cisdiJournal

Computing, Information Systems & Development Informatics Journal
Vol. 11 No. 2, 2020 - www.isteams.net/cisdijournal

Smile Prediction from Facial Images Using Convolutional Neural Network (CNN)

Ojo Abosede Ibironke
Department of Computer Science
Ogun State Institute of Technology
Igbesa, Ogun State, Nigeria.
Email: ronkujoyme@gmail.com
Phone: +23481646227444

ABSTRACT

This research work is aimed at developing a model where smile can be predicted using Convolutional Neural Network (CNN) with the objectives of implementing the model and carried out several evaluation methods. Several research done on this area made use of some well-known datasets for their classifications and prediction. The backbone of this paper relied on Convolutional Neural Network with several layers that filter and extract facial features from human images. The Convolutional Neural Network provides a multitasking classifier with high competitive standard than other traditional algorithms. The facial images used in this work was a local dataset with the total number of 490 images; where 245 images denoted as 1 (smiling) and the other 245 images denoted as 0 (not smile). All these images were cropped to the required dimension 227by227by3 as acceptable by the AlexNet Pre-trained Algorithm and five different augmentation was performed on them. Multi-task CNN Algorithm was used for the development of the model while MATLAB 2018a was used as the Simulation tool. The system was tested and performance evaluation such as Precision, Recall, Accuracy and F1-score were carried out and their scores were highlighted in bold. Precision (94.7388), Recall (94.7388), Accuracy (92.9273) and F1 Score (94.7360) with the Average performance of 94.7388 for both Precision and Recall.

Keywords: Smiling, Convolutional Neural Network, Model, Prediction

CISDI Journal Reference Format
Ojo, A.I. (2021): Smile Prediction from Facial Images Using Convolutional Neural Network (CNN). Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 11 No 2, Pp 15-.28
DOI Affix - https://doi.org/ 10.22624/AIMS/CISDI/V12N1P2. Available online at www.isteams.net/cisdijournal

1. INTRODUCTION

Face recognition system is one of the ways people display their emotion and it is a fundamental biometric system that proves effective, efficient and highly authenticated (Dahghan, Ortiz, Shu and Masood, 2017). Emotion classification types which include joy, sadness, anger, fear etc. are now identified by machine. The machine will capture individual emotions from face-to-face interaction Gan, (2018) and predict accordingly. Smiling which is known as an expression of joy may not be easily predicted by machine, because it is not all the people that smile that are joyful, deception may probably not be noticed by machine which can be identified by human being except in a rare occasion. But, due to the complexity and progress made in the development of Convolution Neural Network has provided an improved, efficient and better performance than humans (Dehghan, Ortiz Shu, Masood, 2017 and Ranjan, Sankaranarayanan, Castillo and Chellappa (2016)

2. LITERATURE REVIEW

Harris and Yen (2002) identify different types of biometrics such as fingerprint which stands out from others due to the loop, whorl and arc to match a print. The Iris scanning examines the iris pattern which evolves round the pupil by opening or widening the pupils which is made of elastic fibre While, the hand geometry biometric type uses scanner to make a shot of the hand in a 3D. A template of the scanned hand is stored and this is used for verification and recognitions. The Retinal scanner scans the blood vessels of the retina by taking lots of different reading with a low light and the information gathered are stored in a database. Voice verification was pinpointed where some principal areas such as palate, teeth, vocal tract etc. of the mouth would be captured. And finally, signature identification that depicts how individual endorses through the way individual hold pen, press on paper, and write strokes and how the signature appears. Jain, Ross, and Prabhakar (2004) itemized fourteen different types of biometrics which are used for different purpose and at different areas of application. While mentioning the fingerprint, Iris, Retinal scan, signature, voice and face etc. as the major and popularly known biometrics.

