e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

FACIAL EMOTION RECOGNITION USING MACHINE LEARNING

K Pavan Kumar[*1], Y Shankar Reddy[*2]

[*1,2]School Of Computer& Engineering, Lovely Professional University, Punjab India.

ABSTRACT

Facial Emotion Recognition system plays a major role in many fields, like social communication, image processing, identifying moods, etc. Detecting and understanding emotions through a camera video. In this study, the imbalanced fer2013 dataset is balanced using Random Over Sampling method to achieve the best accuracy. Convolution Neural Network (CNN) proposed methods deals with image data easily through layers and these layers play a vital role after balancing the data. The model is obtaining an accuracy of 83.0% on test data.

Keywords: Facial Emotion Recognition, Cnn, Conv2d, Max Pooling, Activation Functions, Optimizers, Confusion Matrix.

I. INTRODUCTION

Facial Emotions play a major role in human communication that helps us to know others' intentions. Generally, human communication involves both verbal and non-verbal. Non-verbal communication means expressing emotions. According to different surveys, Non-verbal communication conveys two-thirds of human communication to verbal conveys one-third [1]. Over the past decades, Facial emotions have been gaining lots of attention with their applications.

Fer2013 dataset was introduced at International Conference on Machine Learning (ICML) in 2013 and after that, it became in comparing model performance by evaluating accuracies. As per literature work [2], maximum accuracy of 65±5% is achieved on Fer2013 dataset. The fer2013 dataset is used in a CSV (Comma Separated Values) format for the facial emotion recognition system. While training CNN, require more volumes of training data. However, there are some challenges in improving the facial emotion recognition system. [3] These problems arise due to the high degree of variations occurring in faces during the expression of emotions. And a repeatedly occurring problem in data field is the lack of large and publicly available imbalanced datasets. Balancing the fer2013 dataset, play a major role in improving the accuracy of the model.

Different methodologies are proposed for distinguishing the human face in images. Common steps involved in facial emotion recognition are Image acquisition, segmentation, feature extraction and classification. Feature extraction contains two types, they are geometrical based and appearance-based. The Classification process seven expressions such as anger, sad, neutral, happy, fear, disgust and surprise [4]. In non-verbal communication, eye contact plays a major role. For each emotion, there will be different eye movements. Angry expression is expressed with squeezed eyebrows and stretched eyelids. Sad expression is expressed as rising eyebrows and frown. Smile expressed as curved as eyes. Disgust emotions are expressed in the creased nose and lowered eyebrows. Surprise is expressed as broad eyes and an opening mouth. Fear expression is expressed as eyebrows raised lips stretched, and in Neutral expression, there are no movements in facial parts [5].

In CNN algorithm, the input file is the image in two-dimensional and assigning some important weights and biases [6]. The main layer in CNN is the Convolutional layer because it performs an operation called 'Convolution'. This CNN neutral network enables machines to see the world as individuals do. For example, picture and video affirmation, Image Processing and Matching Design, Fingerprint matching, etc [7].

II. LITERATURE REVIEW

Nowadays, Facial Emotion Recognition (FER) has become more popular in research. To recognize emotions on the face, people take the help of computer vision (CV) and deep learning or machine learning technologies.

This research study [1] gives a brief overview of the FER system over the last few decades. In this research, describes deep-learning models and helps to understand deep networks. Focusing on a hybrid deep-learning approach combining a convolutional neural network (CNN) for the spatial features of an individual frame. Instead of optimizing the cross-entropy loss, learning decreases a margin-based loss.

Implementing multiple deep learning models for facial expression recognition [2]. Their work focuses not only to increase the accuracy but also to apply results to the real world. In this study, utilizing numerous techniques from recent research, they achieve 75.8% accuracy by using ensemble learning methods on the FER2013 test set, outperforming all existing publications. Additionally, they developed a mobile web app that runs and utilizes our FER models on devices in real-time.

