**Faculty of Computers &Artificial Intelligence**

**Computer Science Department**

**2022-2023**

**Facial Expression Recognition Using Artificial Neural Networks**

Prepared by:

1. 202000076 احمد محمد محمد امام عبدالفتاح
2. 202000899 مصطفى حماده جمال محمد
3. 202000081 احمد محمود حمادة محمود
4. 202000812 محمد كمال عبدالحفيظ محمد
5. 202000051 احمد عصام عبدالفتاح العدل
6. 202000022 احمد جمال حسيني علي
7. 202000544 عبدالله أشرف السيد احمد

**FACIAL EXPRESSION RECOGNITION**

**USING**

**ARTIFICIAL NEURAL NETWORKS**

**1- Project Idea :**

Analysis and recognition of human facial expressions from images and video forms the basis for understanding image content at a higher semantic level.

Expression recognition forms the core task of intelligent systems based on human–computer interaction (HCI). In this paper, we explore the use of Artificial Neural Networks in performing expression recognition. We analyze seven basic types of human expressions – neutral, happy, sad, disgust, anger, surprise and fear. The expression recognition task is divided into two steps – an automatic facial feature extraction step, and a classification step that employs Multi Layer Perceptrons and Radial Basis Function Networks. In many face recognition systems the important part is face detection. The task of detecting face is complex due to its variability present across human faces including colour, pose, expression, position and orientation. So using various modeling techniques it is convenient to recognize various facial expressions. In the field of image processing it is very interesting to recognize the human gesture by observing the different movement of eyes, mouth, nose, etc. Classification of face detection and token matching can be carried out any neural network for recognizing the facial expression. Facial expression provides vital cues about the emotional status of a person.

**2- Main Functionalities :**

In the Automated Facial Expression Recognition , the goal is to detect face reaction : determining which reaction is on the face in the image

Facial Emotion Recognition (FER) is the technology that analyses facial expressions from static images in order to reveal information on one’s emotional state. The complexity of facial expressions, the potential use of the technology in any context, and the involvement of new technologies such as artificial intelligence raise significant privacy risks.

**3- Similar applications in the market :**

1. In Education :

Monitor students’ attention

Diagram

Description automatically generated with low confidence

High school students in one Chinese school may want to think twice before dozing off in class. Artificially intelligent cameras with facial recognition tools will be watching.

The Hangzhou No. 11 Middle School has installed a “smart classroom behavior management system”, which captures students’ expressions and movements, analyzing them with big data to make sure they’re paying attention, media reported on Thursday.

1. **In Public safety :**
   * lie detectors and smart border control
   * predictive screening of public spaces to

identify emotions triggering potential terrorism threat

* + analyzing footage from crime scenes to indicate

potential motives in a crime



1. In **Crime detection :**

* detect and reduce fraudulent insurance claims
* deploy fraud prevention strategies
* spot shoplifters



1. In Health care **:**

* detect autism or neurodegenerative diseases
* predict psychotic disorders or depression to

identify users in need of assistance

* suicide prevention
* detect depression in elderly people
* observe patients conditions during treatment



1. **In Employment :**

* help decision-making of recruiters
* identify uninterested candidates in a job interview
* monitor moods and attention of employees

Two people sitting at a table

Description automatically generated with low confidence

1. In **Provision of personalized services :**

* analyze emotions to display personalized messages

in smart environments

* provide personalized recommendations e.g. on

music selection or cultural material

* analyze facial expressions to predict individual

reaction to movies

A picture containing person, person, wearing, posing

Description automatically generated

1. **Customer behavior analysis and advertising :**

* analyze customers’ emotions while shopping

focused on either goods or their arrangement

within the shop

* advertising signage at a railway station using

a system of recognition and facial tracking for

marketing purposes

Two people in a grocery store

Description automatically generated with low confidence

**-4 An initial literature review:**

In a research field of emotion detection, there is a contribution of several domains like machine learning, natural language, neuroscience, etc. In previous works, they individually rummaged facial expressions, voice features, and textual data as universal indicators of emotions. Emotion can be classified into several static classifications like happiness, sadness, disgust, anger, fear, and surprise. In later works are improved by combining the image, voice, and textual data. The fusion of this data gives the maximum accurate result. This type of fusion can be done in three ways early, late, or hybrid. Other ethos features the elements of emotion and the collaborations between emotional processes and other intellectual

procedures.

