FACIAL EXPRESSION DETECTION AND RECOGNITION

Aradhana singh parihar
M.Tech. Scholar SAGE University, Indore

Dr. Shweta Agrawal
Professor, CSE SAGE University, Indore

Abstract

In recent years, there has been a huge amount of research on facial expression detection and recognition. Facial expression recognition software is useful in a variety of circumstances. In recent years, there has been a lot of research on facial expression detection and recognition. Facial expression recognition software is useful in a variety of circumstances such as, security, camera surveillance, criminal investigations,smart card applications, database management systems, and in modern devices for identity verification etc. This paper shows how deep learning algorithms may be utilized to correctly identify and recognize face expressions. The purpose of facial recognition is to recognize and validate facial traits. However, Haar cascade detection is used to captured facial features in real time. In three different phases the sequential process work can be define, In the first step, a camera detects a human face, and the acquired input is processed based on features with the help of the Keras convolutional neural network model database.Human faces are validated in the third stage to categorize human emotions as happy, neutral, furious, sad, and surprised.This suggested study is broken down into two aspirations: face detection and expression identification.This work will comes under computer vision field. We will use opencv, keras and python programming in this project. An experiment will be conducted for numerous students to assess their emotions and find physiological changes for each face in order to verify real-time applicability. The testing result illustrates the system's perfection in detecting and recognizing facial expressions. Finally, we will be able to obtain accurate facial expression detection and identification results.

Keywords:
Facial expression detection, Haar cascade, Keras, CNN

1)Introduction

Human-computer interaction is a popular trend these days, and it has the intrinsic ability to distinguish between various faces. In the past, computer vision problems were difficult to solve, but with the advancement of modern technologies, problems such as varying light, age, hair, and other accessories are no longer a problem. Face recognition software, on the other hand, is used to make it easier to identify and authenticate people based on their facial features. As a result, evaluating facial features as well as changes in their expression and movements is essential. These characteristics and expressions aid in the classification of human facial emotions. The technology advancement in the recent years has resulted in the use of Artificial intelligence system as these systems are capable enough to understand and realize the emotion recognition through facial features[1][2]. Here this is an attempt to prove the existence of new technological developments for human-computerinteraction (that comes under computer vision field)using deep learning or Convolution neural network models .To recognize and classify human faces, various approaches are necessary, but deep learning beats conventional methods due to its massive dataset capabilities and quick computation skills. Preprocessing, detection, orientation, feature extraction, and emotion classification are all steps in the facial expression identification and classification process. The deep learning[3][4] keras model, which outperforms the convenient computations, makes these steps simple.The representation of working of human brain with neurons is done by Deep learning technique . This learning usually takes the form of a neural network model, in which neurons serve as inputs and are coupled to move as outputs.

Similar to machine learning, deep learning algorithms are interconnected, but there are multiple levels of these algorithms, each of which interprets the data they introduce differently. Because their action is a source of inspiration, or we may say; an attempt to emulate the role of the human neural networks in the brain, this network of algorithms is known as the network of artificial neurons. Deep neural networks can examine data functions in the so-called functional hierarchy thanks to several hidden levels, which allow simple functions, such as two pixels, to integrate from one level to the next, generating, for example, more complicated functions. Low-level networks are less capable of processing mathematical operations than multilevel networks, which can process large amounts of data. Deep learning models a high degree of characteristics for computing intensive mathematics and is in high demand in the form of a little chip known as a Graphical Processing Unit (GPU). This study is divided into three aspects: facial expression detection, recognition, and emotion classification.

Computer Vision is the field of study which enables computers to see and identify digital images and videos as a human would. The challenges computer vision faces largely follow from the very limited understanding of biological vision.Computer vision is the process of capturing, processing, analysing, and comprehending digital images in order to extract high-dimensional data from the real environment and generate symbolic or numerical data that may be used to make choices.Object recognition, video tracking, motion estimation, and image restoration are all common techniques used in this procedure.

Convolution Neural Networks

In recent years, the development of Convolution neural networks has had a substantial impact on the computer vision domain, as well as a crucial step and capacity to recognize objects or real-world objects. Because of the increasing value of the computer and the amount of data available to create a neural network, the technique has gained in popularity.

Kernels are used to detect boundary functions or outline for an image in convolutional neural network (CNN) models. Weights are organized in an array of values in this model to form and get desired qualities. Every CNN model allots space to determine the picture control to be identified. The values in the image show the degree to which the convolution operation is dependent on the location, and so the product is computed and determined with the location.

