FACIAL EXPRESSION EMOTION DETECTION FROM FACIAL IMAGES

Dr. R.N Devendra Kumar[1], Arunprakash C[2], Balaji N.S[3], Jenifer S[4]

[1] *Assistant Professor, Computer Science and Engineering, Sri Ramakrishna Institute Of Technology, Tamil Nadu, India*

[2] *Student, Computer Science and Engineering, Sri Ramakrishna Institute Of Technology, Tamil Nadu, India*

[3] *Student, Computer Science and Engineering, Sri Ramakrishna Institute Of Technology, Tamil Nadu, India*

[4] *Student, Computer Science and Engineering, Sri Ramakrishna Institute Of Technology, Tamil Nadu, India*

ABSTRACT

Emotion care for human well-being is important for all ages. Automatic emotion recognition plays a crucial role in various fields such as healthcare, human-computer interaction (HCI) and security and defense. Capturing the dynamics of facial expression progression in video is an essential and challenging task for facial expression recognition (FER). In this project, we propose a new low-cost and multi-user framework for emotion care system based on big data analysis for patient feelings, where emotion is detected in terms of facial expression. The system works with deep learning techniques on emotional big data to extract emotional features and recognize six kinds (e.g., angry, disgust, fear, happy, sad, surprise, and neutral) of facial expressions in real-time and offline. In addition, a new dataset for emotion recognition is collected to train the DCNN model. A deep convolutional neural network (DCNN) is further applied to the whole facial observation to learn the global characteristics of six different expressions. And the performed facial expression and predicted results can be saved into device to help cares to observe facial expressions of patients at any time and provide better suggestions to patients.

Keyword - *Expression detection, Health monitoring, Emotion detection.*

1. INTRODUCTION

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. It is an interdisciplinary field spanning computer science, psychology, and cognitive science. While the origins of the field may be traced as far back as to early philosophical inquiries into emotion (“affect” is, basically, a synonym for “emotion.”), the more modern branch of computer science originated with Rosalind Picard’s 1995 paper on affective computing. A motivation for the research is the ability to simulate empathy. The machine should interpret the emotional state of humans and adapt its behavior to them, giving an appropriate response for those emotions.

A major area in affective computing is the design of computational devices proposed to exhibit either innate emotional capabilities or that are capable of convincingly simulating emotions. A more practical approach, based on current technological capabilities, is the simulation of emotions in conversational agents in order to enrich and facilitate interactivity between human and machine. While human emotions are often associated with surges in hormones and other neuropeptides, emotions in machines might be associated with abstract states associated with progress (or lack of progress) in autonomous learning systems. In this view, affective emotional states correspond to time-derivatives in the learning curve of an arbitrary learning system. Two major categories describing emotions in machines: Emotional speech and Facial affect detection.

1.1 Background History

During the 1990ies, there was a wave of new research on the role of emotion in diverse areas such as psychology (e.g. Ellsworth and Scherer, 2003), neurology (e.g. LeDoux, 1996), medicine (e.g. Damasio, 1995), and sociology (e.g. Katz, 1999). Prior to this new wave of research, emotions had, as I mentioned, been considered to be a low-status topic of research, and researchers had mainly focused on how emotion got in the way of our rational thinking. Results at that point focused on issues like when getting really scared, pilots would get tunnel vision and stop being able to notice important changes in the flight's surroundings. Being upset with a colleague and getting angry in the middle of a business meeting could sabotage the dialogue. Or giving a presentation and becoming very nervous could make you lose the thread of the argument. Overall emotions were seen the less valued pair in the dualistic pair rational – emotional, and associated with body and female in the "mind – body", "male – female" pairs.

This dualistic conceptualization goes back as far as to the Greek philosophers. In Western thinking, the division of mind and body was taken indisputable and, for example, Descartes looked for the gland that would connect the thoughts (inspired by God) with the actions of the body.

