FACIAL EXPRESSION RECOGNITION USING CNN

Siddavatam Siva Jyothi[1], Dr. A. Rama Mohan Reddy[2]

[1]Academic Consultant, Dept of CSE, YV University, Kadapa

[2]Professor, Dept of CSE, Krishna Teja Institutes,Tirupathi

Abstract:

Automatic expression recognition based on facial expression is a fascinating study area that has been presented and utilized in a variety of fields, including safety, health, and human-machine interactions. Researchers in this subject are interested in developing strategies to understand, code, and extract facial expressions in order to improve computer prediction. Facial emotion recognition is a subset of facial recognition that is becoming in popular as the need for it grows. Though there are methods for identifying expressions using data science techniques like machine learning and deep learning, this work seeks to recognize expressions and classify them based on photographs utilizing deep learning and image classification methods. In this case we are going to use CK+48 dataset and going to build deep learning models. First we do image processing after doing image processing split the data in 80:20 or 70:30 ratio. By using train we train the model then test it on test data. In this paper we build CNN and VGG-16 model. The goal of this project based on images we need to predict face emotion. In this by using deep learning model we can predict with 98% accuracy.

Keyword: Deep learning, Image processing, CNN, VGG-16, Transfer learning.

1. Introduction: -

Traditional learning and lecture delivery methods can be improved by automating instructor expression recognition. Feedback is beneficial to instructors, but it is costly to do extensive human classroom observations, therefore feedback is infrequent. Typically, feedback is focused on evaluating performance rather than correcting outdated procedures [17]. "Student evaluation of teachers (SETs)" [19], a survey-based valuation in which student's rate separate teachers on numerous factors on a preset scale range, is one classic method. The characteristics include the instructor's knowledge of the course material, ability to present a lecture, engagement with students, delivery of lecture materials, and punctuality. Physicalcalculations may not be as consistent as they appear because students are mainly concerned with their marks, subsequent in artificialresponse. Aside from that, the procedure is time consuming, and the legality of the data obtained remains a mystery [17]. Marsh [19] wants to use self-recorded speech recognition to automate instructor feedback while presenting courses. This method makes use of the instructors'dissertation variables and encourages students learning by giving instructors objectives feedbacks for improvement. [20] Describes a real time student engagements system that delivers tailored support from teachers to students who are on the verge of dropping out. It assists instructors in allocating time to students who require the most assistance as well as enhancing instructor classroom methods. [21] Describes an intelligent tutoring system (ITS) that uses real-time teacher analytics to bridge the break in learning results for student with varying prior abilities. Lumilo is a new system that combines mixed reality smart glass with the ITS. This notifies coaches when student require assistance that the training system is impotent to offer.

A mission in computer vision and robotics' is automated face expression recognition. This is a new area of study, particularly in public signal processing and emotional calculating. The difficulty in automatic facial emotions detection is recognizing each different facial emotions and categorise into its proper emotion classifications [2]. This issue has a widespread range of application areas, such as in [2] Entertaining, teaching, bargain hunting, healthiness, and safety.

The prediction of action units and the finding of facial points are two methods for facial emotion acknowledgment [3–9]. The first technique employs a framework known as FACS's (Facial Actions Coding Systems). The framework assesses human facial emotion by analyzing changes in facial influence when an expression is evoked [10]. FACS characterizes facial muscle movement around 44 locations of the face, known as action units (AUs). As a result, the presence and strength of numerous AUs can be used to recognise face expression. There are two primary processes in facial expression: AU detection and AU recognition.

2. Relateed work

G. Cao et al. [1] used a CNN demonstrate to identify human feeling from an ECG dataset, and same demonstrate may too be utilized brain signals. On testing, the framework had an exactness of around 83 percent. The DNN demonstrate displayed by G. Yang et al [2] utilizes vectorized confront characteristics as input. With an exactness of 84.33 percent, the calculation can expect different feelings. Liu et colleagues [3] recognize particular facial expressions utilizing the for 2014 dataset and a two-layer ANN. They moreover compared it to four other existing models, finding that the recommended show had a test precision of 48.8%. S Suresh's et al. [5] displayed a signs dialect acknowledgment framework that employments a DNN to classify six distinctive sign dialects.