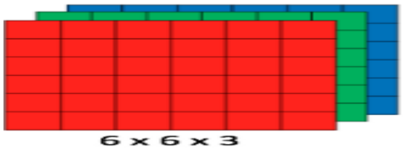


Figure 1. Array of RGB Matrix

CNN's goal is to train and test the created system. The image is used as an input, with many layers acting as filters that are connected with weights as needed. Softmax is a function that is used to acquire and classify images between 0 and 1. As a result, figure 2 depicts the CNN model processing based on the input image.

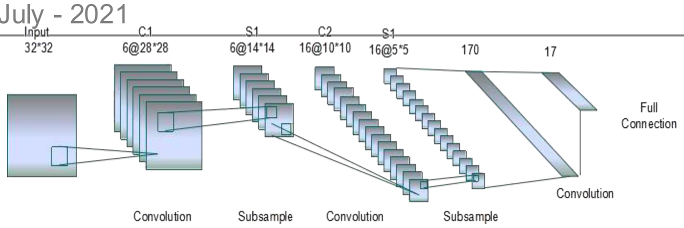


Figure 2.convolutional neural network

Figure 3 depicts the basic first layer of a CNN model, which extracts features from the input picture. The features are then sent to the convolution layer, where the weights functions are multiplied to get the product of each layer. The image matrix, two inputs, and kernel values are used in the mathematical processing. Figure 1 shows the produced image matrix with dimensions and volume as (h x w x d), the filter as (fh x fw x d), and the volume dimension outputs as (h-fh+1) x (w-fw+1) x 1. Section II goes over the network layers, weights, and function values in greater depth.

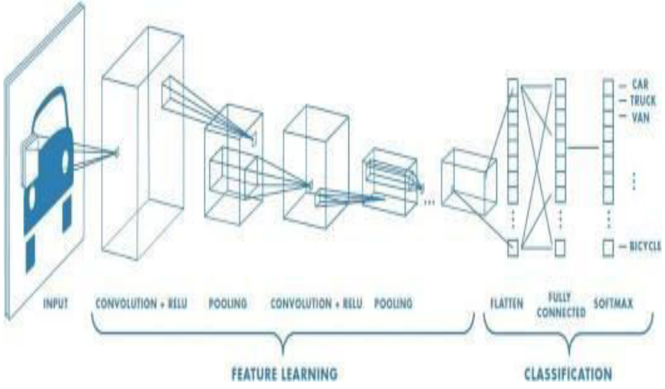
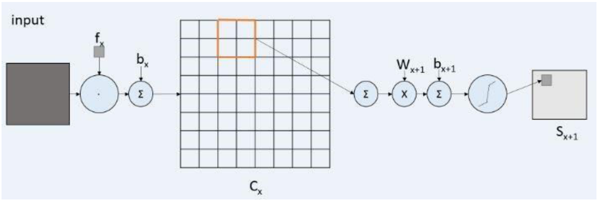


Figure 3. Main Process of CNN

The most important technologies in CNN are local field of perception, weight distribution, and subsampling of time or space to isolate functions and reduce the number of training parameters.The CNN algorithm has the advantage of avoiding a clear separation of functions as well as indirect learning of training data. The use of a subsample structure in time or space can give size, speed and stability; superior network entry and topology can be achieved; and speech understanding and image processing can provide unique benefits.Based on a 96*96 input grey value in the pre-processing stage, which yields 32.

Figure 4.stages ofConvolution neural network (CNN) model



2)Related work:

Facial expression detection by viola jones face detector algorithm
"The Viola-Jones face detector is used to detect faces.
To decrease noise, images are pre-processed. In addition, photos in the dataset must have the same amount of exposure, lighting, and brightness. Histogram equalisation techniques are used to improve the image quality of such photos. The Viola- Jones approach, which employs the Ada-boost algorithm, is used in the face detection module.
The Viola- Jones approach, which employs the Ada-boost algorithm, is used in the face detection module. It iteratively combines a large number of weak classifiers to build a strong classifier, where a weak classifier lowers the weighted error rate per iteration". [5].