A major area in affective computing is the design of computational devices proposed to exhibit either innate emotional capabilities or that are capable of convincingly simulating emotions. A more practical approach, based on current technological capabilities, is the simulation of emotions in conversational agents in order to enrich and facilitate interactivity between human and machine. While human emotions are often associated with surges in hormones and other neuropeptides, emotions in machines might be associated with abstract states associated with progress (or lack of progress) in autonomous learning systems. In this view, affective emotional states correspond to time-derivatives in the learning curve of an arbitrary learning system. Two major categories describing emotions in machines: Emotional speech and Facial affect detection.

1.2 Problem Statement

Human beings are mostly emotional, and our social interaction is measured by taking into consideration our ability to communicate emotions and to perceive the emotional states of others. Affective computing provides computing systems with mechanisms that emulate and/or interpret human emotions. Its main objective is to make communication with computing systems easier and more natural. combination, starting from facial expressions, oral intonation, psycho-physiological information, or even the texts used. To make them known to users, it is usual to employ avatars and speech synthesis, frequently combining the two. Although, in general, people are experts in recognizing and expressing emotions, sometimes there is misunderstanding when transmitting them. This may be caused by ambient issues (noise, lighting, or distance between interlocutors), or even personal issues (concentration or the behavior or confidence with the interlocutor). This is why emotional resources are frequently validated by people, in order to ascertain whether they really express the correct emotion or if the interlocutors are able to perceive them adequately. Many times, resources are not very expressive or not correctly understood by humans; therefore, computing Facial expression is one of the most natural and immediate means for human beings to communicate their emotions, as the human face can express emotions sooner than people verbalize or even realize their feelings. Automatic facial expression recognition (FER) has become an increasingly important research area that involves computer vision, machine learning, and behavioral sciences.

Much progress has been made in building computer systems to understand and use this natural form of human communication, although most of these systems attempt to recognize only a small set of prototypical emotional expressions. FER can be used for many applications, such as security, human-computer interaction, driver safety, and health care.

1.3 Scope

A new low-cost and multi-user framework for emotion detection is proposed. The system is based on a Web application that uses CNN model to classify facial expressions and works in real time and online.

Using facial features extracted with publicly available facial landmarks, action unit detection tools, and emotional video databases, we show that the proposed method for categorizing emotional photograms allows a valid set of emotionally labeled photograms that can then be used for emotion recognition.

1.4 Applications

As emotion AI technology can be introduced to many different industries with a variety of applications, tech giants and startups have started to invest in either computer vision or voice analysis to recognize human emotions. As a result, the technology has grown rapidly in the past two years and expanded into various new areas and industries to help businesses offer better customer experience and achieve real cost savings. Some examples

- Marketing
- Customer Service
- Human Resources
- Healthcare
- Insurance
- Retail
- Autonomous driving
- Education

1.5 Exiting System

As an important way to emotion recognition, there are many diverse FER methods that achieve a good performance SVM, Linear Discriminant Analysis (LDA), Bayesian Network (BN), Neural Network (NN), Gaussian mixture model (GMM).AdaBoost, PCA.

1.6 Proposed System

This section presents the proposed deeply learned classifiers for facial emotion classification of unfiltered real-life face images

- We propose a model that uses DCNN architecture to predict the emotion of human's faces from unfiltered real-world environments. The novel CNN approach addresses the Emotion labels as a set of discrete annotations and train the classifiers that predict the human's Expressions.
- We design a quality and robust image pre-processing algorithm that prepare and pre-process the unfiltered images for the CNN model and this greatly has a very strong impact on the performance accuracy of our facial Emotion classifiers.
- We demonstrate that pretraining on large-scale datasets allows an effective training of our Facial Emotion CNN model which enable the classifiers to generalize on the test images and then avoid over fitting.
- Health Care Recommendation System using predicted facial expressions.

