IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

EMOTION DETECTION THROUGH FACIAL FEATURE RECOGNITION

[1]ATLURI SREEJA, [2]KHAJJAYAM TEJA SAI SRINIVAS, [3]SHAIK HEENA

[1]UNDERGRADUATE, SRKR ENGINEERING COLLEGE, BHIMAVARAM, ANDHRA PRADESH, INDIA
[2]UNDER GRADUATE, SRKR ENGINEERING COLLEGE, BHIMAVARAM, ANDHRA PRADESH, INDIA
[3]UNDER GRADUATE, SRKR ENGINEERING COLLEGE, BHIMAVARAM, ANDHRA PRADESH, INDIA
atlurisreeja@gmail.com, kajjayamsrinivas@gmail.com, heena.shaik010@gmail.com

ABSTRACT: Automated emotion recognition (AEE) plays an important role in areas such as robotics, marketing , education, and the entertainment industry in-order to improve the human computer interaction, with increased industrialisation and rapid advancements in smart technologies, the need of hardware/software capable of assessing the needs and provide appropriate solution is also increasing.Some of the real world use cases are : improved learning process, designing service robots, suggesting most appropriate entertainment for the target audience.This paper tries to classify and present the most suitable methods for classifying emotion and their intensity using Convolutional Neural Networks (CNN).

KEY WORDS: Emotion recognition, classification, deep learning, image procrssing, neural networks

I. INTRODUCTION

It is important to acknowledge that a wide-array of emotions exist according to research performed by Feidakis, Daradoumis and Cabella. There are 66 emotions which are classified into two groups : Basic Emotions and Secondary Emotions. It is extremely difficult to find a pattern to differentiate, as most of them have overlapping symptoms or features. Therefore understanding and implementing basic emotion detection is the primary obstacle that is addressed in this paper based on the Russell Circumplex model of emotions [figure 1] .

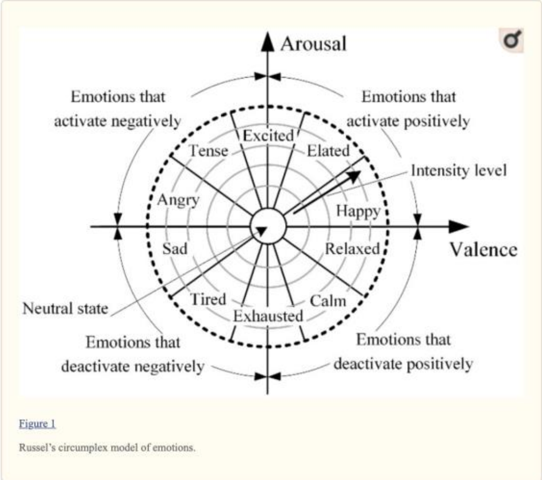

Arousal
Emotions that activate negatively
Emotions that activate positively
Excited
Tense
Elated
Intensity level
Angry
Happy
Valence
Sad
Relaxed
Neutral state
Tired
Calm
Exhausted
Emotions that deactivate negatively
Emotions that deactivate positively
Figure 1
Russel's circumplex model of emotions.

Fig.1 Russell Circumplex model of emotions.

Based on the model it is derived that the basic emotions have following key features which can be used for classification:

Emotion: Response to the particular stimulus.

Affect: Effect caused by the emotion and dynamic interaction

Feeling: Experienced in relation to a particular object of which the person is aware.

Mood: tends to be less intensive, long lasting, and can affect the state of a person in a positive or negative direction.

The above model makes classification and evaluation much more clear, yet there are so many issues related to the evaluation and strength assessment of the emotion.Emotion recognition and strength assessment are related to psychology sciences and the measurement and evaluation of human body parameters are related with medical sciences.

Based on the literature emotion evaluation methods are categorised into two types:

Self reporting based on various questionnaires and machine assessment techniques based on measurements of various parameters of human body.