Feature extraction by LBP technique
"Local Binary Pattern (LBP) is a simple and reliable feature extraction technique that is unaffected by image illumination differences. A 3x3 matrix is created, and each pixel in the matrix is given a binary value based on the value of the centre pixel. Except for the centre pixel value, which is transformed to decimal value [3], these 8-bit binary values create an 8-bit binary number.

[5] gives the LBP code for a pixel at (xc, yc).

$\mathrm{LBP}_{P,R}\,(x_c,y_c) = \sum_{(P=0,7)} S(g_p - g_c)\, 2^P$, $S(x) = \{1, x >= 0 \text{ and } 0, x < 0\}$
where,

$g_{c=}$ gray value of center pixel
$g_{p=}$ gray value of neighboring pixel of g_c
P = 8 (maximum 8 neighbors for center pixel in a 3x3 matrix)

Hence a pixel can have 256 different values as $2^P = 2^8 = 256$".

Using HoG features based on a dynamic grid
"The histogram of oriented gradient (HoG) is an edge-based object recognition algorithm. Edge orientations can be used to define each detection window. The face's region of interest is clipped into the image. The detection window, which is broken into even smaller sections called cells, is cropped like this. A magnitude of edge gradient for each orientation bin is computed in each cell, for each pixel, generating the local histogram of oriented gradients"[6][7].

Geometrical facial features extraction
A set of 19 traits that are more linked to human expressions was chosen empirically by observing landmark positions on the face. These 19 features are a subset of the 66 2D features in an existing marker-less system for landmark identification and localization. [8] .

Eccentricity features
The concept of ellipses is used to define eccentricity traits. The eccentricity of an ellipse is the distance the ellipse deviates from being a circle. For ellipses, eccentricity ranges from 0 to 1, with 0 being the case if the ellipse is a circle. Excessiveness will be larger than 0 when smiling, but it will be closer to 0 when expressing astonishment. [9].

Machine Learning Algorithms

The next critical step is to employ a decent classification method once the dataset has been produced with the needed attributes. In practically all examples of multi-class classification of human expressions, Support Vector Machines (SVM) are used[1][4][5]. They are used in conjunction with one or more feature extraction techniques.[5].

Support Vector Machines (SVM)

“One of the most powerful classification methods is the SVM. The goal is to find an ideal hyper plane that appropriately divides the two classes. There is also the concept of margin, which is supposed to be the maximum from both classes in order to avoid any overlap between them [8]. To produce better classification results, data that is not linearly separable is mapped into a higher dimension. For non-linear data, kernel functions such as radial basis function (rbf) and polynomial are used[1].

In the case of emotion detection, a multi-class system is commonly used. Instead of using a binary system to detect emotions like anger, contempt, disgust, fear, happiness, sadness, and surprise, SVM is used [1]. K-fold cross-validation is used to remove any variances in the database and evaluate other machine learning algorithms [6]. The dataset is partitioned k times into k slices in k-fold cross validation, and the prediction results are averaged over all rounds”.

Loconsole et al. [5] employed Principal Component Analysis (PCA) to minimise the size of their feature set before feeding it to SVM. The image feature space is translated to eigen space using an eigen matrix in the PCA technique [2]. SVM includes methods for tweaking parameters like C and [4] in addition to kernel definition. The penalty function for misclassification is C, and gamma aids in the optimization of the decision boundary. Both of these parameters influence the classifiers' accuracy and can be tweaked to achieve the best results in binary and multi-class classification.

Hidden Markov Models (HMM)

“Hidden Markov Models (HMM) are based on statistics and can be used to uncover hidden data structures. They're also popular for detecting emotions in speech . There are hidden states matching to sequential events, and the input is a sequence of observed features. The HMM is written as follows [14]:

$\lambda = (A,B,\pi)$

where,

A = (aij) transition probability matrix between the hidden states B = (bij) observation symbols probability from a state

Π = initial probability of states.”

The process of generating Code – HMM is described in Paper [13]. It attempts to improve on the existing HMM by combining features from different classifiers. HMMs are sometimes used in conjunction with algorithms like k-Nearest Neighbor [13]. The benefit of employing both methods is that HMM can perform sophisticated computations, whilst k-NN only needs to classify amongst the given samples. HMM decisions are based on the highest output probability, which may be contaminated by noise, but K-NN can provide a second layer of classification, boosting accuracy.

To achieve the best results for speech emotion recognition, HMM is combined with SVM as a Serial Multiple Classifier System [12]. HMMs may be used to train the samples and SVM can be used to classify them because SVM returns a classification rather than a score. Boosting can be used in conjunction with many classifiers to create a strong classification system by combining two or more weak classifiers to create a strong classifier [15].

