Real-Time Emotion Recognition System using Facial Expressions and Soft Computing methodologies

[1] S.Arun Inigo, [2]V.Rajesh Kumar, [3]P.Ashokram
[1,2,3]Assistant Professor, [1]Department of Information Technology, [2,3]Department of Computer Science and Engineering, Sethu Institute of Technology, Tamil Nadu, India
[1]barcyinix1527@gmail.com, [2]rvrajee0212@gmail.com, [3]ashok.inaiyam@gmail.com

Abstract:

Facial Expression conveys non-verbal cues, which plays an important role in interpersonal relations. The Cognitive Emotion AI system is the process of identifying the emotional state of a person. The main aim of our study is to develop a robust system which can detect as well as recognize human emotion from live feed. There are some emotions which are universal to all human beings like angry, sad, happy, surprise, fear, disgust and neutral. The methodology of this system is based on two stages- facial detection is done by extraction of Haar Cascade features of a face using Viola Jones algorithm and then the emotion is verified and recognized using Artificial Intelligence Techniques. The system will take image or frame as an input and by providing the image to the model the model will perform the preprocessing and feature selection after that it will be predict the emotional state.

Key words: Emotional Quotient, Emotional Intelligence, Ethnicity.

Introduction:

In One of the important ways humans display emotions is through facial expressions. Facial expression recognition is one of the most powerful, natural and immediate means for human beings to communicate their emotions and intensions.

Humans can be in some circumstances restricted from showing their emotions, such as hospitalized patients, or due to deficiencies; hence, better recognition of other human emotions will lead to effective communication Artificial Intelligence is the science of 21st century. Artificial Intelligence (AI) is defined as the ability for a machine to “think or act humanly or rationally”.

Corresponding Author: *S.Arun Inigo*
Department of IT, Sethu Institute of Technology
Tamil Nadu, India
Mail: barcyinix1527@gmail.com

Machines are now able to process the vast amount of data in real time and respond accordingly However, these machines with high IQ (Intelligence Quotient) were always lacking Emotional Intelligence (EI)/ Emotional Quotient (EQ). As technology progresses and the world becomes more and more virtual, there is a fear that we will lose the human connection and communication; but what if our devices could replace those interactions? The question of the era is whether we can build machines that can recognise human emotions.

The study and analysis of human emotions is not new. In 1872, Charles Darwin wrote in his book, The Expression of the Emotions in Man and Animals, “Facial expressions of emotion are universal, not learned differently in each culture”. The seminal research into the topic came from Paul Ekman, a psychologist. He was the first one to classify the emotions.

1. **Joy (Happiness)** - symbolized by raising of the mouth corners (an obvious smile) and tightening of the eyelids
2. **Surprise** - symbolized by eyebrows arching, eyes opening wide and exposing more white, with the jaw dropping slightly
3. **Sadness** - symbolized by lowering of the mouth corners, the eyebrows descending to the inner corners and the eyelids drooping
4. **Anger** - symbolized by eyebrows lowering, lips pressing firmly and eyes bulging
5. **Disgust** - symbolized by the upper lip raising, nose bridge wrinkling and cheeks raising
6. **Fear** - symbolized by the upper eyelids raising, eyes opening and the lips stretching horizontally
7. **Contempt** - symbolized by half of the upper lip tightening up and often the head is tilted slightly back.

These emotions are said to be the universal emotions. Developers and researchers have been advancing artificial intelligence to not only create systems that think and act like humans, but also detect and react to human emotions. Humans show universal consistency in recognizing emotions but also show a great deal of variability between individuals in their abilities. Enabling the devices around us to recognize our emotions can only enhance our interaction with machines, as well as among the family of humanity. The point of this project research is to develop personalized user experiences that can help improve lives. The fig1,2,3 shows the model facial emotions in samples [1-5].

(a) (b) (c) (d)

Figure 1: Example of subjects in the same class disgust in the CK+ dataset having different head shapes and ethnicities.

(a) (b) (c) (d)

Figure 2: Example of various illuminations ((a) and (b) in the CK+) and head orientations ((c) and (d) in the JAFFE) for subjects in the same dataset.