A. Emotion Detections Through Facial Feature

Recognition

This work deals with the emotion recognition with the Machine learning using support vector machine (SVM). Some principles are work to detection, extraction, and evaluation of facial expressions of image. These are:

1. Viola-Jones cascade object detectors and Harris corner key-points to extract faces and facial features from images.
2. Histogram of oriented gradients (HOG) feature extraction.

iii) Support vector machines (SVM) to train a multi-class predictor for classifying the sevenfundamental human facial expressions such as: (Anger, Contempt, Disgust, Fear, Happiness, Sadness, Surprise).

Computers can easily recognize facial expressions and can find out the motive of a person including in entertainment, social media, content analysis, criminal justice, and healthcare. Here is discussed mainly two-approach such as: (Zhang's approach and Gabor wavelet coefficients). Zhang has shown that lower resolution (64x64) is adequate, we will resize the extracted faces to 100x100 pixels. When using the HOG and SVM classifier only, the accuracy for detection is 81%, much better than a Fisher's face. Only approach. When using the dual-classifier method, the accuracy is the same as HOG only at 81%, but the testing process is 20% faster

B. SVM Point-based Real-time Emotion Detection

This work deals with the emotion recognition with Machine learning using a cascade of a multi-class support vector machine (SVM) and a binary SVM. This algorithm is developed to extract emotions based on the movement of 19 feature points. These feature points are located in different regions of the countenance such as the mouth, eyes, eyebrows, and nose. It mainly works non-changeable rigid points on the nose. Its divide into facial recognition and action unit (AU). Computers can easily recognize facial expressions and can find out the motive of a person including in entertainment, social media, content analysis, criminal justice, and healthcare. A final suggestion for improvement is the fact that in the

real-time application the user needs to stay on the same distance concerning the camera from which the neutral frame was taken. Otherwise, the theory behind the displacement ratios is no longer valid. Rescaling the neutral distances based on the movement of the user can be a solution to this problem

The first major challenge was the confined measure of information for preparing a broad framework. Which needs to defeat for framework in nature. Move learning is the most prevalent response to this. In this methodology that was begun from pre-prepared strategy and calibrated this model with the put-away information which is gathered from a genuine world. A progression of starter investigations affirmed the presumption that face acknowledgment would serve better in highlight extraction. There are models where such systems are effectively utilized.

work well on the datasets that have a couple of hundred highlights or segment. The

algorithm successfully classifies an image and classify the sentiment of the image and choose the match emotion for the image. The reason behind choosing the deep learning classifier is that the classifier runs data through several layers. And a deep learning algorithm can be useful for less unpredictable issues since they gain admittance to an immense measure of information to be compelling. For pictures, the regular benchmark for preparing profound learning models for broad picture acknowledgment approaches more than 14 million pictures. For perfect visualization of emotion detection pattern analysis, it used a decision tree. In the decision tree, the character is represented by the nodes and layers, and also the outcome of the experiment is represented by the branch. The advantage of the decision tree is that it is very helpful and easy to visualize the emotion and interpret the result. The working process of a decision tree is easy to understand. If

it has been classified the data according to their movement, reactions, and order which ideally different types of emotions. This also has been classified into trees and sub trees which reflects that whether the person is sad, angry or happy, etc. if this could find something that can categorize their using these methods more simply. To do this it has been used retrain method that memorized the pattern and

