iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.34900

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Real Time Facial Gesture Recognition using Deep Learning

Abaikesh Sharma[1], Jasbir Singh[2]
[1, 2]*University of Jammu, Department of Computer Science & IT, Jammu, J&K, India*

Abstract: ***The human faces have vibrant frequency of characteristics, which makes it difficult to analyze the facial expression. Automated real time emotions recognition with the help of facial expressions is a work in computer vision. This environment is an important and interesting tool between the humans and computers. In this investigation an environment is created which is capable of analyzing the person's emotions using the real time facial gestures with the help of Deep Neural Network. It can detect the facial expression from any image either real or animated after facial extraction (muscle position, eye expression and lips position). This system is setup to classify images of human faces into seven discrete emotion categories using Convolutional Neural Networks (CNNs). This type of environment is important for social interaction.***
Keywords: Convolutional Neural Network, Facial Gestures, Animated Face, Deep Neural Network, Emotion recognition

I. INTRODUCTION

From the beginning of this technology oriented world, when computers were first introduced, scientists and engineers found a need for artificially intelligent systems that can compete to human abilities up to some extent. And by time, they became successful in producing various software that helped in development of fast learning machines. Adding on to it, the internet provided enormous amounts of data for training. Both these developments facilitated increment in research on self-learning systems consisting of neutral networks providing fast learning techniques.

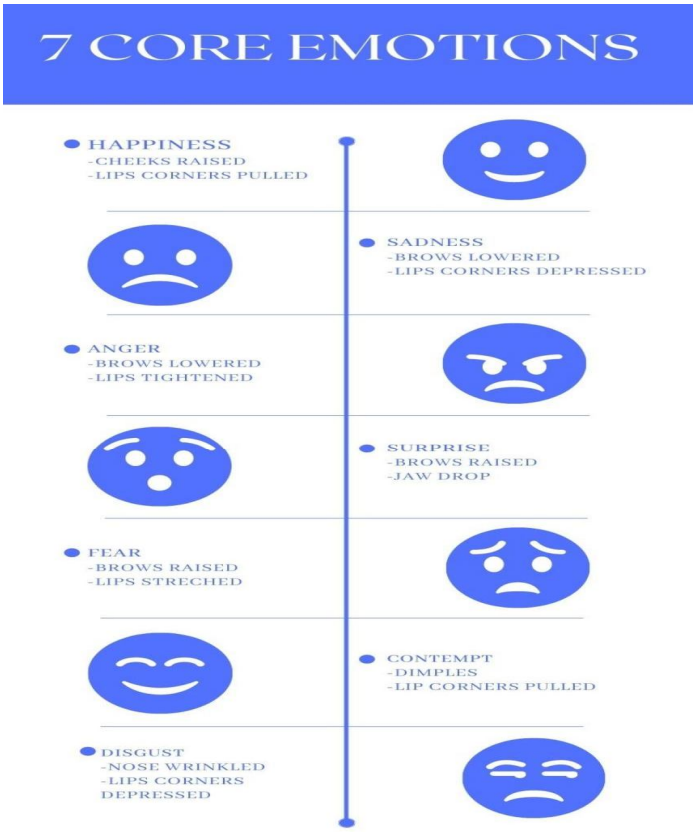


Fig. 1 Seven emotions and expressions

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

The best way humans can communicate with each other is through speech, emotions and gestures. The co-relation between emotions and the gestures is very important and useful. The environment which can recognize it can be used in many different field.**Error! Reference source not found.**

Fig. 1 shows this relationship and the different 7 emotions of human. A Model can be used to make the interaction more reliable between the system and the users.

Humans physiological states are perfectly related to emotions. Usually, emotions are autonomous bodies responses to certain internal and external events. Emotions came up with the thought process likewise, Fig. 2 shows the build of an expression.

The face recognition in photos and videos is one of the loftiest applications of artificial intelligence. Visual data is most prominently used to search and generate general patterns present in human faces. Face recognition is also well acclaimed in surveillance purposes by law enforcers as well as in crowd management. Also, automatic blurring of faces on Google Street view footage and automatic recognition of Facebook friends in photos are widely used in today's application. The advancements in this technology has also included emotion recognition in a meticulous manner. On the flip side, the most promising applications involve the humanization of artificial intelligent systems. If computers are able to keep track of the mental state of the user, robots can react upon this and behave appropriately. Therefore, Emotion recognition plays a vital role in improving human-machine interaction.

