Facial Expression Evaluation to assess mental health through Deep Learning

1. Prajwal Gaikwad, Computer dept,AISSMS IOIT, Pune
2. Sanskruti Pardeshi, Computer dept, AISSMS IOIT, Pune
3. Shreya Sawant, Computer dept, AISSMS IOIT, Pune
4. Shrushti Rudrawar, computer dept, AISSMS IOIT, Pune
5. Ketaki Upare, Computer dept, AISSMS IOIT, Pune

Abstract – Expressions have been one of the most primary forms of non-verbal communication for humans. This is very natural and has formed an irreplaceable part of our lives. The facial expressions have also been highly useful parameters in the understanding of human behavior and analysis of the human psyche. The psychotherapists have been engaged in this practice for a long time and the rise of awareness about mental health issues have been leading to an increase in the number of patients. Therefore, there is a need for an accurate analysis tool for achieving the facial expression detection. To provide a solution to this problem, this research article illustrates a facial expression technique through the use of Feature Extraction along with Convolutional Neural Networks and Decision tree. The approach has been effectively evaluated for its accuracy which has resulted in highly satisfactory results.

Keywords: *YC_BC_R Model, Convolution Neural Network, Decision Tree, Expression Evaluation.*

I INTRODUCTION

The expressions on one's face and the changes related to the facial patterns gives the information about the emotional state the subject is in and it helps regulate a healthy conversation with the subject. The overall mood of the subject is understood from these expressions, eventually developing a better understanding of the subject. In non-verbal communications and human interactions, an important role is played by facial expressions. The analysis of facial expression deals with analyzing and visually recognizing facial features changes and different facial motions.

The most important aspect of mood detection is to understand the subject. If a subject cannot be understood their problem cannot be solved being able to detect the mode of a person helps one understand the subject in a better way. Using this method, a doctor can easily be able to understand the patient better. Similarly, a psychologist can easily understand the current mood of the subject and give them the right treatment that would benefit them.

The expressions on one's face plays a vital role in understanding the person, in facilitating communication with humans as well as interacting with them. They also play a major role in medical rehabilitation of a patient and become a base for behavioral studies.Mood detection based on the technique of capturing facial images main provide a very practical approach towards detection of mood non-invasively.

In this system live streaming of the video of the subject is done via a camera placed in front of the subject and processed further. The process of frame grabbing comes next. Here, the frames of the video are extracted and then these frames are processed to extract the region of interest. The region of interest is the area of the image on which further processing is to be done.

In feature extraction and image processing parameters extracted from the images were of two types binary parameters and real value parameters. Depending upon the measured distance the real valued parameters have a definite value which is measured in the number of pixels. Whereas, the binary measures returned either an absent (= 0) or present (= 1) value. On a whole, 7 binary measures and 8 real valued measures are obtained.

To identify a certain facial expression a number of parameters both, binary and real-valued are analysed and extracted to decide their effectiveness. Elimination of the features which do not provide any significant information of the expression portrayed is vital. The following binary parameters and real-valued parameters are considered in image processing:

Raised eyebrow distance: It is the distance between the lower and upper eyelid and the lower central point of the eyebrow. Eyebrow to upper eyelid to distance: It is the distance between the eyebrow surface and upper eyelid. Inter-eyebrow distance: It is the distance between the lower central points of the eyebrows. Lower eyelid to upper eyelid distance: It is the distance between the lower and upper eyelid. Top lip thickness: It is the thickness of the upper lip. Lower lip thickness: It is the thickness of the lower lip. Mouth width: It is the distance between the tips of the lips. Mouth opening: It is the distance between the upper surface of the bottom lip and the lower surface of the upper lip.

Upper teeth visible: It is a parameter that indicates whether the upper teeth are visible or not. Lower teeth visible: It is a parameter that indicates whether the lower teeth are visible or not. Forehead lines: It is a parameter that indicates

whether wrinkles are present in the forehead's upper side or not. Eyebrow lines: It is a parameter that indicates whether wrinkles are present in the area above the eyebrows or not. Nose lines: It is a parameter that indicates whether wrinkles are present in the area between the eyebrows and over the nose or not.

Chin lines: It is a parameter that indicates whether lines or wrinkles are present in the area below the chin or under the lips or not. Nasolabial lines: It is a parameter that indicates whether thick lines are present on both sides of the nose extending till the upper lip or not. These parameters real-value based and binary, play a major role in creating the base for our calculations to be given to the neural network to compare and analyse the facial expression of the subject whose image has been captured from the video.