When two copies with diverse optimizers (Adam and SGD) are associated, it is found that the show with the Adam optimizer is more precise. K. Bouaziz et al [5] have displayed an analytics pipeline that joins picture acknowledgment apparatuses and techniques. The recommended show employments CNN engineering to classify diverse sorts of penmanship. For PolSAR pictures, F. Zhouset colleagues [8] used a profound convolutional neural arranged demonstrate to recognized is patch movement. To distinguish ships of different sizes, the demonstrate utilizes a speedier region-based CNN (Faster-RCNN) procedure. The AIRSAR dataset from NASA/JPL is utilized to confirm the show.

[9] Examines an exhaustive consider on confront feeling acknowledgment, which uncovers dataset and facial feeling recognizable proof ponder classifier characteristics. In [10], visual characteristics of pictures are explored, and certain classifier approaches are depicted, which helps within the future examination of feeling distinguishing proof frameworks. Utilizing a few classes of classifiers, this inquire about [11] examined future responses expectation from pictures based feeling distinguishing proof. [11] Employments classification strategies such as K-Nearest Neighbour and Arbitrary Woodland feelings. The utilize of neural systems to handle challenges in information science has detonated. [12] Employments a profound LSTM and a bi-directional of LSTM outlined for sound and visual characteristics.[13] Compares a assortment of CNNs that have been planned and prepared for confront expression distinguishing proof. Expressions on the confront within the domain of inquire about, acknowledgment is picking up in significance. All consider areas look at and examine facial feeling acknowledgment. Emotion's recognized from confront pictures utilizing max pooling and CNN [16], which contains a great exactness score, driving us to accept that profound learning may be used for feeling forecast as well.

3. Methodology

Deep Learning [4] is a neural network approach that models the data in order to perform a certain job. CNN and VGG-16 offers a large range of API in image identification, classification task, decision making task, pattern recognition task, and other fields [5]. For feature selection, picture recognition, and other Deep learning tasks, multimodal deep learning techniques is utilized

Dataset: -The Extended Cohn-Kanade (CK+) dataset includes 593 video sequences from 123 different people ranging in age from 18 to 50 years old, as well as gender and ethnicity. Each movie displays a face transitioning from neutral to a certain peak emotion, filmed at 30 frames per second (FPS) in 640x490 or 640x480 pixel resolution. Three hundred twenty-seven films have been categorised to one of seven expression categories: fury, disdain, disgust, fear, pleasure, sadness, and surprise. The CK+ database is widely regarded as the most commonly used laboratory-controlled facial expression classification database, and it is used in the vast majority of facial expression classifications.

Data preprocessing: -Images come in a variety of shapes and sizes. They also come from many sources. Some photographs, for example, are what we term "natural images," which implies they were captured in color in the real world. As an example:

1. A photograph of a person's face is a realistic image.
2. An X-ray image is not the same as a natural image.

Taking all of these variables into account, we must conduct some pre-processing on any image data. RGB is the most widely used encoding format, and the majority of "natural images" we encounter are in RGB. Making photos of the same size is also one of the first steps in data pre-processing.

Class Imbalance: - This is highly imbalanced data. When we have an imbalanced data accuracy can't use as performance matric. We can handle imbalanced data in different ways like under sampling, over sampling, and by changing performance metric. In this project we have only less amount data that why we tried oversample technique. By doing over sample we balance the data and use accuracy as a performance metric.

Data Splitting: - Data splitting also one of the part in data preprocessing. If we have temporal data we can go with time based splitting but in this case we don't have time stamp feature we go with random split. We randomly divided the data in 70:30 ratio as train and test. We train the models by using train data and test it on test data based on test accuracy we can conclude how the model working.