satisfies the condition

**Conclusion :**

An experienced human can often identify another human’s emotions by analyzing and looking at him or her. However, in this modern age machines are becoming more intelligent. For the time been machines are trying to act more like humans. If the machine has been trained on how to react on behalf of the human sentiment at that time. Then the machine can behave and act like a human. On the other hand, if the machine can identify the emotion it can prevent lots of occurrences too. With increased proficiency and errorless computation emotion, data mining can facilitate accurate expression patterns enabling machines to find and act more like humans effectively. To determine the emotion expression patterns this thesis is created or framework with comprehensive research and field works. This followed the framework step by step to get the expected outcome. To follow the framework and to identify the emotion expression patterns more effectively and used deep learning CNN algorithm along with Keras, Tensorflow, and retraining concepts. With these techniques, it was possible to identify emotions, type of emotion in the real image. To delineate the result and procedures more visually and this has also introduced decision tree techniques which helps to decide which emotions percentage is high and which emotions percentage is low. Now the high percentage of emotions get the most possible accurate emotions. And the low percentage of emotions get the low chance of existence. With this discovery, it is now possible to determine accurate emotions. And machines can identify emotion more accurately and on behalf of that, they can give a proper reaction and also can help to prevent the same unwonted occurrence. This machine can also become the replacement of a human.

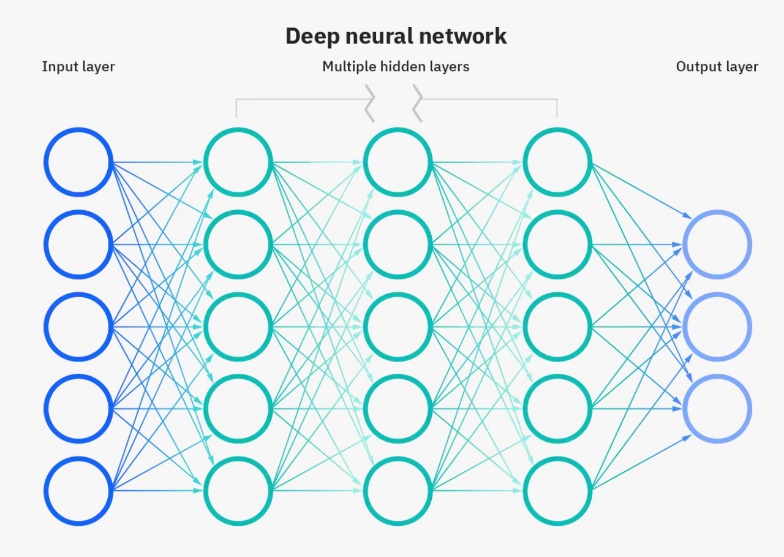
**5- Dataset employed:**

<https://www.kaggle.com/datasets/msambare/fer2013?resource=download>

**6- Algorithm will be used :**

Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, and it transfer data in the way that biological neurons signal to one another

Artificial neural networks (ANNs) are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network.



Neural networks rely on training data to learn and improve their accuracy over time. However, once these learning algorithms are fine-tuned for accuracy, they are powerful tools in computer science and artificial intelligence, allowing us to classify and cluster data at a high velocity. Tasks in speech recognition or image recognition can take minutes versus hours when compared to the manual identification by human experts.

Think of each individual node as its own linear regression model, composed of input data, weights, a bias (or threshold), and an output. The formula would look something like this:

∑ Wi Xi = bias +W1X1+W2X2 ………

And, the OUTPUT = 1 if (∑ Wi Xi + b >= 0 )

Or OUTPUT = 0 if (∑ Wi Xi + b < 0 )

1- FORWARD PROPAGATION :

Once an input layer is determined, weights are assigned. These weights help determine the importance of any given variable, with larger ones contributing more significantly to the output compared to other inputs. All inputs are then multiplied by their respective weights and then summed. Afterward, the output is passed through an activation function, which determines the output. If that output exceeds a given threshold, it activates the node, passing data to the next layer in the network. This results in the output of one node becoming in the input of the next node and so on until we reach the output layer .