(a) Disgust (b) Angry (c) Sadness (d) Fear

Figure 3: Example of high similarity between facial expressions in two different classes

Proposed Methodology

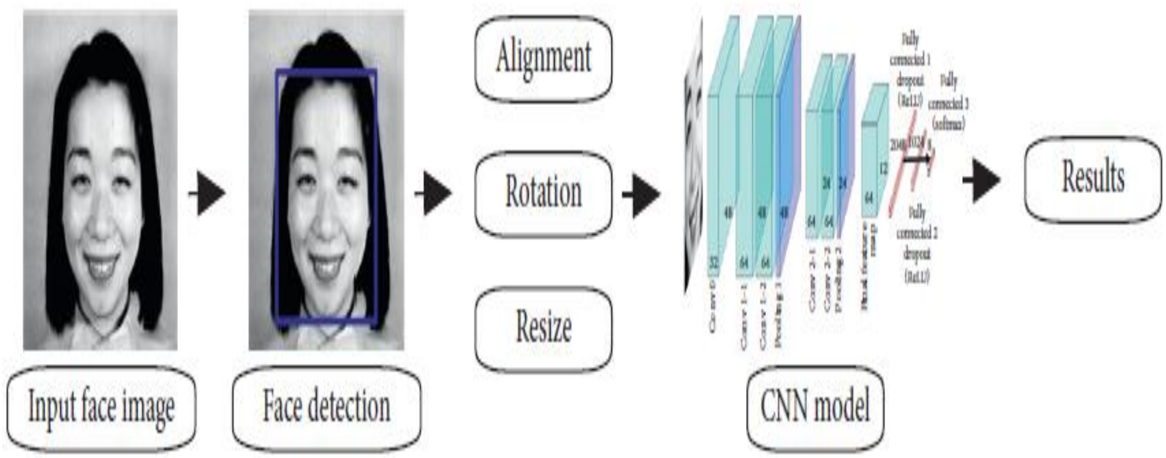

Figure 4 Proposed Overall Architecture

After the face detection, an efficient CNN-based architecture is used for the training and testing of face mask detection. There are several already available architectures for the training purpose, which were already discussed in the literature review section. In this work, we presented a custom architecture to detect whether the person wears a face mask or not. In the proposed work, face mask detection is performed by facial feature analysis.

The proposed lightweight CNN model consists of four convolutions, one fully connected and the output layer, and after each convolution layer, a nonlinear ReLU activation function is used to perform thresholding operation. The ReLU activation function drops neurons from the network whose values are less than zero, and the neuron with positive values are unchanged. In the proposed model, we employ two maximum pooling layers for dimensionality reduction, which affects the training duration of the network [6-11].

During training the model, the problem of overfitting can be occurred due to the model and data simplicity. To reduce the overfitting problem, we used the dropout and batch normalization mechanism. The output neurons in the last fully connected layer are equal to the number of classes recognized by the network, and lastly, the softmax classifier is used to classify the given input into the corresponding class such as face mask detected and no face mask detected.

IMAGE ACQUISITION (DATA SET DESCRIPTION)

The output of the first convolution layer is the input of the second convolutional layer after normalization and pooling layers. There is a total of 64 kernels with a size of 3 _ 3 in the second convolutional layer. The rest of the convolutional layers are connected without any intervening subsampling or batch normalization. In the third convolutional layer, 128 kernels are used, where each kernel has a size of 3 _ 3 _ 3. The final convolutional layer has 256 kernels having of size 3 _ 3 _ 3. The fully connected layers have a total of 128 neurons. The output of the fully connected layer is fed into a softmax layer, which produces a distribution over the 2-class labels, namely, mask detected and mask not detected. The complete detail of the proposed system is shown in Figure 3 along with subsequent layers descriptions.

Figure 5 Image Data Set Descriptions

IMAGE PRE-PROCESSING

As our input images are grayscale images, the pixel value ranges between 0 and 255 (256 Levels of grey). 0 representing white and 255 representing black colours. Having such big numbers as input, the neural network computation of optimal weights and bias will be too difficult. So we need to pre-process these image pixel values between -1 and 1. So we divide each of these pixel values by 255 and then map to a value between -1 and 1.

