ZKG INTERNATIONAL

ISSN: 2366-1313

Analysis of Facial Expressions using ResNet and VGG

Guide : Mr. M. Sri Ramulu,M.Tech

By : G. Meghana, B. Bhagya Sri, A. Sai Krishna, G. Kavya Sri

[1]**Assistant professor, Dept. of CSE, Lendi Institute of Engineering and Technology, Vizianagaram,**
Abstract— : **Human facial expressions are an integral means of displaying sentiments. Automatic analysis of these unspoken sentiments has been an interesting and challenging task in the domain of computer vision with its applications ranging across multiple domains including psychology, product marketing, process automation etc. This task has been a difficult one as humans differ greatly in the manner of expressing their sentiments through expressions. Previously various machine learning techniques like random forest, SVM etc. were used to predict the sentiment using converted images. Deep learning has been instrumental in making breakthrough progress in many fields of research including computer vision. We implement a ResNet and VGG based model for facial expression detection.**

Keywords Deep Learning, Facial Sentiment Analysis, Convolution Neural Network, Face Detection, Network Architecture.

I. INTRODUCTION

Any interpersonal contact must take into account feelings. These can be expressed in a variety of ways, including voice, gesture, attitude, and even facial expressions. The most noticeable and information-rich of these options are facial expressions, which can be taken advantage of to analyze feelings. Additionally, collecting and processing faces is easier than other forms of expression. A facial expression is a complex movement of the facial muscles that communicates the subject's feelings or emotional state to anyone watching it. Expressions are essentially a message about how someone is feeling on the inside. For these reasons, researchers in psychology, animation, and human computer interaction have become more interested in developing a human-computer interaction system for automatic face recognition or facial sentiment analysis. An emerging field today is computer-assisted facial expression analysis.

Associating an emotion with a facial image is the basis of sentiment analysis. Therefore, the goal is to infer a person's inner feelings from their face. Automatic facial expression is an analysis system is crucial for streamlining human-machine interactions. But this is not an easy task. For some time now, many aspects of facial expressions have been retrieved and analysed for effective sentiment analysis using deep learning and ResNet and VGG networks, which were inspired by the study of biology. This has inspired us to do this research with the goal of developing a deep learning-based model for face sentiment analysis.

II. RELATED WORKS

Chu, William Wei-Jen Tsai, Hui-Chuan, YuhMin Chen and Min-Ju Liao. “Facial expression recognition with transition detection for students with high-functioning autism in adaptive e-learning.” Soft Computing: 1-27, 2017. Emotions deeply affect learning achievement. In the case of students with high-functioning autism (HFA), negative emotions such as anxiety and anger can impair the learning process due to the inability of these individuals to control their emotions. Attempts to regulate negative emotions in HFA students once they have occurred, subsequent regulation to HFA students is often ineffective because it is difficult to calm them down. Hence, detecting emotional transitions and providing adaptive emotional regulation strategies in a timely manner to regulate negative emotions can be especially important for students with HFA in an e-learning environment. In this study, a facial expression-based emotion recognition method with transition detection was proposed.

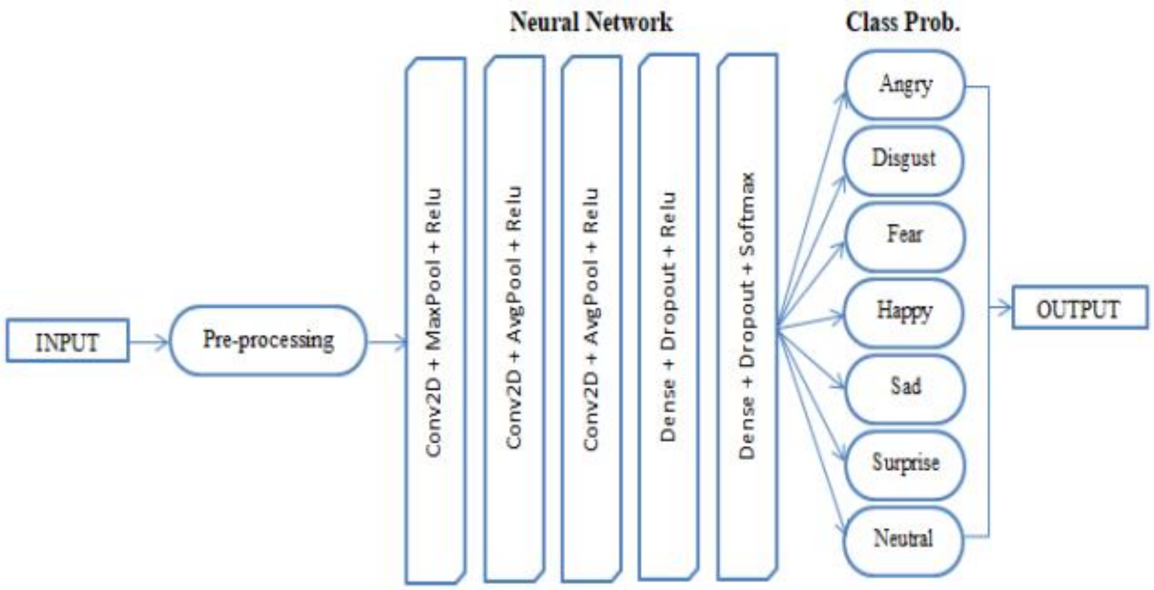
Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Panagiotis, Bjrn W. Schuller, and Stefanos Zafeiriou. ”End-to-end multi-modal emotion reco. using neural networks.” IEEE Journal of Topics in Signal Processing 11, no. 8: 13011309, 2017.