2. LITERATURE REVIEW

“Spatial Augmented Reality Based Customer Satisfaction Enhancement and Monitoring System”Author(s) Udaya Dampage; D.A.Egodagamage; A.U.Waidyaratne; D.A.W.Dissanayaka; A.G.N.M.SenarathneUdaya Dampage; D.A.Egodagamage; A.U.Waidyaratne; D.A.W.Dissanayaka; A.G.N.M.Senarathne, Overall, the research outcome proposed a comprehensive scheme to monitor, analyze and enhance customer satisfaction with the marketing and logistic extensions, utilizing spatial augmented reality for smart restaurants.

“Multimodal Emotion Recognition Fusion Analysis Adapting BERT With Heterogeneous Feature Unification”Author(s) Sanghyun Lee; David K. Han; Hanseok Ko. The proposed HFU-BERT integrated visual and acoustic modalities into heterogeneous features and was successfully fine-tuned using BERT.

“Recognition of Teachers’ Facial Expression Intensity Based on Convolutional Neural Network and Attention Mechanism”Author(s) Kun Zheng; Dong Yang; Junhua Liu; Jinling Cui, Through the detection of the frequency and intensity of teachers’ facial expression in the classroom, one can understand the positive degree of teachers’ emotion in the teaching situation. It can not only provide an objective reference for teaching assessment, but also be used to analyze students’ interest in teaching content.

“Multimodal Attention Network for Continuous-Time Emotion Recognition Using Video and EEG Signals”Author(s) Dong Yoon Choi; Deok-Hwan Kim; Byung Cheol Song, Experimentally demonstrated that the proposed method improves emotional recognition performance over single-modality networks.

“WGAN-Based Robust Occluded Facial Expression Recognition”Author(s) Yang Lu; Shigang Wang; Wenting Zhao; Yan Zhao, The experimental results show that the proposed method has satisfactory complementing effect for monocular occlusion, binocular occlusion, mouth occlusion, nose occlusion, half face occlusion, and random occlusion area are less than 40%. In addition, the expression recognition rates of the images with or without occlusion are improved much more than the state-of-the-art methods.

3. REQUIREMENTS SPECIFICATION

3.1 Software requirements

- Processors: Intel® Core™ i5 processor 4300M at 2.60 GHz or 2.59 GHz (1 socket, 2 cores, 2 threads per core), 8 GB of DRAM
- Disk space: 320 GB
- Operating systems: Windows® 10, macOS*, and Linux*

3.2 Hardware requirements

- Server Side : Python 3.7.4(64-bit) or (32-bit)
- Client Side : HTML, CSS, Bootstrap
- Framework : Flask 1.1.1
- Back end : MySQL 5.
- Server : Wampserver 2i
- DL DLL : TensorFlow, Pandas, SiKit Learn

4. METHODOLOGY

For the system, image preprocessing is necessary before an image is fed into the CNN model. The image preprocessing mainly consists of two stages: face detection, data augmentation. A face detector is adopted for face detection in our system. If faces are detected, the four coordinates of region of interest (ROI) of the faces would be returned Project Flow Description to the system, the system would crop the faces and discard irrelevant background. Data augmentation is used to process the detected face images and increase the quantity of data, because training process of deep learning model usually needs huge amounts of data. The images are cropped by the random bounding boxes that have different cropped ranges from 0.85 to 1. Then the data are randomly flipped and rotated. After importing the haar cascade file we will have written a code to detect faces and classify the desired emotions. We have assigned the labels that will be different emotions like angry, happy, sad, surprise, neutral. As soon as you run the code a new window will pop up and your webcam will turn on. It will then detect the face of the person, draw a bounding box over the detected person, and then convert the RGB image into grayscale & classify it in real-time. The DCNN model is developed using TensorFlow platform, which is an end-to-end open-source platform for machine learning.

