Project Report on:

FACIAL EXPRESSION BASED EMOTION RECOGNITION USING DEEP CONVOLUTIONAL NEURAL NETWORK

Project Report Submitted for Fulfilment of the 8th Semester of B.Tech Course

Under the supervision of

PROF. SIRSHENDU HORE

Assistant Professor, Dept. of CSE

Submitted by:

Student Name	University Roll No.
SAIKAT DEY	17600116029
PRITAM MONDAL	17600116036
SUDIPTA SEN	17600116013
ANINDA GANGULY	17600116062

HOOGHLY ENGINEERING & TECHNOLOGY COLLEGE

Approved by AICTE| Recognized by UGC | Affiliated to MAKAUT

2016-2020

CONTENTS

EMOTION RECOGNITION

INTRODUCTION

FACIAL EMOTION

Emotions and facial expressions:

Emotions are an inevitable part of any inter-personal communication.

Facial expressions contribute to a major part of recognizing human emotions.

How Deep Learning can be used to recognize emotions:

Most of the clues for recognizing emotions come from a few parts of the face. This means, ideally, a machine learning approach can be use to focus only on those important parts of the face.

A deep learning based frame work for facial expression recognition can be built to recognize emotions of human faces.

OBJECTIVE

Recognition of Emotion is subject of interest for both psychologists and engineers. The objective of the study is to build a model with high accuracy and low validation loss to classify human emotions from videos using Deep Convolutional Neural Networks.

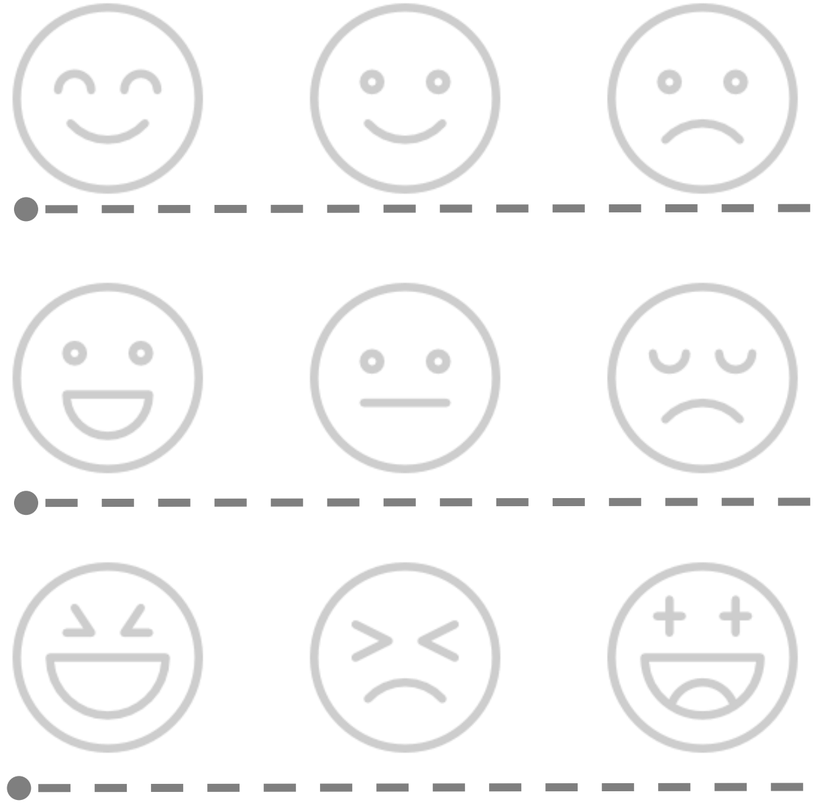
In the previous report, we had worked on Frequency Domain by recognizing emotions from EEG channels in the DEAP dataset using the Valence Arousal Model. Here, we have worked on Spatial Domain by working on image dataset in the FER 2013 dataset.

7 Different types of emotions (Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise) were identified on human faces. Supervised Deep Learning approach was used to train the model and test it against video samples.

