Emotion Detection

Aditya Mittal
dept. of Computer Science
Indraprastha Institute of
Information Technology
Delhi, India
aditya18061@iiitd.ac.in

Arjun Tyagi
dept. of Computer Science
Indraprastha Institute of
Information Technology
Delhi, India
arjun18023@iiitd.ac.in

Piyush Dhyani
dept. of Computer Science
Indraprastha Institute of
Information Technology
Delhi, India
piyush18131@iiitd.ac.in

I. INTRODUCTION

Every emotion is carried out by the contraction or expansion of some face muscles. Facial expressions can be used as an efficient way of emotion detection. Each facial muscles can be captured in the form of Facial Action Units(AUs). In this project we would be working on different facial action units like eyes, cheeks and lips, which are responsible for different expression. Emotion detection can also be useful for Intelligent Human-Computer Interaction (HCI). The set of action unit represent a particular expression. For example AU6 and AU12 together indicates happy Emotion. Some feature extraction algorithms would be applied on action units relative to the emotion and classification will be perform on that using different classifiers. The result of different feature extraction algorithm will be compared with different classifiers. This work plays a predominant role in fields such as psychology, indirect customers feedback in different places like malls and also helpful in trading.

II. LITERATURE REVIEW

The first known work on facial recognition is by Darwin in 1878 [1] after that a subtle amount of research has been done on this area. A facial Action coding System (FACS) for the categorization of facial expression is uses by the Ekman et al. [2] which is used extensively in the areas of analysis of facial expression. In FACS system the facial attributes are represented. In their study Wu et al. uses Gabor filters for feature extraction in frame by frame analysis of videos. It mainly used as edge detection algorithm and helps to reduce variability due to its different orientation and frequency. Lyons et al in their work used LDA as classifier. LDA allows fast and simple training from data. The input features are easily interpreted from this. Kim et al. uses Convolution neural network (CNN) in their study due to evolution of deep learning features. In our project we will use state-of-art techniques to obtain the results.

III. DATA SET DISCRIPTION

In our project we used two Dataset CK+ [3] and FER2013[4]. In CK+ there were eight emotions anger, contempt, disgust, fear, happiness, neutral, sadness, surprise. This dataset is imbalance(Fig 1) with neutral having high frequency of images. There are total 1256 images in the dataset. All subjects in the images having still head posed. We used 1004 for training and 252 for testing purpose. The FER dataset used was having more images and was challenging also. There dataset consist of 7 emotions anger, disgust, fear, happy, sad, surprise, neutral. Each image is 48*48 pixel grayscale image. The training dataset consists of 28709 examples and test consist of 7178 examples. Which are divided into two parts, i.e. Public Test and Private Test, Each consisting of 3589 images. The main challenge in this dataset is that the subjects in the images having head posed in different orientation.

