**Live Class Monitoring System**

**(Face Emotion Recognition )**

**Aniket Nichat, Rohit Thawali, Aniket Deulkar**

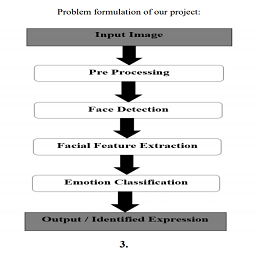
**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

This project aims to classify the emotion on a person's face into one of **seven categories**, using deep convolutional neural networks. The model is trained on the **FER-2013** dataset which was published on International Conference on Machine Learning (ICML). This dataset consists of 35887 grayscale, 48x48 sized face images with **seven emotions** - angry, disgusted, fearful, happy, neutral, sad and surprised.

**1.Problem Statement**

Humans are well-trained in reading the emotions of others, in fact, at just 14 months old, babies can already tell the difference between happy and sad. But can computers do a better job than us in accessing emotional states? To answer the question, We designed a deep learning neural network that gives machines the ability to make inferences about our emotional 

* Destination\_Type: Sigma Cabs divides any destination into one of the 14 categories.
* Customer\_Rating: Average of lifetime ratings of the customer till date CancellationLast1Month: Number of trips cancelled by the customer in last 1 month
* Var1, Var2 and Var3: Continuous variables masked by the company. Can be used for modelling purposes
* Gender: Gender of the customer

Facial expression recognition is a process performed by humans or computers, which consists of:

1. Locating faces in the scene (e.g., in an image; this step is also referred to as facedetection),

2. Extracting facial features from the detected face region (e.g., detecting the shape of facialcomponents or describing the texture of the skin in a facial area; this step is referred to asfacial feature extraction),

3. Analyzing the motion of facial features and/or the changes in the appearance of facialfeatures and classifying this information into some facial-expressioninterpretativecategories such as facial muscle activations like smile or frown, emotion (affect)categories like happiness or anger, attitude categories like (dis)liking or ambivalence, etc.(this step is also referred to as facial expression interpretation).

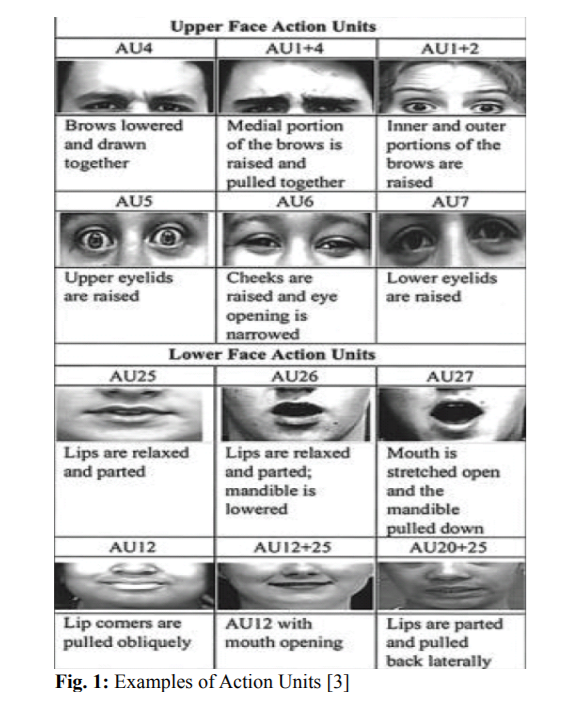
Several Projects have already been done in this fields and our goal will not only be to develop a Automatic Facial Expression Recognition System but also improving the accuracy of this system compared to the other available systems.

**2. Introduction**

## The Indian education landscape has been undergoing rapid changes for the past 10 years owing to the advancement of web-based learning services, specifically, eLearning platforms. Global E-learning is estimated to witness an 8X over the next 5 years to reach USD 2B in 2021. India is expected to grow with a CAGR of 44% crossing the 10M users mark in 2021. Although the market is growing on a rapid scale, there are major challenges associated with digital learning when compared with brick and mortar classrooms. One of many challenges is how to ensure quality learning for students. Digital platforms might overpower physical classrooms in terms of content quality but when it comes to understanding whether students are able to grasp the content in a live class scenario is yet an open-end challenge. In a physical classroom during a lecturing teacher can see the faces and assess the emotion of the class and tune their lecture accordingly, whether he is going fast or slow. He can identify students who need special attention. Digital classrooms are conducted via video telephony software program (exZoom) where it’s not possible for medium scale class (25-50) to see all students and access the mood. Because of this drawback, students are not focusing on content due to lack of surveillance. While digital platforms have limitations in terms of physical surveillance but it comes with the power of data and machines which can work for you. It provides data in the form of video, audio, and texts which can be analysed using deep learning algorithms. Deep learning backed system not only solves the surveillance issue, but it also removes the human bias from the system, and all information is no longer in the teacher’s brain rather translated in numbers that can be analysed and tracked.

## **3. Psychological Background**

## In 1978, Ekman et al. [2] introduced the system for measuring facial expressions called FACS – Facial Action Coding System. FACS was developed by analysis of the relations between muscle(s) contraction and changes in the face appearance caused by them. Contractions of muscles responsible for the same action are marked as an Action Unit (AU). The task of expression analysis with use of FACS is based on decomposing observed expression into the set of Action Units. There are 46 AUs that represent changes in facial expression and 12 AUs connected with eye gaze direction and head orientation. Action Units are highly descriptive in terms of facial movements, however, they do not provide any information about the message they represent. AUs are labeled with the description of the action (Fig.1)..



