Discriminating Real from Fake Smile Using Convolution Neural Network

Rajesh Kumar G A[1], Ravi Kant Kumar[2], Goutam Sanyal[3]
Department of Computer Science and Engineering
National Institute of Technology
Durgapur, India
E-mail: {rajuloki046, vit.ravikant, nitgsanyal}@gmail.com

Abstract: **In our society, sometime we hide our genuine feeling and emotion and purposely express different emotion in front of our surrounding folks. But as it's not actually a natural emotion, hence, it is more or less, predictable by others. Human vision system has enormous capability to recognizing genuine and fake smile of an individual. Discriminating genuine and fake smile is very thought-provoking task and even though very smaller amount of research has been carried out in this topic. In this paper, we are exploring a method to distinguish real from fake smile with high precision by using convolution neural networks (CNN). System has been train with FERC-2013 dataset having seven types of emotions namely happy, sad, disgust, angry, fearful, surprised and neutral. Emotions percentages of real and fake face are recorded by the emotion detection system. Based on recorded score, we investigate the effect of various percentages of emotions presented on both faces and then we are going to classify the smile on the face is real or fake.**

***Keywords*— Facial expressions, Facial Emotions, Non-Verbal Communication, Face Detection, Convolution Neural Network (CNN), Deep Learning, real-fake smile.**

I. INTRODUCTION

An actual truthful smile is a actually convincing reflection of happiness. People like to poster such confident smiles for interacting and trusted mutual communication. However, an 'Unreal' or "Fake" smile reflects less self-confidence in such tasks, and even at the same time such facial reply leave a doubtful vision on them.

Recognizing genuine and fake expression, seem on human face is one of the hardest job for once brain. Humans vision system, have a remarkable capacity to recognize genuine and fake smile of an individual. Even though, countless times our brain is also not talented enough to distinguish it clearly. But how computer vision system can differentiate between genuine and fake emotions? There is no appropriate reply for such questions, till date. Still, in order to discover solutions of such difficult problems at some extent, quite a few computational techniques have been demonstrated. To make these things understandable, primary challenge has been reserved by a well-known French physician named Guillaume Duchenne, from the 19th century to Distinguish genuine and fake smile founded on the muscles that are involved in generating facial expressions[1]. In [2], writers claim about eyes as an evidence for the finding of real and fake smiles. In order to find out social impact of truthful smiles a research has been conducted in [3]. In this research, examination discovered that; associated to involved and control members, excluded folks exhibited a better preference to work with folks displaying "actual" as opposed to "fake" smiles [3]. Some scientist proposed how the adaptive responses to social exclusion work and social refusal improves the finding of genuine and fake smiles [4]. Eye movements based real and fake smiles judgment has been studied in [5].

In associate of exact emotion investigation (Like genuine and fake smile), in the zone of generic emotion recognition, more exploration has been accomplished. Some of those are as: emotions recognition in people with amygdale damaged [7], geometric feature-based approach and holistic template matching [8], Local binary patterns (LBP) grounded emotion classification [6, 10], Emotion recognition by means of Hidden Markov Model (HMM) [11], Several hybrid methods has also been explored as; facial emotion classification by means of NN and HMM [12], Emotion calculation with joint visual and audio cues [13], Emotion classification after combining multiple kernel methods [14], Facial study with convolution neural networks(CNN) [15] etc. In this research work we have used convolution neural network for discriminating real and fake smile. First, we are going to compute percentages of emotions on specified face and then based on percentages analysis we classify smile is genuine or fake.

Further, the paper is organized as: In section II, gives complete system architecture ,The data set description described in section III. We explained Image pre-processing algorithm and step by step working process of convolution neural network algorithm in section VI and section V gives complete results and validation information. Finally, section VI describes the concluding remarks.

II. SYSTEM ARCHITECTURE

Overall proposed system architecture has been depicted as below. We divided the algorithm in to two parts training and testing. Before testing, we need to train the system to detect the emotions of given face. First, we check whether the trained data is available or not. If trained data are not available, we will train the system before the testing process. If trained database is available then we can use the system for testing and further steps of this algorithm are explained in further sections.