Five main components of emotion are Behavioural tendencies, physiological reactions, motor expressions cognitive appraisals and subjective feelings based on Scherer et.al.First four components can be assessed naturally but subjective feelings as the name suggests it is subjective to person's attitude and psychological state and can only be assessed based on the self evaluation and these patterns are different for everyone.

Therefore this model tries to stay out of the medical or psychological classification and evaluation, but tries to stick with the facial features to evaluate mood and in turn getting to classify Emotion accurately

It is important to understand the role of emotion detection in Human Computer Interaction. Emotions are a key semantic component of human communication as evaluation of emotion is part of imbibing the meaning of the context. Robotics is one example where service robots are installed with Automated Emotion Detection System inorder to understand and cater the needs and to provide and provide suitable solution.First the data is processed to eliminate the inanimate object as this increases the difficulty of the classification. It is hard to understand the feelings of the animals or other living organisms compared to humans, This is mainly due lack of communication, we humans are the only one who broke the barrier of communication and tried to excel.Therefore a large data set is considered for evaluation and processed it into a model to understand and characterize the key features for classification into the basic emotions.

II. RELATED WORK

Face and Emotion recognition has been an actively monitored and developed technology. It is due to the intra class classification and multiple feature involved in diversity of the physical features and expressions the development is being delayed.Most of the existing systems such as Scale Invariant Feature Transform(computer vision algorithm to detect, describe, and match local features in images), Histogram of Oriented Gradients (feature descriptor with a linear SVM machine learning algorithm to perform face detection) and Local Binary Pattern Histogram (labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number). These methods works very well on the datasets but lag behind in the real world test cases.Therefore the newest technologies such as deep learning are being employed.

III. METHODOLOGY

The scope of this article is to use CNN (Convolutional Neural Networks) to classify the given data sets (FER2013) .Different stages involved are

Preprocessing

Registration of face

Feature extraction of face

Classification of expressions.

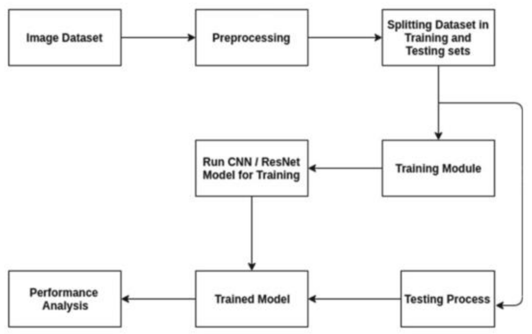


Fig.2 Flowchart of Image Classification

A webcam is utilized to capture the subjects face and sent to the prepossessing stage.As the image is sent to the software live there appears a rectangular object popping right above the face of the subject. This is done using the Viola Jones Algorithm, LBPH Face Recognizer algorithm with a Haarcascade frontal face dataset.Around 60 Images of a person are captured, preprocessed and stored in a folder with

his/her ID and Name (Labels are stored in face.csv file) These images are trained using LBPH algorithm and the trained dataset is stored as Trainer.h5py.The classification of the facial features is done using the CNN by training with the dataset FER2013.
The highest probability of emotion acquired with respect to the features, is shown as the facial emotion. One out of seven emotions are displayed with the detected image of the person.

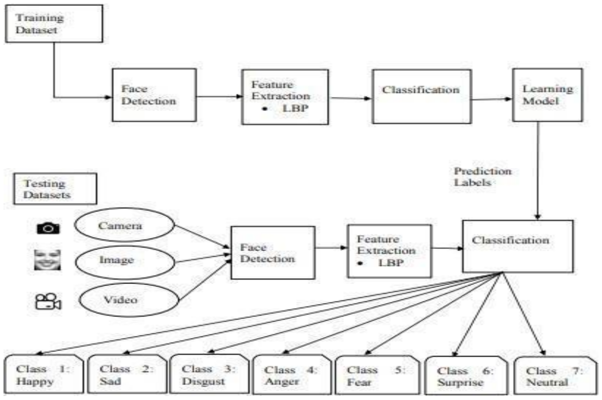


Fig.3 System Design diagram for emotion detection from facial expressions.