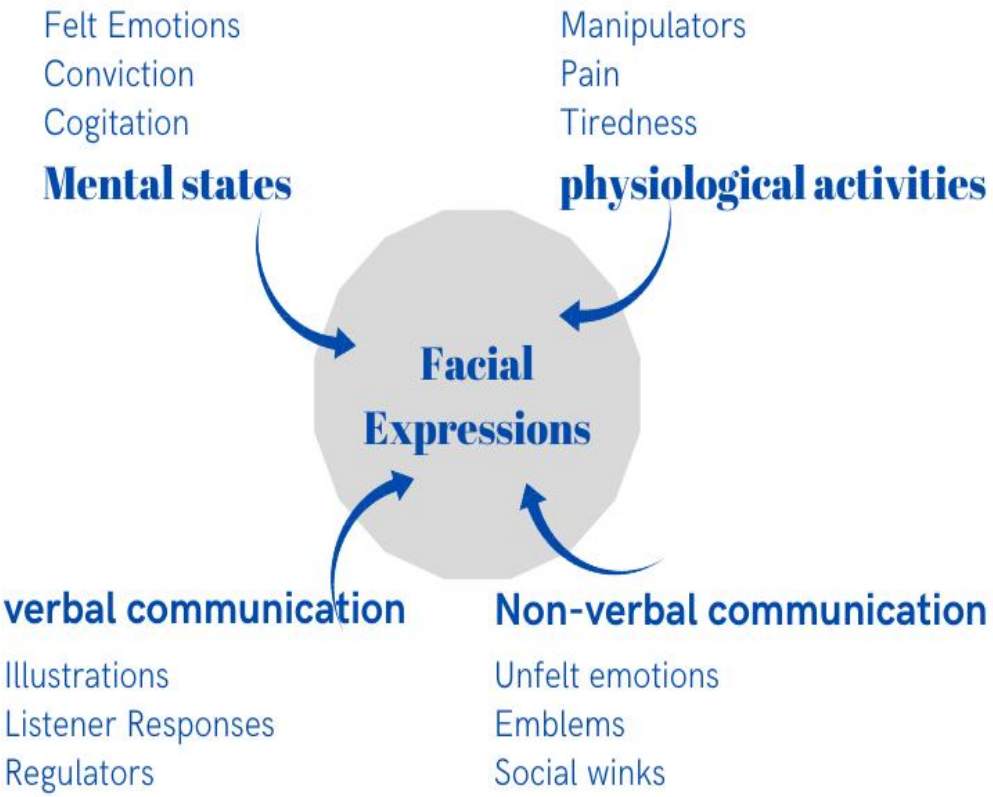


Fig. 2 Facial expression structure

The environment can be used in different fields, where the emotions of the users are the fundamental way of operating like:

1) *Entertainment World:* Difficulty of the game for the user or gamer changes with the emotion he/she is carrying.
2) *Feedback Mechanism:* System will make it easy to take the feedback from the users if one is not having the same native language, here the emotions can help.
3) *Cameras:* Will create a conjunction with the technology of the cameras to take photos only when user smiles.

©IJRASET: All Rights are Reserved

There are limitations over the working conditions i.e. problem related to data set. Most of data set is made of images which were taken in very stable environment which is not universally suitable for the system to work. Also, sometime the situation may arise to determine the emotion of the image or real time human with half of his face present which is difficult to perform.

II. LITERATURE REVIEW

For developing a system enabled to recognize emotions through facial expressions, thorough research, specifically on the way humans reveal emotions along with theory of automatic image categorization, is necessary.The Table I briefs about the existing work over the same environment:

Table I

AUTHOR	TITLE	JOURNAL AND PUBLICATION YEAR	SUMMARY
A. Gudi.	Recognizing semantic features in faces using deep learning.	*arXiv preprint arXiv:1512.00743*, 2015.	Pros: explores the effectiveness of the system to recognize various semantic features present in any face. Cons: semantics are not easily defined for special landmarks on the faces and for the contraction of specific face muscles.
J. Nicholson, K. Takahashi, and R. Nakatsu.	Emotion recognition in speech using neural network.	*Neural computing & applications*, 9(4): 290–296, 2000.	Pros:emotion recognition for speech could serve as a kind of “emotional translator”.
P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews.	The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression.	*IEEE Computer Society Conference on*, pages 94–101. IEEE, 2010.	Pros: provided an extended version of CK dataset which overcome the limitations of the previous.
Y. Lv, Z. Feng, and C. Xu	Facial expression recognition via deep learning	In *Smart Computing (SMARTCOMP), 2014 International Conference on*, pages 303–308. IEEE, 2014.	Pros: a comprehensive study on deep FER which includes the datset and the algorithm to use. Cons: the prototypical expressions covers only a small portion of specific categories and cannot capture the full repertoire of expressive behaviors for realistic interactions.
Shashak, jaiswal, michael,	Deep learning the dynamic appearance and	"Deep learning the dynamic appearance and shape of facial action units," *2016 IEEE Winter Conference on Applications of*	Pros : approach to detect the facial action unit detection with the combination of convolutional and bi-

©IJRASET: All Rights are Reserved

vlalstar	shape of facial action units	*Computer Vision (WACV)*, 2016, pp. 1-8, doi: 10.1109/WACV.2016.747762	directional long short term memory neural network.