Shervin and Amirali [3] propose a deep learning approach based on an attentional convolutional network (ANN), which is able to focus on important regions of the face, and achieves significant improvement over previous models on different datasets, including FER-2013, FERG, JAFFE, and CK+. A visualization technique is used that can find important face regions for detecting different emotions, based on the classifier's output. Through experimental results, [3] show that different emotions seem to be sensitive to different parts of the face.

Trinh and Chee sun [4] suggested how to overcome the problem of training CNNs with facial action units (AUs) and the imbalance problem of the training datasets for facial emotion classification. In this paper, they describe how to deal with imbalanced datasets and what are the problems faced while training, for example, yields a learning bias toward the major classes and it leads to a drop in the classification accuracy, it is required to increase the number of training images for the minority classes. The experiments were conducted on three imbalanced facial image datasets, RAF-DB, FER2013, and ExpW. These results demonstrate that the CNNs trained with the AU features improve the classification accuracy by 3%–4%.

Zahara [5] proposed a facial image threshing (FIT) machine. It incorporates sophisticated characteristics of trained facial recognition and Xception algorithm training. In addition to the data-augmentation methodology, the FIT machine deletes duplicate images.

In this research [6] briefly explains how emotions are identified through a video processing system. Analysing facial landmarks and their positions to identify expressions on the face. This paper describes three CNNs and combines these networks to obtain maximum accuracy. They adapted the networks to work with white and black containing facial landmarks.

Millan Tripathi [7], getting an accuracy of 91.8% on the fer2013 dataset. The maximum accuracy was achieved using the VGG19 model. The confusion matrix shows predictions based on six emotions. Various optimizers are used, among them, the best optimizer is considered based on obtaining maximum accuracy. In this study [8], proposed an architecture, that is recognizing the expression on the face dynamically. Based on movements of the eyes and mouth, emotions were recognized. The output displays how much percentage is obtained from the expression. This paper briefly explains deeply connected neural networks and how it works through computer vision.

III. MODEL AND INVOLVED STEPS

3.1 FER2013 Dataset:

FER2013 dataset has been used in competitions like ICML, Kaggle facial recognition challenge and many research papers [9]. It is a more challenging dataset with having human-level accuracy of 65 ± 5%. The highest test accuracy achieved is 75% on the FER2013 dataset. This dataset is downloadable on Kaggle and has 35,887 images in grayscale. Fer2013 is an imbalanced dataset, as it contains images of 7 emotions with the distribution of Neural (6,198), angry (4,953), fear (5,121), disgust (547), happy (8,989) and sad (6,077) [10].

3.2 Pre-Processing:

The FER2013 dataset is taken in CSV file format that consists of pixels of the images and labels. The pixel values of each image are stored in an array [11]. Fer2013 dataset is an imbalanced dataset because each class (emotion) has a different number of images, so balance the dataset using balancing multi-class methods like oversampling, and under-sampling [13].

Balancing The dataset using an oversampling method, oversampling technique increases the images in minority classes like disgust, fear and anger [14]. Increasing images based upon the number of images contained in major class here neutral having the highest number of images. Setting auto method to sampling strategy parameter, it will automatically select the major class. After balancing the dataset, normalize the pixel values from 255 to 0 – 1. While training with the direct pixels or raw data will be more complex, occupies large memory and it took more time when executing. To overcome this, normalizing the pixels to 0 and 1 [15].

3.3 Convolution Layer (Conv2D):

The convolution 2D layer is the simplest operation where you begin with a kernel, which consists of a small matrix of weights. This kernel slides by single step each time over the input data and performs matrix multiply with feature matrix, and after that summing up the results into a single output pixel [16]. In order to obtain output in the same dimension as the input image, utilize the zero-padding technique.

3.3.1 Batch Normalization:

Batch Normalization could be a layer that permits each layer of the network to do learning more freely. It is utilized to normalize the output of the past layers. The activations scale the input layer in normalization. Utilizing batch normalization learning gets efficient moreover it can be utilized as regularization to avoid overfitting of the model [18]. The layer is included to the sequential model to standardize the input or the outputs.