A use case diagram is the representation of a user's interaction with the system and depicts the specifications of a use case. A use case diagram can portray different types of system users and different in which they interact with the system. This type of diagram is typically used in conjunction with the textual use case and will often be accompanied by other types of diagrams as well.

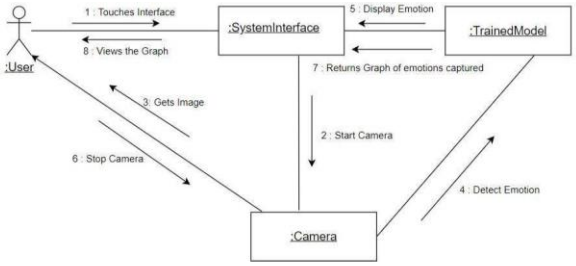


Fig.4 Collaboration diagram for emotion detection from facial expressions

Collaboration diagram provides the interactions among the objects in terms of sequenced texts which are represented in a combination of data obtained from class, sequence, UML describing system's static structure and dynamic behavior.

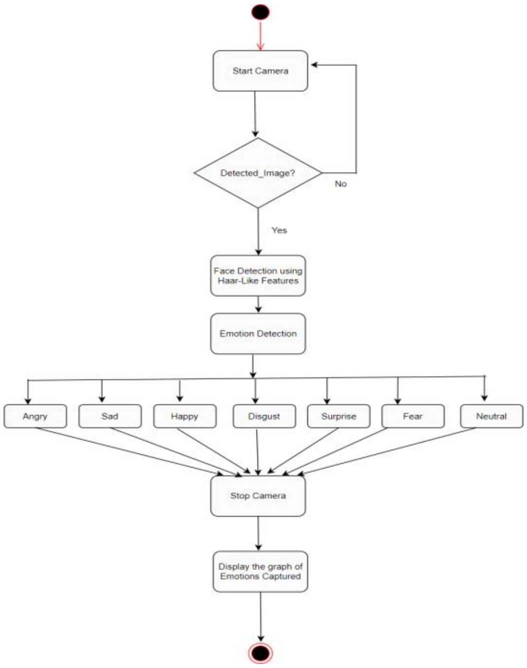


Fig.5 State Chart diagram for emotion detection from facial expressions.

1. Webcam is utilised to capture the subjects face and sent to prepossessing.
2. The obtained photo is stored in the format of .h5py, which is comprised of facial information.
3. 60 photos are taken from each subject for deeper study of facial feature classification
4. .h5py record helps us to speedup handling of the caught pictures.
5. The ontained photos are sent to training with CNN
6. The faces are perceived using the LBPH calculations. It will recognize the faces with the face ID and NAME that are all previously stored.

7. Biometric markers are used to identify the inhumane appearances. Classification is done through CNN

For the purpose of detection of faces from images using features,Haar Cascade Classifier is used. The haarcascade_frontalface_default.xml [17] is used for detecting the frontal face. It was developed by Viola and Jones on the basis of a proposal by Papa Georgiou et. al in 1998[2].

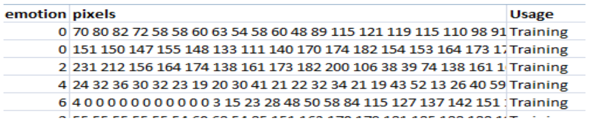
To ensure that the extracted faces are positioned in the same location, we have used an additional classifier, from the same OpenCV library, called 'haarcascade_eye.xml'[18]. This detects the area of the eyes, which is then used to adjust the right and left margins of the face window, to ensure equal distance between eyes and sides of the face. In this way,unwanted information (such as background,hair, ears) is discarded and the extracted faces will have normalized positions.