The region of interest is analysed by the Convolutional Neural Network and a Convolution Rate is calculated by the Convolutional Neural Network. The Convolution Rate is analysed by the Entropy Estimation block where the Entropy List is calculated using the Convolution Rate. This Entropy List is then sent to the Decision Tree where it determines a response about the facial expression of the subject. This expression is then sent further to the Facial Analysis Module where it derives a detailed facial analysis of the subject.

The obtained expression is used by the doctor's application to evaluate the patient's mood in a more technical way. This research mainly focuses on identification of the patient's expression in a different time and helps the doctor to evaluate his/ her mental status on each of the visits.

This research article dedicates section 2 for the evaluation of the past works under the name literature survey. And the implemented technique is broadly described under the section Proposed methodology which is numbered as 3. The section 3 discusses the obtained results from the experimental process. And finally, the section 5 concludes this paper along with the scope for the future enhancements.

II LITERATURE SURVEY

M. Pantic explains that there has been an increased interest in the human facial characteristics as they are useful for achieving an identification of the inner emotional state of individual [1]. The human beings are highly dependent on such emotional cues which allow for effective communication. These visual cues for the facial expression identification are quite easily identified and extracted by the human beings due to the inherent detection mechanisms in the human brain. This is highly difficult for a computer interface to achieve this identification due to the lack of a useful technique to achieve it. For an improvement in the recognition of the facial expression, an effective approach is determined using temporal, spatial and rule-based reasoning.

Yongqiang Li expresses that there have been multiple techniques that have been instrumental in driving the paradigm of image processing and computer vision forward [2]. This has been crucial in the development and significant improvement in the underlying technique for the detection for the purpose of enabling a useful and effective image processing guidelines. The paradigm of achieving a highly useful facial feature extraction approach for the purpose of analyzing the expression of the individual and recognizing it accurately. The approach implements Bayesian networks to achieve the considerable improvements in the facial expression detection mechanism.

Yongmian Zhang elaborates on the concept of facial expression and its use for the purpose of achieving effective interactions with one another [3]. This is highly crucial as human are social beings that need to interact with each other to maintain their mental health in a better position. The effective realization of the automatic detection of the facial expression is needed for utilization in various implementations. This is achieved by the technique presented in this research article through the use of an Image sequences along with dynamic and active fusion of information for the facial expression detection.

Yuta Kihara states that there has been an increased interest in the evaluation of the facial characteristics that have been useful in various implementations that have been usually performed manually [4]. The process of paralysis cannot be determined effectively and discussed by the doctors. The paralysis is highly debilitating and can only be diagnosed by a professional by thorough analysis that takes into account various changes to achieve the accurate evaluation. Therefore, there is a need for an approach that can effectively identify facial parameters in achieving the facial paralysis detection using a database containing dynamic facial expressions using quantitative analysis.

Nazil Perveen discusses that the facial expression evaluation through an image processing approach is useful in achieving various implementations accurately. The facial expression recognition has been realized to allow for greater understanding of the non-verbal communication approaches of human beings [5]. This is highly useful as it can be effective in achieving an enhancement in the various realizations of the detection approach. This publication outlines an effective methodology for the purpose of portraying an improvement in the facial expression recognition through the use of the Gini Index along with the Facial Characteristic points.

Catherine Soladie introduces the concept of assisting the elderly individuals in achieving an effective living alone and achieved other goals. There have been large amounts of interest in achieving an improvement in the detection of the changes in the behavior of the elderly individuals living alone [6]. The most effective approach for the purpose of facial expression detection has been devised taking into account the invariant representation which is useful for achieving an effective and useful implementation. The implementation of such a system is highly useful in achieving an improvement in the recognition of blended expressions.

Su-Jing Wang narrates that the concept of micro expression is one of the most important and useful indications of concealed or obscured emotions. These emotions have been

effective in realizing the emotion of an individual in a high stakes environments. These emotions and the facial expressions, especially micro-expressions are very difficult to detect and identify even for a trained eye [7]. This is due to the fact that these expressions are short lived and can be very fragmented. Therefore, this research article has been utilized to devise an effective micro-expression detection through the use of color Spaces.

Jiannan Yang explains that the process of facial expression recognition has been extremely basic as a large number of individuals can easily do this with a relative ease. This is problematic for the computer to detect expression as it cannot understand the various nuances of the facial expression [8]. The detection of the facial expression has been highly useful for determination of the various implementations that can be highly useful. Therefore, the authors in this approach provide a system that utilizes the facial action unit for the purpose of facial expression detection.