The work [15] also discusses embedding HMM, which entails creating a two-dimensional HMM with super and embedded states. Super states and embedded states are used to model the data in two directions. Features from the top to the bottom of a facial image can be super states, while features from the right to the left can be embedded states.

"Random Forest Classifiers have also been shown to be superior to SVM in some circumstances [9][10]. Random forests are based on decision trees, but instead of using just one classifier, they employ several forests or classifiers to determine the target variable's class. Some of the techniques used for categorization and prediction of emotion include K-Nearest Neighbor, Linear Discriminant Analysis, and Neural Networks (ANN)".

3)Approach

Face Detection, Face Recognition, and Face Classification are three consecutive processes in the proposed work. The first stage involves using a video camera to capture a human face and determining its exact location using bounding box coordinates for the face recognized in real-time. Face detection is performed using Haar cascade detection with the open CV library in this step.To detect human faces, the Viola Jones algorithm and haar cascade characteristics are integrated. Shapes, objects, and landscapes are among the images recognized. Human faces are recognized in this phase, and face features are retrieved and stored in a database for face recognition. Figure 4 shows how the CNN model uses VGG 16 to match a face from the database and recognizedit with the name associated with it.Faces are recognized in the database and compared using embedding vectors to identify or detect the face. In order to process face detection, identification, and classification, the distribution platform employs Anaconda and Python 3.5 software.The image can be found in the database and other libraries. Face detection is done first, followed by database feature recognition and matching utilizing the CNN model training and testing database. Finally, the recognized human face is categorized as angry, ,sad, joyful, neutral, or surprise based on the expression in real time.For massive database recognition and classification, the VGG 16 network design uses the CNN paradigm. The honeycomb 3 x 3 layer network model includes 4096 nodes with Softmax classification in the two connected layers.For detecting human faces, the local binary model histogram is employed as an open CV library. Setting a threshold identifies the image pixels, and the end result is expressed as a binary integer. LBPH uses four factors to do this: radio, neighbours, Grid X, and Y.For the binary model, the radio button specifies the local radius in a circular fashion. The number of sample points as neighbours is represented by the circular local binary model. The number of horizontal and vertical direction cells is determined by the grid's X and Y dimensions. To accept test data from the machine, the circular binary model is constructed to the indicated areas with appropriate labels.Create a folder and name it, then write two python files to do the procedural outcome of face detection. Set and run complete files from the Integrated Development Environment of Python, including create data.py, face recognize.py, and the haarcascadefrontalface default.xml library.

Proposed Functional Design

First of all makeba folder and give name , then code two python files for procedural outcome of face detection. Set and run complete files from the Integrated Development Environment of Python, including create data.py, face recognize.py, and the haarcascadefrontalface default.xml library.To construct and train the dataset for identification and emotion classification, the collected image is modelled using network architecture. The information is presented as a flow chart in which the processes are given in order.

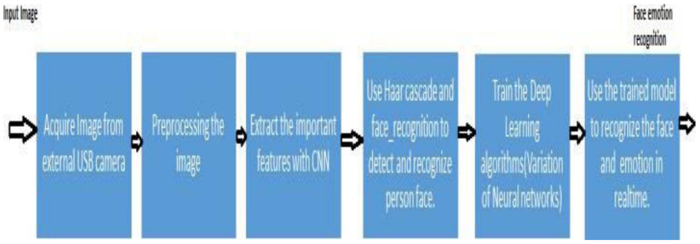

Input Image
Acquire Image from external USB camera
Preprocessing the image
Extract the important features with CNN
Use Haar cascade and face_recognition to detect and recognize person face.
Train the Deep Learning algorithms(Variation of Neural networks)
Use the trained model to recognize the face and emotion in realtime.
Face emotion recognition

Figure 5.Block diagram

<u>Process in the form of Flowchart and Design Specifications</u>

Figure 5 shows the real-time face detection, recognition, and classification.The acquired image is bounded in a box, transformed to a binary pattern, and saved in a database as a feature vector.The images are given training to replicate the input image as well as characterise facial expressions as neutral, happy, angry, sad ,and surprise.seven steps in the training process: importing the dataset, preprocessing the data, supplementing the data as a feature vector, developing and compiling the design model, training and storing the feature vector, and testing the test model.