Face recognition is the process of identifying and verifying individual using facial extraction features such as ear, nose, eyes, and mouth. It is the frontal recognition of individual by extracting the biological trait of a person. Haider, Bashir, Sharif, Sharif and Wahab (2014) defined face recognition system as the process that identify and verify individual face from captured images based on the stored dataset, where the stored images in the database must be monitored (Lai, Kumar, Arain, Maitlo, Ruk, 2018)). Naeem et al. al (2015) reported that the algorithm for face recognition was dated to 1960s where geometric features were used for face detection and recognition of identity. They did mention that Kenade seminal in 1973 launched the traditional face recognition that used template match. Different researchers have described the face recognition operation. Lai et al (2018) described face recognition operation has a process that involved three modules which are;

Fig 2.1 Face Recognition Operation described by Lai et al (2018)

Where, Naeem et al. al. (2015) stated that three operations related to face recognition are:

Fig 2.2 Face Recognition Operation described by Naeem et al. al. (2015).

Barnouti, Mahmood , and Matti (2016) and Mishra, Swain, and Dash (2012) in their opinions affirmed that for every face recognition, there should be an input to the system which is normally the digital images or video stream, where the processing is the face detection and feature extraction and the output is the face recognition. They concluded by ascertained that there are three steps required in the face recognition process

s which are:

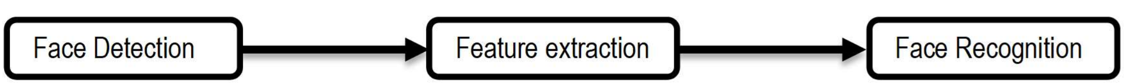


Fig 2.3: Face Recognition Operation described by Barnouti. et al. (2016)

3. METHODOLOGY

Each input image is cropped to size 227x227x3 pixels, which is the required size for input into the pre-trained convolutional network. AlexNet is the pre-trained algorithm used in this study. Augmentation techniques of rotation, shearing, horizontal flip, different lighting and shifts were performed on the cropped image to generate 5 different variations of images.

Smile Detection: The MTL-CNN model is to detect whether the facial image is smiling or not. The loss function used for smile attribute, L_s, is the cross-entropy loss given as:

$$L_S = -(1-S)\log(1-P_S) - S\log(P_S) \qquad (1)$$

where $S = 1$ for a detected smile and 0 for not smiling. The P_S is the predicted probability that the facial image is smiling.

The figure 3.1 below depicts the algorithm for the image pre-processing steps. These steps are;

1. Load the image – this is the stage where the images are loaded one by one
2. Crop image to required size – the image loaded is cropped to the required size for input image in CNN which is 227by227by3.
3. Image augmentation - After the image is inputted, an augmentation is performing on the cropped image in other to generate additional images which will enhance feature extraction.
4. Store image - At the end of the augmentation the image is stored.
5. Check for the end of the image- if the image has not reached its end it will start from step1 otherwise
6. End Image pre-processing steps.