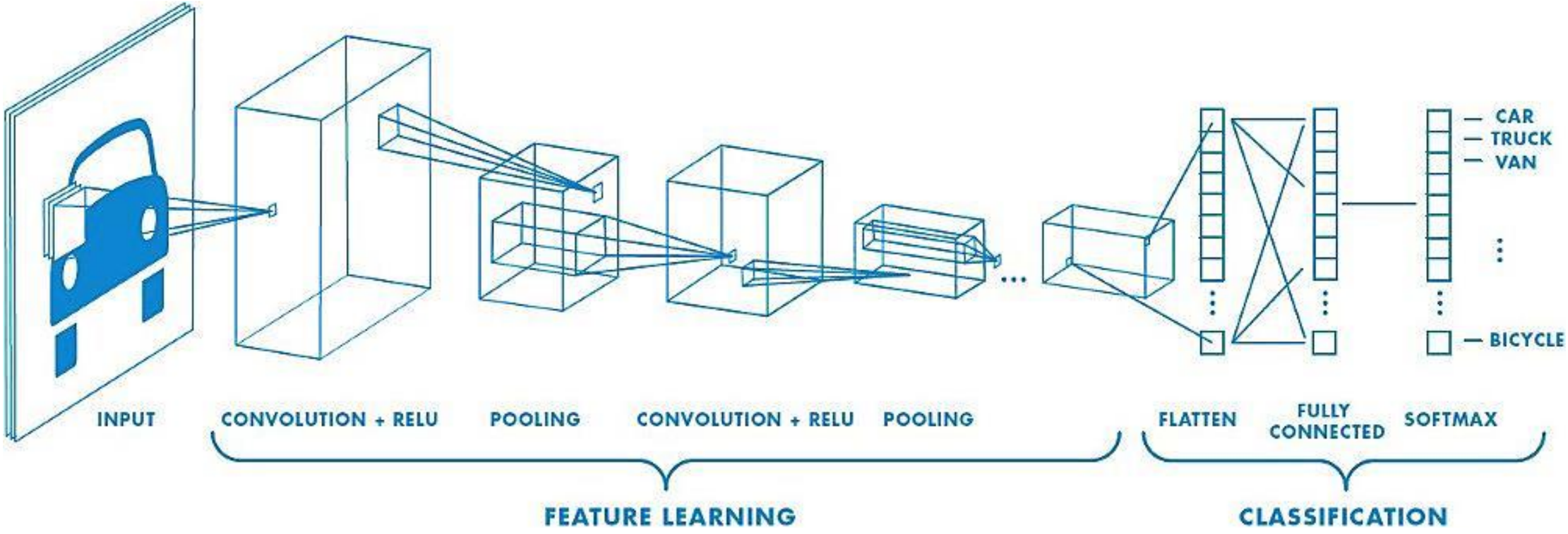
01

02

03



CONVOLUTIONAL NEURAL NETWORK

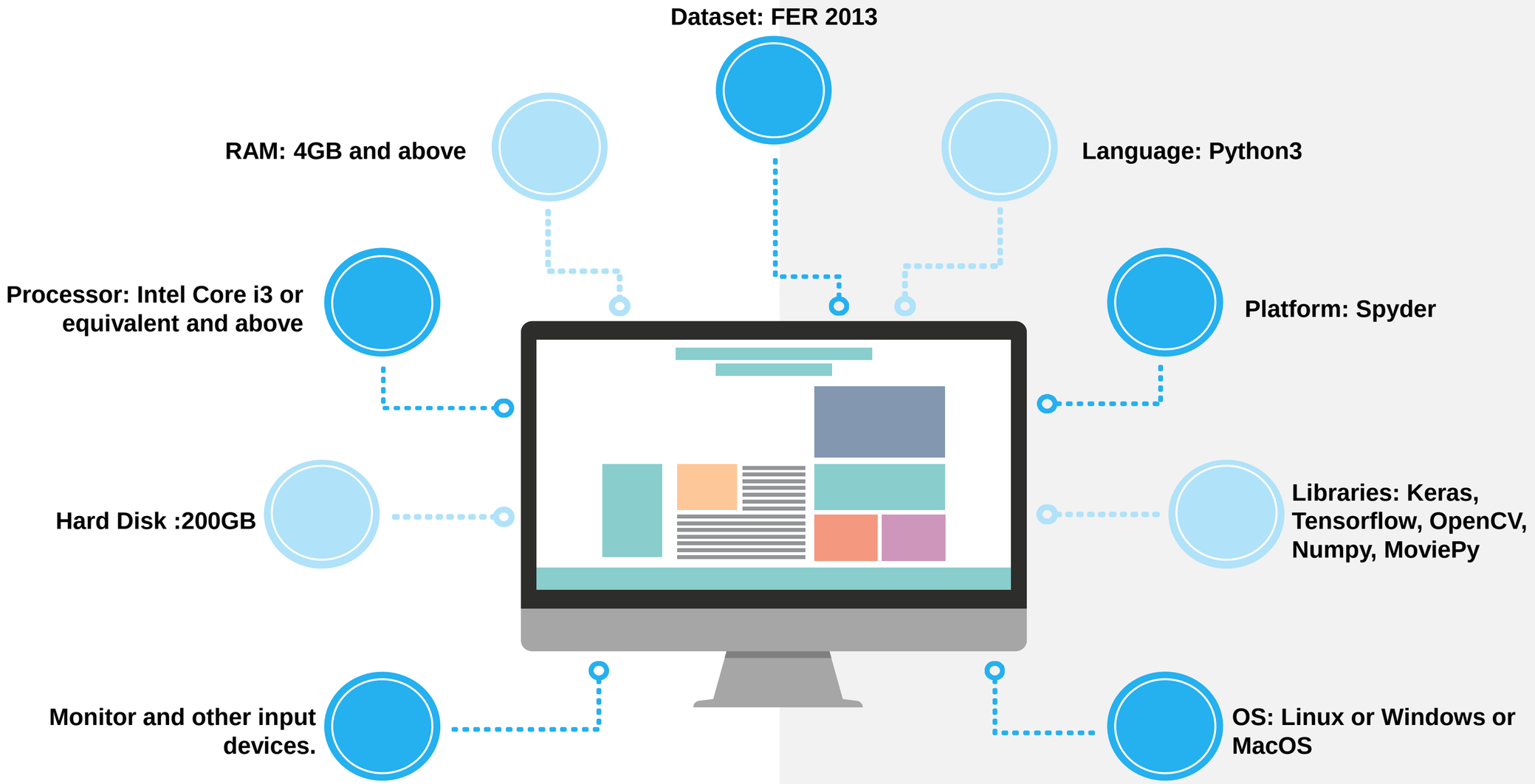


Propagation is uni-directional where CNN contains one or more convolutional layers followed by pooling and bidirectional where the output of convolution layer goes to a fully connected neural network for classifying the images.

Filters are used to extract certain parts of the image. In MLP the inputs are multiplied with weights and fed to the activation function.

Convolution uses RELU and MLP uses nonlinear activation function followed by softmax. Convolution neural networks show very effective results in image and video recognition.