3.3.2 Max Polling:

In max pooling, kernel size n*n is moved over the matrix and for each position the greatest value is taken and kept at the comparing position of the output matrix.

3.4 Model Construction:

Building a CNN using five Conv2D layers, ReLU, with three 2 × 2 Max Pool layers, and with a fully connected layer and softmax activation function used at last. And adding batch normalization and 60% dropout to see desired output.

Table 1: CNN Architecture

Layer	Output Shape
Input Layer	(None, 48, 48, 1)
Conv2D	(None,46,46,32)
Conv2D	(None,46,46,64)
MaxPool2D	(None,23,23,64)
Conv2D	(None, 21,21,64)
Conv2D	(None, 21, 21, 128)
MaxPool2D	(None,10,10,128)
Conv2D	(None,8,8,128)
MaxPool2D	(None,4,4,128)
Flatten	(None,2048)
Dense	(None,200)
Dense	(None,7)

3.4.1 Optimizers

Optimizers are utilized to reduce loss function by upgrading attribute values of neural networks [19]. The optimizer utilized in our model is the **Adam** optimizer. Adam stands for Adaptive moment Estimation. Adaptive moment Estimation is used to determine adaptive learning rates for each attribute.

3.4.2 Loss Function

The loss function is utilized to estimate the difference between our prediction and the actual value present in validation data. Loss function utilized in this model is categorical cross-entropy. **Cross-entropy** loss demonstrates the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss function output value converts as the predicted probability differs from actual output.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

3.4.3 RELU Activation Function

To decrease overfitting of data activation functions are utilized. In this model, ReLu activation function is used. The most advantage of ReLu is its gradient is continuously equal to 1. Negative values in matrix input are continuously changed to zero and all other positive values stay constant.

3.4.4 SoftMax Activation Function

This function is used in the output layer to predict a multinomial (polynomial) probability distribution. Softmax is used in the case of the multi-class classification problem.

3.4.5 Dropout

Dropouts act as a regularization procedure. It decreases the chance and reliance on single neurons terminating for certain emotions. It brings in regularization by randomization. Because it is very simple to overfit with neural systems, dropouts with appropriate dropout rates were connected.

IV. RESULTS

After completing 32 epochs, Model getting an accuracy of 83% on test data and a Loss of 0.8. Implementing class weighting, which increases accuracy on misclassified emotions. Plotting graph, model accuracy to epochs in fig 1

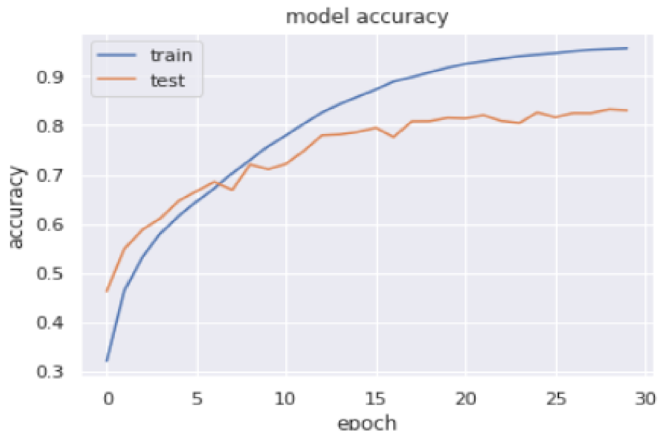

Figure 1: Model Accuracy

In fig 1, The number of epochs is increasing the accuracy of the model also increasing. Similarly, the test accuracy also increases with respect to epochs

In fig 2, Represents loss curves of train and test with respect to epochs. Loss value going to decrease towards down with respect to epochs.