Automatic affect recognition is a challenging task due to the various modalities emotions can be expressed with. Applications can be found in many domains including multimedia retrieval and human computer interaction. In recent years, deep neural networks have been used with great success in determining emotional states. Inspired by this success, we propose an emotion recognition system using auditory and visual modalities. To capture the emotional content for various styles of speaking, robust features need to be extracted.

III. METHODOLOGY USED

Model Diagram:

This structure shows whole process how does it works from starting from input to the output of the project.



Algorithms Used:

In this work we have used two deep learning algorithms called ResNet and VGG because of their deep layers. As many of the projects have already used CNN algorithms. So we are going to use CNN model algorithms called ResNet and VGG.

VGG16:

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual computer vision competition. Each year, teams compete on two tasks. The first is to detect objects within an image coming from 200 classes, which is called object localization. The second is to classify images, each labelled with one of 1000 categories, which is called image classification. VGG 16 was proposed by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group Lab of Oxford University in 2014 in the paper "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION". This model won 1st and 2nd place in the above categories in the 2014 ILSVRC challenge.

The ImageNet dataset contains images of fixed size of 224*224 and have RGB channels. So, we have a tensor of (224, 224, 3) as our input. This model processes the input image and outputs the vector of 1000 values. This vector represents the classification probability for the corresponding class. Suppose we have a model that predicts that the images belong to class 0 with probability 1, class 1 with probability 0.05, class 2 with probability 0.05, class 3 with probability 0.03, class 780 with probability 0.72, class 999 with probability 0.05 and all other class with 0. So, the classification vector for this will be: To make sure these probabilities add to 1, we use softmax function.

Residual Networks (ResNet):

After the first CNN-based architecture (AlexNet) that win the ImageNet 2012 competition, every subsequent winning architecture uses more layers in a deep neural network to reduce the error rate. This works for less number of layers, but when we increase the number of layers, there is a common problem in deep learning associated with that called the Vanishing/Exploding gradient. This causes the gradient to become 0 or too large. Thus when we increases number of layers, the training and test error rate also increases.

In order to solve the problem of the vanishing/exploding gradient, this architecture introduced the concept called Residual Blocks. In this network, we use a technique called skip connections. The skip connection connects activations of layer to further layers by skipping some layers in between. This forms a residual block. Resnet are made by stacking these residual blocks together. The approach behind this network is instead of layers learning the underlying mapping, we allow the network to fit the residual mapping. So, instead of say H(x), initial mapping, let the network fit,

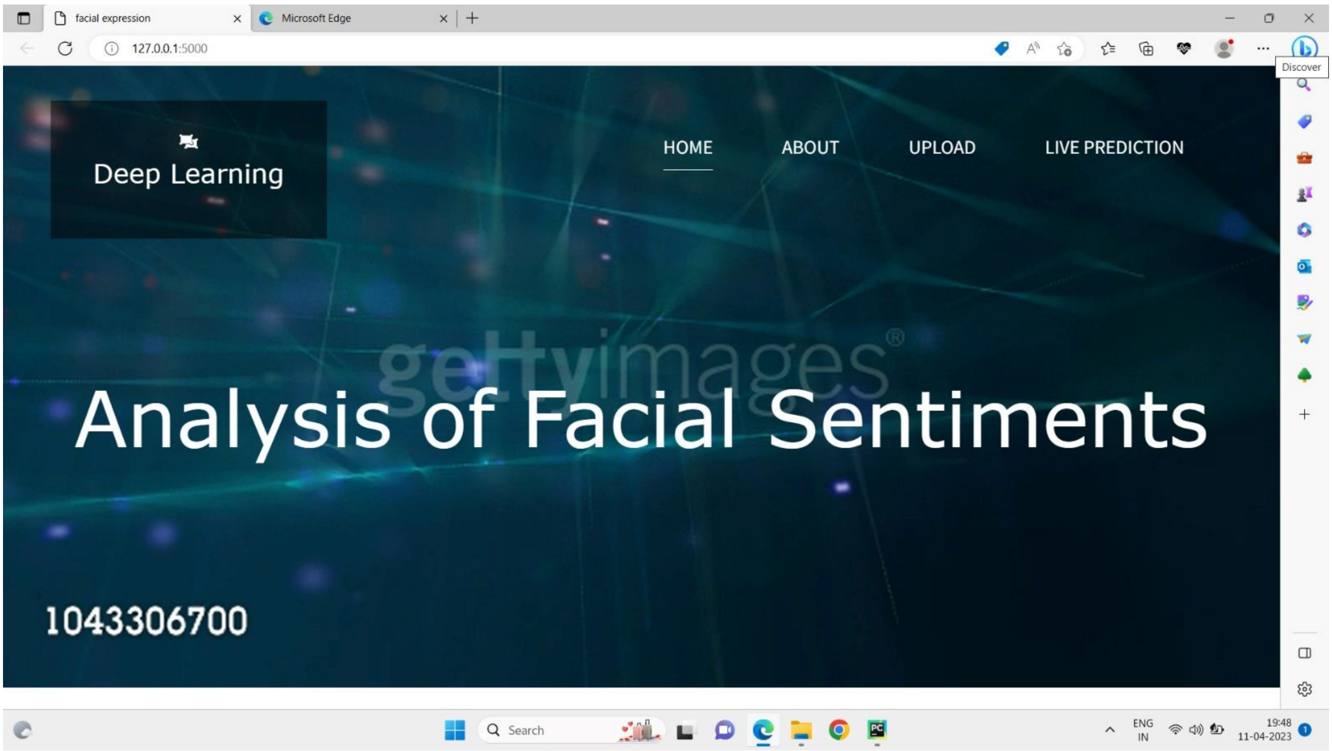
$$F(x) := H(x) - x \text{ which gives } H(x) := F(x) + x.$$