FER2013[15] is an open-source dataset which was created for an ongoing project by Aaron Courville and Pierre-Luc Carrier , then shared publicly for a Kaggle competition [15]. Moreover,FER2013 db was introduced during the ICML 2013 Challenges in Representation Learning. FER2013 is an unconstrained and large-scale database collected by the Google image search API automatically. All images have been registered and resized to 48*48 pixels after rejecting incorrect labelled frames and adjusting the cropped area. This dataset consists of 35,887 grayscale, 48x48 sized face images with various emotions -7 emotions, all labelled.

Emotion labels in the dataset:

0: -4593 images- Angry
1: -547 images- Disgust
2: -5121 images- Fear
3: -8989 images- Happy
4: -6077 images- Sad
5: -4002 images- Surprise
6: -6198 images- Neutral

During the competition, 3,589 and 28,709 images were shared with the participants as public test sets and training sets respectively and the remaining 3,589 images were considered as private test sets to find the winner of the competition. The dataset was set to be accessed by everybody once the competition was finished.The FER-2013 dataset was developed by collecting the results of Google image search of each emotion and synonyms of the emotions. The images in FER-2013[15] consist of both posed and un-posed headshots.



emotion	pixels	Usage
0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121 119 115 110 98 91	Training
0	151 150 147 155 148 133 111 140 170 174 182 154 153 164 173 17	Training
2	231 212 156 164 174 138 161 173 182 200 106 38 39 74 138 161 1	Training
4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 19 43 52 13 26 40 59	Training
6	4 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84 115 127 137 142 151	Training

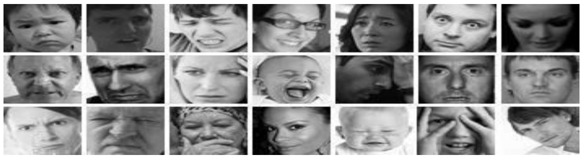


Fig.5 Example images from the FER2013 dataset

Fig.6 A Snapshot of FER2013 dataset

The algorithm used in this project is Convolutional Neural Network (CNN). There are three layers that constitute the CNN. They are convolutional layer, pooling layer, and fully-connected (FC) layer. Along with these 3 layers, there are two more important parameters called the dropout layer and the activation function.

CONVOLUTION LAYER

It is used to extract the different features from the input images. In this layer, the convolutional mathematical operation is performed between a filter of particular size MxM and input image. By sliding the filter over the input image, the dot product is calculated between the filter and the input image parts with respect to filter size (MxM).

The output is called the Feature map which gives the information about the image like corners and edges. Later, this feature map is passed to other layers to learn several other features of the image.

The 2-D convolution between image A and Filter B can be given as:

(□, □) = ∑□□−1 ∑□□−1 (□, □) ∗ □(□ − □, □ − □)
□=0 □=0

where size of A is (Ma x Na), size of B is (Mb x Nb), 0 ≤ □ < □□ + □□− 1 □0 ≤ □ <□□+ □□− 1

The Feature Map size (Convolved Feature) is controlled by 3 parameters

Depth : It is the number of filters used for the convolution operation. Stride : It is the filter size, if the filter size is 4x4 then stride is 4.

Zero-padding : Sometimes, it is better to pad the input matrix with zeros around the border. So, filter can also be applied to bordering elements of the input image matrix. Feature map can be controlled by using this.

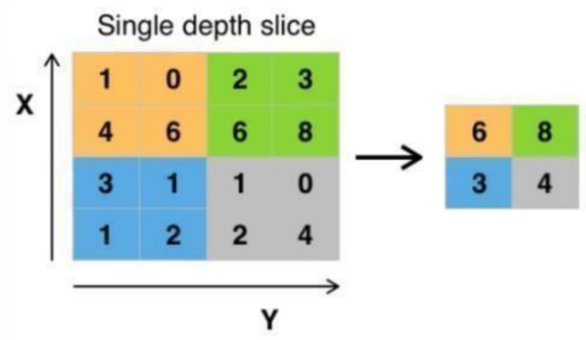
ReLU UNIT

The activation function used is ReLU. An extra operation called ReLU is used after every Convolution operation. A Rectified Linear Unit (ReLU)is a linear function that will give the input directly if it is positive, otherwise, it will give zero.