Algorithm: *complete project flowchart*

```
Step 1: if (trained database is not presented)
Step 2:         run Algorithm1
Step 3:         run Algorithm 2
Step 4:         save trained database
Step 5: else (load trained database)
Step 6:         Get input image from webcam or system folder
Step 7:         run Algorithm1
Step 8:         run Algorithm2
Step 9:(result 1) display the emotions with percentage of each emotion.
Step10:(result2) Analyses of real and fake emotions.
```

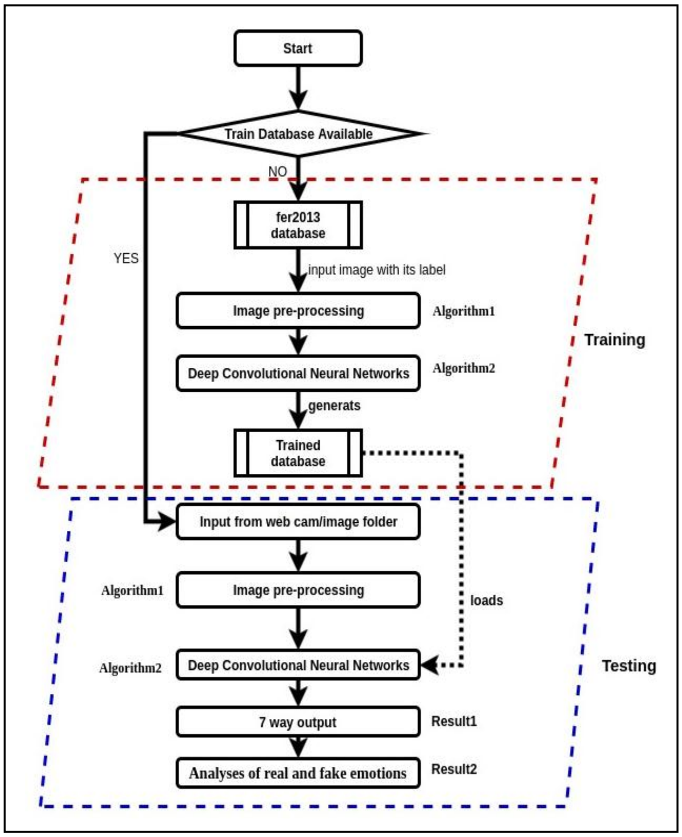


Figure 1. Complete Project Flowchart

III. DATABASE DESCRIPTION

We used (FERC-2013) database[19] for training and testing of a system because the dataset holds around 2600 happy images. Therefore, database gives more accurate results then other database. Total images in database are around 32000 low resolution images and exhibits emotions in the great variety and, sometimes emotions are not and therefore, harder to interpret (image distribution shown in Figure 10). FERC-2013 is a massive size which contains various range of emotions. So, varieties of range can be appreciated for the robustness of system. So, as soon as our experiment is trained with FERC-2013 dataset, in the testing phase, from datasets having clear emotions, can be effortlessly classified, but not vice versa. Henceforth, the networks are trained by means of FERC-2013 dataset in this experiment.

The dataset holds greyscale images of faces having dimension of [48x48] pixel . The faces are atomically resized, so that they occupied same space in each image. The foremost purpose is to classify each face based on the exposed emotion on the faces into one of seven classes (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The trained file includes two columns, "emotion" and "pixels". The "emotion" column comprises a numbers ranging from 0 to 6, both inclusive, for the emotion. The "pixels" column holds a string of pixel array of every face. The contents of this string is a space-separated pixel values in row major order. Test file contains only the "pixels" column and our job is to guess the emotion column. so initially we converted the trained file to two files one comprises numpy array of pixels and alternative file contains emotions vector of size[1x7].The training set consists of 28,709 samples. The public test set used for the leader board consists of 3,589 samples. The concluding test-set, which was used to determine the champion of the race, contains of another 3,589 samples.