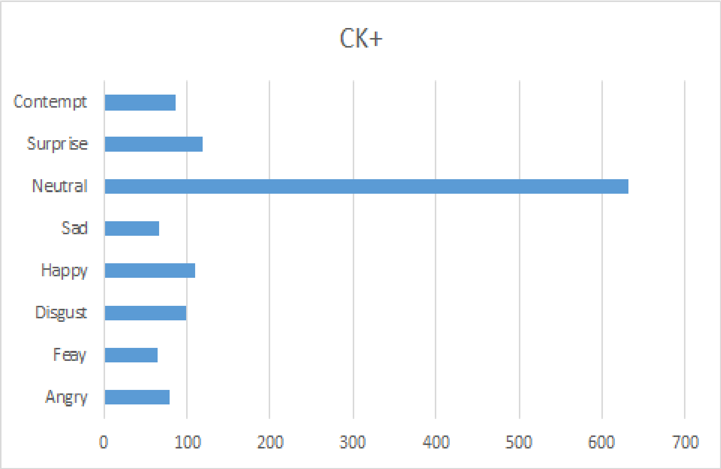


FIG 1: CK+ data distribution across different classes

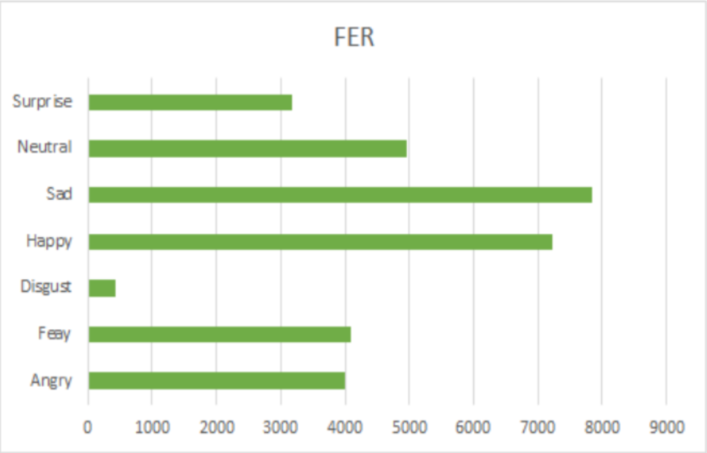


FIG 2: FER data distribution across different classes

IV. PROPOSED ARCHITECTURE

In this project, we proposed two different architectures for both CK+ and FER dataset. These architecture consist of both deep learning and non-deep learning approaches. For CK+ dataset we proposed a model based on low-level features like HOG and LBP. For FER dataset we proposed a deep learning based approach.

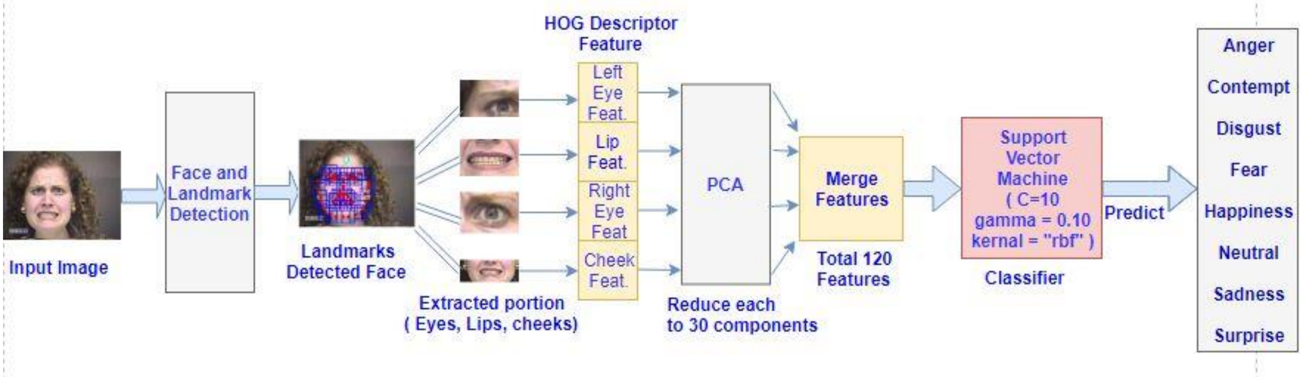


FIG 3 : Proposed architecture of CK+ dataset

A. *CK+ Architecture*

The following pipeline is used in CK+ Dataset to get an accuracy around 90 percent. The Facial Landmarks from the input image is generated. With the help of facial landmarks, parts of facial action units were extracted such as lips, left eye, right eye, cheeks. The hog feature of each extracted parts is calculated. 30 components of each hog features of extracted parts were taken with the help of PCA algorithm.

B. *FER Architecture*

From experiments we can infer that the low level features are not performing well on FER2013[4]. So we use features from pertained convolutional neural networks AlexNet,VGG-16, Inceptionv3, Densenet161 among which Densenet161 give the best result. These deep convolutional neural networks are trained on ImageNet dataset that do not have any facial expression class. So, we trained our own 9 layer network on that help us outperforming the approach of classification using features from these features.