Lei Pang expresses that the human facial expression forms an important aspect of communication between two different individuals [9]. These have been effective in realization of the state and the emotional wellbeing of the individual while performing the communication. Therefore the use of the effective realization of the automatic recognition of the facial expression needs the implementation of an efficient neural network that have been combined with the use of Gabor Features by the researchers to achieve highly accurate facial expression identification with high accuracy. The neural network used by the authors is the BP neural networks.

Lutfiyatul Fatjriyati Anas elaborates on the concept of non-verbal communication that is one of the major communication mechanisms that are used by humans for the purpose of conveying their message [10]. This consists of emotions that are effectively identified by the target individuals through the highly visible facial expressions. These expressions are difficult to identify by the computer and marks a major problem in computer vision. Therefore, the authors in this research article have proposed the use of Fashion images and landmarks along with the K Nearest Neighbor to achieve the accurate determination of the facial expressions.

Zhang Nan[11] states that there has been an important and highly challenging problem of detection of facial expression in computer vision and image processing. The detection of facial expressions comes naturally by individuals due to the large amount of non-verbal communication that can be extremely challenging for the computer to determine. This problem needs to be solved to implement an effective approach that can be extremely useful in a large selection of implementations. The authors in this research article have proposed the use of Local Facial regions for the purpose of achieving the effective realization of the facial expression.

III PROPOSED SYSTEM FOR FACIAL EXPRESSION EVALUATION

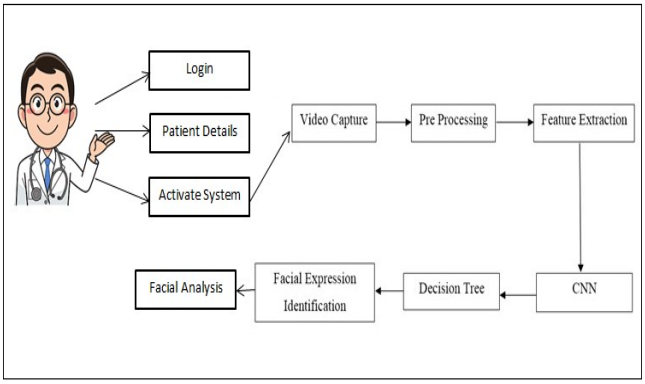

Figure 1: Facial Expression evaluation System overview

The proposed model for facial expression evaluation for the patient's mental health assessment is being depicted in the figure 1. The steps that are included to develop the model are detailed in the below mentioned steps.

Step 1: Frame extraction and preprocessing – This is the initial step of the proposed model where the doctor is logged into the system and access the camera to get the picture of the patient. For this purpose an opencv library is installed for the Java programming language, with the help of this an image of a patient's face is being captured and preprocessed in the image object to store it in jpg format in a destined folder.

Step 2: Feature extraction – This is the important step of the proposed model, where a python program iteratively visit the destined folder and checked for the presence of any image. This image was captured by the doctor through the interactive user interface designed by the java swing framework. If the image is present in the destined folder, then the image is processed to detect the face using the haar cascade properties stored in a .xml file. The HAAR features are used extensively in this xml file which eventually helps to identify the face position in the image. Once the face position is identified in the image, then the coordinates of this facial position are identified to extract only the face in the image object. This image object is subjected to estimate the skin of the face using the YC_BC_R model.

In the YC_BC_R model initially each pixel of the image is read to extract the RGB values. These RGB values are extracted to estimate the blue chroma and red chroma components with the below mentioned equations 1 and 2.

$$Cb=-0.169 * R - 0.332 * G + 0.500 * B + 128 \quad (1)$$

$$Cr=0.500 * R - 0.419 * G - 0.081 * B + 128 \quad (2)$$

This blue chroma component is evaluated in the range 137 and 177 for the skin pixel. And then again red chroma component is evaluated in the range 77 to 127. Then the combination of these two red chroma and blue chroma components are used to estimate the pixel is skin or not. If the pixel satisfies all the conditions of blue and red chroma

components, then it is considered as the skin pixel and marked with the white pixel color value. On the other hand, if it fails to satisfy the condition then it is considered as the non-skin pixel and marked with the black color value. Once this process is iterated for all the pixels of the image the binary image is yielded with skin in white pixel values and non-skin in black pixel values. This process is depicted in the below shown algorithm 1.