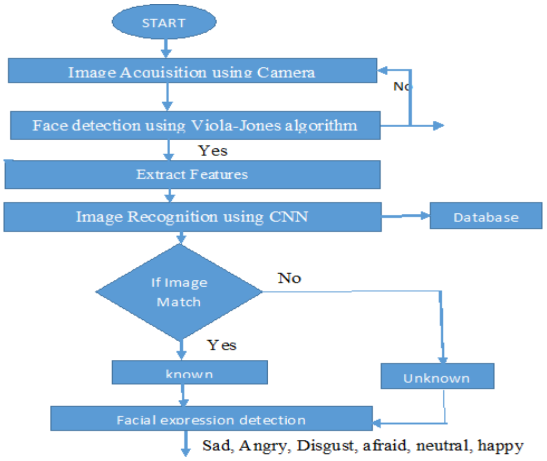

START
Image Acquisition using Camera
No
Face detection using Viola-Jones algorithm
Yes
Extract Features
Image Recognition using CNN
Database
If Image Match
No
Yes
known
Unknown
Facial expression detection
Sad, Angry, Disgust, afraid, neutral, happy

Figure 6 flow chart

To convert the collected image into a binary vector, the work uses a local binary pattern histogram (LBPH). Using the viola jones algorithm, this processing aids in the recognition of faces. For the face feature vector, the picture pixels are stored with a given threshold.Figure 7 shows how these vectors are combined with weights to construct a network architecture model for face expression categorization using the VGG 16 CNN model. VGG 16 is usually trained using an image net database, which has 16 layers of network divided into 1000 categories.To measure maximum pooling, the 3x3 convolution layers are placed on top of one another. As a softmax classifier, these network connections are linked to 4096 nodes.

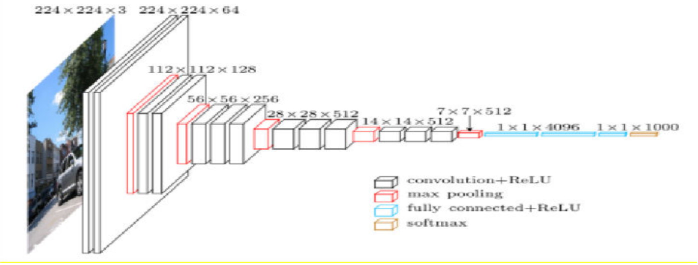


Figure 7 Net model VGG16

4)Implementation

We must first obtain a dataset from Google. This dataset will be used to train the machine learning algorithm. The dataset is divided into two folders, one for training and the other for validation. Each folder contains photos of emotions such as angry, sad, happy, natural, and surprise, resulting in a total of five classes. To import dense,dropout,activation,flatten, and batchnormalization, we'll utilisekeras.layers.

Creating a database that includes an open CV library as well as other resources

a) Create a folder and name it, then create two python files, facial expression recognition.py and classification tiny vgg. py. To check for problems, the code is copied into the resulting source file and executed. To support the facial features, copy the xml file into the project directory. Haarcascadefrontalface default.xml is the file that has to be duplicated.

c) Keras: Keras is a Python high-level library that runs on top of the Tensorflow framework. It was created with the goal of better comprehending deep learning approaches, such as layering neural networks while keeping shape notions and mathematical intricacies.
The types of framework creation are as follows:

- Sequential API
- Functional API

To develop a deep learning model in Keras, we took the following 8 steps into consideration:

- Loading the data
- Preprocess the loaded data
- Definition of model
- Compiling the model
- Fit the specified model
- Evaluate it
- Make the required prediction
- Save the model

c) VGG model: This model is used to support and train Convolution neural networks as solvers and models, as well as data preparation and training. During the creation of Python scripts, this model is used to anticipate the unneeded data. The CNN Caffe model in Python programming is shown here.

d) Get-pip, numpy: This is an open source library tool for the Python programming language that is used as an installation package.

e)openCV: This is an open source computer vision library. This library will be used to do image transformations. The implementation of algorithms in open cv is diverse.

f) Python : python is the most powerful programming language, and it is particularly good for handling statistical problems involving machine learning techniques and deep learning.

Training procedure for recognition and classification

The training and testing data is separated into two models, which are loaded using keras.models. The image utilised is 48x48 pixels, which is rescaled and retrieved as a feature vector before being used to train the classifier. The following is the main code for extracting features to train the classifier.