IV. RESULT:

The output of the project shows the different expression of the face that had been uploaded.

Page that opens after running the code:

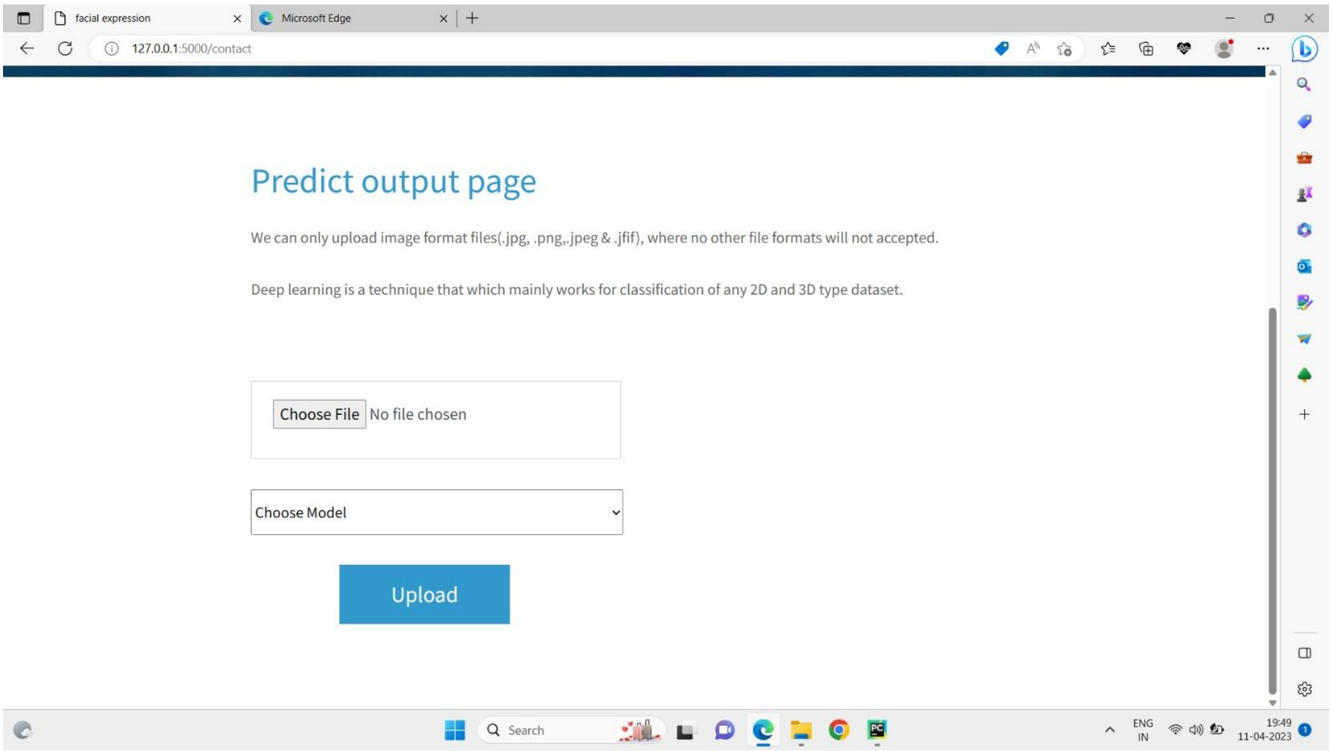
The page that opens after running our code is that shows the name of our project. The front end had been done using the xml language and CSS for page colours and style. The introduction page shows our title in home component and about components gives the description of our project and coming to the upload is where we can upload the picture and expect the output of the picture the live prediction is a component where the live video would come.



1 The Upload page:

The upload page is where you upload the image and gets the output according to the pages that had been uploaded.

At first we select the image and then we click on the upload. After clicking the upload we the expression of the image that had been uploaded by using pattern of particular expression.