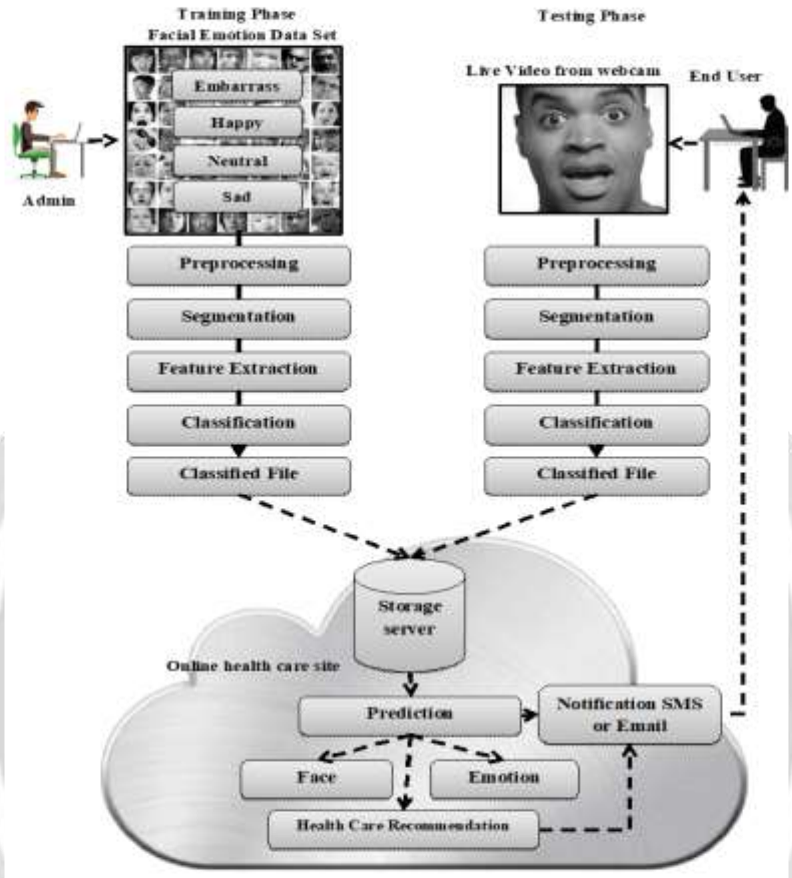

Fig -1: System Architecture

4..1 TRAINING PHASE

Facial Data Set Annotated

Facial Expression Recognition 2013 (FER-2013) dataset was prepared in Challenges in Representation Learning: Facial Expression Recognition Challenge, which is hosted categories (e.g., angry, disgust, fear, happy, sad, surprise, and neutral) and three different sets such as training set (28.709 images), validation set (3.589 images), and test set (3.589 images). All images in this dataset are grayscale with 48 X 48 pixels, thus corresponding to faces with various poses and illumination, where several faces are covered by hand, hair, and scarves. Because of FER-2013 is collected from the Internet and has various real-world conditions, it becomes one of the largest and most challenging databases for facial expression recognition.

4.2 TESTING PHASE

Live Video Annotation

Cameras should be deployed in critical areas to capture relevant video. Computer and camera are interfaced and here webcam is used. For every participant, one video with six kinds of facial expressions is collected and processed. A haar cascade classifier proposed by Viola and Jones is used to detect the face from video frame by frame. When a face is detected, the face image is saved into the database and labeled according to the facial expression the participant shows. Because the database has a lot of similar images due to the successive frames, we use difference hash (dhash) algorithm to select representative images from the dataset. The difference hash is one of image fingerprint algorithms, and it creates a unique hash value by calculating the difference between adjacent pixel values. To select images, form the dataset, we use difference hash to compute our image fingerprints because of its speed and accuracy.

4.3 FRAME EXTRACTION

Frames are extracted from video input. The video must be divided into sequence of images which are further processed. The speed at which a video must be divided into images depends on the implementation of individuals. From we can say that, mostly 20-30 frames are taken per second which are sent to the next phases.

4.4 PREPROCESSING

In that we will enhance the different features of images we get for example its intensity, contrast, saturation for different image processing. Low pass-filters a grayscale image that has been degraded by constant power additive noise. It uses a pixel wise adaptive Wiener method based on statistics estimated from a local neighborhood of each pixel.