2- COST FUNCTION :

As we start to think about more practical use cases for neural networks, like image recognition or classification, we’ll leverage supervised learning, or labeled datasets, to train the algorithm. As we train the model, we’ll want to evaluate its accuracy using a cost (or loss) function. This is also commonly referred to as the mean squared error (MSE). In the equation below,

(Y' – Y)^2 1/2m ∑ = COST FUNCTION

* *i* represents the index of the sample
* y-hat is the predicted outcome
* y is the actual value
* *m* is the number of samples.

Ultimately, the goal is to minimize our cost function to ensure correctness of fit for any given observation. As the model adjusts its weights and bias, it uses the cost function and reinforcement learning to reach the point of convergence, or the local minimum. The process in which the algorithm adjusts its weights is through gradient descent, allowing the model to determine the direction to take to reduce errors (or minimize the cost function). With each training example, the parameters of the model adjust to gradually converge at the minimum.

A picture containing diagram

Description automatically generated

3- BACK PROPAGATION :

that is, move in the opposite direction from output to input. Back propagation allows us to calculate and attribute the error associated with each neuron, allowing us to adjust and fit the parameters of the model(s) appropriately, it's done by using the delta of the last layer(by computing the diferrence between the real and expected value) then get the delta of the previous layer and so on , then we get the partial derivative, then we get the triangle delta

4- GRADIENT DECSENT :

It's done by calculating the difference between numerical and analytical gradient , the difference should be less than 10^-7

"this step won't be done in every iteration but every N times"

\* then we update the values of the weights and itrate again doing the steps in the same sequence then update weights…..and so on

\* after a number of itrations we get the best fit values for weights so we minimize the cost function

\*we test the algorithm with test dataset and see if it is work or there is a problem happened

--TYPES OF NEURAL NETWORK :

Neural networks can be classified into different types, which are used for different purposes.

But we will focus on one type ,CONVOLUTIONAL NEURAL NETWORK (CNN).

--CONVOLUTIONAL NEURAL NETWORKS (CNNs) :

They are usually utilized for image recognition, pattern recognition, and/or computer vision. These networks harness principles from linear algebra, particularly matrix multiplication, to identify patterns within an image.