Here we have timeline chart for the same in Fig. 3:

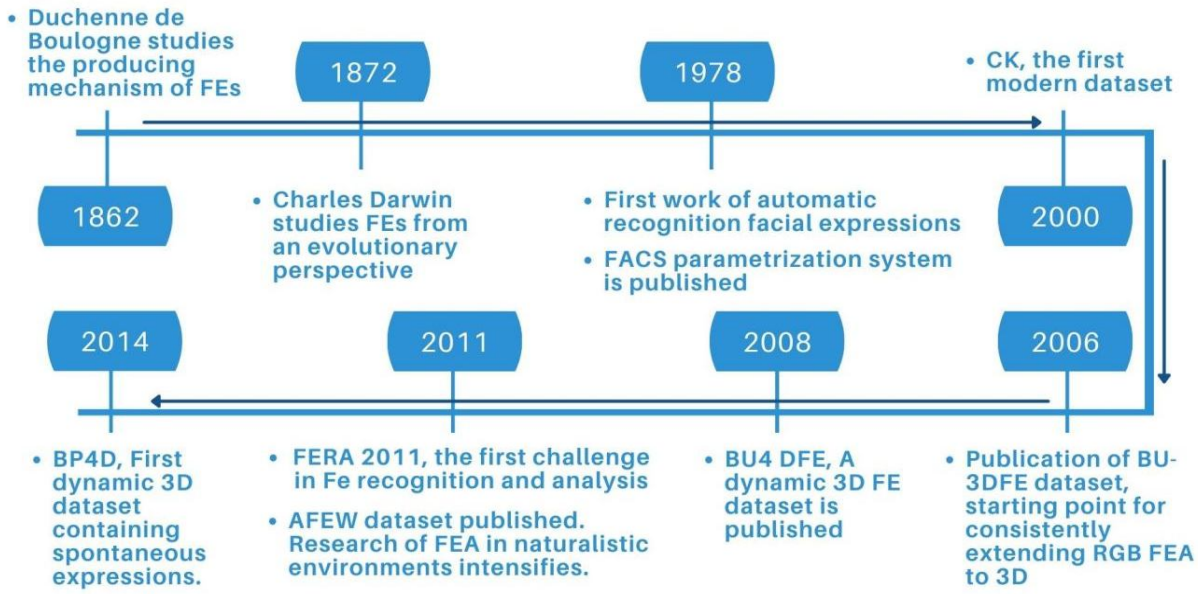


Fig. 3 Timeline Chart

III.METHODOLOGY

The work is mainly focused on neural network based artificially intelligent systems that are capable of generating the emotion of the person through their faces.

A. Network

The use of the TFLearn library on top of TensorFlow is done to program the networks, running on Python. This minimizes the complexity of code, as just the neuron layers have to be created and not every neuron. The real-time feedback on training progress and accuracy is provided in the program, which makes it easy to save and reuse the model post training. The reference provides for more details [6].

1) The firstly tested network is based on the research by Krizhevsky and Hinton [2]. This one is the smallest of the three networks, resulting in lowest computational demands. Fast working algorithms are beneficial as one of the future applications might be in the form of live emotion recognition in embedded systems. The network has three convolutional layers and two fully connected layers, combined with max pooling layers to reduce the image size and a dropout layer to eradicate the chance of overfitting. The hyper parameters are chosen in a way that the number of calculations in each convolutional layer remains almost the same. It ensures the security of the information throughout the network. Different numbers of convolutional filters are used for training to evaluate their effect on the performance.

2) The AlexNet convolutional network was developed in 2012, to classify images in more than 1000 different classes, using 1.2 million sample pictures from the ImageNet dataset. Because of having just 7 distinguished emotions and limited computing resources, the size of the original network is considered to be too large. Normalization is done over the network to speed up the training.