4.5 FACE DETECTION

The Background subtraction approach is mostly used when the background is static. The principle of this method is to use a model of the background and compare the current image with a reference. The foreground objects present in the scene are detected.

It attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion. We will use the static background for the image subtraction which will give us the human we have to track. This step detects objects of interest as they move about the scene. The action detection process is independently applied to all the static cameras present in the scene. For human recognition is feature extraction and representation where the important characteristics of image frames are extracted and represented in a systematically way as features.

4.6 FEATURE EXTRACTION

Deep neural network (DNN) is a popular deep learning (DL) structure that consists of multiple-layered models of inputs. The DCNN architecture that we used to train and build the classifier model. The model can predict Facial Emotions. Hence, the DCNN model running at a fog node detects and labels the images with the name of the Facial Emotions having the highest probability, and saves those images.

4.7 CLASSIFICATION

In Classification stage, Convolutional neural networks algorithm is used for classification of Face Expression images. It is a non-parametric method which is used for both classification and regression.

Deep Convolution Neural Network Classifier: The Deep Convolution Neural Network (CNN) classifier is used mainly for image and video recognition. The CNN is able for automatically learning the respective feature for data itself. The CNN follows few steps like receiving different inputs, calculating the sum of their weights, forward output to activation function and respond with the desired output. Based on CNN classification, the Liver CT images important features like lines, edges, and object etc. complex features automatically able to identify with more accurately.

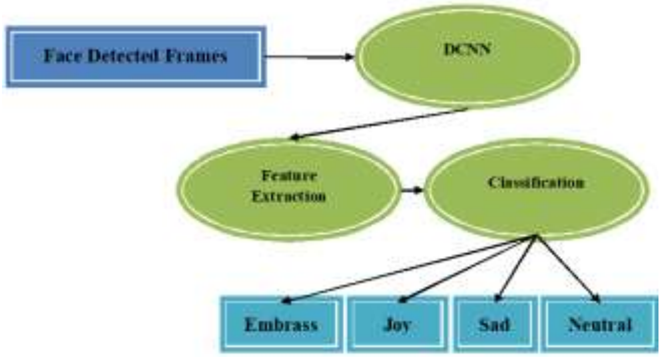

Fig -2: Classification

4.8 PREDICTION

In this module the matching process is done with trained classified result and test Live Camera Captured Classified file. Hamming Distance is used to calculate the difference according to the result the prediction accuracy will be displayed.

4.9 RECOMMENDATION SYSTEM

Facial Expression based recommend the health care systems where suggestions are based on an influence about a user's emotion and based on a degree of domain expertise and knowledge. Rules are defined that set context for each recommendation.

4.10 PERFORMANCE ANALYSIS

In this module, the outcome is counted as true positive (TP); if the same outcome is incorrectly classified as negative, it is counted as a false negative (FN). If the valid diagnosis is CHD absent and it is correctly classified as negative, the outcome is counted as true negative (TN); if the same outcome is incorrectly classified as positive, it is counted as a false positive (FP).