**Block Digram :**

Diagram

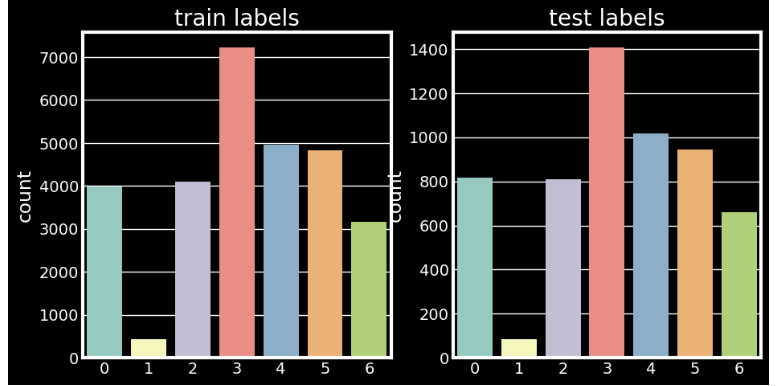
Description automatically generated

Diagram

Description automatically generated

# PHASE 2

1. **Is cross-validation used? Yes**



**2.Hyperparameters used:**

**1. Learning Rate = 0.1**

**2. Optimizer : Adam "tensorflow"**

**3. Regularization : Overfit the Dataset :**

**a. Traning Not Big**

**Number of Training : 28709 = 80%**

**Number of Testing : 7178 = 20%**

**3.Hyperparameters used:**

**1. Learning Rate = 0.1**

**2. Optimizer : Adam "tensorflow"**

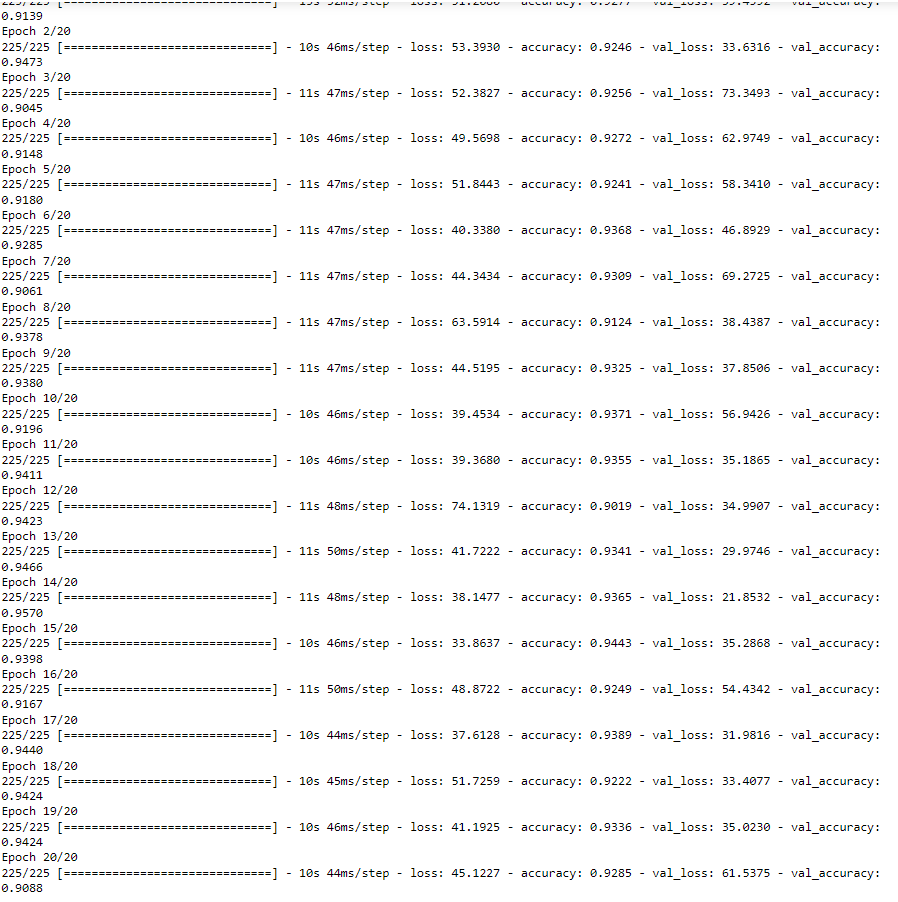
**3. Regularization : Overfit the Dataset :**