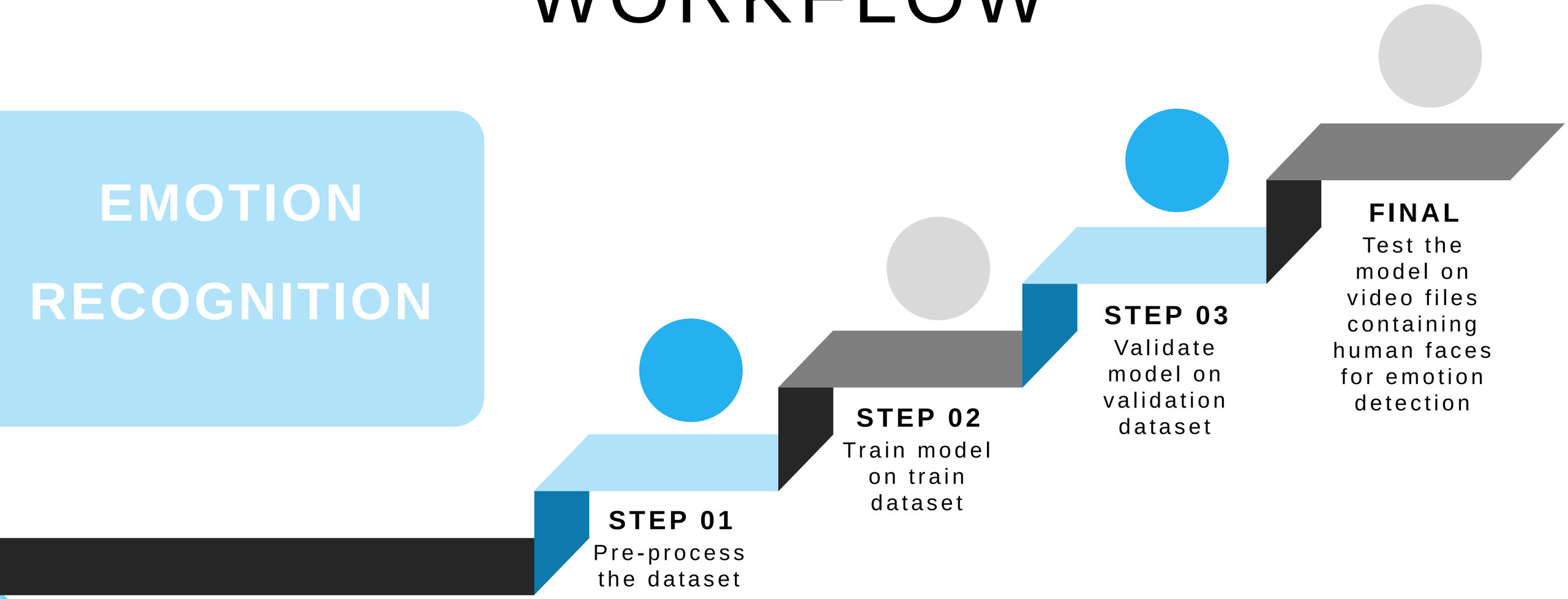
HARDWARE & SOFTWARE REQUIREMENTS



EXPERIMENTAL METHODOLOGY

WORKFLOW

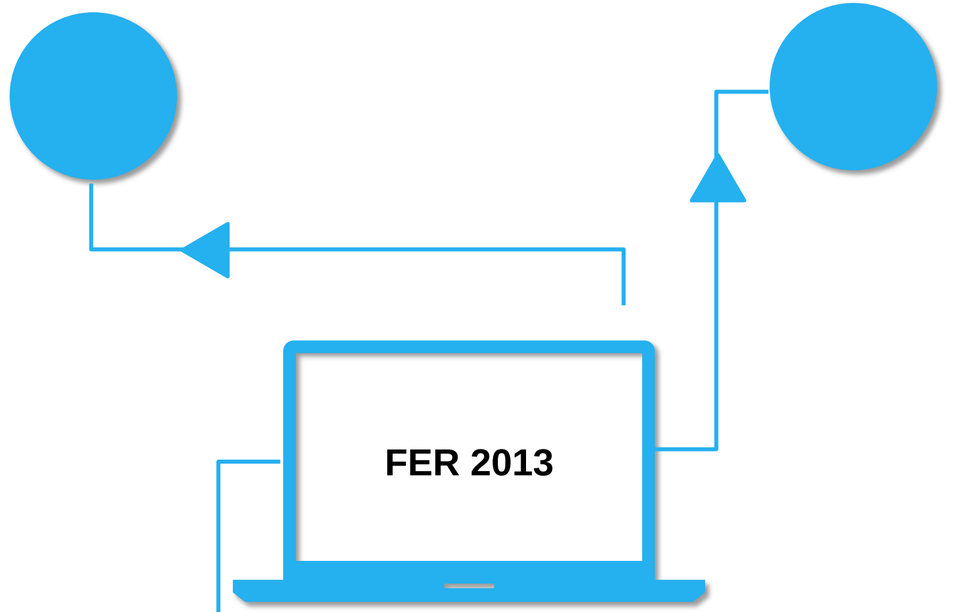
EMOTION RECOGNITION



DATASET

Content

FER 2013 or the Facial Expression Recognition 2013 data consists of 48x48 pixel grayscale images of faces.



Description

The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image.

Categories

(0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral)

Size

The training set consists of 28,736 examples. The validation set used consists of 3,589 examples.

PRE-PROCESSING

2 sets of data used, one for training, the other for validation

Train data set was normalized and size was increased using image augmentation

Dataset was converted to grayscale

Categorical class mode used, and dataset was shuffled.

Image Augmentation

Rescale range 1-255

Rotation range:30%

Shear and zoom range: 30%

Width and height shift range: 30%

Horizontal flip

TRAINING THE MODEL

2D CNN

4 blocks, with each block containing 2 2D CNNs used.
Kernel size 3x3.
Feature detectors from block 1 to 4 were 32,64,128,256 respectively.

ACTIVATION FUNCTION

Exponential Linear Unit (ELU) was used in the CNNs.

ELU becomes smooth slowly, and can produce negative outputs.

BATCH NORMALISATION

To increase the stability of the neural network, batch normalization normalized the activations of the previous layers at each batch.

MAX POOLING

Maxpooling down sampled the input representation by taking maximum value over the window defined by pool size for each dimension along feature axis.

DROP OUT

Dropout helps in generalizing by randomly turning off a given neuron to avoid over-fitting.

20% dropout was used.

DENSE LAYER

It is a regularly connected NN layer.

7 neurons, each corresponding to an emotion class was used.

TRAINING THE MODEL (contd.)

SOFTMAX

Softmax converts a real vector to a vector of categorical probabilities.

It helped our model in categorically predicting the emotion classes.

OPTIMISER

Optimizers were used to obtain high accuracy and low validation loss.

Adam Optimizer was used since it converges rapidly.

CHECKPOINT

Checkpoints were used during training to save the best model with least validation loss, and resume training or fine tune the model later.

EARLY STOPPING

Early stopping was used to stop the training once the model performance stopped improving to avoid over fitting and under fitting.

REDUCE LEARNING

Learning rate was reduced by 20% once the learning stagnates, and validation loss stopped improving.