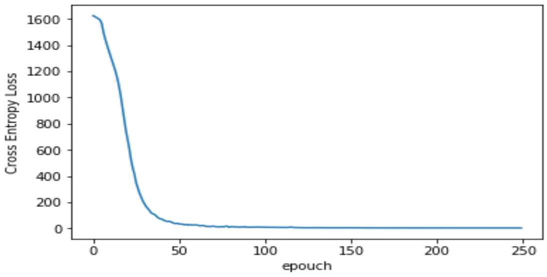


FIG 4: Train error curve of CNN

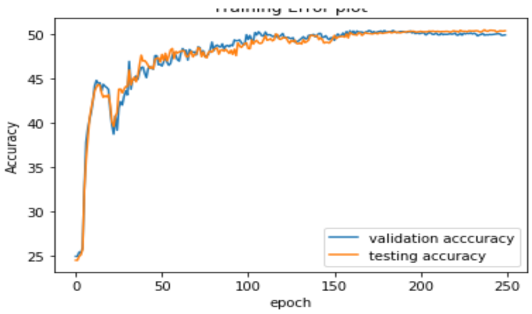


FIG 5: Accuracy curve on validation and testing

V. Data Visualization

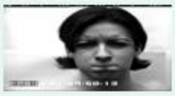
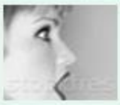
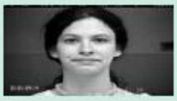
Emotion	CK+	FER
	Images	Images
Angry		
Fear		
Disgust		
Happy		
Sad		
Neutral		
Surprise		
Contempt		—

Table 1: Sample images of both datasets

FER dataset is very challenging as compared to CK+ dataset as shown in Table 1. FER dataset has images with different orientation, angles and occlusion.

Fig 6 shows the visualization (Gaussian curve and scatter plot) of FER dataset for two different components of PCA.

Fig 7 shows the curve between principal components and Cumulative Explained Variance. It explains the number of features involved for best representation of FER dataset.

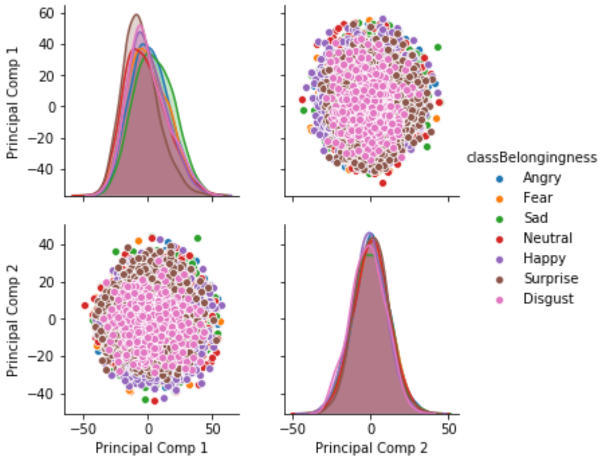


FIG 6: Data Visualization of FER

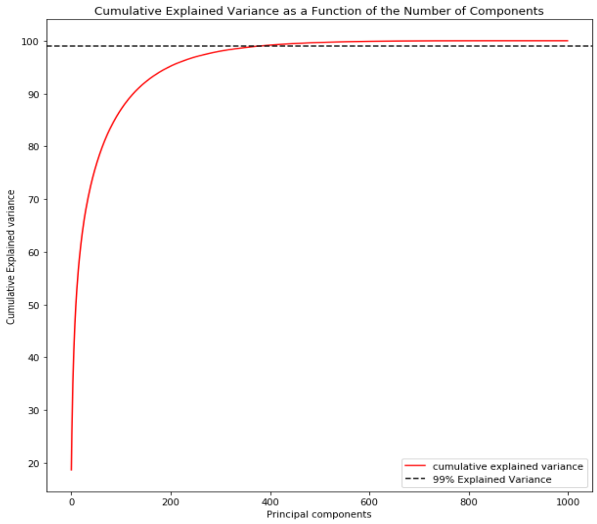


FIG 7: PCA eigen energy conservation plot.