3) This tested network is based on the work done by Gudi [3]. This network starts with an input layer of 48 x 48 which is as same as the size of the input followed by a convolutional, normalization, max pooling layer respectively. At last, the whole model is connected with a two more convolutional layers and also a fully connected layer with a dropout, which is further connected to a soft max output layer.

In the current investigation New network is proposed, as a modification. In order to reduce the computational intensity of the network a second max pooling layer is applied.This also shifts the performance but only to some little extent. In our investigation,

©IJRASET: All Rights are Reserved

without reducing the learning rate as done Gudi [3], the momentum is used to faster the learning rate when gradient works in the same direction.

The assessment of the three approaches on capability of emotion recognition is being done by the development of three networks based on the concepts from [1], [2], and [3].

B. Dataset

Large amounts of training data require neural and deep networks significantly. Also, the image choice used for training plays a major role in the performance of the final model. It highlights the requirement for both quantitative and qualitative datasets. To recognize emotions, various datasets are available for research, starting from a few sets to tens of thousands smaller images. Further, the data set we choose to work on is RaFD[4] and another set of animated images.

The Fig. 4 shows the human expressions from FERC-2013 and RaFD dataset[10] and the Fig. 5 shows the samples of the animated expressions[11].



Fig. 4 Human expressions

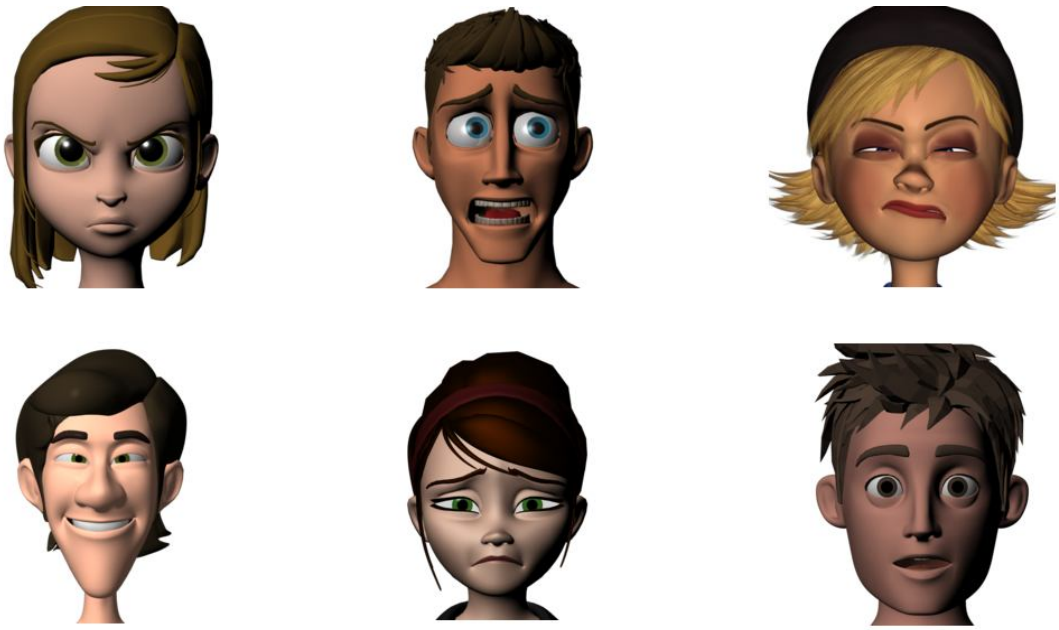


Fig. 5 Animated expressions

©IJRASET: All Rights are Reserved

Training for the three networks is done with 9000 faces from FARC-2013 dataset with 1000 fresh samples for validation. And testing is done with RaFD dataset. For the proposed network, 20000 faces are used for training from FARC-2013 and validation set and test set are taken from both FARC-2013 and RaFD.

Here, Analysis of distribution of data for the FERC-2013 and RaFD is dipected in Fig. 6. Also,the Fig. 7 shows the Animated face distribution.