EPOCH

25 Epochs or iterations were used to train the model.

MODEL SUMMARY

EMOTION CLASSES

7 classes of emotions: [Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise]

DATASET SIZE

Train Data: 28736 images belonging to 7 classes

Validation Data: 3589 images belonging to 7 classes

MODEL INFO

Model architecture: Custom VGG Model

Model type: Keras Sequential Model

PARA METERS

Total Parameters: 1,328,167

Trainable Parameters: 1,325,991

Non-trainable Parameters: 2,176

MODEL SUMMARY (contd.)

8 CONVOLUTION 2D LAYERS

11 ACTIVATION LAYERS

10 BATCH NORMALIZATION LAYERS

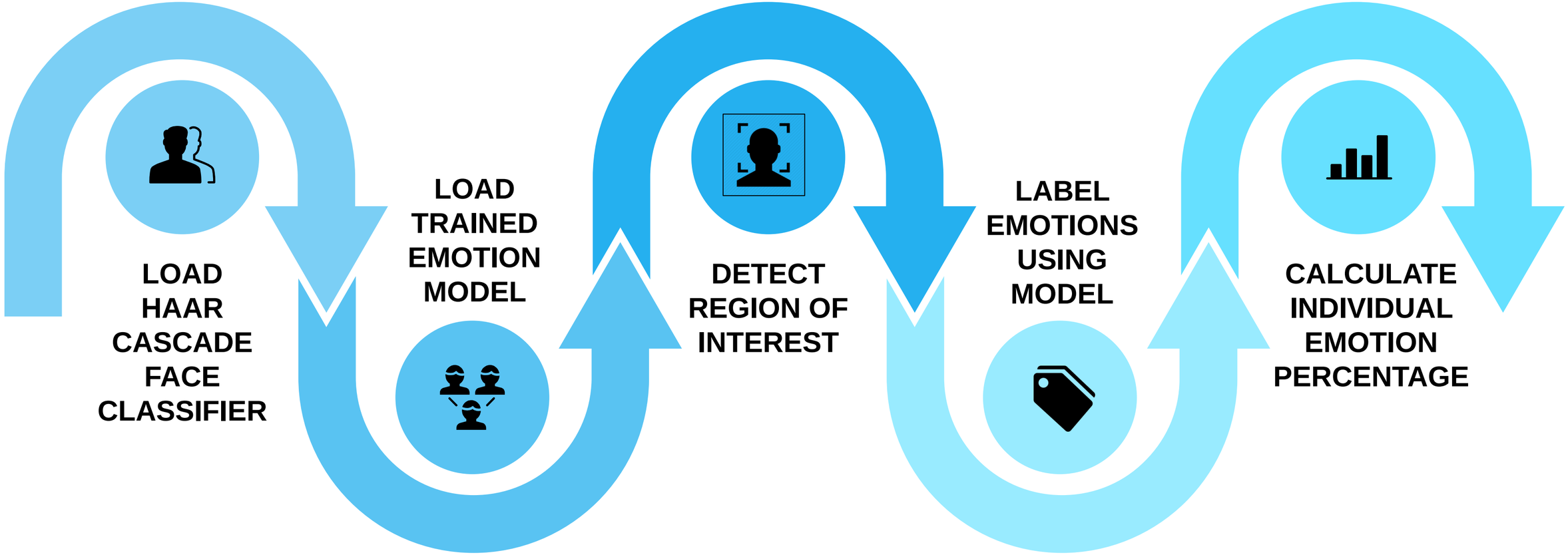
4 MAX POOLING 2D LAYERS

5 DROPOUT LAYERS

1 FLATTEN LAYER

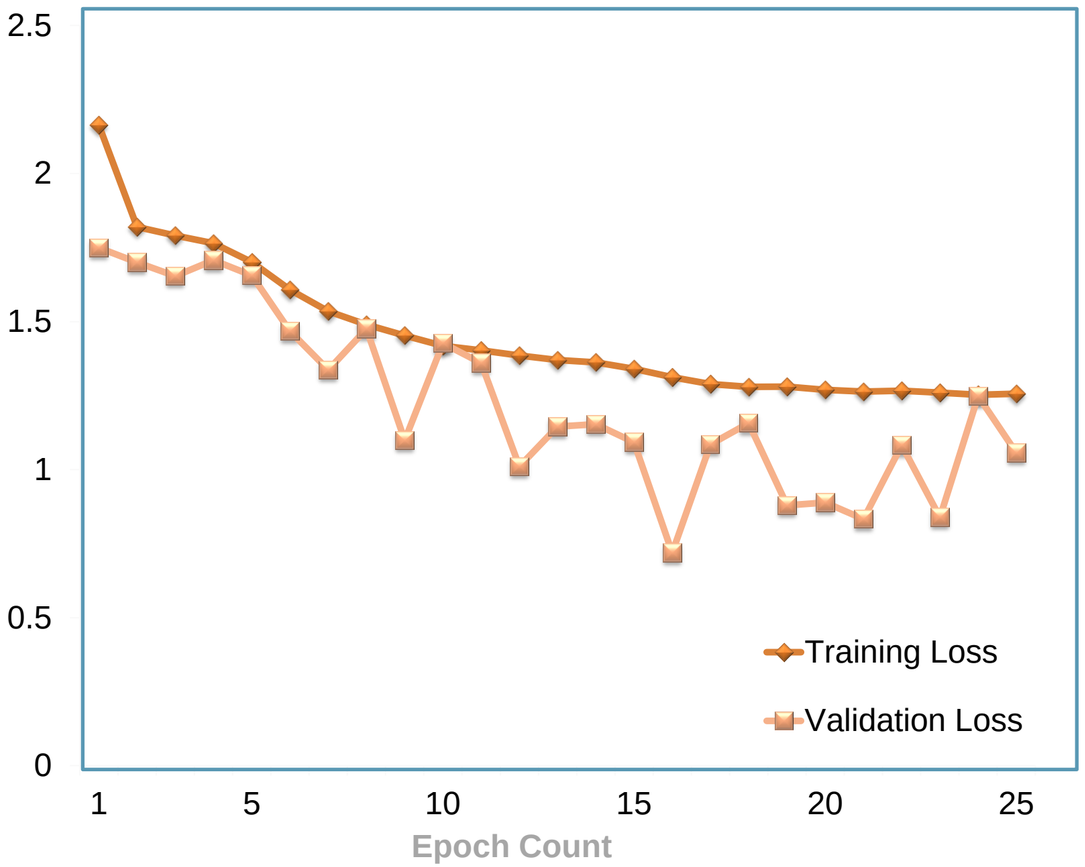
3 DENSE LAYERS

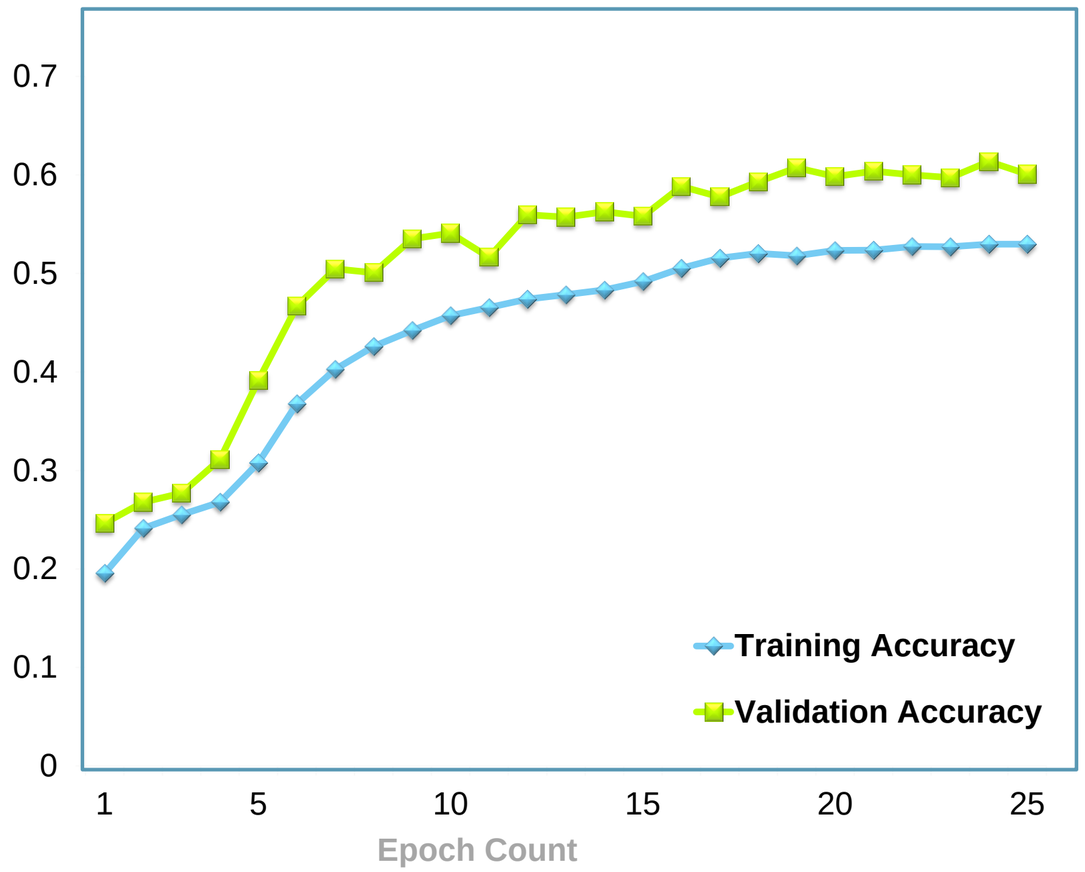
TESTING THE MODEL



RESULTS AND DISCUSSION

LOSS AND ACCURACY RESULTS





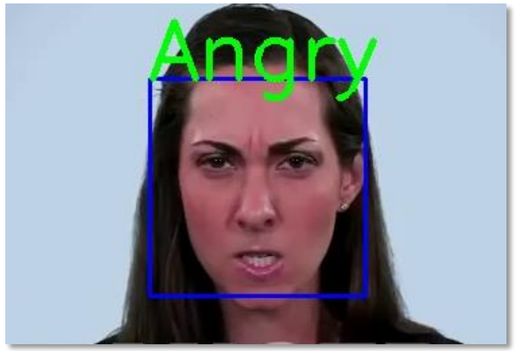