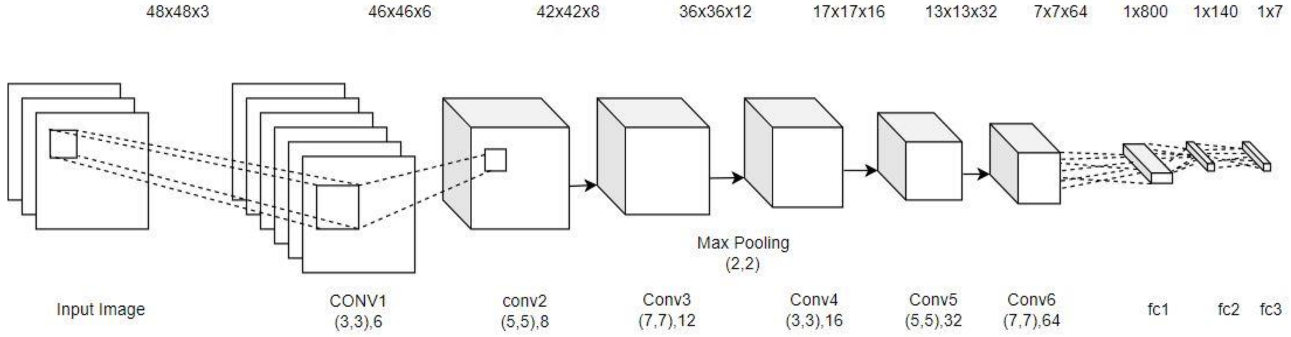
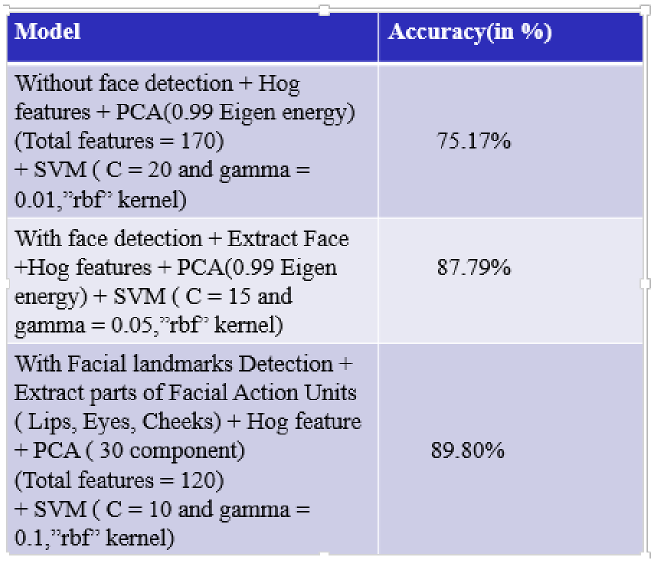


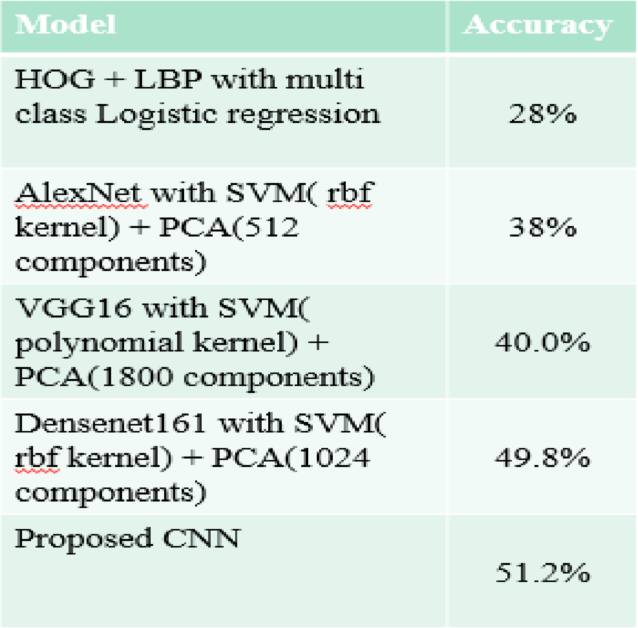
FIG 8: CNN architecture for FER dataset

VI. Results



Model	Accuracy(in %)
Without face detection + Hog features + PCA(0.99 Eigen energy) (Total features = 170) + SVM (C = 20 and gamma = 0.01,"rbf" kernel)	75.17%
With face detection + Extract Face +Hog features + PCA(0.99 Eigen energy) + SVM (C = 15 and gamma = 0.05,"rbf" kernel)	87.79%
With Facial landmarks Detection + Extract parts of Facial Action Units (Lips, Eyes, Cheeks) + Hog feature + PCA (30 component) (Total features = 120) + SVM (C = 10 and gamma = 0.1,"rbf" kernel)	89.80%

Table 2: Accuracies of different models on CK+ dataset



Model	Accuracy
HOG + LBP with multi class Logistic regression	28%
AlexNet with SVM(rbf kernel) + PCA(512 components)	38%
VGG16 with SVM(polynomial kernel) + PCA(1800 components)	40.0%
Densenet161 with SVM(rbf kernel) + PCA(1024 components)	49.8%
Proposed CNN	51.2%