ALGORITHM 1: Skin Detection through YC_BC_R model

```
//Input : Input image I_IMG
//Output: Skin Detected image S_IMG
// function: skindetection(I_IMG)
1: Start
2: SIMG = ∅ , count=0
3:          for i = 0 to size of breadth of I_IMG
4:          for j=0 to size of Height of I_IMG
5:              col = I_IMG [i,j] ( PIX)
6:              R= col[0]
7:              G=col[1]
8:              B=col[2]
9:         Cb = -0.169 * R - 0.332 * G + 0.500 * B +128)
10:        Cr = 0.500 * R - 0.419 * G - 0.081 * B + 128
11:    if (Cr > 137 && Cr < 177), then
12:    if (Cb < 127 && Cb > 77), then
13:         t = Cb + 0.6 *Cr;
14:         if (t > 190 && t < 215),then
16:             SIMG [i,j] ( PIX )=[255,255,255]
17:         end if,else
18:           S_IMG [i,j] ( PIX )=[0,0,0]
19:        end if,else
20:           S_IMG [i,j] ( PIX )=[0,0,0]
21:       end if,else
22:           S_IMG [i,j] ( PIX )=[0,0,0]
23:       end if
24:         endfor
25:         endfor
26: return S_IMG
27: stop
```

Step 3: Convolution Neural Network – The original captured image along with the skin object is subjected to estimate the user expression based on the learned details stored in .h5 files through the CNN model. This trained details are extracted with the help of the dataset downloaded through the URL : https://drive.google.com/file/d/1X60B-uR3NtqPd4oosdotpbDgy8KOfUdr/view.

This dataset contains two subfolders like train and test. These two subfolders contains many expression folders like angry, disgusted, fearful, happy, neutral, sad and surprised. Each of these expression folders contains many images belong to that expression. Once these dataset folders are feed to the CNN training model, then they are subjected to set the rescale ratio and then they are rescaled to the width and height of 48 X 48. After this a training batch size of 64 and epochs of 50 is set to train the model. The CNN model is trained by using the keras and tensor flow libraries built for the python programming language. This CNN model is built with the architecture as mentioned below in figure 2.

Layer	Activation
CONV 2D 32 X 3 X 3	Relu
CONV 2D 64 X 3 X 3	Relu
MaxPooling2D 2 X 2	
Dropout 0.25	
CONV 2D 128 X 3 X 3	Relu
MaxPooling2D 2 X 2	
CONV 2D 128 X 3 X 3	Relu
MaxPooling2D 2 X 2	
Dropout 0.25	
Flatten	
Dense 1024	Relu
Dropout 0.25	
Dense 7	Softmax
Adam Optimizer	

Figure 2: CNN network Architecture

Once the CNN Model is compiled for the given number of epochs, then the trained data is stored in an .h5 model file. This h5 model file is used to test the input image from the camera which was preprocessed in the step 2.

Step 4: Decision Tree- This is the final step of the proposed system, where trained data is used to predict the emotions for the captured image. For this purpose another CNN neural network model is created which is used to read the trained data from .h5 file. And then by using the predict method of the model an integer value is obtained which eventually indicates the different emotions with respect to the input directory of the trained data. Based on this integer value a particular an expression is extracted from the parallel array created for the expressions. The obtained expression is stored in the database for the further analysis by the doctor himself.

IV. RESULTS AND DISCUSSIONS

The proposed methodology for the purpose of detecting facial expression has been achieved through the use of a combination of java and python programming language. The interfacing has been achieved through the NetBeans IDE and the facial expression detection mechanism has been achieved using the Spyder IDE in python. The laptop for development of the technique has been equipped with an Intel core i5 processor assisted by 4GB of RAM and 500 GB of storage. The MySQL database server is being used to handle the storage requirements.

The performance of the technique has been achieved through the realization of intensive experimentation of the

methodology. The assessments have been performed to determine accuracy of the facial expression detection approach in thorough detail. This is necessary to attain the performance metrics of the technique to identify if the methodology has been deployed correctly and functions appropriately. The performance evaluation has been elaborated below.

Performance Evaluation based on Precision and Recall

The Precision and Recall metrics are being utilized for the purpose of enabling an effective evaluation of the accuracy of the facial expression detection. The precision and recall are essential components to measure the accuracy of the detection mechanism as they are able to determine the real performance of the methodology. The detailed evaluation of the approach has been depicted as follows. The in-depth evaluation of the approach determined through the use of precision and recall metric requires the analysis of a number of parameters. The precision is the measure of the relative accuracy of the detection whereas the recall metric determines the absolute accuracy of the detection approach.

The precision in this system is the division of the correctly detected facial expressions by the total number of facial expression detections. On the other hand the Recall is measured as the division of total number of facial expression detections by the total number of expected facial expression detections.

The mathematical depiction of the precision and recall metrics has been illustrated below.