Validation loss was best on Epoch 16 and did not improve further.

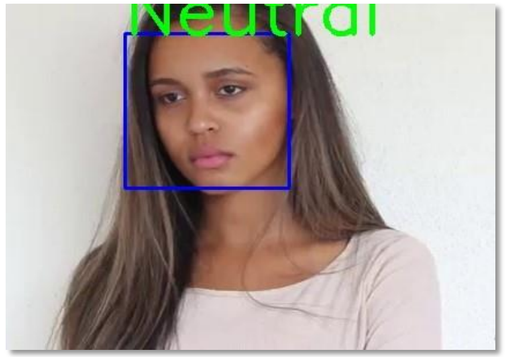
Learning rate was reduced in epochs 15,19,22 and 25, because no improvement in validation loss was seen.

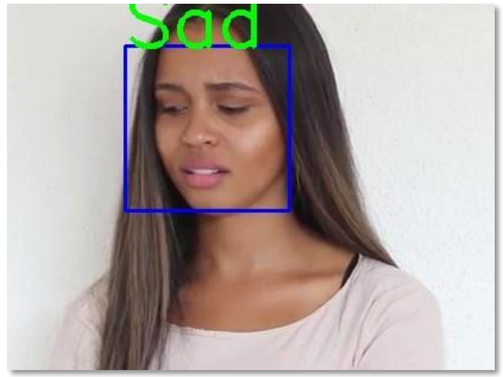
Early stopping was observed in epoch 25 because the validation loss didn't improve for 9 continuous epochs.

Validation accuracy obtained from the best epoch was 0.5884 or 58.84%.