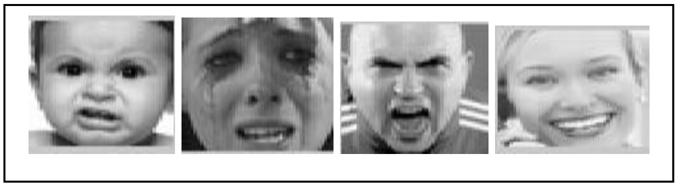


Figure 2. Some Valid Samples of FERC-2013 Database

Before training we pre-processed (described in section IV-A) the FERC-2013 database images. In the pre-processing, we used the Viola-Jones algorithm [13], [14] on the dataset in which out of 28,709 samples for pre-processing and validation, we got 11246 valid samples for training. Images in which face has been detected by Viola Jones are considered as the valid samples.

Due to drawback of Viola-Jones algorithm [13], [14] many samples got fail during validation. Some failed images are shown below:

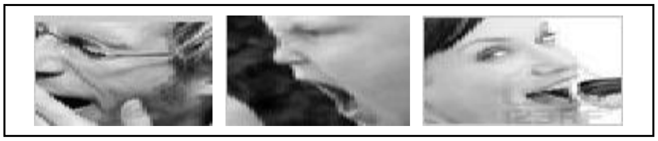


Figure 3. Some Failure Images to Detect Faces using [13,14]

IV. PROPOSED ALGORITHMS

A. *IMAGE PRE-PROCESSING*

Algorithm 1: *Image pre-processing*

```
step1: Input from webcam frame or selected image.
step2: Face-detection using Viola Jones algorithm [5].
step3: Maximum area faces detection among available faces.
step4: Crop the face from image.
step5: Resize the cropped face into to 48x48 images
```

As shown in above algorithm we first get the image as input and face detection is the primary and most important step in any emotion recognition system for that we used Viola-Jones algorithm [13], [14], which is most successful frontal face recognition existing algorithm. However, it has a few disadvantages which cannot be accounted for in other types of face pictures. The face indicator is most effective only on frontal faces images. It cannot manage with more than 45 degree of face rotation both about the horizontal and vertical axis. The algorithm is pretty sensitive to lighting circumstances. After face detection among all the faces we are going to get max area face and then crop that face and resize the face into the fixed size i.e. [48x48] for further processing.

Image pre-processing algorithm flow chart shown below

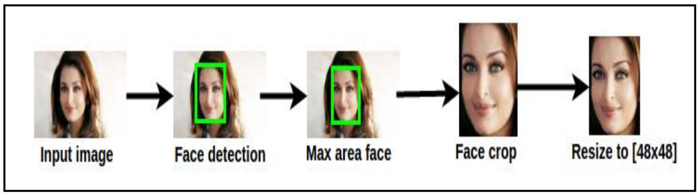


Figure 4. Image Pre-Processing

C. CONVOLUTIONAL NEURAL NETWORKS

In recent times, convolutional neural networks (CNN)[15],[16] have confirmed lifting performance in abundant computer vision tasks. However, vast performance hardware is obviously essential for the usage of CNN models due to the colossal computation complexity, which prohibits their further extensions. Our principal objective is to classify CNN architectures that have erratic parameters whereas upholding competitive accuracy. To attain this, we employ nine foremost layers while designing CNN architecture. And CNNs were announced first to identify handwritten digits [17] in the early 90s, but a main revolution was achieved 2012 with the publication of the AlexNet [18]. The basic principle can be understood as a superior case of the multilayer perceptron (MLP) where each neuron is individually linked to a receptive field in front of it. Furthermore, all neurons of a specific layer share the equal weights. The weighted input of a neuron with N inputs formerly applying the activation function is:

$$v = \sum_{i=1}^{N} a_i w_i$$

where a, w signifies the input from the preceding layer and weights respectively

Algorithm 2: Deep convolutional neural network

Step by step description of Convolutional networks

Step 1: First we set entirely weights and filters with arbitrary values.
Step 2: Training image is input to the network, then goes to forward propagation phases (Convolution layer, ReLU layer and pooling layer which goes through forward propagation in the fully connected layer) and it gives output probabilities of each class. Let’s assume the output probabilities are given as [0, 0.3, 0.1, 0.2, 0.4, 0, and 0]. With the random weight, obtained output probabilities are also random values.
Step 3: Total error is calculated at the output layer and is given as Total Error = ∑ (target probability – output probability) ^{2})
Step 4: To compute the gradients of the error we use backward run as for all weights in the system and utilize gradient drop and refresh all channel weights and parameter values to limit the total output error. The weights are balanced in extent to their commitment to the total error. At the point when a similar picture is input once more, output probabilities may now be [0, 0.2, 0.7, 0.1, 0.2, 0, 0] which is nearer to the objective vector [0, 0, 1, 0, 0, 0, 0]. This implies the system has learnt to group this specific picture accurately by conforming its weights/channels to the end goal that the output error is decreased. Parameters like number of channels, channel sizes, engineering of the system and so forth have all been settled before Step 1 and don't change during training process – just the estimations of the training network and association weights get refreshed during the process.