The pre-train program uses a set of labeled facial images to train on a CNN deep learning training model. Figure 3 shows the required input and the output files type for the pre-train program. Images collected from user is option and chapter 4 will discuss this option. The recognizer uses an RGB camera to obtain real time video. The facial detector detects the largest face and crops a facial image from the frame.

The facial image is used as input for the facial expression recognizer. The facial expression recognizer processes the facial image and shows the result on the command line. The inputs of the recognizer are an RGB real time video from a camera and checkpoint file from the pre-train program. The output of the facial detector is a fixed size image which is an input for the facial expression recognizer that uses it for the identification of the facial expression. The facial detector reads frames from the live video and uses OpenCVs cascade classifier to detect faces from the frames. While faces are being detected, a square facial image of the largest face is extracted from frame. Figure 4 shows an example detection of the facial detector[12-19].

In our work, this Haar cascade classifier is adopted for face detection both in our offline and real-time systems, and we ignore the non-frontal situation since the primary concentration is on the FER part. The input images are loaded and converted into grayscale mode. If the classifier finds the faces, it returns the four coordinates of the rectangular region of interest (ROI) of the faces. Once the locations of this ROI are obtained, these four vertices are used to crop the faces, and the irrelevant backgrounds are deleted; image processing can be addressed later. Examples of detected ROI for faces of the two datasets (CK+ and JAFFE) are shown in Figure 3.3. Concerning the FER-2013 dataset, the low resolution of its images (48_48 pixels) and the various head orientations prevent reliable face detection, so no face detection was applied to this dataset. Even without face detection, the performance is remarkably enhanced when FER-2013 is applied to the first stage of fine-tuning, as shown in below figure.