DETECTING EMOTIONS IN VIDEO







LIVE EMOTION DETECTION

Click play to view the emotion detection using our model in action

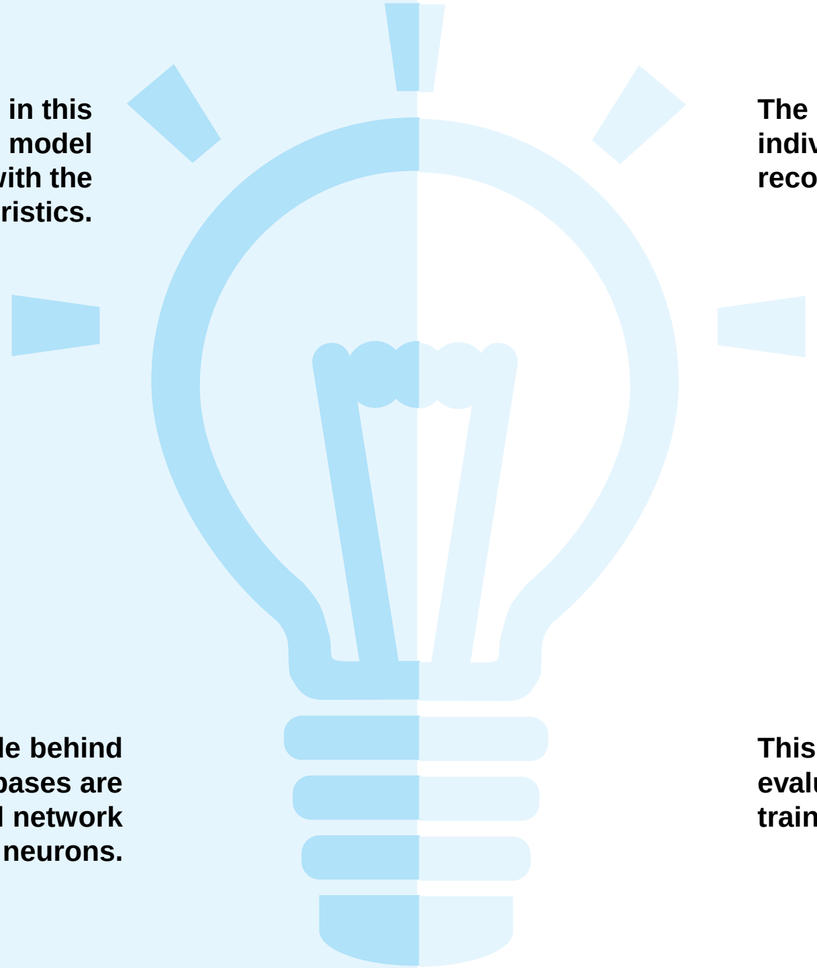
CONCLUSION & FUTURE SCOPE

CONCLUSION

The facial expression recognition system presented in this project work contributes a resilient face recognition model based on the mapping of behavioural characteristics with the physiological biometric characteristics.

The physiological characteristics of the human face with relevance to various expressions such as happiness, sadness, fear, anger, surprise and disgust are associated with geometrical structures which restored as base matching template for the recognition system.

The behavioural aspect of this system relates the attitude behind different expressions as property base. The property bases are alienated as exposed and hidden category in neural network algorithmic neurons.



The model evaluates the expressional uniqueness of individual faces and provide a resilient expressional recognition model in the field of emotion recognition.

The training of facial expression-based face recognition system supervised by FER-2013 emotion data set and resulting an accuracy of 52.98% on training set and 60.08% on validation set.

This system is able to perform real time facial expression evaluation on any video completely independent from the train and validation set.

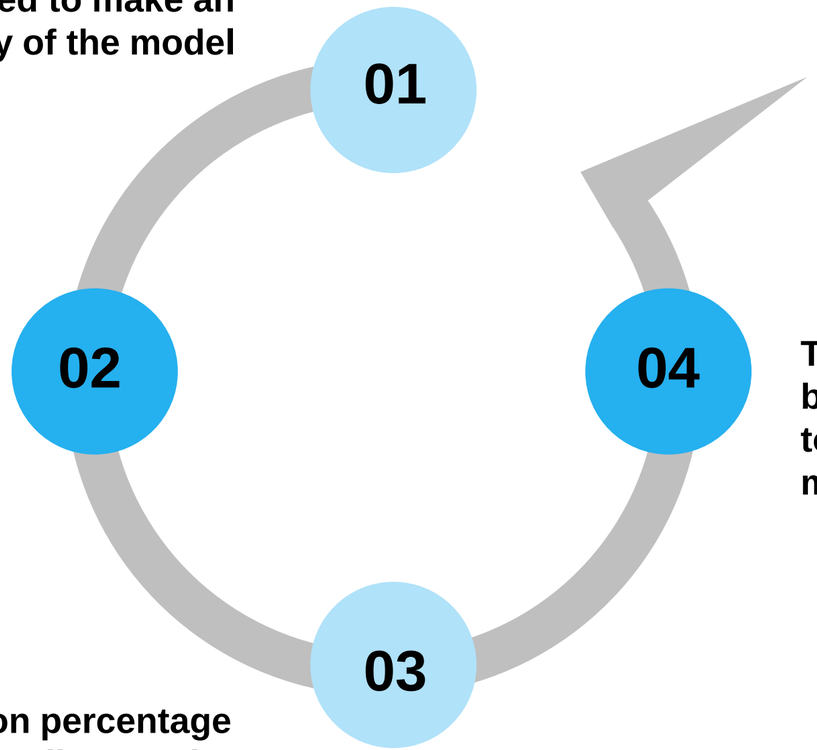
FUTURE SCOPE

01 Alternative optimizers may be used to make an attempt at improving accuracy of the model

02 To increase efficiency, fast facial expression recognition system may be used by application of new feature called W_HOG where W indicates discrete wavelet transform and HOG indicates histogram of oriented gradients feature

03 Simultaneous emotion percentage calculation using multi-threading can be done later.

04 Testing script can be implemented into a backend of a web-app to move processing to server and reduce load on client machine.



BIBLIOGRAPHY

01

Think Stats probability and Statistics for Programmers by Allen B. Downey, published by O'Reilly Media.

02

Mining of Massive Datasets by Jure Leskovec Stanford Univ, Anand Rajaraman Milliway Labs, Jeffrey D. Ullman Stanford Univ.

03

Bayesian Reasoning and Machine Learning by David Barber

04

Kaggle Simulations

05

Hands-On Convolutional Neural Networks with TensorFlow by Iffat Zafar, Giounona Tzanidou, Richard Burton, Nimesh Patel, Leonardo Araujo

THANK YOU