- ✓ A = The number of correctly detected facial expressions
- ✓ B= The number of incorrectly detected facial expressions
- ✓ C = The number of facial expressions not detected

So, precision can be defined as

Precision = (A / (A+ B)) *100

Recall = (A / (A+ C)) *100

The extensive evaluation of the presented approach has been performed through the realization of the equations given above. The evaluation outcomes have been listed in the table 1 below.

No. of expected Detections	The number of correctly detected facial expressions (A)	The number of incorrectly detected facial expressions (B)	The number of facial expressions not detected (C)	Precision	Recall
38	27	5	6	84.375	81.81818182
67	48	9	10	84.21052632	82.75862069
87	63	10	14	86.30136986	81.81818182
119	89	19	11	82.40740741	89
139	98	20	21	83.05084746	82.35294118

Table 1: Precision and Recall Measurement Table for the performance of Facial Expression Detection

Figure 3: Comparison of Precision and Recall for the performance of Facial Expression Detection

The achieved outcomes for the experimental evaluation through the use of experimentation using the precision and recall have also been converted into a graphical representation given in figure 2 above. As it is evident from the line graph, the approach has been effective in realization of the accurate facial expression detection. The precision and recall performance metrics attained in this experimental assessment are 84.06% and 83.54% respectively. These values depict the correct implementation of the facial expression detection approach which achieves satisfactory accuracy measures as expected.

V. CONCLUSION AND FUTURE SCOPE

The presented system for the facial expression detection approach has been achieved through the implementation of Convolutional Neural Networks and Decision Tree. The facial expression detection is a highly useful measure of the actual state of the person's mental status accurately, and can be a valuable tool for the psychoanalysts and psychotherapists. The presented system uses the video frames containing the face of the subject are captured from a camera into the system as an input. These frames are then first preprocessed by resizing and normalization of the frames. These preprocessed images are then provided to the next step for extraction of the features such as, skin detection and Region of interest Evaluation. These features are then provided to the Convolutional Neural Networks for the evaluation of the convolution and the fully connected layers which achieves the convolution rate. These values are then classified using the Decision Tree module. The Decision Tree approach implements if-then rules to achieve accurate determination of the facial expression. The experimental evaluation of the approach yielded satisfactory levels of precision and recall.

The future research directions for this facial expression detection approach can be applied to initiate this approach into an API for effective use in a variety of scenarios. And also this can be implemented as mobile application for the fun and effective usage in many real time scenarios.

REFERENCES

[1] Maja Pantic, and Ioannis Patras," Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences" IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, Vol. 36, No. 2, April 2006.

[2] Yongqiang Li, Shangfei Wang, Yongping Zhao, and Qiang Ji, "Simultaneous Facial Feature Tracking and Facial Expression Recognition" IEEE Transactions On Image Processing, Vol. 22, No. 7, July 2013.

[3] Yongmian Zhang, and Qiang Ji, "Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 5, May 2005.

[4] Yuta Kihara, Guifang Duan, Takeshi Nishid, Naoki Matsushir, and Yen-Wei Chen," A Dynamic Facial Expression Database for Quantitative Analysis of Facial Paralysis" 6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT) 2011.

[5] Nazil Perveen, Shubhrata Gupta, and Kesari Verma, "Facial Expression Recognition Using Facial Characteristic Points and Gini Index" Students Conference on Engineering and Systems 2012.

[6] Catherine Soladi´e, Nicolas Stoiber, Renaud S´eguier," A New Invariant Representation Of Facial Expressions: Definition And Application To Blended Expression Recognition" 19th IEEE International Conference on Image Processing 2012.

[7] Su-Jing Wang, Wen-Jing Yan, Xiaobai Li, Guoying Zhao Chun-Guang Zhou, Xiaolan Fu, Minghao Yang, Jianhua Tao," Micro-expression Recognition Using Color Spaces" IEEE Transactions On Image Processing 2015.

[8] Jiannan Yang, Fan Zhang, Bike Che3, Samee U. Khan," Facial Expression Recognition Based on Facial Action Unit" Tenth International Green and Sustainable Computing Conference (IGSC) 2019.

[9] Lei Pang, Nianqiang Li, Li Zhao, Wenxiu Shi, Yunpan Du, "Facial expression recognition based on Gabor feature and neural network " International Conference on Security, Pattern Analysis, 2018.

[10] Lutfiyatul Fatjriyati Anas, Nana Ramadijanti, Achmad Basuk," Implementation of Facial Expression Recognition System For Selecting Fashion Item Based on Like and Dislike Expression" International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC) 2018.

[11] Zhang Nan, Geng Xue" Facial Expression Recognition Based on Local Facial Regions" 4th IET International Conference on Wireless, Mobile & Multimedia Networks (ICWMMN 2011).

*****