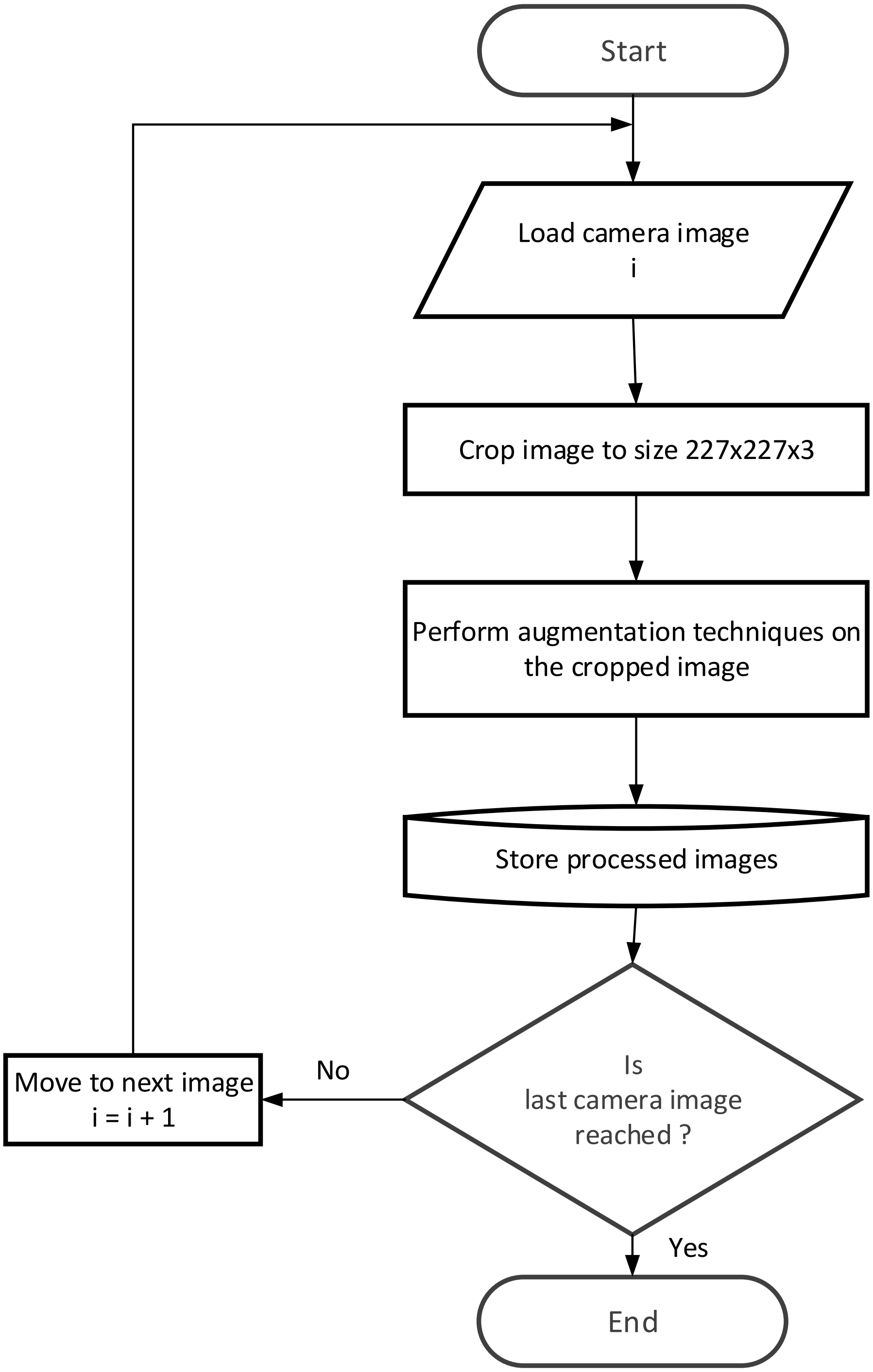


Fig. 3.1: Image pre-processing Model

4. RESULT AND DISCUSSION

In this section we present results in Tables and discus them.

Table 4.1: CNN training parameters for smiling

Parameter	Specification
Size of input image	227-by-227-by-3
Number of convolution layers	5
Number of fully connected layers	4
Activation function	Softmax
Optimizer	Stochastic gradient descent
Momentum	0.9
Maximum epoch	27
Learning rate	0.0001
PC used for simulation	64-bit OS, Core i5-5200U CPU @ 2.2GHz, 4GB RAM

Table 4.2: Testing of the Created CNN Model for Smiling Detection

Test Image ID	Smiling (Actual or ground truth) Yes/No	Smiling (Predicted by CNN) Yes/No
1	No	No
2	No	No
3	No	No
4	No	No
5	No	No
6	No	No
7	No	No
8	No	No
9	No	No
10	No	No
11	No	No
12	No	No
13	No	**Yes**
….	….	….
495	Yes	Yes
496	Yes	Yes
497	Yes	**No**
498	Yes	Yes
499	Yes	Yes
500	Yes	Yes
….	….	….
507	Yes	Yes
508	Yes	Yes
509	Yes	Yes

Table 4.3: Results of the Created CNN Model for Smiling Detection

Smiling	Number of images tested	Correct classification	Misclassification	Precision (%)	Recall (%)	Accuracy (%)	F1-score (%)
Yes (1)	225	212	13	94.2222	95.2555	92.9273	94.7360
No (0)	284	261	23	95.2555	94.2222	92.9273	94.7360
Average				**94.7388**	**94.7388**	**92.9273**	**94.7360**