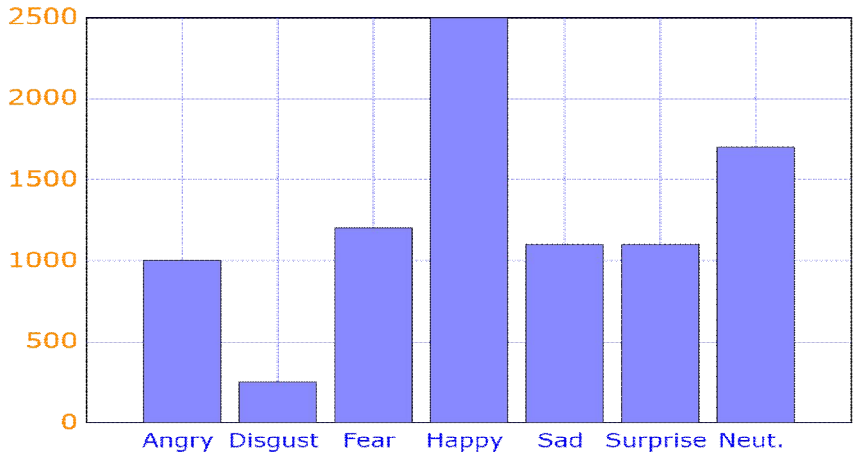


Fig. 6 Distribution of FERC-2013, RaFD

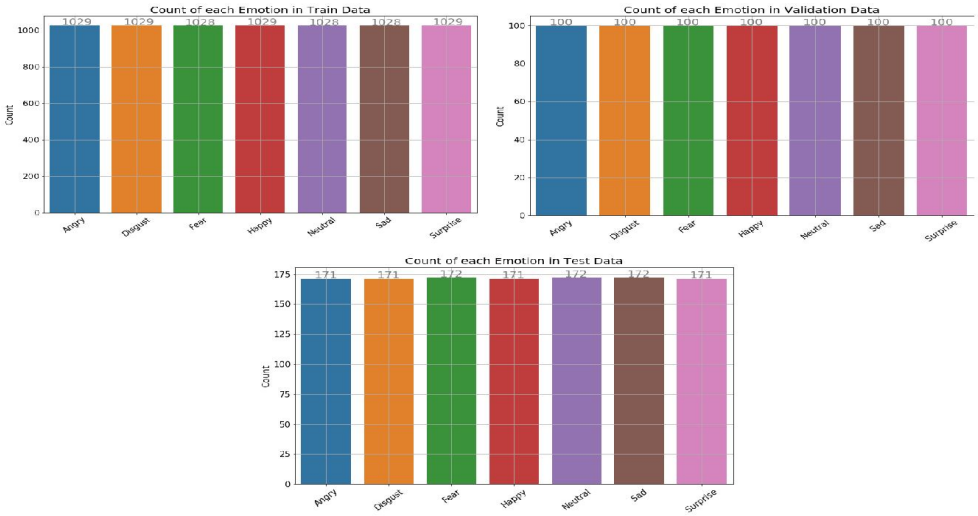


Fig. 7 Distribution of Animated images

C. Training

The training is done over 20000 images from FERC-2013 dataset and the network is trained with 100 epochs.

©IJRASET: All Rights are Reserved

```
Training Step: 22486  | total loss: ?[1m?[32m0.45681?[0m?[0m | time: 901.628s
| Momentum | epoch: 100 | loss: 0.45681 - acc: 0.8466 -- iter: 10550/11214
Training Step: 22487  | total loss: ?[1m?[32m0.45617?[0m?[0m | time: 905.829s
| Momentum | epoch: 100 | loss: 0.45617 - acc: 0.8499 -- iter: 10600/11214
Training Step: 22488  | total loss: ?[1m?[32m0.45360?[0m?[0m | time: 910.198s
| Momentum | epoch: 100 | loss: 0.45360 - acc: 0.8489 -- iter: 10650/11214
Training Step: 22489  | total loss: ?[1m?[32m0.45593?[0m?[0m | time: 914.992s
| Momentum | epoch: 100 | loss: 0.45593 - acc: 0.8420 -- iter: 10700/11214
Training Step: 22490  | total loss: ?[1m?[32m0.44042?[0m?[0m | time: 919.108s
| Momentum | epoch: 100 | loss: 0.44042 - acc: 0.8498 -- iter: 10750/11214
Training Step: 22491  | total loss: ?[1m?[32m0.44757?[0m?[0m | time: 923.462s
| Momentum | epoch: 100 | loss: 0.44757 - acc: 0.8429 -- iter: 10800/11214
Training Step: 22492  | total loss: ?[1m?[32m0.48312?[0m?[0m | time: 927.571s
| Momentum | epoch: 100 | loss: 0.48312 - acc: 0.8346 -- iter: 10850/11214
Training Step: 22493  | total loss: ?[1m?[32m0.47198?[0m?[0m | time: 931.740s
| Momentum | epoch: 100 | loss: 0.47198 - acc: 0.8391 -- iter: 10900/11214
Training Step: 22494  | total loss: ?[1m?[32m0.46501?[0m?[0m | time: 935.906s
| Momentum | epoch: 100 | loss: 0.46501 - acc: 0.8392 -- iter: 10950/11214
Training Step: 22495  | total loss: ?[1m?[32m0.45688?[0m?[0m | time: 940.505s
| Momentum | epoch: 100 | loss: 0.45688 - acc: 0.8453 -- iter: 11000/11214
Training Step: 22496  | total loss: ?[1m?[32m0.44715?[0m?[0m | time: 944.724s
| Momentum | epoch: 100 | loss: 0.44715 - acc: 0.8508 -- iter: 11050/11214
Training Step: 22497  | total loss: ?[1m?[32m0.45724?[0m?[0m | time: 948.831s
| Momentum | epoch: 100 | loss: 0.45724 - acc: 0.8477 -- iter: 11100/11214
Training Step: 22498  | total loss: ?[1m?[32m0.44647?[0m?[0m | time: 952.962s
| Momentum | epoch: 100 | loss: 0.44647 - acc: 0.8489 -- iter: 11150/11214
Training Step: 22499  | total loss: ?[1m?[32m0.44031?[0m?[0m | time: 957.410s
| Momentum | epoch: 100 | loss: 0.44031 - acc: 0.8480 -- iter: 11200/11214
Training Step: 22500  | total loss: ?[1m?[32m0.43021?[0m?[0m | time: 972.135s
| Momentum | epoch: 100 | loss: 0.43021 - acc: 0.8572 | val_loss: 1.16665 - val_acc: 0.6598 -- iter: 11214/11214
--
[+] Model trained and saved at Gudi_model_100_epochs_20000_faces
PS C:\Users\Vipul\documents\emotion-recognition-neural-networks-master>
PS C:\Users\Vipul\documents\emotion-recognition-neural-networks-master>
```