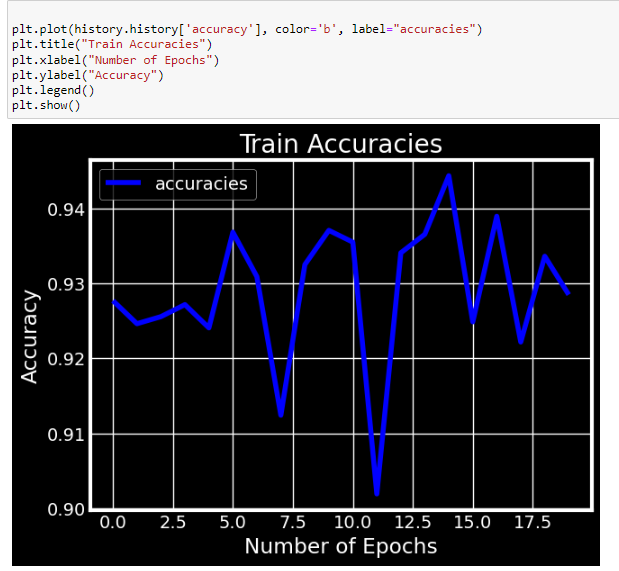
**4. Batch Size = 128**

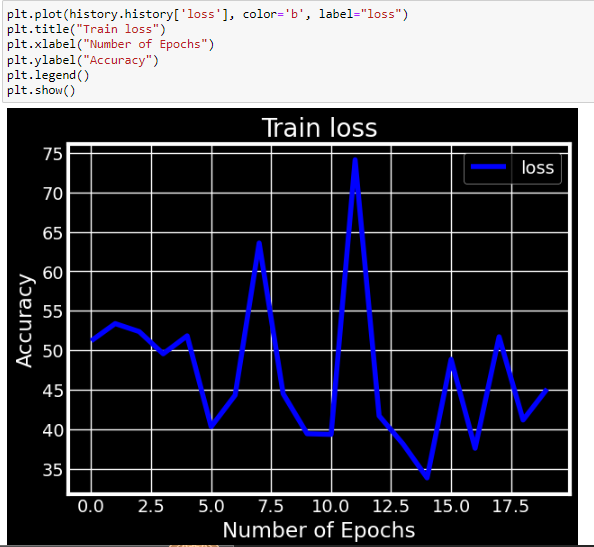
**5. Number of epochs = 20**

****

**" We use 20 epochs, stopped in 10 epochs with accuracy = 91.96% Then we increase epochs to 20 epochs with accuracy = 90.88%"**