Table 4.4: Comparison of the Training Set and Testing Set

	Training set	Testing set
Number of images	2058	509
Testing time (s)	3786	33.8265

Table 4.5: Number of Epoch for Training Vs. Testing Accuracy and Training Vs. Testing Loss

Epoch	Training Accuracy	Testing Accuracy	Training Loss	Testing Loss
0	50.0000	47.5442	0.6955	0.6944
1	62.5000	55.5992	0.6815	0.6888
2	56.2500	55.7957	0.6834	0.6804
3	60.9375	63.6542	0.6574	0.6502
4	67.1875	69.1552	0.6102	0.5834
5	64.0625	80.3536	0.5755	0.5084
6	78.1250	84.4794	0.4341	0.3706
7	79.6875	83.8900	0.4174	0.3365
8	75.0000	87.6228	0.4725	0.2843
9	82.8125	83.4971	0.3232	0.3284
10	84.3750	92.3379	0.3341	0.2269
11	85.9375	92.5344	0.2579	0.1926
12	89.0625	90.3733	0.2179	0.2235
13	98.4375	91.5521	0.0897	0.1890
14	85.9375	87.0334	0.2633	0.2478
15	90.6250	89.1945	0.2202	0.2184
16	87.5000	92.9273	0.3128	0.1777
17	90.6250	91.9450	0.2021	0.1796
18	95.3125	92.9273	0.1614	0.1763
19	95.3125	92.5344	0.1365	0.1718
20	96.8750	93.5167	0.1417	0.1716
21	87.5000	89.9804	0.3943	0.2310
22	96.8750	93.3202	0.1144	0.1825
23	96.8750	93.5167	0.0919	0.1601
24	93.7500	93.1238	0.107	0.1566
25	89.0625	92.5344	0.2418	0.1732
26	98.4375	93.5167	0.0858	0.1584
27	95.3125	92.9273	0.1664	0.1725