POOLING LAYER

The Convolutional Layer is followed by a Pooling Layer. It is used to reduce the size of the convolved feature map to reduce the computational costs. This is done by decreasing the connections between layers and operating independently on each feature map. There are several types of Pooling operations.

The greatest element is taken from the feature map in Max Pooling. Average Pooling calculates the average of all elements in a Image section of predefined size. The total sum of predefined section elements is computed in Sum Pooling. The Pooling Layer usually acts as a bridge between the Convolutional Layer and the FC Layer.



1	0	2	3
4	6	6	8
3	1	1	0
1	2	2	4

6	8
3	4

Fig.7 Example of Max Pooling

FULLY CONNECTED LAYER

The FC layer is a traditional Multi-Layer Perceptron and it uses softmax activation function in the output layer. Here, every neuron in the previous layer is connected to every other neuron in the immediate layer. The output of convolutional and pooling layers give high-level features of the image. The Fully Connected layer uses these features for classification of the input image into different classes based on the training dataset.

Softmax is utilized for activation function. It shows the output as a score for each class. Other than classification, a fully-connected layer is also used as a cheap way for learning the non-linear combinations of features. Many of the features from convolutional and pooling layers may be good for the classification, but combinations of those features will be even better.

DROPOUT LAYER

This layer drops the input to a hidden layer if the previous layer is also provided with the same input.

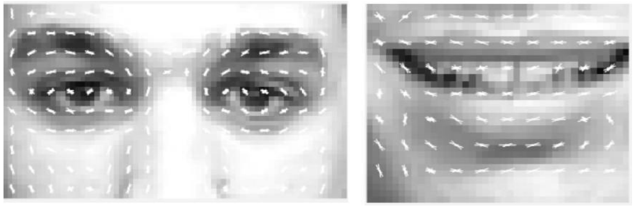


Fig. 8 CNN Model

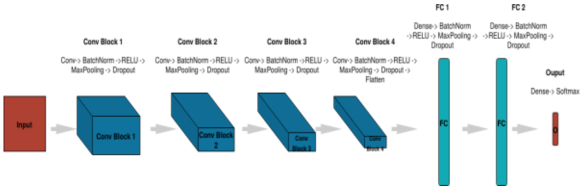

FC 1
FC 2
Conv Block 1
Conv Block 2
Conv Block 3
Conv Block 4
Dense-> BatchNorm ->RELU -> MaxPooling -> Dropout
Dense-> BatchNorm ->RELU -> MaxPooling -> Dropout
Conv-> BatchNorm ->RELU -> MaxPooling -> Dropout
Conv-> BatchNorm ->RELU -> MaxPooling -> Dropout
Conv-> BatchNorm ->RELU -> MaxPooling -> Dropout
Conv-> BatchNorm ->RELU -> MaxPooling -> Dropout -> Flatten
Ouput
Dense-> Softmax
Input
Conv Block 1
Conv Block 2
Conv Block 3
Conv Block 4
FC
FC
O

Fig.9 Plotted visualizations of HOG features on extracted eye and mouth regions.

IV. RESULTS AND DISCUSSION

The completed training implementation uses Viola-Jones's Haar-like feature cascade detector to detect faces as well as eyes and mouths. Detected faces are cropped, resized, and mean subtracted, then PCA is performed. Using the reduced-dimensionality training dataset Fisher LDA is performed to extract Fisherfaces on which we can project test data. Also during training, eye and mouths are detected using Haar-like features, or using a Harris corner based approach is Haar-like features fail. The detected eye and mouth regions are then extracted and resized. HOG features are extracted from each region, and a SVM is trained using a combined eye-mouth HOG vector and training labels.

