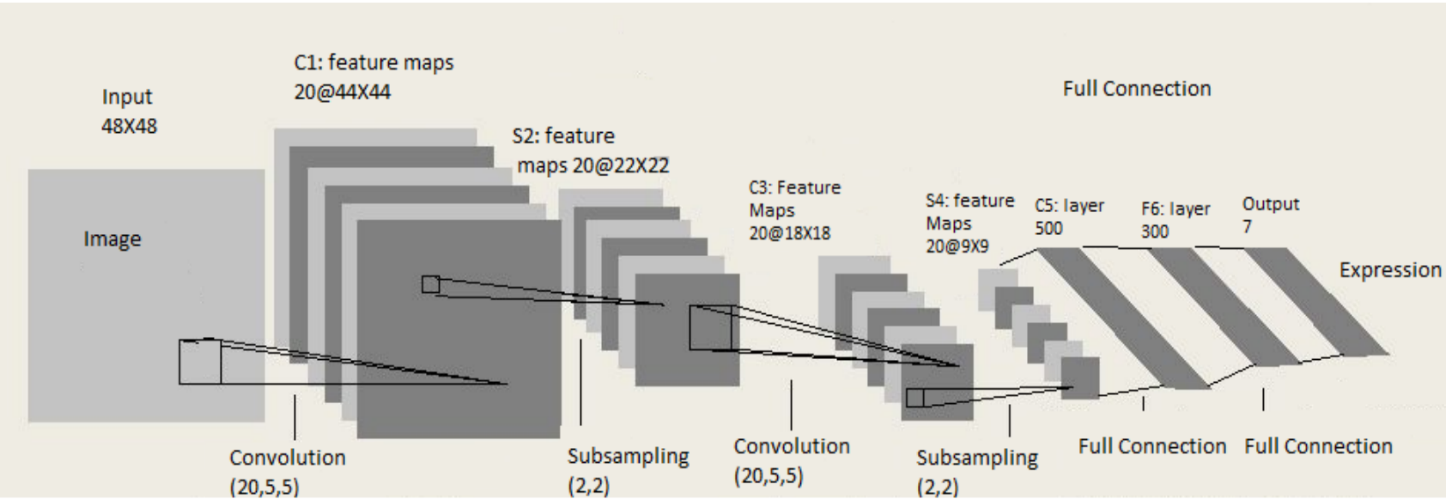


**Fig3.** Sample of images that are in the FER2013 dataset.

**VI**

**3.3. ARCHITECTURE OF CNN**

A typical architecture of a convolutional neural network contains an input layer, some convolutional layers, some fully-connected layers, and an output layer. CNN has 6 layers without considering input and output. The architecture of the Convolution Neural Network used in the project is shown in the following figure.



**Fig.4**. Architecture of our CNN model

**3.3.1. Input Layer**

The input layer has pre-determined, fixed dimensions, so the image must be pre-processed before it can be fed into the layer. Normalized gray scale images of size 48 X 48 pixels from Kaggle dataset are used for training, validation and testing. For testing propose laptop webcam images are also used, in which face is detected and cropped using OpenCV Haar Cascade Classifier and normalized.

**3.3.2. Convolution and Pooling (ConvPool) Layers**

Convolution and pooling is done based on batch processing. Each batch has N images and CNN filter weights are updated on those batches. Each convolution layer takes image batch input of four-dimension N x Color-Channel x width x height. Feature map or filter for convolution are also four dimensional (Number of feature maps in, number of feature maps out, filter width, filter height). In each convolution layer, four-dimensional convolution is calculated between image batch and feature maps. After convolution only parameter that change is image width and height. New image width = old image width – filter width + 1 New image height = old image height – filter height + 1 After each convolution layer down sampling / subsampling is done for dimensionality reduction. This process is called Pooling. Max pooling and Average Pooling are two famous pooling method. In this project max pooling is done after convolution. Pool size of (3x3) is taken, which splits the image into grid of blocks each of size 2x2 and takes maximum of 4 pixels. After pooling only height and width are affected. Two convolution layer and pooling layer are used in the architecture.

**VII**

At first convolution layer size of input image batch is Nx1x48x48. Here, size of image batch is N, number of color channel is 1 and both image height and width are 48 pixels. Convolution with feature map of 1x20x5x5 results image batch is of size Nx20x44x44. After convolution pooling is done with pool size of 2x2, which results image batch of size Nx20x22x22. This is followed by second convolution layer with feature map of 20x20x5x5, which results image batch of size Nx20x18x18. This is followed by pooling layer with pool size 2x2, which results image batch of size Nx20x9x9.