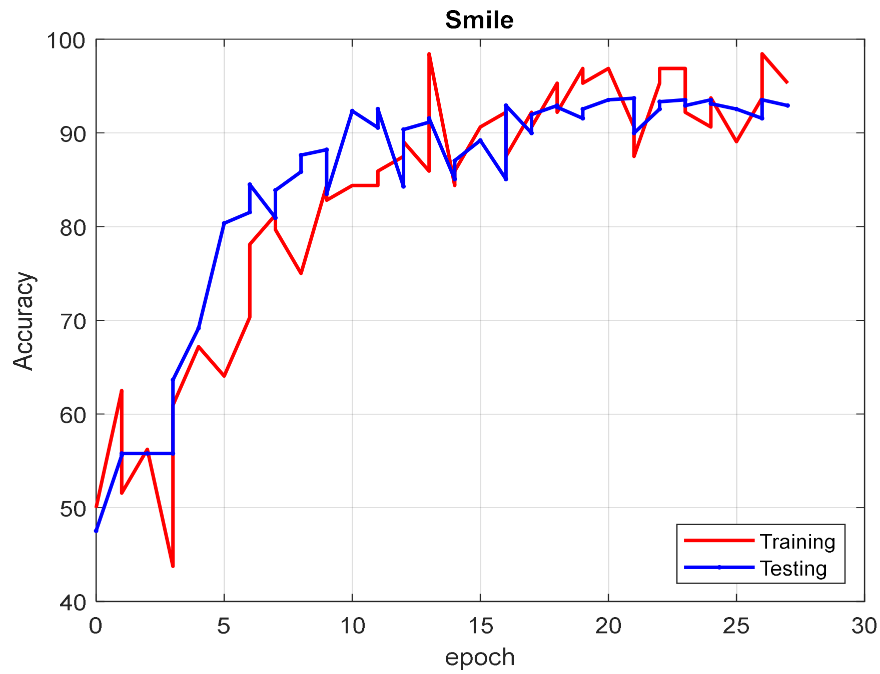


Figure 4.1: Average Smiling Classification Accuracy vs. Epoch

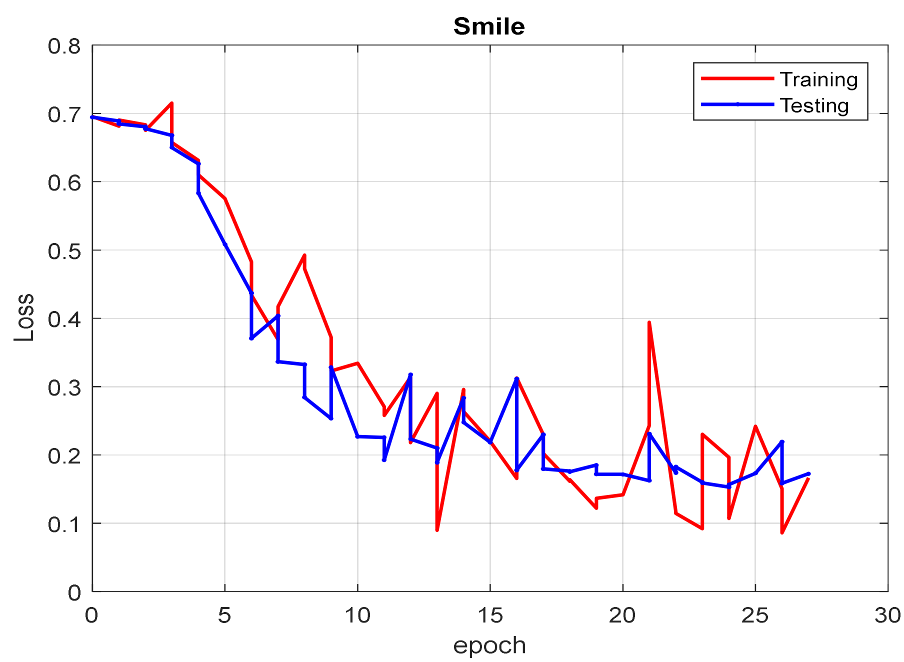


Figure 4.2: Average Smiling Classification Loss vs. Epoch

Table 4.6: Precision versus Accuracy and True Positive Rate versus False Positive Rate

Recall	Precision	False Positive Rate	True Positive Rate
0.4222	1.0000	0.0000	0.4222
0.4711	1.0000	0.0000	0.4711
0.5156	1.0000	0.0000	0.5156
0.5689	1.0000	0.0000	0.5689
0.6133	1.0000	0.0000	0.6133
0.6578	1.0000	0.0000	0.6578
0.6978	0.9937	0.0035	0.6978
0.7378	0.9881	0.0070	0.7378
0.7733	0.9775	0.0141	0.7733
0.8178	0.9787	0.0141	0.8178
0.8578	0.9747	0.0176	0.8578
0.8889	0.9615	0.0282	0.8889
0.9156	0.9450	0.0423	0.9156
0.9333	0.9211	0.0634	0.9333
0.9467	0.8950	0.0880	0.9467
0.9600	0.8710	0.1127	0.9600
0.9778	0.8527	0.1338	0.9778
0.9822	0.8246	0.1655	0.9822
0.9822	0.7950	0.2007	0.9822
0.9911	0.7743	0.2289	0.9911
0.9911	0.7483	0.2641	0.9911
0.9911	0.7240	0.2993	0.9911
0.9956	0.7044	0.3310	0.9956
0.9956	0.6829	0.3662	0.9956
1.0000	0.6657	0.3979	1.0000
1.0000	0.6466	0.4331	1.0000
1.0000	0.6285	0.4683	1.0000
1.0000	0.6114	0.5035	1.0000
1.0000	0.5952	0.5387	1.0000
1.0000	0.5799	0.5739	1.0000
1.0000	0.5653	0.6092	1.0000
1.0000	0.5515	0.6444	1.0000
1.0000	0.5383	0.6796	1.0000
1.0000	0.5257	0.7148	1.0000
1.0000	0.5137	0.7500	1.0000
1.0000	0.5022	0.7852	1.0000
1.0000	0.4913	0.8204	1.0000
1.0000	0.4808	0.8556	1.0000
1.0000	0.4707	0.8908	1.0000
1.0000	0.4611	0.9261	1.0000
1.0000	0.4518	0.9613	1.0000
1.0000	0.4429	0.9965	1.0000