CNN model:-

Figure 1 shows the building of the proposed face expression recognition model. The model employs two convolution layers, with dropouts between them. The first convolution layer receives the input picture, which is scaled to 48 by 48 pixels.

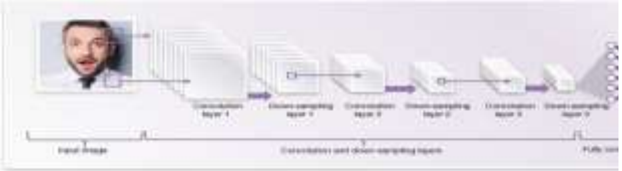
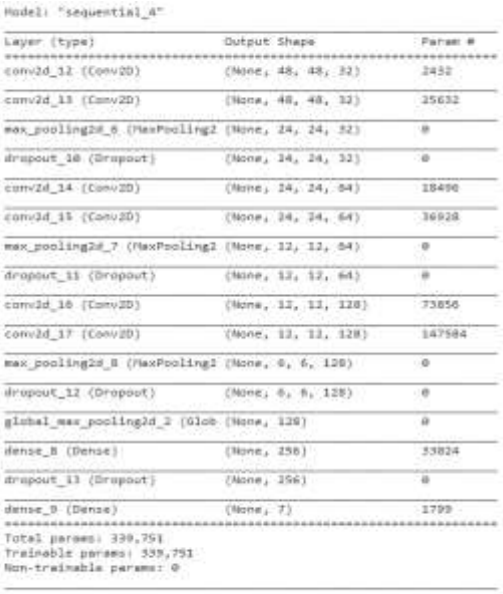


Fig: - 1 CNN architecture

The feature map is created by applying an activation function to the output of the convolution layer. The ReLU (Rectified Linear Unit) activation function is used here, which makes the negative values zero while keeping the positive values constant. This feature map is applied to a pooling layer with a pool size of 2 2 in order to reduce the size without compromising any information. A dropout layer is used to reduce over fitting. This method is repeated for the following convolution layer as well. Finally, a two-dimensional array of feature values is generated.The flatten layer is utilised to transform these two-dimensional arrays to a single-dimensional vector for use as input to the neural network, which is represented by the dense layers. In this case, a two-layer neural network is deployed, one for input and one for output. Because there are five classes to classify, the output layer contains five units. The output layer's activation function is softmax, which generates probabilistic output for each class. Figure 2 illustrates a snapshot of the proposed system's model summary, which was created using the Keras DL Library.



```
Model: "sequential_4"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_12 (Conv2D)           (None, 48, 48, 32)        2432
_________________________________________________________________
conv2d_13 (Conv2D)           (None, 48, 48, 32)        25632
_________________________________________________________________
max_pooling2d_6 (MaxPooling2 (None, 24, 24, 32)        0
_________________________________________________________________
dropout_10 (Dropout)         (None, 24, 24, 32)        0
_________________________________________________________________
conv2d_14 (Conv2D)           (None, 24, 24, 64)        18496
_________________________________________________________________
conv2d_15 (Conv2D)           (None, 24, 24, 64)        36928
_________________________________________________________________
max_pooling2d_7 (MaxPooling2 (None, 12, 12, 64)        0
_________________________________________________________________
dropout_11 (Dropout)         (None, 12, 12, 64)        0
_________________________________________________________________
conv2d_16 (Conv2D)           (None, 12, 12, 128)       73856
_________________________________________________________________
conv2d_17 (Conv2D)           (None, 12, 12, 128)       147584
_________________________________________________________________
max_pooling2d_8 (MaxPooling2 (None, 6, 6, 128)         0
_________________________________________________________________
dropout_12 (Dropout)         (None, 6, 6, 128)         0
_________________________________________________________________
global_max_pooling2d_2 (Glob (None, 128)               0
_________________________________________________________________
dense_8 (Dense)              (None, 256)               33024
_________________________________________________________________
dropout_13 (Dropout)         (None, 256)               0
_________________________________________________________________
dense_9 (Dense)              (None, 7)                 1799
=================================================================
Total params: 339,751
Trainable params: 339,751
Non-trainable params: 0
_________________________________________________________________
```