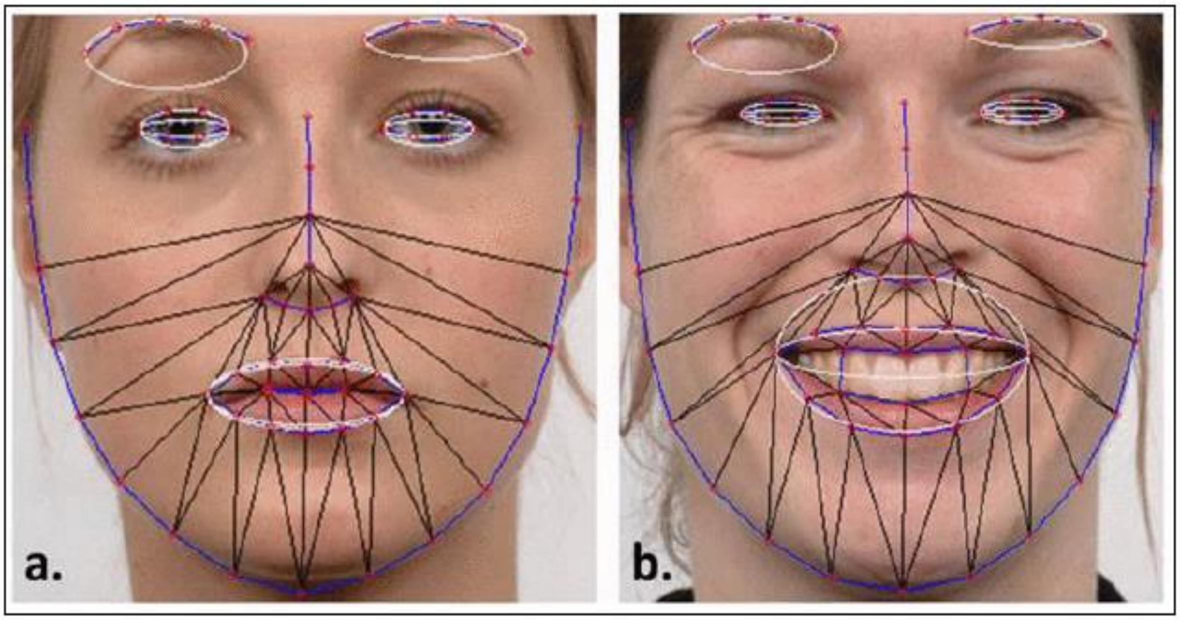

Figure 6. Face Detection Using Distance Metrics

Distance	Description of the distances
D1 and D2	Distance between the upper and lower eyelid of the right and left eyes
D3	Distance between the inner points of the left and right eyebrow
D4 and D5	Distance between the nose point and the inner point of the left and right eyebrow
D6 and D8	Distance between the nose point and the right and left mouth corner
D7 and D9	Distance between the nose point and the midpoint of the upper and lower lip
D10	Distance between the right and left mouth corner
D11	Distance between the midpoint of the upper and lower lip
D12	Mouth circumference

Table 1 Distance Metrics

FEATURE EXTRACTION

Facial Features extraction is an important step in face recognition and is defined as the Process of locating specific regions, points, landmarks, or curves/contours in a given 2-D image or a 3D range image. In this feature extraction step, a numerical feature vector is generated from the resulting registered image. Common features that can be extracted are-

1. Lips 2. Eyes 3. Eyebrows 4. Nose tip

DATA AUGMENTATION

Data augmentation is often employed during the training of the CNN since the process itself incorporates a large quantity of data. In the training scheme of this work, the cropped faces are first distorted with a lightweight library in Tensor Flow [20-27] before feeding them into the CNN.

Figure 7: Examples of face detection. Examples of detected ROIs for faces in the CK+ and JAFFE datasets.

RESULT AND DISCUSSION

Accuracy is the overall number of the correct predictions fractionated by the whole number of predictions created for a dataset. It can inform us immediately if a model is trained correctly and by which method it may perform in general. Nevertheless, it does not give detailed information concerning its application to the issue. Precision, called PPV, is a satisfactory measure to determination, whereas the false positives cost is high. Recall is the model metric used to select the best model when there is an elevated cost linked with false negative. Recall helps while the false negatives' cost is high. F1-score is required when you desire to seek symmetry between both precision and recall. It is a general measure of the accuracy of the model. It combines precision and recall. A good F1-score is explained by having low false positives and also low false negatives [28-33].

True Positives (TP):- These are the correctly predicted positive values, which mean that the value of the actual class is yes and the value of the predicted class is also yes.

True Negatives (TN):- These are the correctly predicted negative values, which means that the value of the actual class is no and value of the predicted class is also no.

False positives and false negatives, these values occur when our actual class contradicts with the predicted class.

False Positives (FP):- When actual class is no and predicted class is yes.

False Negatives (FN):- When actual class is yes but predicted class in no.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TN} + \text{FN}) + (\text{FP} + \text{TP})}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{FN} + \text{TP})}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{FP} + \text{TP})}$$

$$\text{F1} - \text{measure} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

Where,

- True positive (TP) = correctly identified
- False positive (FP) = incorrectly identified
- True negative (TN) = correctly rejected
- False negative (FN) = incorrectly rejected

Precision

Precision means to determine the number of positive class predictions that actually belong to the positive class.

Precision = TP/TP+FP

Recall

Recall means to determine the number of positive class predictions made out of all positive samples in the dataset.