TESTING

The purpose of testing is to find errors. Testing is the way of trying to find each possible issue or faults in a work product. It provides an approach to check the functionality of components, sub-assemblies, assemblies and additionally a completed product. It is the way of exercising software with the goal of ensuring that the errors can be solved.

Software system lives up to its necessities and user's expectations and doesn't fail in an unacceptable manner. There are different types of tests. Each test type addresses a particular testing requirement.

White Box Testing

White Box Testing is a testing where the software tester knows about the inner operations, design and language of the software, or at least its purpose. It is used to test regions that can't be reached from a black box level.

Black Box Testing

Black Box Testing is testing a software with no knowledge on the internal workings, design or language of the module being tested. Black Box tests, as most different kinds of tests, should be written from a definitive source document, for example, specification or requirements document. It is a testing technique where the software under test is treated as a black box you can't "see" into it. The tester gives inputs and reacts to outputs without thinking about how the software works.

Functional testing

Functional tests give systematic validation that functions tested are accessible as determined by the business and technical requirements, system documentation, and client manuals.

Functional testing is focused on the following items:

Valid Input : identified classes of valid input should be accepted. Invalid Input : identified classes of invalid input should be rejected. Functions : identified functions should be worked out.

Output : identified classes of application outputs should be worked out. Systems/Procedures: interfacing systems or methods should be invoked.

Organisation and preparation of useful tests is focused around prerequisites, key functions, or special test cases. Furthermore, systematic coverage pertaining to identifying Business process flows; data fields, predefined processes, and successive processes should be considered for testing. Before functional testing is finished, additional tests are recognised and the effective value of current tests is resolved. Functional testing techniques are:

Unit testing

Integration testing

System testing

Acceptance testing

Unit testing

Unit testing includes the design of test cases that approve that the internal program logic is working appropriately, and that program inputs produce valid outputs. All the decision branches and internal code flow should be verified. It is the testing of individual programming units of the application. It is done after the fulfillment of an individual unit before integration. This is a structural testing that depends on information on its development and is invasive. Unit tests perform fundamental tests at unit level and test a particular business cycle, application, as well as system configuration. Unit tests guarantee that every unique path of a business cycle performs precisely to the documented specifications and contains clearly defined inputs and expected outcomes.Unit testing is normally conducted as a part of a consolidated code and unit test phase of the software lifecycle, despite the fact that it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing to be performed manually and functional tests will be written in detail.

Test objectives

All field entries should work properly.

Pages should be activated from the identified link.

The entry screen, messages and responses should not be delayed.

Features to be tested

Verify that your face is detected or not.

All links should take the user to the correct page.

Test Results: All the test cases are passed successfully. No errors were identified.

Integration testing

Integration tests are intended to test integrated software components to decide whether they actually run as one program. Testing is occasion driven and is more concerned about the essential result of screens or fields. Integration tests exhibit that although the components were independently satisfied, as displayed by successful unit testing, the set of components is correct and reliable. Integration testing is explicitly aimed towards exposing the issues that arise from the set of components.

Software integration testing is the incremental integration testing of at least two integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, for example components in a software system or – one step up – software applications at the organisation level – associate without error.

Test Results: All the test cases are passed successfully. No errors were identified.