$$Accuracy\,(\%) = \frac{TP+TN}{TP+TN+FP+FN} \; x\,100$$

$$Sensitivity\,(\%) = \frac{TP}{TP+FN} \; x\,100$$

$$\text{Specificity}\,(\%) = \frac{TN}{TN+FP} \; x\,100$$

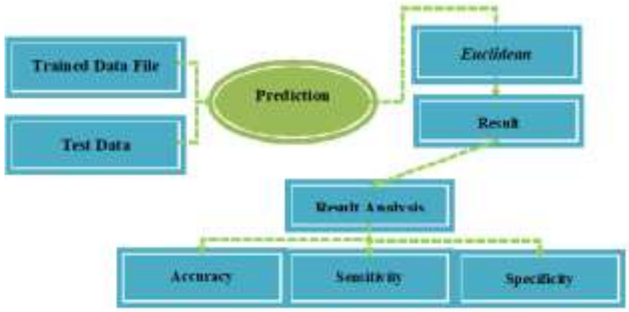

Fig -3: Performance Analysis

5. RESULTS AND DISCUSSION

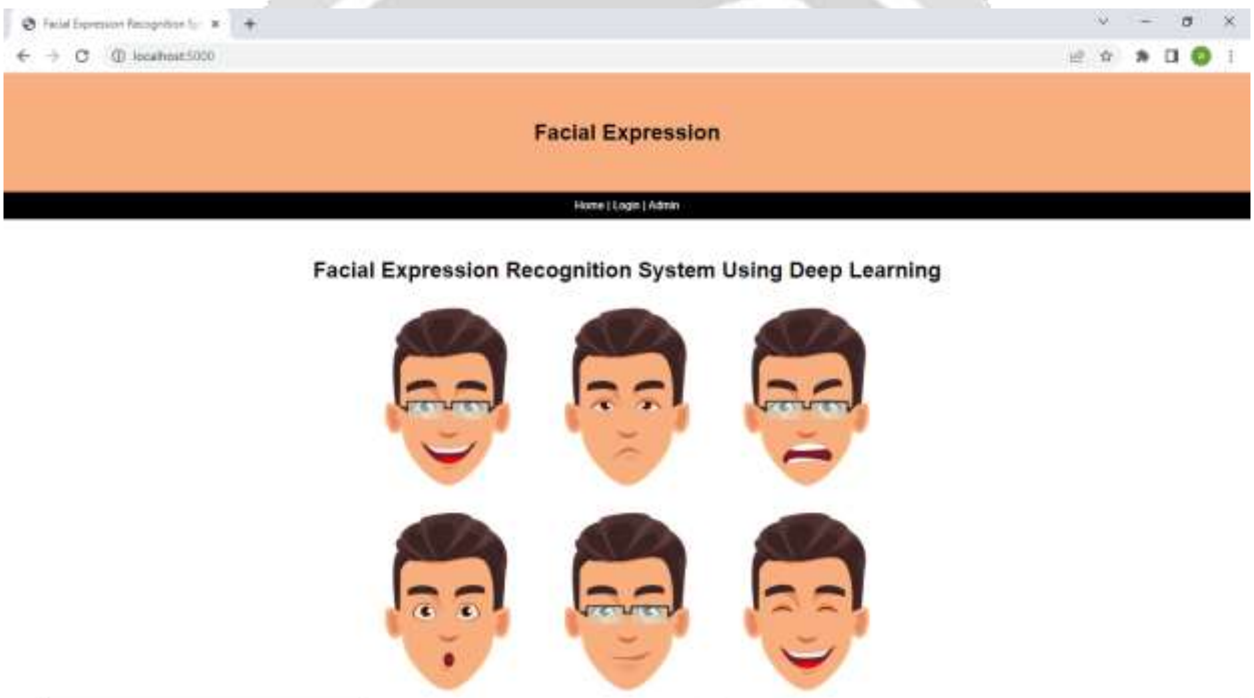

Fig -4: Home Page

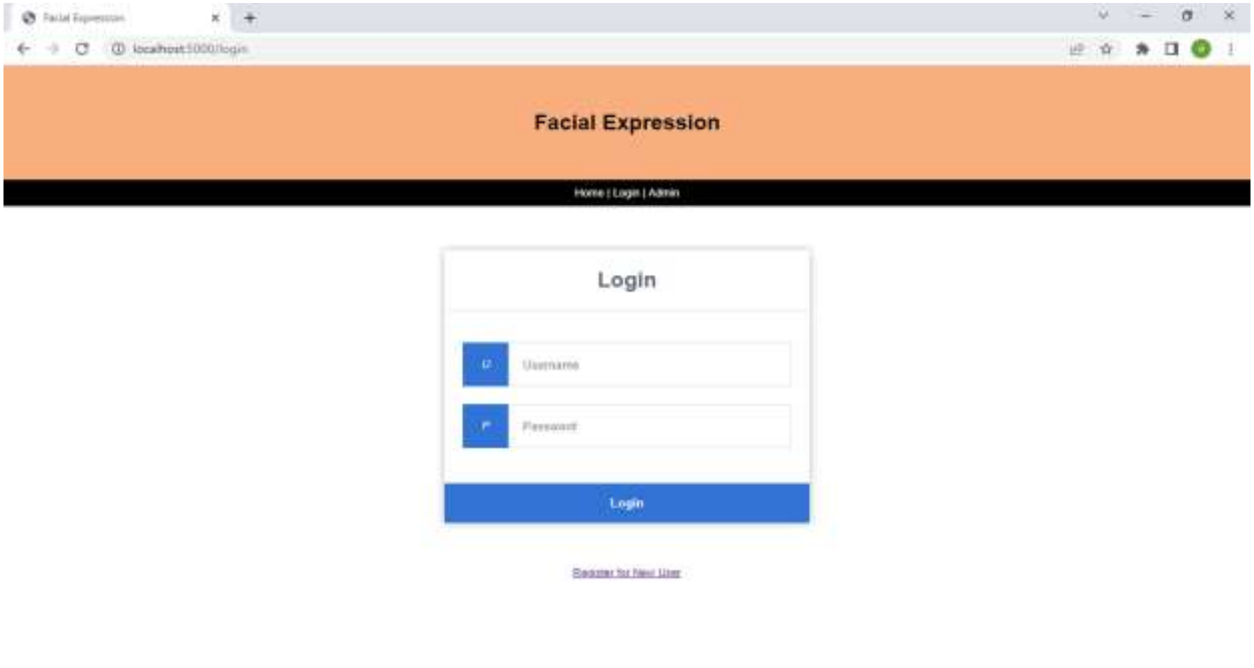

Fig -5: Login Page

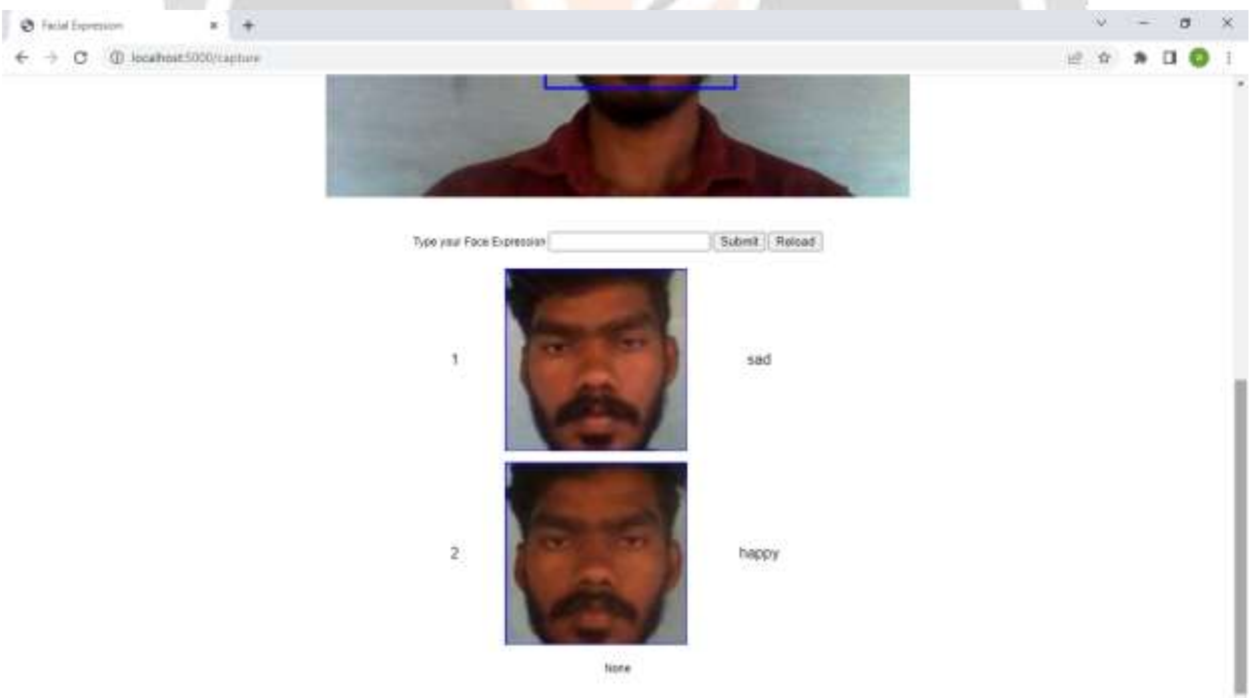

Fig -6: Capture Images

Fig -7: Emotion Detection

6. CONCLUSIONS

6.1 Summary

In this project, we adopt a deep learning technique to process emotional big data and develop an emotion care system using facial expression recognition system. We propose an algorithm that recognizes emotional changes in continuous-time domain by using video. In addition, emotion recognition has received much attention in the field of artificial intelligence. Conventional emotion recognition algorithms distinguished emotion categories by detecting changes in facial expressions. Recently, various emotion recognition mechanisms based on convolutional neural network (CNN) which are trained in an end-to-end manner have been developed and showed