The outputs of the expression:

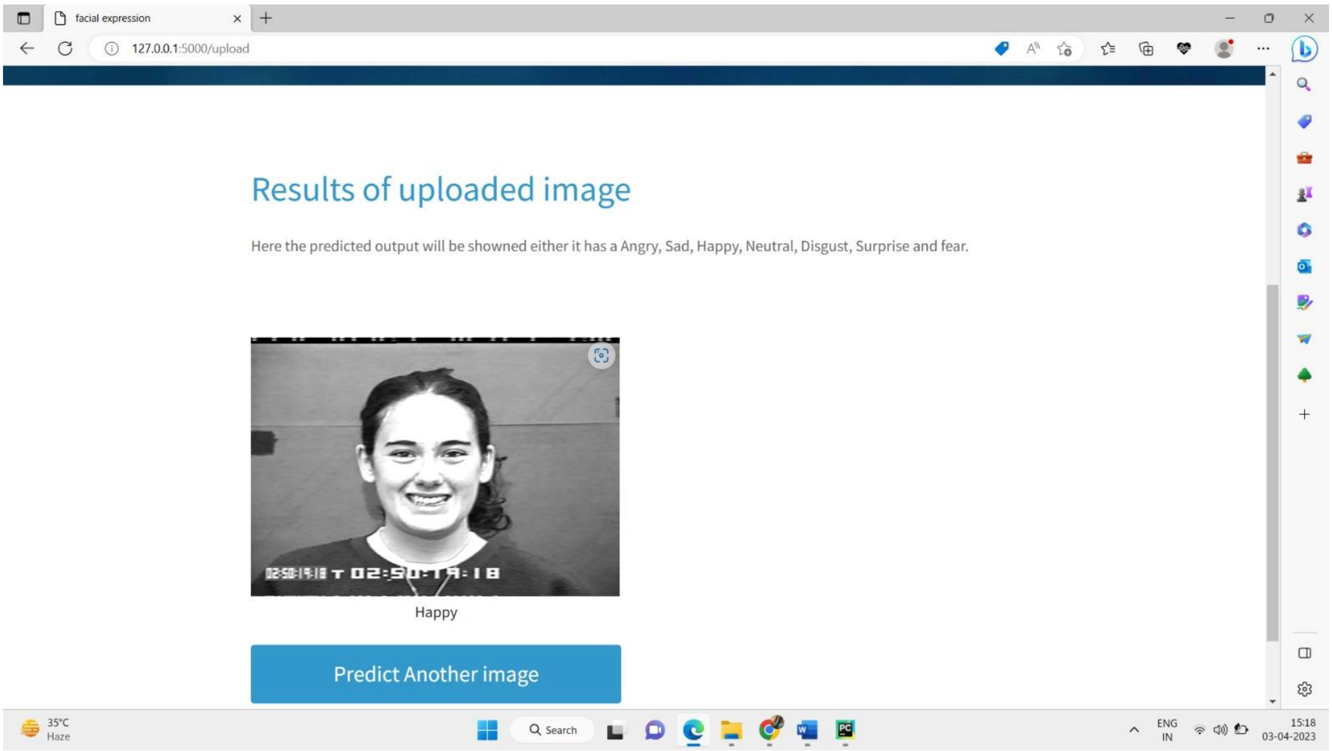


Fig: output as happy expression