Recall = TP/TP+FN

Fig.8 Prediction Result Screen Shot for Camera – Happy Face with Accuracy level

Fig.9 Prediction Result Screen Shot for Camera– Sad Face with Accuracy level

Fig 10.Prediction Result Screen Shot for Camera – Angry Face with Accuracy level

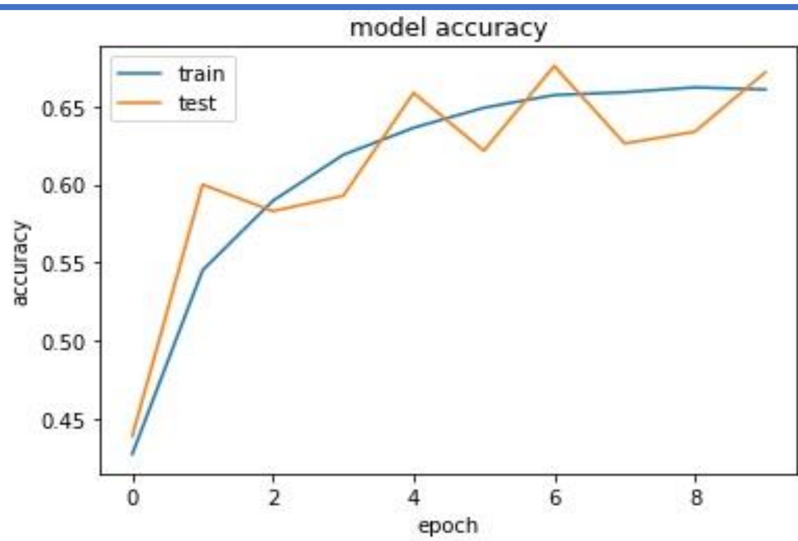

Fig.11 Plot for Predictive Model Accuracy

Conclusions:

Prediction of human facial expression using soft computing techniques are important in human communication and interactions. Also, they are used as an important tool in behavioral studies and in medical rehabilitation. Facial image detection techniques provide a fast and practical approach for non-invasive detection. The purpose to present this concept was to develop an intelligent system for facial image expression classification using neural networks.

Reference:

1. Solecki RS (1975) Shanidar IV, a Neanderthal flower burial in northern Iraq. Science 190:880–881.
2. Chan, K.-P., Zhan, X., and Wang, J. Facial expression recognition by correlated topic models and bayes modeling. In Image and Vision Computing New Zealand (IVCNZ), 2015 International Conference on (2016), IEEE, pp. 1–5.
3. Cheng, F., Yu, J., and Xiong, H. Facial expression recognition in jaffe dataset based on gaussian process classification. IEEE Transactions on Neural Networks 21, 10 (2010), 1685–1690.
4. Coelho, G., Kougianos, E., Mohanty, S. P., Sundaravadivel, P.,and Albalawi, U. An iot-enabled modular quadrotor architecture for real-time aerial object tracking. In Nanoelectronic and Information Systems (iNIS), 2015 IEEE International Symposium on (2015), IEEE, pp. 197–202.
5. Ekman, P., and Friesen, W. V. Manual for the facial action coding system. Consulting Psychologists Press, 1978.

6. Gavrilescu, M. Proposed architecture of a fully integrated modular neural network-based automatic facial emotion recognition system based on facial action coding system. In Communications (COMM), 2014 10th International Conference on (2014), IEEE.
7. Guo, Y., Tao, D., Yu, J., Xiong, H., Li, Y., and Tao, D. Deep neural networks with relativity learning for facial expression recognition. In Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on (2016), IEEE, pp. 1–6.
8. Punarselvam, E., and P. Suresh. "Non-Linear Filtering Technique Used for Testing the Human Lumbar Spine FEA Model." Journal of medical systems 43, no. 2 (2019): 1-13.
9. Punarselvam, E., Hemalatha, E., Dhivahar, J., Gowtham, V., Hari, V., & Thamaraikannan, R. PREDICTING WIRELESS CHANNELS FOR ULTRA-RELIABLE LOW-LATENCY COMMUNICATIONS.
10. Punarselvam, E., et al. "Segmentation of Lumbar spine image using Watershed Algorithm." International Journal of Engineering Research and Applications, ISSN (2013): 2248-9622.
11. Burkert, P., Trier, F., Afzal, M. Z., Dengel, A., and Liwicki, M. Dexpression: Deep convolutional neural network for expression recognition. arXiv preprint arXiv:1509.05371 (2015).
12. Chakraborty, A., Konar, A., Chakraborty, U. K., and Chatterjee, A. Emotion recognition from facial expressions and its control using fuzzy logic. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 39, 4 (2009), 726–743.
13. Punarselvam, E., et al. "Different loading condition and angle measurement of human lumbar spine MRI image using ANSYS." Journal of Ambient Intelligence and Humanized Computing 12.5
14. Punarselvam, E., and P. Suresh. "Edge detection of CT scan spine disc image using canny edge detection algorithm based on magnitude and edge length." 3rd International Conference on Trendz in Information Sciences & Computing (TISC2011). IEEE, 2011. (2021): 4991-5004.
15. Punarselvam, Dr E., and S. Gopi. "Effective and Efficient Traffic Scrutiny in Sweet Server with Data Privacy." International Journal on Applications in Information and Communication Engineering 5.2 (2019): 1-5
16. Punarselvam, E., and P. Suresh. "Investigation on human lumbar spine MRI image using finite element method and soft computing techniques." Cluster Computing 22, no. 6 (2019): 13591-13607.
17. Karthick, R., and M. Sundararajan. “A Reconfigurable Method for Time Correlated Mimo Channels with a Decision Feedback Receiver” International Journal of Applied Engineering Research 12.15 (2017): 5234-5241.
18. Karthick, R., and M. Sundararajan. “A novel 3-D-IC test architecture-a review” International Journal of Engineering and Technology (UAE) 7.1.1 (2018): 582-586.
19. Karthick, R., and M. Sundararajan.,”PSO based out-of-order (ooo) execution scheme for HT-MPSOC” Journal of Advanced Research in Dynamical and Control Systems 9 (2017): 1969.
20. Punarselvam, Dr E., and S. Gopi. "Effective and Efficient Traffic Scrutiny in Sweet Server with Data Privacy." International Journal on Applications in Information and Communication Engineering 5.2 (2019): 1-5.