Facial expression described by Action Units can be then analyzed on the semantic level in order to find the meaning of particular actions. According to the Ekman's theory [2], there are six basic emotion expressions that are universal for people of different nations and cultures. Those basic emotions are joy, sadness, anger, fear, disgust and surprise.

The Facial Action Coding System was developed to help psychologists with face behavior analysis. Facial image was studied to detect the Action Units occurrences and then AU combinations were translated into emotion categories. This procedure required much effort, not only because the analysis was done manually, but also because about 100 hours of training were needed to become a FACS coder. That is why, FACS was quickly automated and replaced by different types of computer software solutions.

# 4. **Reason For Face Emotion Recognition**

* Student emotion during live class
* Entering in offices with face detection
* To help access personal device
* Face attendance based system and so on

# **5. How Face Emotion Recognition Work**

Facial emotion recognition is the process of detecting human emotions from facial expressions. The human brain recognizes emotions automatically, and software has now been developed that can recognize emotions as well. This technology is becoming more accurate all the time, and will eventually be able to read emotions as well as our brains do.

AI can detect emotions by learning what each facial expression means and applying that knowledge to the new information presented to it. Emotional artificial intelligence, or emotion AI, is a technology that is capable of reading, imitating, interpreting, and responding to human facial expressions and emotions.

**6. Steps involved:**

# **Realtime Local Video Face Detection**

We created one patterns for detecting and predicting single faces and as well as multiple faces using OpenCV videocapture in local. For Webapp , OpenCV can’t be used. Thus, using Streamlit-Webrtc for front-end application.

.

* **Real time video detector usi ng haarcascade file**

Haar Cascade is a machine learning-based approach where a lot of positive and negative images are used to train the classifier. Positive images – These images contain the images which we want our classifier to identify. Negative Images – Images of everything else, which do not contain the object we want to detect

**7.1. Model:**

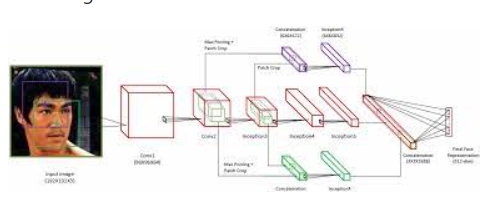
1. **Deep Face:**

DeepFace is a deep learning facial recognition system created by a research group at Facebook. It identifies human faces in digital images. The program employs a nine-layer neural network with over 120 million connection weights and was trained on four million images uploaded by Facebook users.The Facebook Research team has stated that the DeepFace method reaches an accuracy of 97.35% ± 0.25% on Labeled Faces in the Wild (LFW) data set where human beings have 97.53%. This means that DeepFace is sometimes more successful than the human beings.



• The actual emotion in the Picture was ANGRY Face and then using DeepFace I found the prediction is ANGRY.

## **CNN**

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural network, most commonly applied to analyze visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation equivariant responses known as feature maps.Counter-intuitively, most convolutional neural networks are only equivariant, as opposed to invariant, to translation. They have applications in image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series.CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "full connectivity" of these networks make them prone to overfitting data. Typical ways of regularization, or preventing overfitting, include: penalizing parameters during training (such as weight decay) or trimming connectivity (skipped connections, dropout, etc.) CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filters. Therefore, on a scale of connectivity and complexity, CNNs are on the lower extreme. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns to optimize the filters (or kernels) through automated learning, whereas in traditional algorithms these filters are hand-engineered. This independence from prior knowledge and human intervention in feature extraction is a major advantage. 

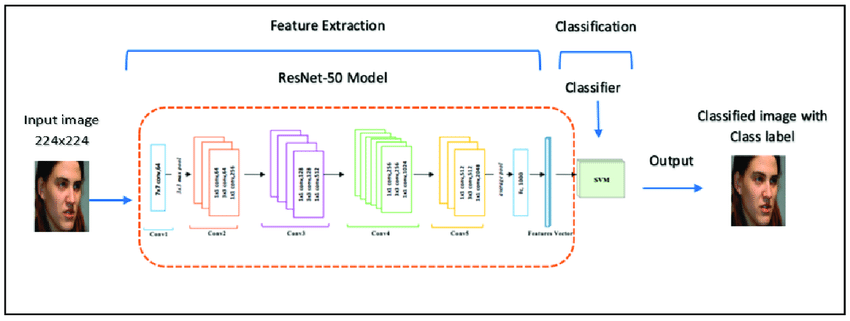
• The training gave the accuracy of 66% and test accuracy of 56%. It seems excellent. So, I save the model and the detection i got from live video was excellent.

• One drawback of the system is the some Disgust faces are showing Neutral .Because less no. of disgust faces are given to train .This may be the reason.

• I thought it was a good score should improve the score.

• Thus I decided that I will deploy the model.

# **3 Using ResNet50**

ResNet50 is a variant of ResNet model which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer. It has 3.8 x 10^9 Floating points operations. It is a widely used ResNet model and we have explored ResNet50 architecture in depth.ResNet-50 is a convolutional neural network that is 50 layers deep.  You can load a pretrained version of the network trained on more than a million images from the ImageNet database [1]. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals

**8. Conclusion:**

* The model which was created by CNN layers gave training accuracy of 66% and test accuracy of 56% .
* We have also included the video of my WebApp working in Local.
* Codes which are deployed are in Github Repository.
* It was such an amazing and interesting project. We learnt a lot from this.

**References-**

1. https://arxiv.org/pdf/1902.01019.pdf
2. Towardsdatascience
3. Analytics Vidhya