**3.3.3. Fully Connected Layer**

This layer is inspired by the way neurons transmit signals through the brain. It takes a large number of input features and transform features through layers connected with trainable weights. Two hidden layers of size 500 and 300 unit are used in fully-connected layer. The weights of these layers are trained by forward propagation of training data then backward propagation of its errors. Back propagation starts from evaluating the difference between prediction and true value, and back calculates the weight adjustment needed to every layer before. We can control the training speed and the complexity of the architecture by tuning the hyper-parameters, such as learning rate and network density. Hyper-parameters for this layer include learning rate, momentum, regularization parameter, and decay. The output from the second pooling layer is of size Nx20x9x9 and input of first hidden layer of fully-connected layer is of size Nx500. So, output of pooling layer is flattened to Nx1620 size and fed to first hidden layer. Output from first hidden layer is fed to second hidden layer. Second hidden layer is of size Nx300 and its output is fed to output layer of size equal to number of facial expression classes.

**3.3.4. Output Layer**

Output from the second hidden layer is connected to output layer having seven distinct classes. Using SoftMax activation function, output is obtained using the probabilities for each of the seven class. The class with the highest probability is the predicted class.

**3.4. Data Pre-processing**

Image pre-processing is an essential step in any computer vision task and involves operations to give the data a format suitable for training. Deep learning models are computationally expensive and require all input images to have the same shape. The dataset used in this research work contains high-dimensional images with different shapes. We reduced the size of images by reshaping them to the same dimension of 48 X 48 X 1. Moreover, the images were normalized by rescaling the pixel values between 0 and 1. Normalizing input images removes the difference in magnitude between different pixels, which benefits learning.

**VIII**

**3.4.1. Data Augmentation**

The images in the dataset were deficient and had imbalanced classes. DL models, if trained with a small dataset, results in over-fitting, and as a result, the models’ generalization capability becomes very poor. In the case of imbalanced data, the models have poor predictive performance, specifically for the minority class. We overcame these issues by using data augmentation techniques. The images in the minority class (non-cancerous) were augmented by two folds applying geometrical transformations such as horizontal-flip and vertical-flip. Since the pathologists can easily interpret the images from different angles, the flipped images are invariant. Similarly, the overall size of the data was increased by augmenting the images during training.

**3.5. Functions and Parameters**

|  |  |
| --- | --- |
| Functions/Parameters | CNN |
| Classification Function | SoftMax |
| Loss function | Adam |
| Optimizer | categorical Cross Entropy |
| Epochs | 50 |
| Horizontal Flip | True |
| Fill\_Mode | Nearest |
| Zoom | 10 |
| Rotation | 10 |
| Height shift Range | 10 |
| Width Shift Range | 10 |

**Table 2:** Functions and parameters used for model during the training.

**IX**

**3.6. Model Evaluation**

We validated the performance of models with 3589 images. To evaluate the performance, we calculated accuracy, precision, Recall, F1-score, specificity, and AUC value for the model. These statistical metrics are based on True Positives (TP), False Negatives (FN), False Positives (FP), and True Negatives (TN). Here, TP and TN represent the number of correctly identified cancerous and normal images, while FP and FN denote misclassified normal and cancerous images, respectively.

The accuracy scores tell how often the models produced correct result

Accuracy = (TP +TN)

(TP+TN+FP+FN)

Precision score determines the ratio of correctly identified cancerous images to all the images predicted by a model as cancerous. In other words, precision reflects a model’s consistency with respect to cancerous outcomes.

Precision = TP

(TP + FP)

Recall calculates the ratio of correctly identified cancerous images to all the cancerous images in the test data.

Recall = TP

(TP + FN)

Specificity performs the same operation as recall but for normal images.

Specificity = TN

(TN + FP)

F 1 score represents a weighted average of precision and recall.

F1 = 2 X (Precision x Recall)

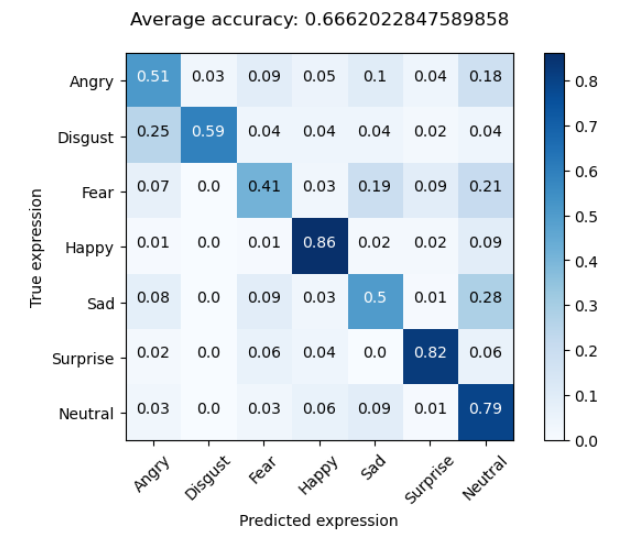
(Precision + recall)

**X**

**CHAPTER: -4**

**4.1. RESULTS AND DISCUSSION**

The testing of the model is carried out using 3589 images. The classifier provided 66.62% accuracy. The confusion matrix for seven facial expression classes is shown below: The confusion matrix shows the performance of the model (fig 5).