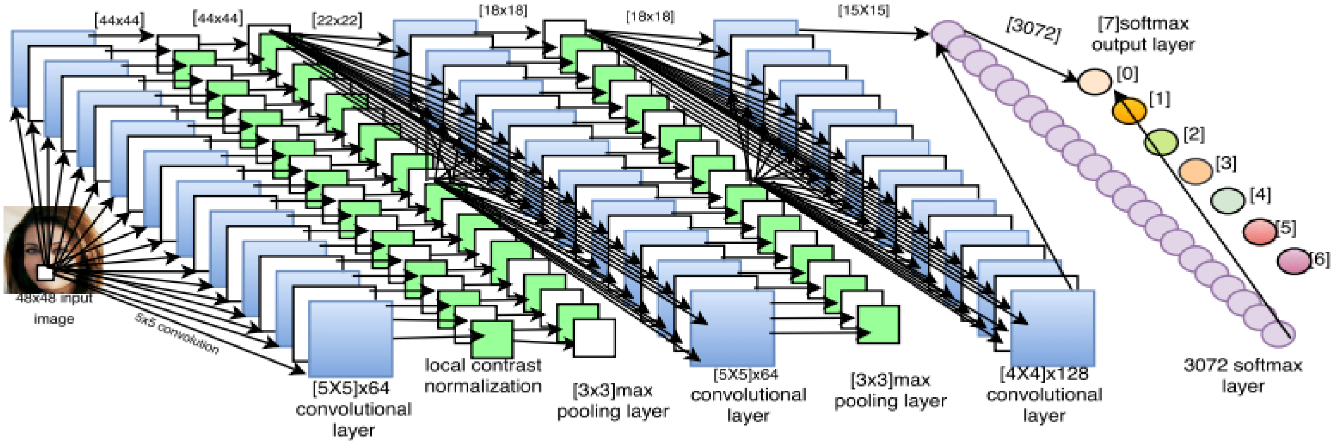


Figure 5. Architecture of Deep Convolutional Neural Network

Layer by Layer Explanation of CNN

- *Layer 0*: *Input layer* Input[48x48x1] will hold the raw pixel values of the image, in this case a face image of width 48, height 48, and with one color channel is considered.
- Layer 1: *Convolutional layer* calculates the output of neurons which are associated to native regions in the input, each calculating a dot product among their weights and a small region they are connected to in the input volume. This may yield result in volume such as [44x44x64] if we decided to use 64 filters. with 64 filters of size 5*5, stride 1, padding 0 , Total Size: [44 x 44 x 64], and (48-5)/1 + 1 = 44 is the size of the outcome 64 depths because 1 set denotes 1 filter and there are 64 filters.
- Layer 2: *RELU layer* will be applied elementwise activation function, such as the *max* (0, *x*) zero. This made the size of the volume unaffected ([44x44x64]),and Batch normalization is done.
- Layer 3: *POOL layer* will achieve a down sampling process along the spatial sizes (height, width), resultant is volume such as [22x22x64]. Max-Pooling with 3×3 filter, stride 2, There four size is [22x22x64], i.e. (44-3)/2+1=22 is output size, depth is same as before, i.e. 64 because pooling is done independently on each layer.
- Layer 4: Convolution with 64 filters, size 5×5, stride 1, now size is [18x18x64],i.e. (22-5)/1+1=18 is size of output 64 depths because of 64 filters.
- Layer 5: Max Poling Layer with 64 filters, size 5×5, stride 1,now size is [18x18x64],i.e. (18+2*1-3)+1=18 original size is restored..
- Layer 6: Convolution with 128 filters, size 4x4, stride 1, padding 0,now size is [15x15x128],i.e. (18-4)/1+1=15 is size of output 64 and depths of 128 filters.
- Layer 7: Fully Connected with 3072 neurons in this later, each of the 15x15x128=28800 pixels are fed into each of the 3072 neurons and weights determined by back-propagation.
- Layer 8: *Fully-connected layer* will compute the class scores, resultant capacity of size [1x1x7], where each of the 7 numbers match to a class scores, such as among the 7 classes of emotions. As with regular Neural Networks and as the name implies, each neuron in this layer will be connected to all the numbers in the previous volume and Soft max layer with 3072 neurons.
- Layer 9: soft max layer with 7 neurons to predict 7 classes output.

V. RESULTS

i. Success fully detected emotions

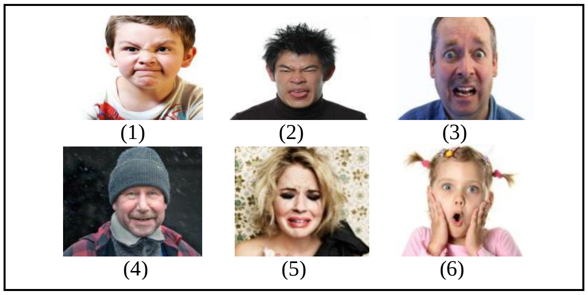


Figure 6. (1) is angry face, (2) is disgusted face, (3) is Fearful face, (4) is happy face, (5) is sad face, (6) is surprised face.

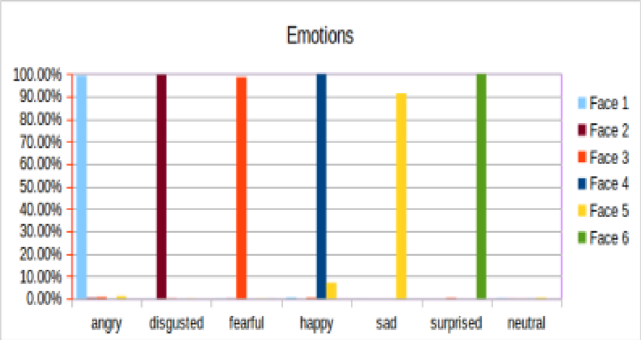


Figure 7. Emotion percentages of successfully detected faces

ii. *Some failure test cases*

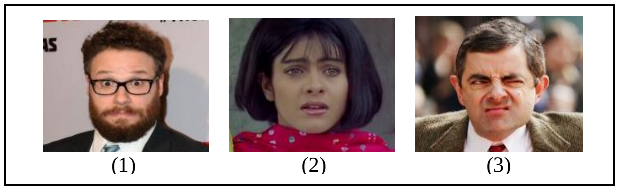


Figure 8. (1) surprised detected as neutral, (2) sad detected as angry, (3) disgusted detected as angry