Fig. 8 Training with 100 epoch

The below graph shows the learning curve for every network:

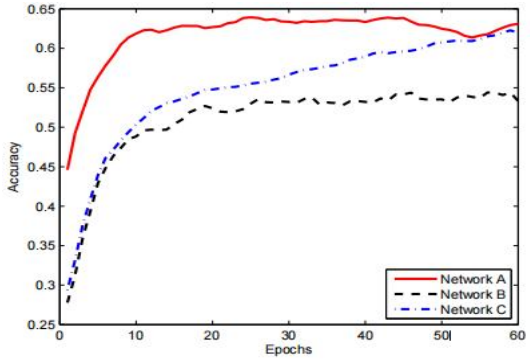


Fig. 9 Learning of Networks

Here, after analyzing the curves shown in above graph we concluded that the network A shows the fast learning whereas the network C came up with the slow learning environment.

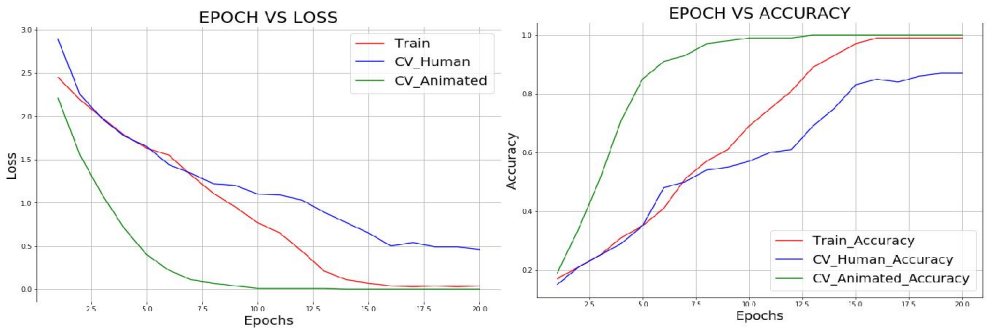
D. Training of Animated Images

After training the animated faces from [11].We ran the model for 20 epochs and the results are as follows as in Fig. 10.