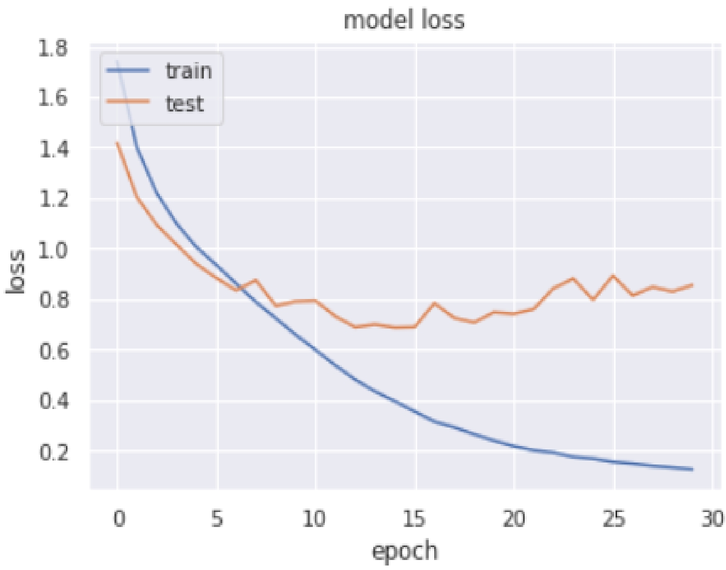

Figure 2: Model Loss

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

In the confusion matrix, Exact predicted images are compared between Truth and Predicted. Measuring the total accuracy of the model using Confusion Matrix.

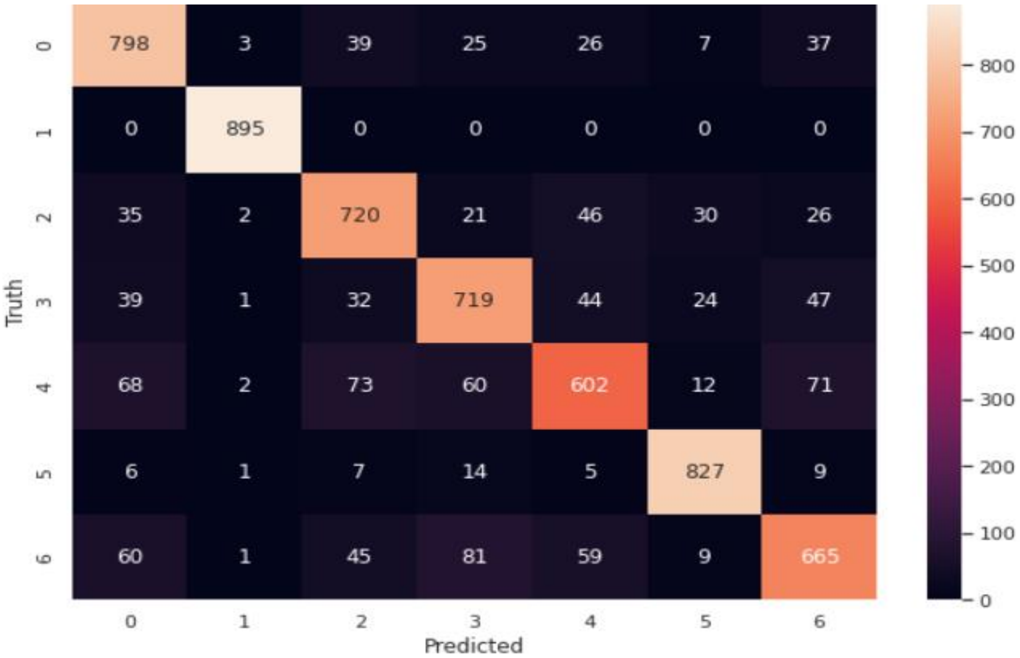

Figure 3: Confusion Matrix

V. CONCLUSION

Aiming on achieving the best accuracy on 7 emotions. In the case of disgust emotion, it contains a minimum number of images, after performing so many operations like data augmentation, rotating images, and changing optimizers there will be a slightly increased or decrease in accuracy. After balancing data with an equal number of images to each emotion using the Random oversampling technique, sudden rise in accuracy and decrease in loss, even disgust emotion also identified easily as compared to the previous. The CNN model achieved 83% test accuracy using the FER2013 dataset. To further improve in accuracy of the model by utilizing facial landmarks detection and alignment. Adapting our model into the real world [20]. Decreasing the number of misclassified emotions by improvement in pipeline models [17] and by utilizing multi-label classification to handle images with multiple possible emotion label.