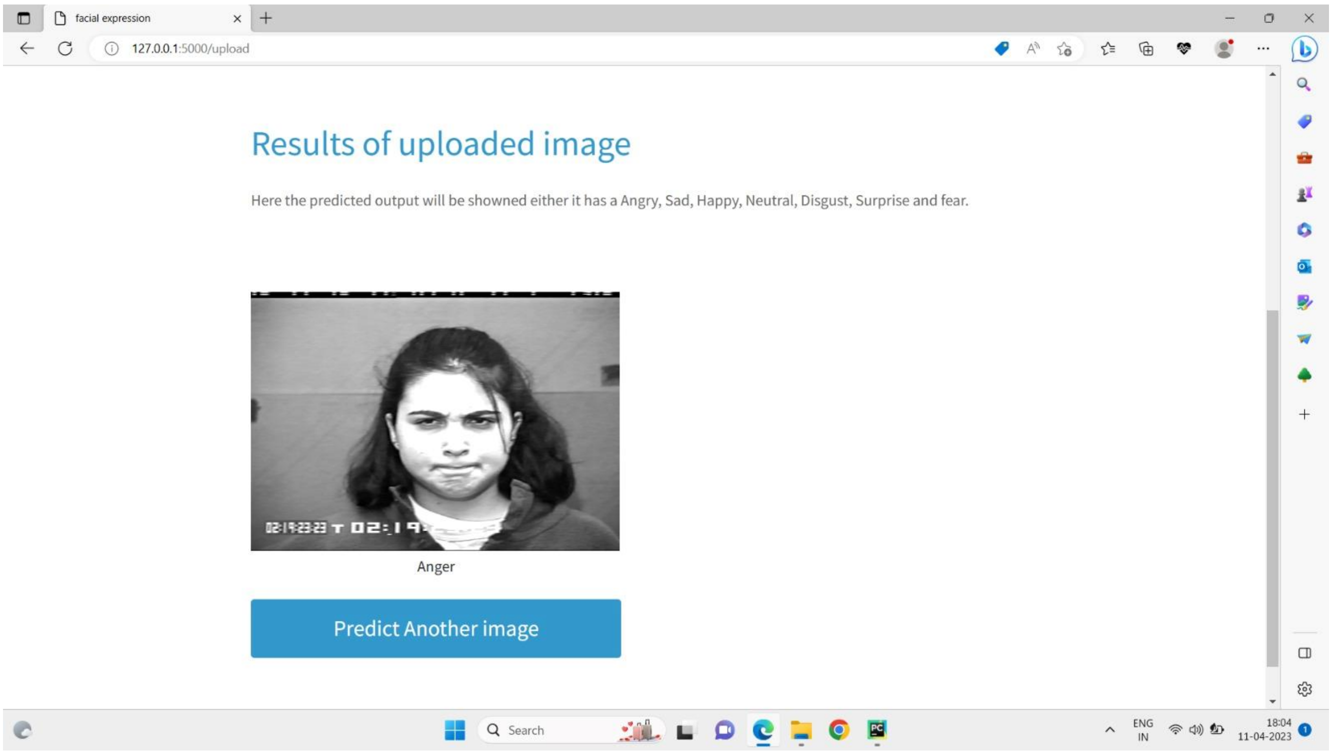


Fig: output as anger expression

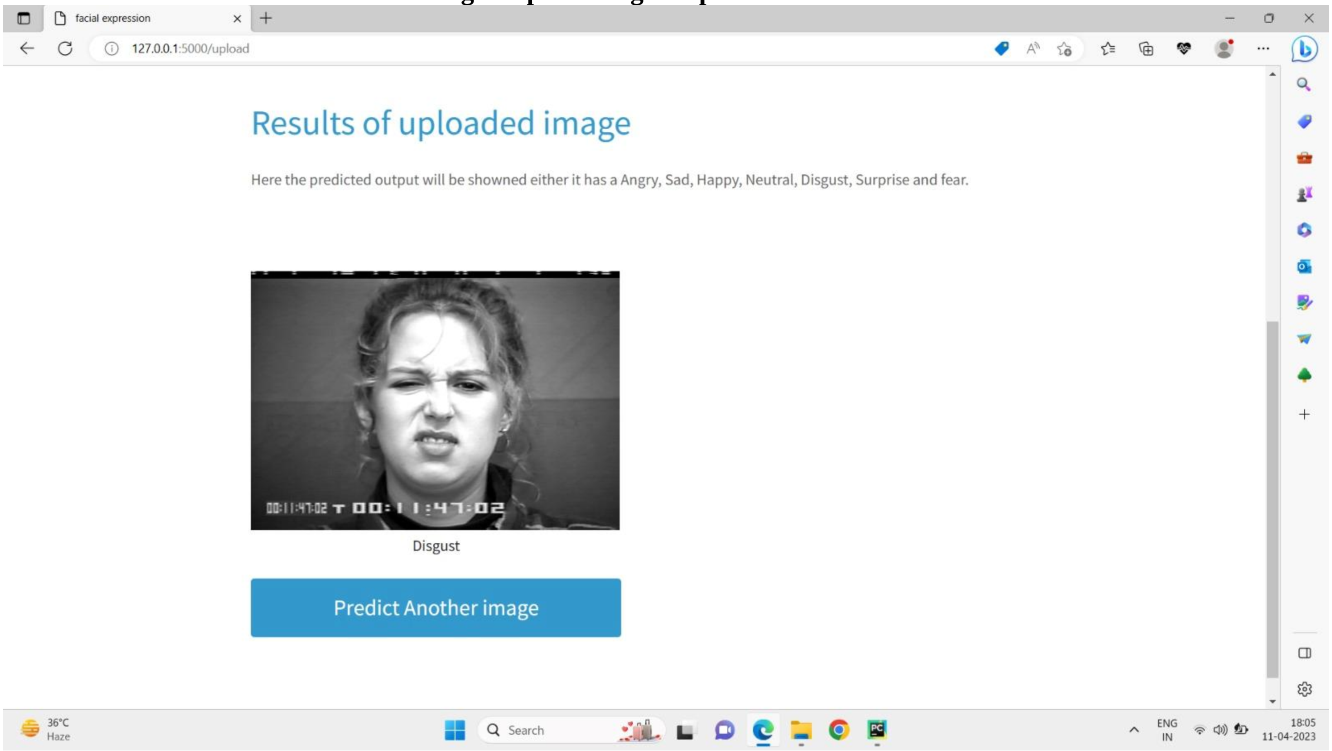


Fig: output as disgust expression