Table 3: Accuracies of different models on FER dataset

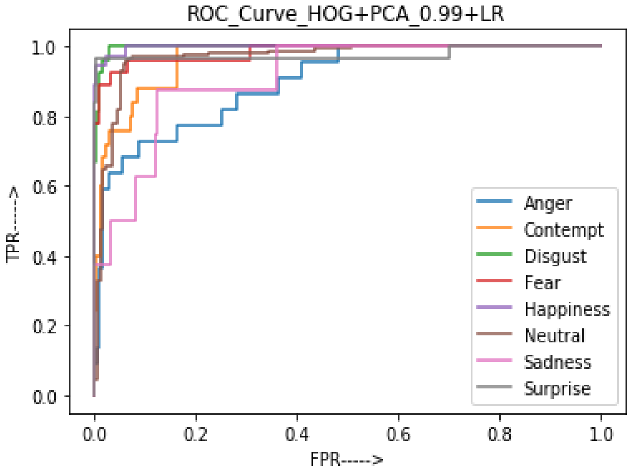


Fig 7: ROC of CK+ dataset for best model

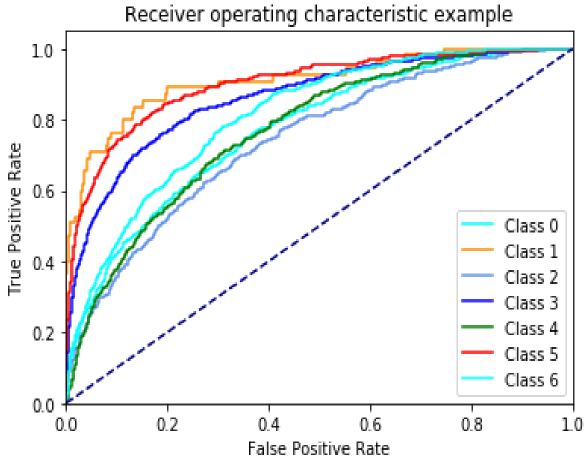


Fig 8: ROC of FER dataset for best model

VII. Analysis

In the CK+ dataset the angle of capturing and the posture of subjects have less variance.

In FER dataset there is a lot of variance in subject postures and the image capturing angle also there is occlusion and varying illumination (Table 1) that hide most important features like lips movement etc. that makes the task more challenging.

In both datasets PCA is performing better then LDA for dimension reduction, it is due to high similarity of images across classes due to similar subjects. So PCA finds the best direction in which there is best distinction of images and LDA tries to maximize distinction across classes.

By visualizing both the datasets we can conclude that they are not linearly separable (Fig 6) so we have to use non-linear classification techniques like Support vector machine with rbf kernel , Neural Networks etc.

From several experiments it is clear that even features of pre-trained deep convolutional neural network are also not very helpful for classification of FER dataset.

By taking motivation from above we trained our own CNN from scratch using FER training data that help us over performing all of our previous attempts.

VIII. Conclusion

- Experimentally we found that combination of HOG and LBP along with PCA is performing best among all other models and give an accuracy of 89.8%.
- For this project we use two datasets i.e. FER and CK+. Both datasets consist of images of different facial expressions.
- FER is more complex dataset then CK+ so we used pre-trained deep convolutional neural networks for feature extraction.
- Finally we trained our own 9 layer convolutional neural network using FER dataset that help us achieving an accuracy of **51.2%**.
- The proposed methodology is limited to classify frontal image only. However, rotation of face or occlusions degrades the performance of the system. In the future, the research will be extended to 3-D face modeling using multiple cameras to improve the proposed facial expression recognition system.

IX. Individual Contribution

Our team consist of three members. Below is the contribution of group members. As FER dataset was Challenging two group members worked on it.

Piyush Dhyani - Worked on CK+ dataset. Get results with preprocessing such as landmarks detection on faces and applying different features with classification techniques to get best possible results.

Aditya mittal - Applied different features like HOG,PCA,LBP on FER Dataset with different classification techniques to results.

Arjun Tyagi - Applied deep learning based features and classification techniques and new CNN based Architecture on FER Dataset.

X. References

[1] Charles. The expression of the emotions in man and animals (p. ekman, ed.), 1872.

[2] Paul Ekman. Facial action coding system (facs). A human face, 2002.

[3] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 94–101. IEEE, 2010.

[4]https://www.kaggle.com/c/challenges-in-representation-learning-facialexpression-recognition-challenge (accessed on 13-02-2019 at 23:50).

[5] G. Levi and T. Hassner, “Emotion recognition in the wild via convolutional neural networks and mapped binary patterns,” in Proceedings of the 2015 ACM on international conference on multimodal interaction. ACM, 2015, pp. 503–510.