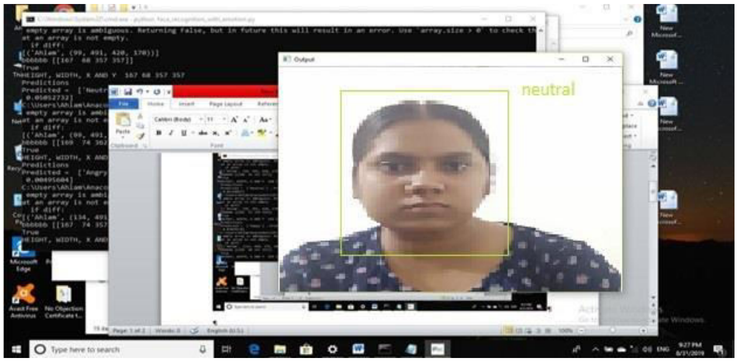
Creating all the files to Python 3.8 and Anaconda 4.8.2 software platform.

Code written using vscode editor in python language. There is two file were created for the whole project the first one is for training the data and the last one is for testing the output or get output.
First we did training After running this code data will trained and one file will be generated that file is emottionlittle.vgg.

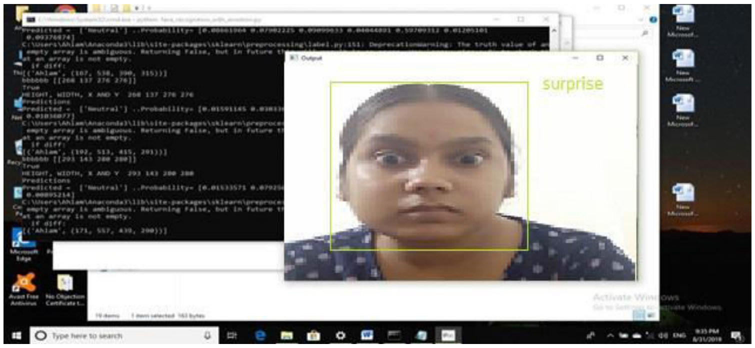
In the second step we did testing, code is written in file Facial expression recognition.py

After running this file we will able to see the desired output. The software platform used in processing the results is Python 3.8 and Anaconda 4.8.2.

Some output are here



```
empty array is ambiguous. Returning False, but in future this will result in an error. Use `array.size > 0` to check th
at an array is not empty.
  if diff:
[('Ahlam', (99, 491, 420, 170))]
bbbbbb [[167  68 357 357]]
True
HEIGHT, WIDTH, X AND Y  167 68 357 357
Predictions
Predicted -  ['Neutr
 0.05052732]
C:\Users\Ahlam\Anaco
 empty array is ambi
at an array is not e
  if diff:
[('Ahlam', (99, 491,
bbbbbb [[169  74 362
True
HEIGHT, WIDTH, X AND
Predictions
Predicted -  ['Angry
 0.00495604]
C:\Users\Ahlam\Anaco
 empty array is ambi
at an array is not e
  if diff:
[('Ahlam', (134, 491
bbbbbb [[167  74 357
True
HEIGHT, WIDTH, X AND
```



```
Predicted -  ['Neutral'] ..Probability- [0.00661064 0.07982225 0.09099633 0.04644831 0.59709312 0.01205101
 0.09376874]
C:\Users\Ahlam\Anaconda3\lib\site-packages\sklearn\preprocessing\label.py:151: DeprecationWarning: The truth value of an
 empty array is ambiguous. Returning False, but in future th
at an array is not empty.
  if diff:
[('Ahlam', (167, 538, 390, 315))]
bbbbbb [[268 137 276 276]]
True
HEIGHT, WIDTH, X AND Y  268 137 276 276
Predictions
Predicted -  ['Neutral'] ..Probability- [0.01591145 0.03033
 0.01036077]
C:\Users\Ahlam\Anaconda3\lib\site-packages\sklearn\preproces
 empty array is ambiguous. Returning False, but in future th
at an array is not empty.
  if diff:
[('Ahlam', (192, 513, 415, 291))]
bbbbbb [[293 143 280 280]]
True
HEIGHT, WIDTH, X AND Y  293 143 280 280
Predictions
Predicted -  ['Neutral'] ..Probability- [0.01533571 0.07925
 0.00895214]
C:\Users\Ahlam\Anaconda3\lib\site-packages\sklearn\preproces
 empty array is ambiguous. Returning False, but in future th
at an array is not empty.
  if diff:
[('Ahlam', (171, 557, 439, 290))]
```