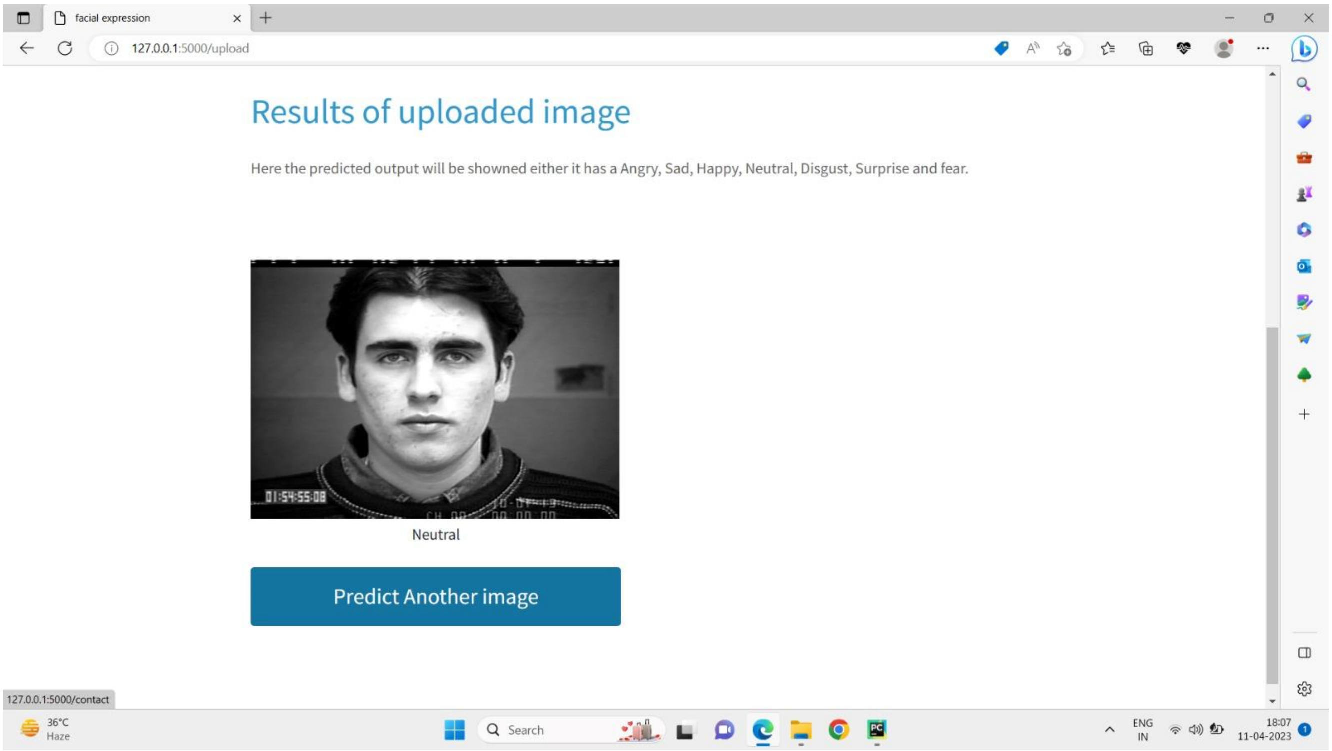


Fig: output as neutral expression

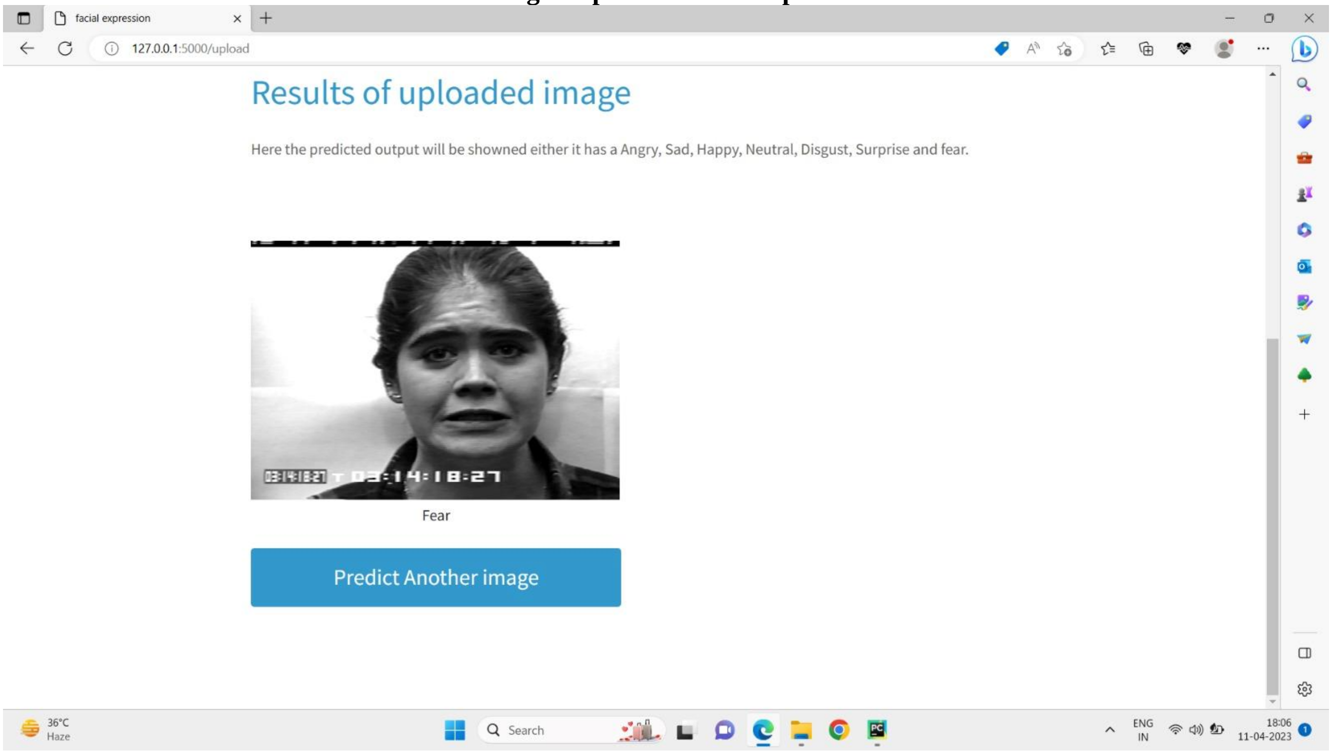