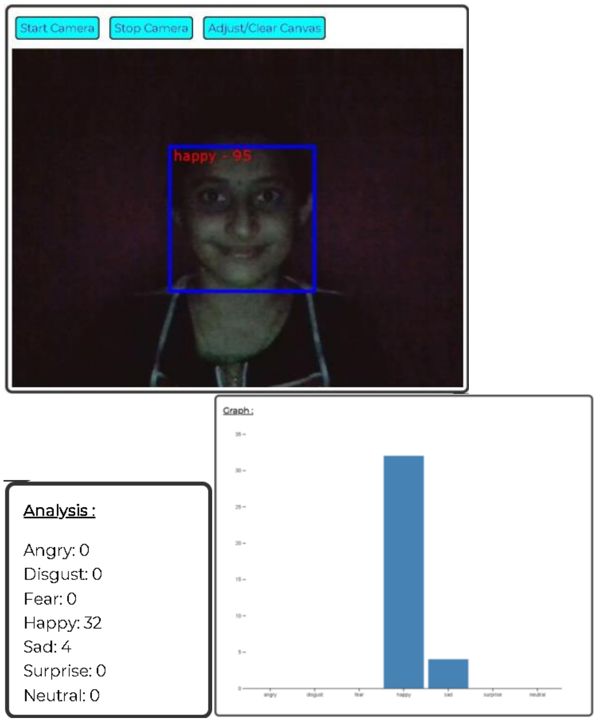
System Testing

System testing verifies that the whole integrated software system meets requirements. It tests the configuration to verify known and predictable outcomes. An example for system testing is the configuration oriented system integration test. System testing depends on process descriptions and flows, emphasising pre-driven process links and integration points.

Acceptance Testing:

User Acceptance Testing is a basic phase of any project and requires significant participation by the end user. In addition, it also ensures that the system meets the functional requirements.

Test Results: All the test cases are passed successfully. No errors were identified.



Analysis :

Angry: 0
Disgust: 0
Fear: 0
Happy: 32
Sad: 4
Surprise: 0
Neutral: 0

Fig.10 Results obtained from the model

TEST CASES AND RESULT ANALYSIS

Test Case ID	Description
001	Determine performance under low light conditions.
002	Determine performance under other face angles
003	To know about Mode distance range

Test Case ID:	001
Test Description :	Determine performance under low lights conditions
Purpose:	Think about how much light conditions influence identification and acknowledgment under 90 lux. of brightening in a 10 mt.sq region.
Input Data:	Yale Face Database which has gray scale pictures plus three more subjects.
Expected Result:	Three new subjects are detected by the system.
Actual Result:	Two subjects were identified in an order towards the camera and perceived in the principal pass, the other subject was distinguished and perceived on the subsequent pass
Pass / Fail:	Pass
Remarks:	N/A

Table 1. Table for Result Analysis

V. CONCLUSION

In this project, a LeNet architecture based six layer convolution neural network is implemented to classify human facial expressions i.e. happy, sad, surprise, fear, anger, disgust, and neutral. The system has been evaluated using Accuracy, Precision, Recall and F1-score. The classifier achieved accuracy of 70.77 % , precision of 7.07, recall 7.07 and F1-score 7.07.In the future work, the model can be extended to color images. This will allow us to investigate the efficacy of pre-trained models such as AlexNet[11] or VGGNet [12] for facial emotion recognition.

VI. REFERENCES

(1) Shan, C., Gong, S., & McOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370). IEEE.

Chibelushi, C. C., & Bourel, F. (2003). Facial expression recognition: A brief tutorial overview. CVonline: On-Line Compendium of Computer Vision, 9.

(2) "Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation". DeepLearning 0.1. LISA Lab. Retrieved 31 August 2013.

(3) Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). Neural Networks. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 17 November 2013.

(4) LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013

(5) C. Zor, "Facial expression recognition," Master's thesis, University of Surrey, Guildford, 2008.

(6) Suwa, M.; Sugie N. and Fujimora K. A Preliminary Note on Pattern Recognition of Human Emotional Expression, Proc. International Joint Conf, Pattern Recognition, pages 408-410, 1978

(7) Recognizing action units for facial expression analysis YI Tian, T Kanade, JF Cohn IEEE Transactions on pattern analysis and machine intelligence 23 (2), 97- 115

(8) Raghuvanshi, Arushi, and Vivek Choksi. "Facial Expression Recognition with Convolutional Neural Networks." Stanford University, 2016

(9) Alizadeh, Shima, and Azar Fazel. "Convolutional Neural Networks for Facial Expression Recognition." Stanford University, 2016

(10) Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

(11) Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large- scale image recognition." arXiv preprint arXiv:1409.1556 (2014).