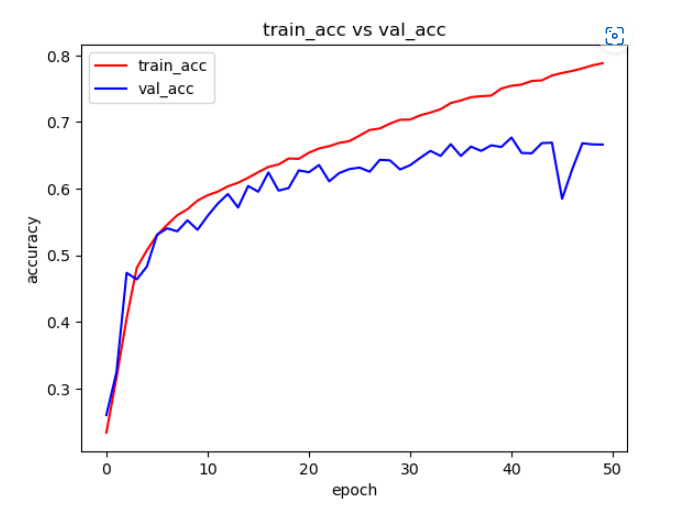
**Figure 5.** Confusion matrix of the model

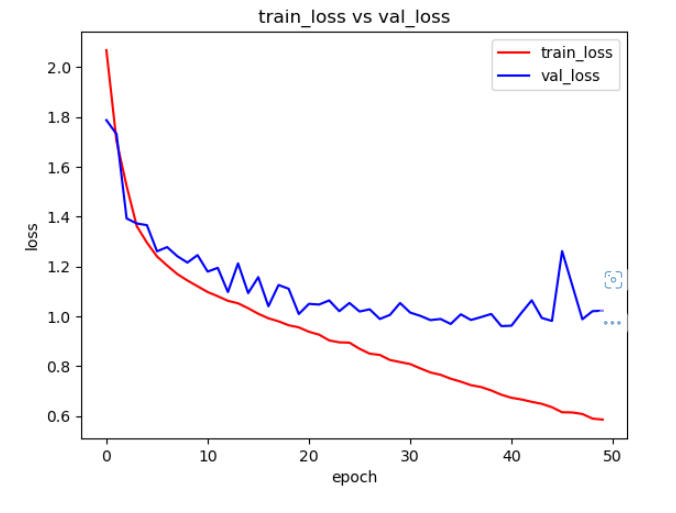
CNN architecture for facial expression recognition as mentioned above was implemented in Python. Along with Python programming language, NumPy, Pandas and CUDA libraries were used. We used a computer with Intel i5 10th processor, 8GB of RAM, 240GB of SSD, 4GB of VRAM which is a Nvidia GTX 1650 graphics card.

Training image batch size was taken as 50, while filter map is of size 20x5x5 for both convolution layer. Validation set was used to validate the training process. In last batch of every epoch in validation cost, validation error, training cost, training error are calculated. Input parameters for training are image set and corresponding output labels.

**XI**

The training process updated the weights of feature maps and hidden layers based on hyper-parameters such as learning rate, momentum, regularization and decay. In this system batch-wise learning rate was used as 10e-5 and decay as 0.99999.



**Fig.6.** Accuracy curve in training and validation phase.

**Fig.7.** Loss curve in training and validation phase.

**XII**

Above graph shows that the in 50 epochs training and validation loss is reduced 0.60 and 1.1respectively while validation cost and error is 1.25 and 0.45 respectively.

Figure-5 showed that this model’s highest accuracy for disgust emotion with 59%, followed by happy with 86 %, surprise with 82%, neutral with 79%, anger with 51%, sad with 50% and lowest accuracy for fear emotion as 41%.

The overall precision and recall are 0.57 and 0.57 respectively. The model performs really well on classifying positive emotions resulting in relatively high precision scores for happy and surprised. Disgust has highest precision and recall as 0.95 and 0.99 as images in this class were oversampled to address class imbalance. Happy has a precision of 0.68 and recall of 0.69 which could be explained by having the most examples (6500) in the training set. Interestingly, surprise has a precision of 0.69 and recall of 0.65 having the least examples in the training set. There must be very strong signals in the surprise expressions.