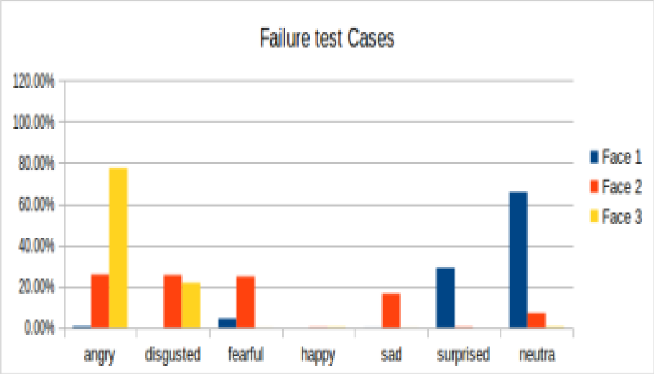


Figure 9. Emotion percentages of above failure test images

The above failures may be due to the dataset imbalance, the (FERC-2013) data set contains non-uniform number of images to different emotions in training set is shown in figure 10. Among 28,709 samples after pre-processing and validation among them we got 11246 valid samples for training Due to drawback of Viola-Jones algorithm [5], [6] most the samples fail during validation.

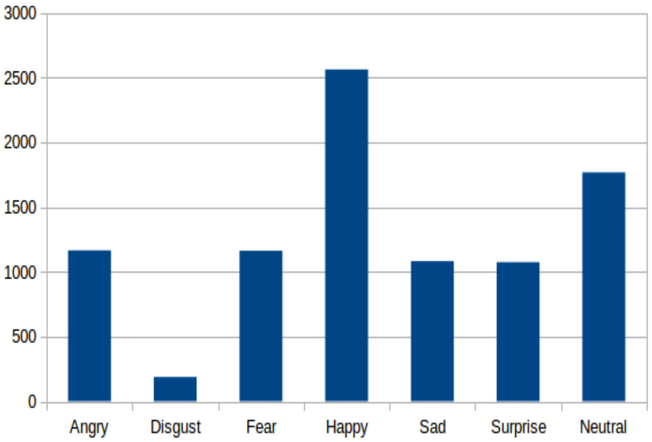


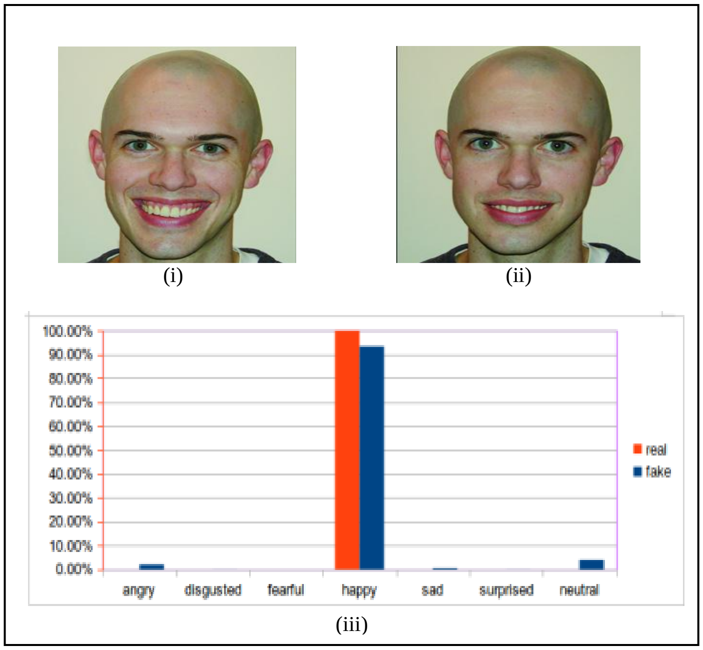
Figure 10. number of sample images for each emotion in FERC-2013 database