21. Hemalatha, E., Dhamodaran, M. and Punarselvam, E., 2019. Robust Data Collection with Multiple Sink Zone in 3-D Underwater Sensor Networks. International Journal on Applications in Basic and Applied Sciences, 5(1), pp.8-14.
22. Punarselvam, E., Suresh, P., & Parthasarathy, R. (2013). Segmentation of CT scan lumbar spine image using median filter and canny edge detection algorithm. Int J Comput Sci Eng, 5, 806-814.
23. GOPI, S., and E. PUNARSELVAM. "HEART ATTACK RECOGNITION AND HEART RATE MONITORING SYSTEM USING IOT."
24. Punarselvam, E., et al. "Segmentation Analysis Techniques and Identifying Stress Ratio of Human Lumbar Spine Using ANSYS." Journal of Medical Imaging and Health Informatics 10.10 (2020): 2308-2315.
25. Karthick, R., et al. “Overcome the challenges in bio-medical instruments using IOT–A review.”; Materials Today: Proceedings 45 (2021): 1614-1619.
26. Suresh, Helina Rajini, et al. “Suppression of four wave mixing effect in DWDM system.” Materials Today: Proceedings 45 (2021): 2707-2712.
27. Soundari, D. V., et al.”Enhancing network-on-chip performance by 32-bit RISC processor based on power and area efficiency.” Materials Today: Proceedings 45 (2021): 2713-2720.
28. Sabarish, P., et al., Investigation on performance of solar photovoltaic fed hybrid semi impedance source converters." Materials Today: Proceedings 45 (2021): 1597-1602.
29. Karthick, R., and M. Sundararajan. “SPIDER-based out-of-order execution scheme for Ht-MPSOC “; International Journal of Advanced Intelligence paradigms 19.1 (2021): 28-41.
30. Karthick, R., and P. Meenalochini. “Implementation of data cache block (DCB) in shared processor using field-programmable gate array (FPGA).” Journal of the National Science Foundation of Sri Lanka 48.4 (2020).
31. Sabarish, P., et al., “An Energy Efficient Microwave Based Wireless Solar Power Transmission System” IOP Conference Series: Materials Science and Engineering. Vol. 937.No. 1. IOP Publishing, 2020.
32. Vijayalakshmi, S., et al. ”Implementation of a new Bi-Directional Switch multilevel Inverter for the reduction of harmonics” IOP Conference Series: Materials Science and Engineering. Vol.937. No. 1. IOP Publishing, 2020.
33. Karthick, R., and M. Sundararajan. “Design and implementation of low power testing using advanced razor based processor” International Journal of Applied Engineering Research 12.17 (2017): 6384-6390.