Fig: output as fear expression

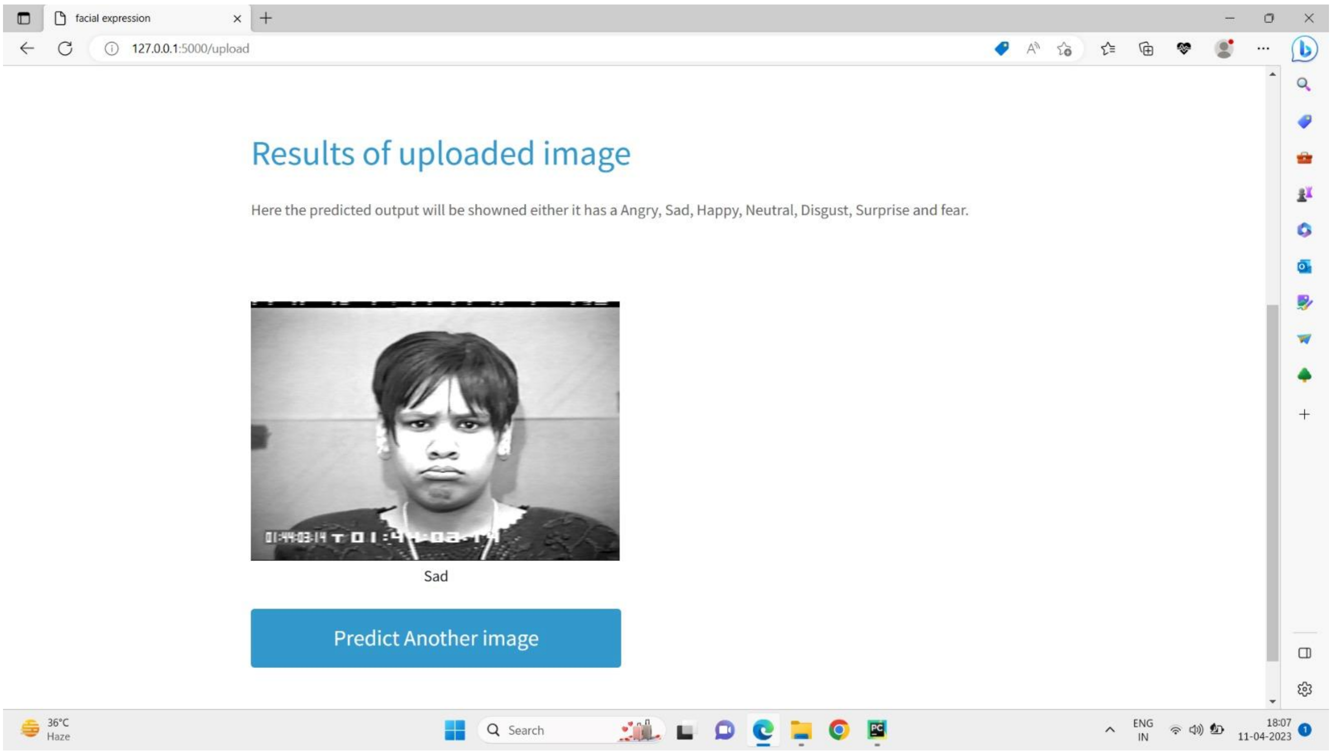
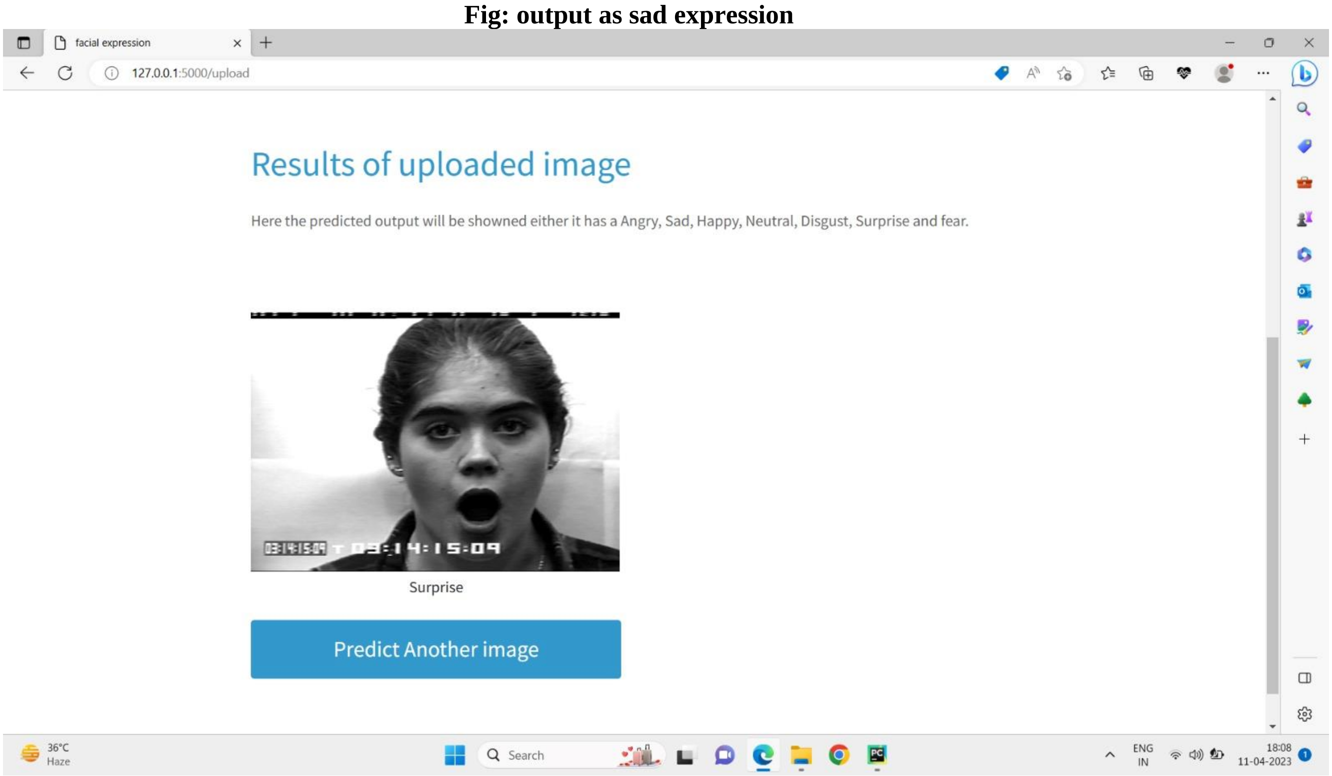