iii. *Discriminating real and fake smiles*

It's very tough duty for human to discriminate genuine and fake smile but some of exports discriminate genuine and fake smiles observing at some facial muscles and dissimilarities in them. All smiles need that we flex muscles around the mouth, but the difference is the way we involve the muscles around our eyes, called the orbicularis oculi. In a genuine smile, we contract those muscles, pulling in the skin next to our eyes. Teeth or no teeth, doesn't he look genuinely happy to see you (and not at all creepy)? Look at the contraction of the muscles around his eyes. That only happens with smiles that reflect true, happy emotions. On the other hand, a fake smile doesn't use those muscles. When forcing a smile, we use a muscle in each cheek, called the risorius, to pull our lips into the right shape, but the eye muscles don't contract. To demonstrate this, Duchenne electrically stimulated the risorius muscles of his tooth-less friend. There are creases on his cheeks but not around his eyes. The orbicularis oculi muscles are not contracted. The skin around the eyes is not pulled in tightly as it is in the first picture. That is the mark of a fake smile. The differences in muscle contraction in genuine versus fake smiles illustrate the separation between the habit and the non-habit systems in the brain. When a smile comes naturally to us, one set of muscles is activated. When we use our conscious powers to feign a smile, we alter the pattern of muscle activation, and people around us can tell.

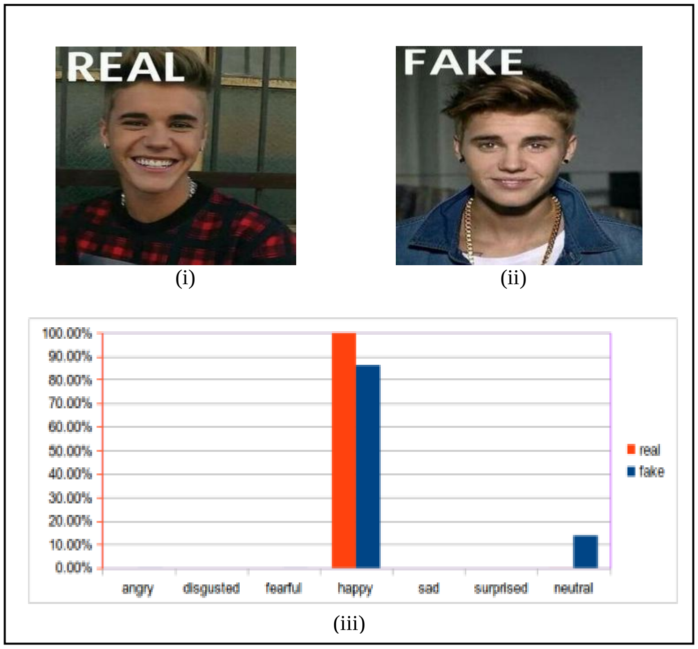
We are giving a system which can distinguish real and fake smile based on their variations in percentage of emotions, we are getting an encouraging accuracy in detection of genuine and fake smiles, the main difficult in this processing is there is no databases are available for this real and fake emotions discriminating, we took some available images from the open sources for testing and we are getting comparatively good results. And these results may helpful in detection systems, human understandability and customer satisfaction feedbacks for social media sites.

Some of results for genuine and fake smile differentiation are shown below


(i) (ii)

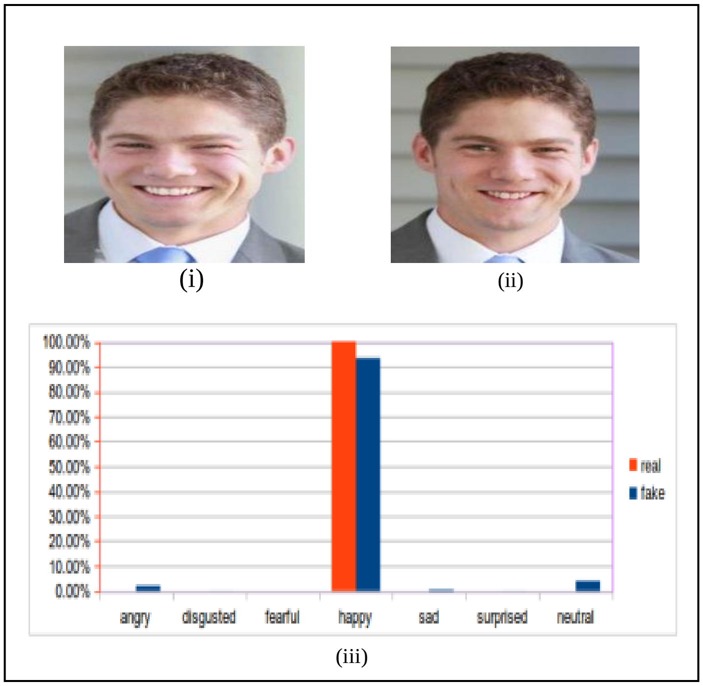
(iii)

Figure 11. (i)genuine smile face (ii) is fake smile face and (iii) graphical representation of emotion percentages on both faces.