Fig: - 2 CNN model summery

Transfer learning: -One of the machine learning approaches is exchange learning, which leverages the data picked up from fathoming one issue to fathom another. Exchange learning does, in reality, address issues in a brief sum of time. When the computation fetched must be diminished and precision must be accomplished with less preparing time, pre prepared models is utilized. Exchange learning could be a broadly utilized approach that includes taking a model's learnt weights (for case, ImageNet) and executing them by settling the other layers to holding the leftover portion of the layers within the whole organize. In this paper, we utilize the pre-trained model from the Kaggle, a enormous dataset, and keep a couple of layers on the CK +48 information set, a littler information set, to attain exchange learning. This strategy was chosen since the CK +48 and Kaggle both give comparable information, to be specific pictures with one of the seven feelings.

VGG-16:- A yearly computer vision competition, the ImageNet Expansive Scale Visual Acknowledgment Challenge (ILSVRC), is held. Groups compete on two challenges each year. Question localization is the introductory step in identifying objects in a picture from 200 diverse classifications. The moment step is picture classification, which includes naming each picture with one of 1000 categories. Karen Simonyan and Andrew Zisserman of Oxford University's Visual Geometry Gather Lab recommended VGG 16 in their article "Exceptionally Profound CONVOLUTIONAL Systems FOR LARGE-SCALE Picture Acknowledgment" in 2014. Within the 2014 ILSVRC competition, this demonstrate took first and moment put within the previously mentioned categories.

Architecture: -The network's input may be a two-dimensional picture (224, 224, 3). The primary two layers have the same cushioning and 64 channels of 3*3 channel estimate. At that point, after a walk (2, 2) max pool layer, two layers of convolution layers of 256 channel measure and channel measure (3, 3). Usually taken after by a walk (2, 2) max pooling layer, which is the same as the going before layer. Taking after that, there are two convolution layers with channel sizes of 3 and 3 and a 256 channel. Taking after that, there are two sets of three convolution layers, as well as a max pool layer. Each has 512 channels of the same estimate (3, 3) and cushioning. This picture is at that point nourished into a two-layer convolution stack.

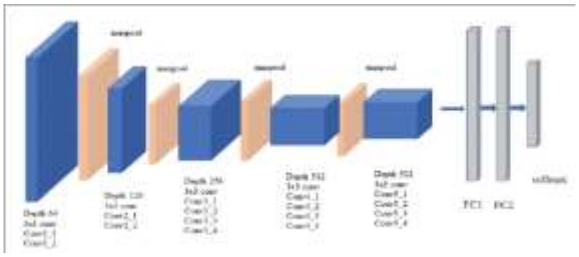


Fig: - 3 Vgg16 Architecture

4. Results and Analysis

The experiment scenario yields a notable outcome in terms of system performance, with an average accuracy rate of 92.81 %. The CNN model has the lowest accuracy rating of 96 %. Each class has a misclassification result, indicating that the system still requires work. To get a better outcome in the future study, we might consider altering the entire architecture. In fig 4 we printed comparison table by vgg16 we got 99% accuracy. Based on that table we can say VGG16 is proposed model with 99% accuracy.