5) conclusion

The goal of this project is to create a real-time system that can detect, recognise, and classify human facial expressions. As indicated in the above data, the categorised expressions are represented in five states. Anaconda and Python 3.8 are the softwares that were utilised to test the functionality.The Haar cascade method was used to detect faces. Similarly, for face recognition and classification, VGG 16 is combined with a convolutional neural network model. Python programming is aided by the keras and other libraries. The CNN model, which has an accuracy of 88 percent, is used to validate the performance measurements.The results, on the other hand, show that the network architecture designed outperforms previous techniques. This programme is widely utilised in a variety of fields, including medicine, manufacturing, education, and electronics. Face categorization and identification are achieved using VGG 16 and the dataset.The five facial expressions shown above depict various aspects of a person's condition. Autism is one of the applications that is directly associated and can be used to read a person's or child's expressions. The proposed study could also be used to assess a student's emotions when using E-learning strategies.

6)Acknowledgments

The research was aided by Sage University, Indore, which provided lab space as well as an academic setting in which to accomplish the necessary results. I am also grateful to all of the department's distinguished teachers who moderated the text at each stage. I am grateful to Dr. SwetaAgrawal for helping me define fresh research findings linked to the planned study.

7)Reference

[1] W. Swinkels, L. Claesen, F. Xiao and H. Shen, "SVM point-based real-time emotion detection," 2017 IEEE Conference on Dependable and Secure Computing, Taipei, 2017.

[2] Medical text and image processing: applications, issues and challenges S Agrawal, SK Jain - Machine Learning with HealthCare Perspective, 2020

[3] Computational Orchestrator: A Super Class for Matrix, Robotics and Control System Orchestration
S Agrawal, R Kamal - International Journal of Computer Applications, 2015

[4] Neerja and E. Walia, "Face Recognition Using Improved Fast PCA Algorithm," 2008 Congress on Image and Signal Processing, Sanya, Hainan, 2008

[5] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based HoG features," Face and Gesture 2011, Santa Barbara, CA, 2011

[6] Saravagi, Deepika, Shweta Agrawal, and Manisha Saravagi. "Indian stock market analysis and prediction using LSTM model during COVID-19." *International Journal of Engineering Systems Modelling and Simulation* 12.2-3 (2021): 139-147.

[4] K. M. Rajesh and M. Naveenkumar, "A robust method for face recognition and face emotion detection system using support vector machines," 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), Mysuru, 2016

[5] C. Loconsole, C. R. Miranda, G. Augusto, A. Frisoli and V. Orvalho, "Real-time emotion recognition novel method for geometrical facial features extraction," 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 2014

[6] J. M. Saragih, S. Lucey and J. F. Cohn, "Real-time avatar animation from a single image," Face and Gesture 2011, Santa Barbara, CA, USA, 2011

[7] G. T. Kaya, "A Hybrid Model for Classification of Remote Sensing Images With Linear SVM and Support Vector Selection and Adaptation," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 6, no. 4, pp. 1988-1997, Aug. 2013

[8] X. Jiang, "A facial expression recognition model based on HMM," Proceedings of 2011 International

Conference on Electronic & Mechanical Engineering and Information Technology, Harbin, Heilongjiang, China, 2011

[9] J. J. Lee, M. Zia Uddin and T. S. Kim, "spatiotemporal human facial expression recognition using fisher independent component analysis and Hidden Markov Model," 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, 2008

[10] Xiaoxu Zhou, Xiangsheng Huang, Bin Xu and Yangsheng Wang, "Real-time facial expression recognition based on boosted embedded hidden Markov model," Image and Graphics (ICIG'04), Third International Conference on, Hong Kong, China, 2004

[11] T. Kundu and C. Saravanan, "Advancements and recent trends in emotion recognition using facial image analysis and machine learning models," 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, 2017, pp. 1-6.

[12] Y. Nijs, ― Children‵s lying behavior towards personified robots: an experimental study,‖ 2016, 10.13140/RG.2.1.4612.5847. [Online] Available: https://www.researchgate.net/publication/297379405_Children's_lying_behavior_towards_personified_robots_an_experimental_study