****

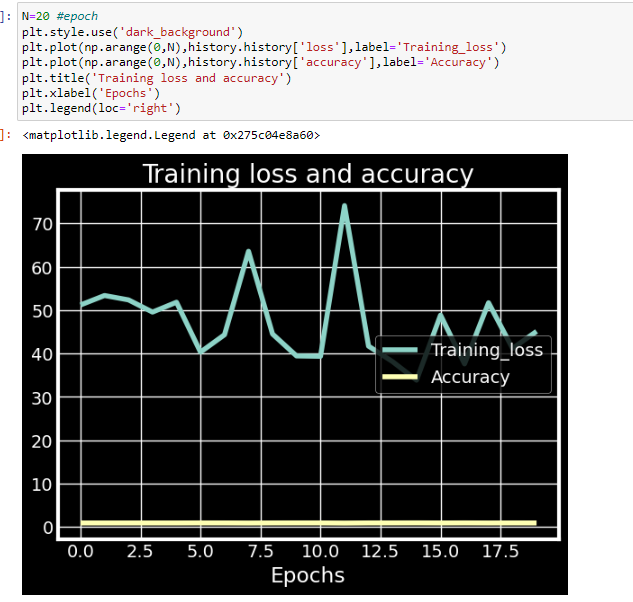
****

****

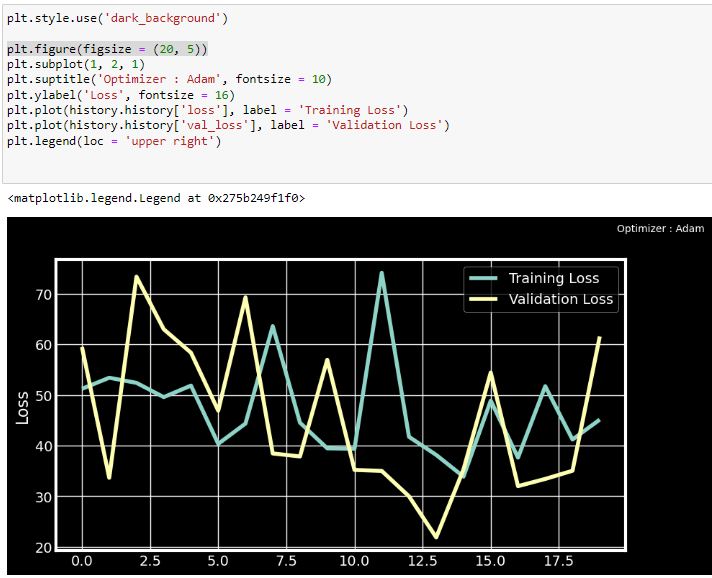
**\*\*Results Details:**

**1. Loss curve :**

**a. Train loss & Accuracy:**

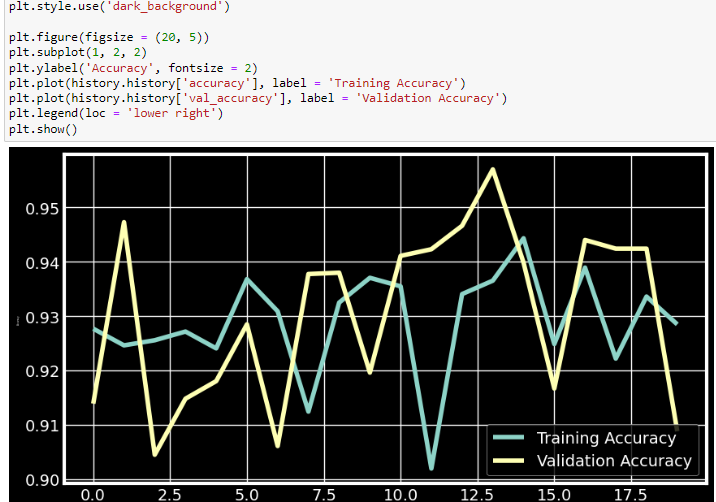
****

**b. Training & Validation loss:**

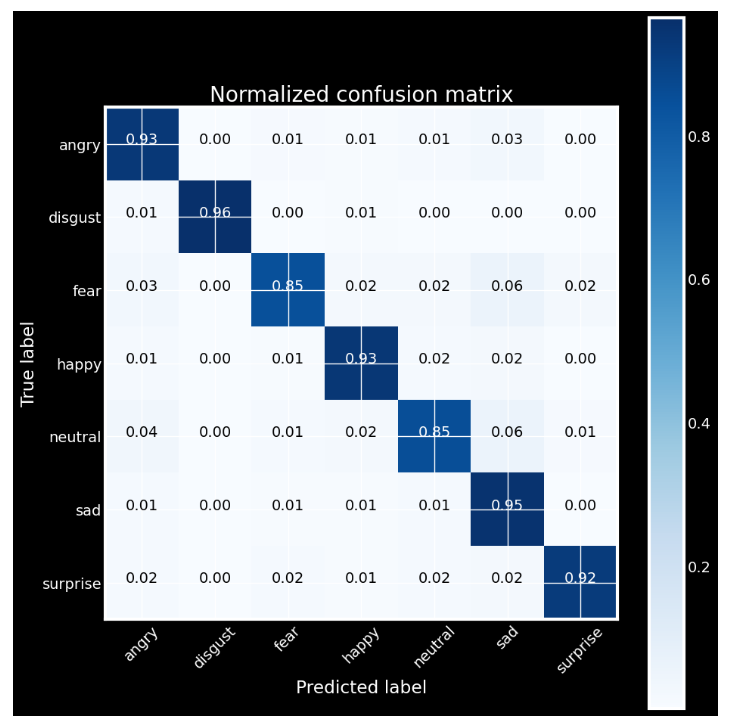
****

**2.** **Accuracy :**

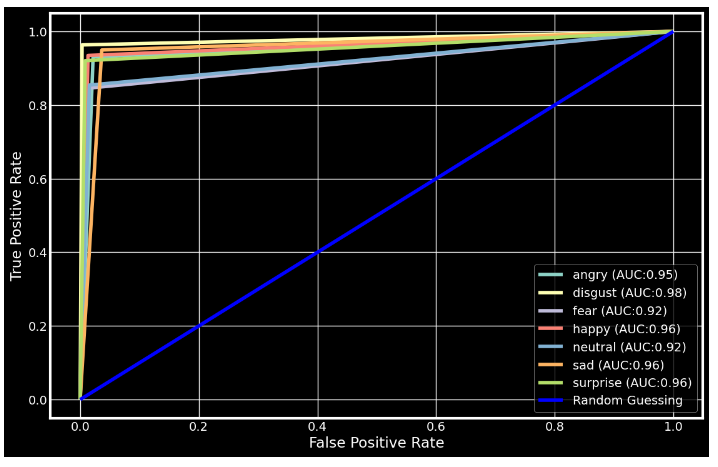
**\* Training & Validation Accuracy:**



**3. Confusion Matrix:**



**4. ROC Curve:**



\*\* Classification Report :