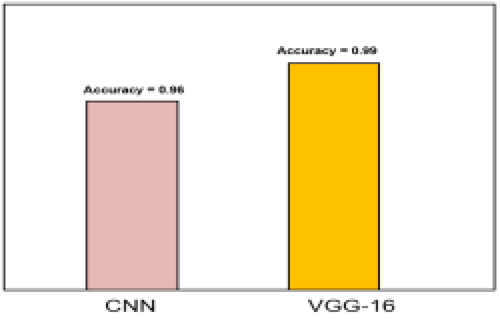


Fig:- 4 Results comparison

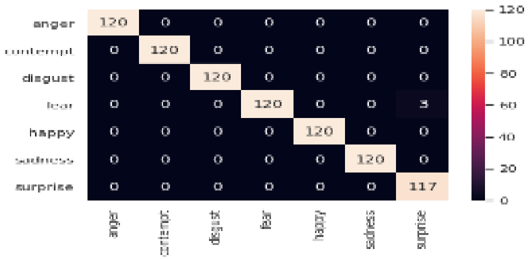


Fig: - 5 Confusion matrix for Vgg16 model

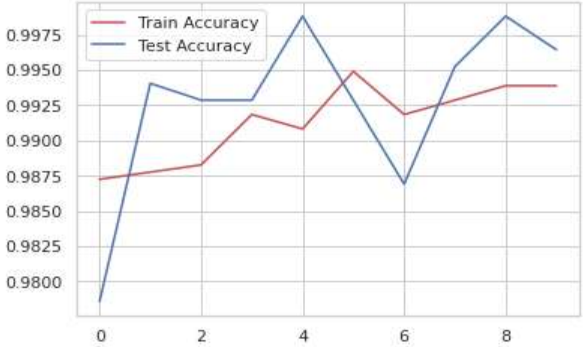


Fig:-6Accuracy plot for Vgg16 model

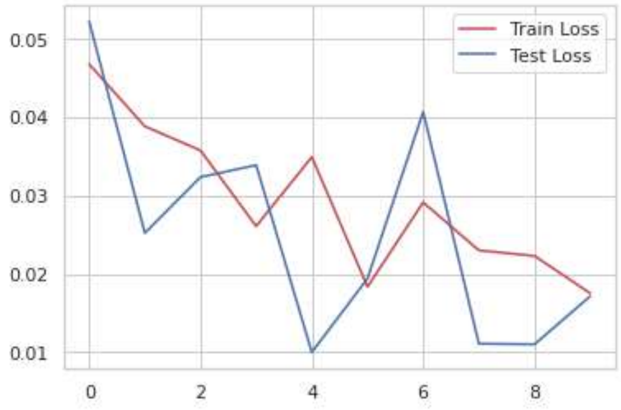


Fig:-7 Loss plot for Vgg16 model

In fig 5 we plotted confusion matrix. It will show hominy data points misclassified and correctly classified per each class. By seeing fig 5 we can conclude in our proposed system by using VGG16 we are getting almost 100% accuracy there is no misclassified data points in confusion matrix. In fig 6 and fig 7 we plotted loss and accuracy plots for test and train data also.

5. Conclusion:

For confront feeling acknowledgment, this think about offers a two-layer convolution arrange demonstrate. The calculation employments the picture collection to classify five diverse confront expressions. The demonstrate has identical preparing and approval exactness, demonstrating that it incorporates a great fit to the information and can be amplified. The exchange learning demonstrate decreases them is fortune work utilizing an Adam optimizer, and it has been tried to have an exactness of 99 %. The work may be long to distinguish variations in feeling employing a film grouping, which can at that point be used for an assortment of real-time applications counting input investigation and so on. For most extreme proficiency, this framework may be combined with other electrical hardware.