VI. REFERENCES

[1] Byoung Chul Ko, "A Brief Review of Facial Emotion Recognition based on Visual Information", Sensors MDPI, Vol. 18, pp. 401 -421,2018.

[2] Amil Khanzada, Charles Bai, F.Turker.celepcikav, "Facial Expression Recognition with Deep Learning" , Stanford University-CS230 Deep Learning , pp.1-5, 2021.

[3] Shervin m, Amiral, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolution Network.", Expedia Group, University of California, Riverside, pp .1-7, Feb-2019

[4] Trinh thi doan pham and Chee sun won, "Facial Action Units for Training Convolutional Neural Networks", Dongguk University, Vol.7, pp. 77816-77822,2019

[5] L. Zahara, p. Musa, E.P.Wibowo, I.Karim, s.musa," The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi" Gunadharma University,pp. 1-8,2020

[6] Alexandru.C, Dan.p, "Facial Expression Detection by Combining Deep Learning Neural Networks" Member IEEE, university Politehnica of Bucharest, ISBN: 978-1-6654-1878, 2021

[7] Milan Tripathi, "Facial Emotion Recognition Using Convolution Neutral Network", Tribuvan University, pp. 2531-2535, 2021

[8] Praveen.R, Benjula Anbu," Emotion Recognition using Convolutional Neural Network", VIIT, ISSN: 2277-3878,

[9] Yousif.K and Zhuofa chen, "Facial Emotion Recognition: State of the Art Performance on FER2013", Boston University, arXiv:2105.03588, 2021.

[10] Jung.H.Kim,Ronald.M, Alwin.P, and Dong.S.H,"A study of a Data Standardization and Cleaning Technique for a Facial Emotion Recognition System", Kyungpook National University, 2020.

[11] I.Michael Revina, W.R sam Emmanuel, "A Survey on Human Face Expression Recognition Techniques.", Journal of King university, pp.619-628,2021.

[12] Christropher Pramerdofer, martin Kampel, "Facial Expression Recognition using Convolutional Neural Networks: State of the Art", Computer Vision Lab,TU Wein, arXiv:1612.02903v1 [cs.CV], 2016.

[13] C. John, R. Llanda, "Video Tutoring System with Automatic Facial Expression Recognition: An Enhancing Approach to E-Learning Environment", Abra State of Sciences and Technology, DOI: https://doi.org/10.1145/3321454.3321462, pp.5-9, 2019.

[14] Arnaud Dapogny, Kevin Bailly," Investigating Deep Neural Forests for Facial Expression Recognition", Sorbonne Universities, DOI 10.1109/FG.2018.0009, pp. 629-633, 2018.

[15] Subodh Lonkar, "Facial Expressions Recognition with Convolutional Neural Networks", SSN No:-2456-2165, Vol. 6, pp. 1125-1130, 2021.

[16] K. Shirisha, Mahidhar Buddha, "Facial Emotion Detection Using Convolutional Neural Network", ISSN 2229-5518, Vol.11, pp. 51-55, 2020.

[17] Shan Li, W. Deng," Deep Facial Expression Recognition: A Survey", IEEE, arXiv:1804.08348v2, pp. 1-25, 2018.

[18] Gede Putra Kusuma, Jonathan, Andreas P.L," Emotion Recognition on FER-2013 Face Images Using Fine-Tuned VGG-16", ASTESJ, Vol.5, No 6, pp. 315-322, 2020.

[19] Y.wang, Yibin Li, Yong song, Xuewen Rong," Facial Expression Recognition on Auxiliary Models" , pp.1-15,2019.

[20] Chao Li, Ning Ma, Yalin Deng,"Multi-Network Fusion Based on CNN for Facial Expression Recognition", CSECE, Vol. 80, pp. 166-169, 2018.