	Epoch	Comb_Train_Loss	Comb_Train_Accuracy	CVHuman_Loss	CVHuman_Accuracy	CVAnime_Loss	CVAnime_Accuracy
0	1	2.45	0.17	2.89	0.15	2.21	0.19
1	2	2.19	0.21	2.25	0.21	1.55	0.34
2	3	1.98	0.25	1.97	0.25	1.09	0.51
3	4	1.78	0.31	1.77	0.29	0.70	0.71
4	5	1.63	0.35	1.65	0.35	0.40	0.85
5	6	1.55	0.41	1.44	0.48	0.22	0.91
6	7	1.32	0.51	1.34	0.50	0.11	0.93
7	8	1.11	0.57	1.22	0.54	0.07	0.97
8	9	0.95	0.61	1.20	0.55	0.04	0.98
9	10	0.77	0.69	1.10	0.57	0.01	0.99
10	11	0.65	0.75	1.09	0.60	0.01	0.99
11	12	0.44	0.81	1.03	0.61	0.01	0.99
12	13	0.21	0.89	0.89	0.69	0.01	1.00
13	14	0.11	0.93	0.77	0.75	0.00	1.00
14	15	0.07	0.97	0.65	0.83	0.00	1.00
15	16	0.04	0.99	0.50	0.85	0.00	1.00
16	17	0.03	0.99	0.54	0.84	0.00	1.00
17	18	0.04	0.99	0.49	0.86	0.00	1.00
18	19	0.03	0.99	0.49	0.87	0.00	1.00
19	20	0.04	0.99	0.46	0.87	0.00	1.00

Fig. 10 result on 20 epoch

©IJRASET: All Rights are Reserved



The above graph shows that loss is reduced as the epochs are increased:
Training loss: 2.45 to 0.24
CV animated loss: 2.21 to 0.00

Graph also shows the increment in the accuracy of the model:
Train accuracy: 17% to 99%
CV animated accuracy: 19% to 100%

IV. RESULTS

The results over the datasets i.e. RaFD is shown in Fig. 11. **Error! Reference source not found.**It is to note that the RaFD test set all different images which shows the accuracy and robustness of the model.

NETWORK	FERC-2013 VALIDATION TEST	RaFD
A	63%	50%
B	53%	46%
C	63%	60%
NEW	66% 63%	70%

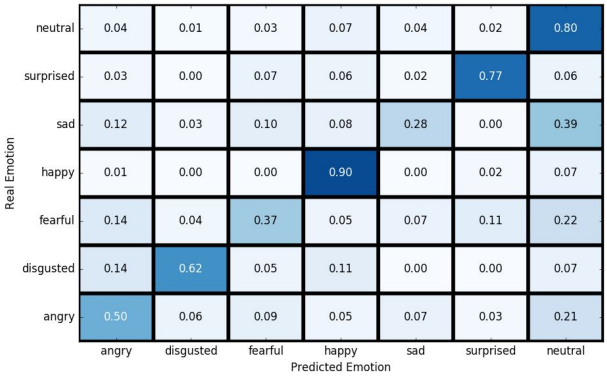


Fig. 11 Result matrix

The results of Testing over the animated images are shown in :

©IJRASET: All Rights are Reserved

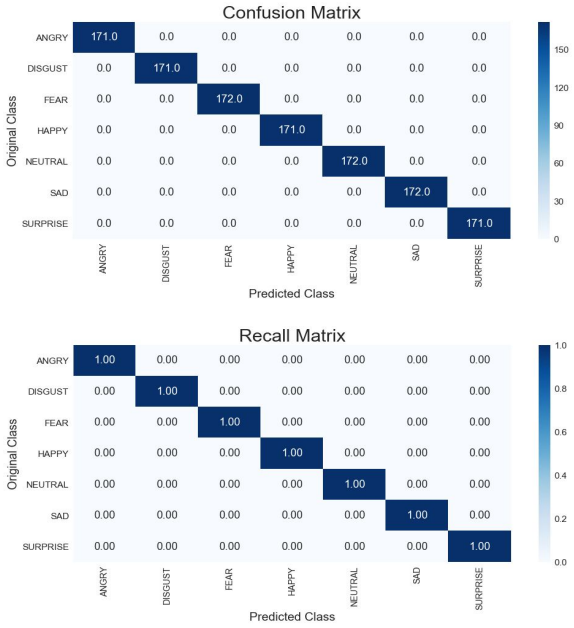


Fig. 12: Result matrix

Accuracy observed is 100%.