(i) (ii)
(iii)

Figure 12. (i)genuine smile face (ii) is fake smile face and (iii) graphical representation of emotion percentages on both faces.



(i)

(ii)

(iii)

Figure 13. (i)genuine smile face (ii) is fake smile face and (iii) graphical representation of emotion percentages on both faces.

VI. CONCLUSION

For different reasons, we smile a lot to hide our discomfort, to react to pain or grief or disgust, or sometimes to show that we're sad. There's only one type of smile that's used to convey happiness i.e. genuine smile, A genuine happy smile is characteristically one that encompasses not just the eyes, but the skin around the eyes and the formation of crow's feet. When someone's giving you a fake smile, they often concentrate too much on what their mouth is doing, and you'll be able to see more teeth than you would during a real smile. Our experiment gives the best results till now and even we can increase the accuracy if we can train the networks with the real and fake database.

VII. REFERENCES

[1] Bulletin of the Transilvania University of Braşov • Vol. 6 (51) - 2009 Series 6: Medical Sciences Supplement – Proceeding of The IVth Balkan Congress of History of Medicine

[2] Mai, Xiaoqin, et al. "Eyes are windows to the Chinese soul: Evidence from the detection of real and fake smiles." *PloS one* 6.5 (2011): e19903.

[3] Bernstein, Michael J., et al. "A preference for genuine smiles following social exclusion." *Journal of Experimental Social Psychology* 46.1 (2010): 196-199.

[4] Bernstein, Michael J., et al. "Adaptive responses to social exclusion: Social rejection improves detection of real and fake smiles." *Psychological Science* 19.10 (2008): 981-983.

[5] Calvo, Manuel G., et al. "Attentional mechanisms in judging genuine and fake smiles: Eye-movement patterns." *Emotion* 13.4 (2013): 792.

[6] Adolphs, Ralph, et al. "Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala." *Nature* 372.6507 (1994): 669.

[7] Russell, James A. "Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies." *Psychological bulletin* 115.1 (1994): 102.

[8] Michel, Philipp, and Rana El Kaliouby. "Real time facial expression recognition in video using support vector machines." *Proceedings of the 5th international conference on Multimodal interfaces*. ACM, 2003.

[9] Shan, Caifeng, Shaogang Gong, and Peter W. McOwan. "Facial expression recognition based on local binary patterns: A comprehensive study." *Image and Vision Computing* 27.6 (2009): 803-816.

[10] Bartlett, Marian Stewart, et al. "Recognizing facial expression: machine learning and application to spontaneous behavior." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE, 2005.

[11] Matsugu, Masakazu, et al. "Subject independent facial expression recognition with robust face detection using a convolutional neural network." *Neural Networks* 16.5 (2003): 555-559.

[12] Yu, Zhiding, and Cha Zhang. "Image based static facial expression recognition with multiple deep network learning." *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 2015.

[13] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE, 2001.

[14] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." *International journal of computer vision* 57.2 (2004): 137-154.

[15] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

[16] LeCun, Yann, and Yoshua Bengio. "Convolutional networks for images, speech, and time series." *The handbook of brain theory and neural networks* 3361.10 (1995): 1995.

[17] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In Advances in neural information processing systems. Citeseer, 1990

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[19] FERC 2013, Form 714 – Annual Electric Balancing Authority Area and Planning Area Report (Part 3 Schedule 2). 2006–2012 Form 714 Database,Federal Energy Regulatory Commission (2013)