Fig: output as sad expression



. **Fig: output as surprise expression**

V. CONCLUSION

This project explores the field of facial sentiment analysis. A convolution neural network is presented for the task of classification of facial images into the seven regular emotions which are, happiness, fear, sadness, anger, surprise, disgust and neutral. The fer2013 dataset has been used for training and testing purposes due to its extensiveness and robustness. The applications of Facial Sentiment Analysis are huge and it is a field that will see many more new and better contributions than ever before.

REFERENCES

[1] A. N. Wiens, R. G. Harper and J. D. Matarazzo, Nonverbal communication: The state of the art. J. Wiley & Sons, 1978.

[2] M. J. Jones and P. Viola, “Robust real-time face detection,” International journal of computer vision, vol. 57, no. 2, pp. 137–154, 2004.

[3] Edwards, Henry J. Jackson, Jane , and Philippa E. Pattison. ”Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review.” Clinical psychology review 22.6: 789-832, 2002.

[4] Chu, William Wei-Jen Tsai, Hui-Chuan, YuhMin Chen and Min-Ju Liao. ”Facial expression recognition with transition detection for students with high-functioning autism in adaptive elearning.” Soft Computing: 1-27, 2017.

[5] Chlo, Clavel, Iona Vasilescu, Laurence Devillers, Gal Richard, and Thibaut Ehrette. ”Feartype emotion recognition for future audio-based surveillance systems.” Speech Communication 50, 2008.

[6] Cowie, Nicolas Tsapatsoulis, Ellen Douglas-Cowie, Roddy, Winfried Fellenz,George Votsis, Stefanos Kollias, and John G. Taylor. ”Emotion recognition in human-computer interaction.” IEEE Signal processing magazine 18, no. 1: 32-80, 2001.

[7] Pooya, Thomas Huang, Khorrami, and Thomas Paine. ”Do deep neural networks learn facial action units when doing expression recognition?.” in IEEE Conference on Comp. Vision. 2015.

[8] Aneja, Gary Faigin, Deepali, Alex Colburn, Barbara Mones, and Linda Shapiro. ”Modeling stylized character expressions via deep learning.” In Asian Conference on Computer Vision, pp. 136-153. Springer, 2016.

[9] Paul, Ekman, and W.V. Friesen. ”Constants across cultures in the face and emotion.” Journal of personality and social psychology, 1971.

[10] LeCun, Y. Generalization & network design strategies. Connectionism in perspective, 143155, 1989.

[11] Carrier, I. J., Mirza, P. L., Courville, A., Goodfellow, M., & Bengio, Y. FER-2013 database. University de Montreal, 2013.

[12] H. Mliki, H. BenAbdallah, M. Hammami, and N. Fourati, “Data miningbased facial expressions recognition system.” in SCAI, 2013, pp. 185– 194.

[13] P. Liu, Y. Tong, Z. Meng, and S. Han, “Facial expression recognition via a boosted deepbelief network,” in the IEEE Conference on Comp. Vision and Pattern Rec., pp. 1805-1812. 2014.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. arXiv preprint arXiv:1409.0575, 2014.

[15] S. Moore and R. Bowden, “Local binary patterns for multi-view facial expression recognition,” Computer Vision Image Understand. vol. 115, no. 4, pp. 541–558, 2011.