reliable performance. The Graphical Web User Interface allows users to do Realtime validation of the system. We have considered seven discrete and unique emotion classes (angry, disgust, fear, happy, neutral, sad and surprise) for emotion classification. So, there is no overlapping among classes.

6.2 Recommendation for future project

In the future work, we will try to implement our method on the videos for other applications and use different kinds of emotion representations (valence, arousal, etc.) for this task.

7. REFERENCES

[1] E. Ahmed, I. Yaqoob, I. A. T. Hashem, I. Khan, A. I. A. Ahmed, M. Imran, and A. V. Vasilakos, “The role of big data analytics in Internet of Things”, Comput. Netw., vol. 129, pp. 459 - 471, Dec. 2017.

[2] K. Lin, F. Xia, W. Wang, D. Tian, and J. Song, “System design for big data application in emotion-aware healthcare” IEEE Access, vol. 4, pp. 6901- 6909, 2016.

[3] Y. Jiang,W. Li, M. S. Hossain, M. Chen, A. Alelaiwi, and M. Al-Hammadi, “A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition” Inf. Fusion, vol. 53, pp. 209 - 221, Jan. 2020.

[4] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, “Understanding emotions in text using deep learning and big data,” Comput. Hum. Behav., vol. 93, pp. 309317, Apr. 2019.

[5] Y. Ma, Y. Hao, M. Chen, J. Chen, P. Lu, and A. Ko²ir, “Audio-visual emotion fusion (AVEF): A deep efficient weighted approach,” Inf. Fusion, vol. 46, pp. 184 - 192, Mar. 2019.

[6] M. S. Hossain and G. Muhammad, “Emotion recognition using deep learning approach from audio-visual emotional big data,” Inf. Fusion, vol. 49, pp. 69 - 78, Sep. 2019.

[7] Y. Miao, H. Dong, J. M. A. Jaam, and A. E. Saddik, “A deep learning system for recognizing facial expression in real-time,” ACM Trans. Multi- media Comput., Commun., Appl., vol. 15, no. 2, pp. 1 - 20, Jun. 2019, doi: 10.1145/3311747.

[8] G. Muhammad, M. S. Hossain, and N. Kumar, “EEG-based pathology detection for home health monitoring,” IEEE J. Sel. Areas Commun., vol. 39, no. 2, pp. 603 - 610, Feb. 2021.

[9] A. E. Saddik, “Digital twins: The convergence of multimedia technologies,” IEEE Multimedia, vol. 25, no. 2, pp. 8792, Apr./Jun. 2018.

[10] M. Chuah and M. Diblasio, “Smartphone based autism social alert system,” in Proc. 8th Int. Conf. Mobile Ad-hoc Sensor Netw. (MSN), Dec. 2012, pp. 6 - 13.