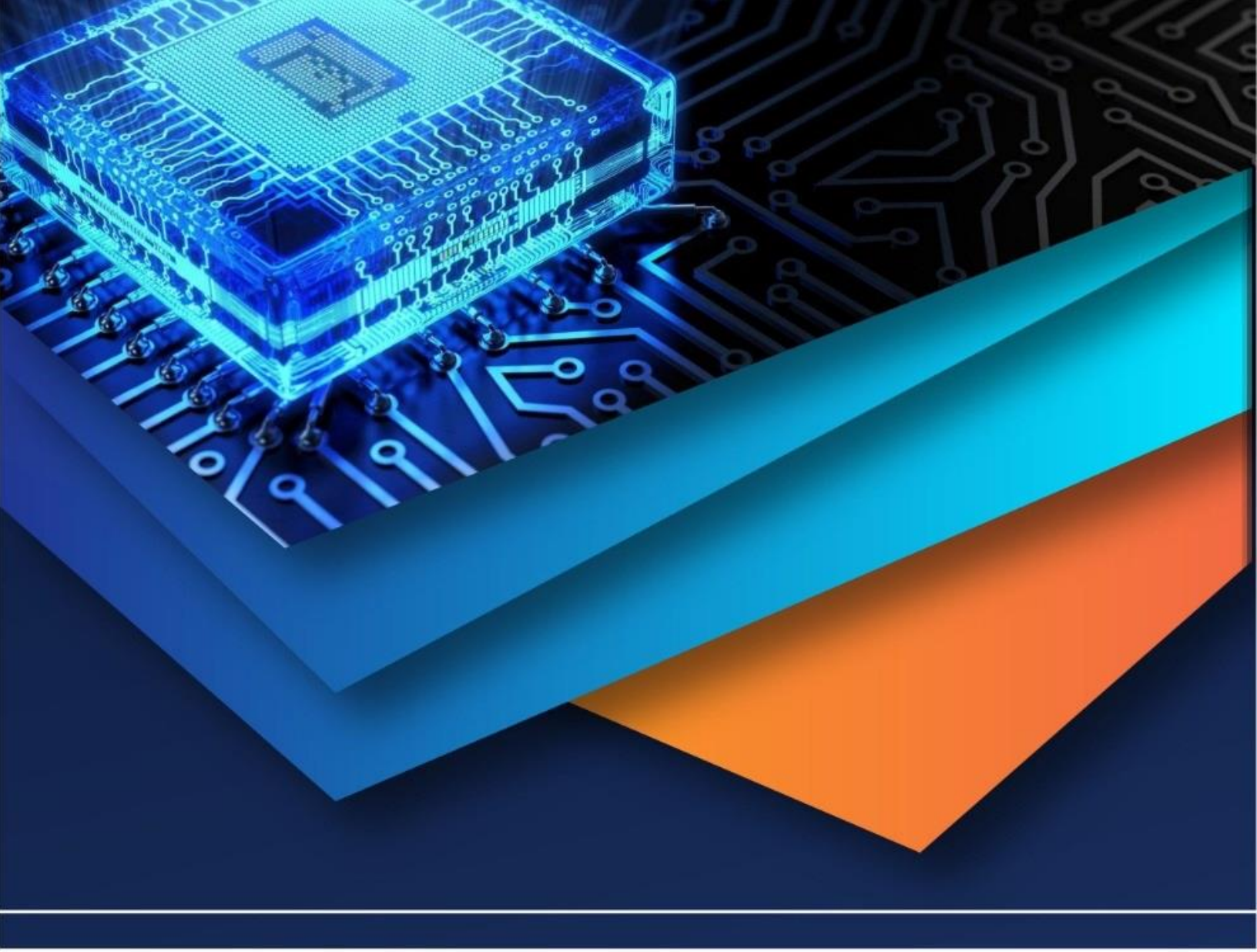
V. CONCLUSIONS

The environment shows the results as 100% accuracy in case of animated images and 70% accuracy in case of human faces. This gap occurred due to the reason that learning the features of human images is hard as compared to animated images. As the animated faces are system generated which results in less variance and also the expression subtlety is small. In order to increase accuracy, more images with high variance can be used and also the possibility of a new CNN model with high computation power can be explored.

REFERENCES

[1] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009.
[2] 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.
[3] A. Gudi. Recognizing semantic features in faces using deep learning. arXiv preprint arXiv:1512.00743, 2015.
[4] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. van Knippenberg. Presentation and validation of the radboud faces database. Cognition and emotion, 24(8):1377– 1388, 2010.
[5] OpenSourceComputerVision. Face detection using haar cascades. URL http://docs.opencv.org/master/d7/d8b/ tutorial_py_face_detection.html..
[6] TFlearn. Tflflearn: Deep learning library featuring a higher-level api for tensorflflow. URL http://tflearn.org/.
[7] J. Nicholson, K. Takahashi, and R. Nakatsu, Emotion recognition in speech using neural network Neural computing & applications, 9(4): 290–2000.
[8] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specifified expression.
[9] Shashak,jaiswal, michael,vlalstar Deep learning the dynamic appearance and shape of facial action units "Deep learning the dynamic appearance and shape of facial action units," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1-8, doi: 10.1109/WACV.2016.747762.
[10] https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge.
[11] http://grail.cs.washington.edu/projects/deepexpr/ferg-2d-db.html.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)