Reference:-

1. M. El Ayaadi, F. Karraeand M. S. Kamal "Survey on emotion recognitions: Features, classification scheme, and database," Pattern Recognit., vol. 44, no. 3, pp. 572–587, 2011.
2. Y. Lecan, Y. Bengeo, and G. Hintin, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
3. J. Schmidhubar, "Deep Learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015.
4. J. Ngiem, A. Khoslaa, M. Kam, J. Nim, H. Lei, and A. Y. Nig, "Multimodal Deep Learning," Proc. 28th Int. Conf. Mach. Learn., pp. 689–696, 2011.
5. S. Lugivic M. Harva and I. Dander "Techniques and applications of emotion recognition," 2016 39th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2016 - Proc., no. November 2017, pp. 1278–1283, 2016.
6. B. Schaller, G. Rigull, and M. Ling, "Emotion recognition combining acoustic features and information in a neural network - belief network architecture," Acoust. Speech, Signal Process., vol. 1, pp. 577–580, 2004.
7. J. Riang, G. Leie, and Y. P. P. Chen, "Acoustic feature selection for automatic emotion recognition from speech," Inf. Process. Manag., vol. 45, no. 3, pp. 315–328, 2009.
8. F. Noruzi, G. Anbarjafaari and N. Akraami "Expression-based emotion recognition and next reaction prediction," 2017 25th Signal Process. Commun. Appl. Conf. SIU 2017, no. 1, 2017.
9. G. Hintin ,Greves and A. Mohemed, "Emotion Recognition with Deep Recurrent Neural Networks," in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649.
10. K.Weint and C.-W. Huaang, "Characterizing Types of Convolution in Deep Convolutional Recurrent Neural Networks for Robust Speech Emotion Recognition," pp. 1–19, 2017.
11. A. M. Yusuf, M. B. Mustaafa, M. and M. Malekzedeh, "emotion recognition research: an analysis of research focus," Int. J. Speech Technol., vol. 0, no. 0, pp. 1–20, 2018.
12. L. Cavidon ,H. M. Fayak, M. Lich,"Evaluating deep learning architectures for Emotion Recognition," Neural Networks, vol. 92, pp. 60–68, 2017.
13. A. M. Badshaah, J. Ahmed and S. W. Baek, "Emotion Recognition from Spectrograms with Deep Convolutional Neural Network," 2017 Int. Conf. Platf. Technol. Serv., pp. 1– 5, 2017.
14. A. Routrey, M. Swaen and P Kabisetpathy, "Database, features and classifiers for emotion

recognition: a review," Int. J. Speech Technol., vol. 0, no. 0, p. 0, 2018.

6. K.-Y. Hueng, C.-H. Wiu, T.-H. Yieng, M.-H. Sha and J.-H. Chiu, "Emotion recognition using autoencoder bottleneck features and LSTM," in 2016 International Conference on Orange Technologies (ICOT), 2016, pp. 1–4.

15. M. N. Sttilar, M. Leich, R. S. Bolie, and M. Skinter, "Real time emotion recognition using RGB image classification and transfer learning," 2017 11th Int. Conf. Signal Process. Commun. Syst., pp. 1–8, 2017.

16. D. Lie and E. M. Provest, "EMOTION RECOGNITION USING HIDDEN MARKOV MODELS WITH DEEP BELIEF NETWORKS."

17. H. Jiaang et al., "Investigation of emotions for identifying depression using various classifiers," Expression Commun., vol. 90, pp. 39–46, 2017.

18. Q. Maio, M. Diong, Z. Huaeng, and Y. Zhaen, "Learning Salient Features for Emotion Recognition Using Neural Networks," IEEE Trans. Multimed., vol. 16, no. 8, 2014.

19. P. Hairár, R. Burgeet, and M. K. Duitta, "Facial Emotion Recognition with Deep Learning," in Signal Processing and Integrated Networks (SPIN), 2017, pp. 4–7.

20. R. Ashrafidust, S. Setaeyeshi, and A. Shaarifi, "Recognizing Emotional State Changes Using Facial Expression," 2016 Eur. Model. Symp., pp. 41–46, 2016.

21. C. Busiso et al., "IEMOCAP: Interactive emotional dynamic motion capture database," Lang. Resour. Eval., vol. 42, no. 4, pp. 335–359, 2008.

22. C. Szeigedy, V. Vanhouicke, J. Shleins, and Z. Woijna, "Rethinking the Inception Architecture for Computer Vision," 2014.