Model performance seems weaker across negative emotions on average. In particularly, the emotion sad has a low precision of only 0.44 and recall 0.38. The model frequently misclassified angry, fear and neutral as sad. In addition, it is most confused when predicting sad and neutral faces because these two emotions are probably the least expressive (excluding crying faces).

The overall F1-score is also 0.57. F1-score is highest for disgust due to oversampling of images. Happy and surprise have higher F1-score as 0.69 and 0.67 respectively. Fear has least F1-score as 0.39 and sad, anger and neutral also have low F1-score.

CNN Classifier is then used to classify image taken from webcam in Laptop. Face is detected in webcam frames using Haar cascade classifier from OpenCV. Then detected face is cropped and normalized and fed to CNN Classifier.

**XIII**

**CHAPTER: -5**

**5.1. CONCLUSION**

In this project, a six-layer convolution neural network is implemented to classify human facial expressions i.e., happy, sad, surprise, fear, anger, disgust, and neutral. The system has been evaluated using Accuracy, Precision, Recall and F1-score. The classifier achieved accuracy of 66.62 %, precision of 0.63, recall 0.64 and F1-score 0.63.

**5.2. FUTURE WORKS**

In the future work, the model can be extended to color images. This will allow to investigate the efficacy of pre-trained models such as Alex Net [11] or VGGNet [12] for facial emotion recognition.

**XIV**

**REFERENCES**

1. E. Sariyanidi, H. Gunes, et A. Cavallaro, « Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition », IEEE Trans. Pattern Anal. Mach. Intell., oct. 2014, doi: 10.1109/TPAMI.2014.2366127.

2. C.-N. Anagnostopoulos, T. Iliou, et I. Giannoukos, « Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011 », Artif. Intell. Rev., vol. 43, no 2, p. 155‑177, févr. 2015, doi: 10.1007/s10462-012-9368-5.

3. L. Shu et al., « A Review of Emotion Recognition Using Physiological Signals », Sensors, vol. 18, no 7, p. 2074, juill. 2018, doi: 10.3390/s18072074.

4. C. Marechal et al., « Survey on AI-Based Multimodal Methods for Emotion Detection », in High-Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet, J. Kołodziej et H. González-Vélez, Éd. Cham: Springer International Publishing, 2019, p. 307‑324.

5. M. H. Alkawaz, D. Mohamad, A. H. Basori, et T. Saba, « Blend Shape Interpolation and FACS for Realistic Avatar », 3D Res., vol. 6, no 1, p. 6, janv. 2015, doi: 10.1007/s13319-015-0038-7.

6. P. V. Rouast, M. Adam, et R. Chiong, « Deep Learning for Human Affect Recognition: Insights and New Developments », IEEE Trans. Affect. Comput., p. 1‑1, 2018, doi: 10.1109/TAFFC.2018.2890471. 7. C. Shan, S. Gong, et P. W. McOwan, « Facial expression recognition based on Local Binary Patterns: A comprehensive study », Image Vis. Comput., vol. 27, no 6, p. 803‑816, mai 2009, doi: 10.1016/j.imavis.2008.08.005.

8. T. Jabid, M. H. Kabir, et O. Chae, « Robust Facial Expression Recognition Based on Local Directional Pattern », ETRI J., vol. 32, no 5, p. 694 Wafa Mellouk et al. / Procedia Computer Science 175 (2020) 689–694 6 Author name / Procedia Computer Science 00 (2018) 000–000 784‑794, 2010, doi: 10.4218/etrij.10.1510.0132.

9. S. Zhang, L. Li, et Z. Zhao, « Facial expression recognition based on Gabor wavelets and sparse representation », in 2012 IEEE 11th International Conference on Signal Processing, oct. 2012, vol. 2, p. 816‑819, doi: 10.1109/ICoSP.2012.6491706.

10. C. Zor, “Facial expression recognition,” Master’s thesis, University of Surrey, Guildford, 2008.

11. Suwa, M.; Sugie N. and Fujimora K. A Preliminary Note on Pattern Recognition of Human Emotional Expression, Proc. International Joint Conf, Pattern Recognition, pages 408-410, 1978

12. Recognizing action units for facial expression analysis YI Tian, T Kanade, JF Cohn IEEE Transactions on pattern analysis and machine intelligence 23 (2), 97- 115

**XV**

13. Raghuvanshi, Arushi, and Vivek Choksi. "Facial Expression Recognition with Convolutional Neural Networks." Stanford University, 2016

14. Alizadeh, Shima, and Azar Fazel. "Convolutional Neural Networks for Facial Expression Recognition." Stanford University, 2016

15. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems. 2012.

16. Simonyan, Karen, and Andrew Zisserman. ”Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).

17. Dataset is fetched from the public distribution [[Dataset](https://www.kaggle.com/datasets/msambare/fer2013)]

**XVI**