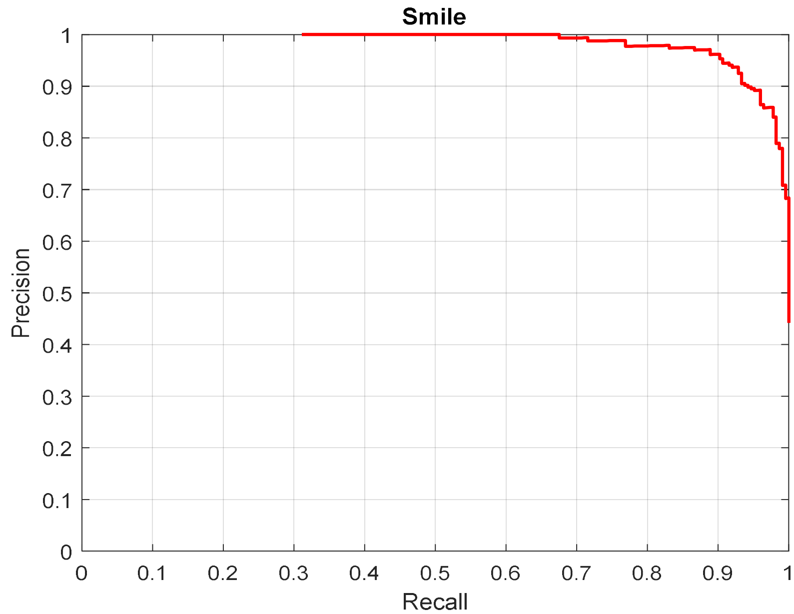


Figure 4.3: Precision vs. Recall on Smile

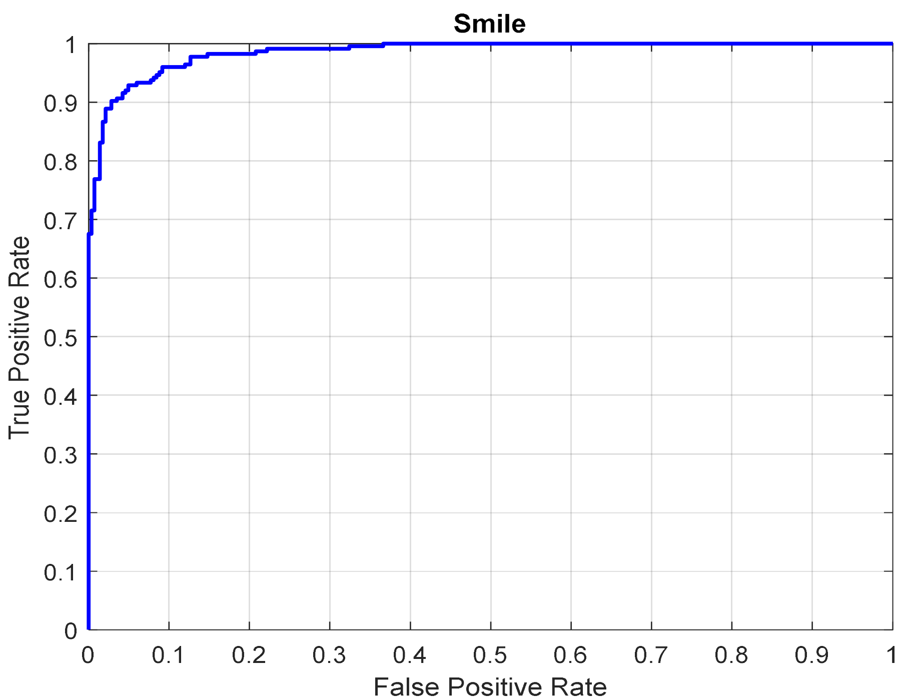
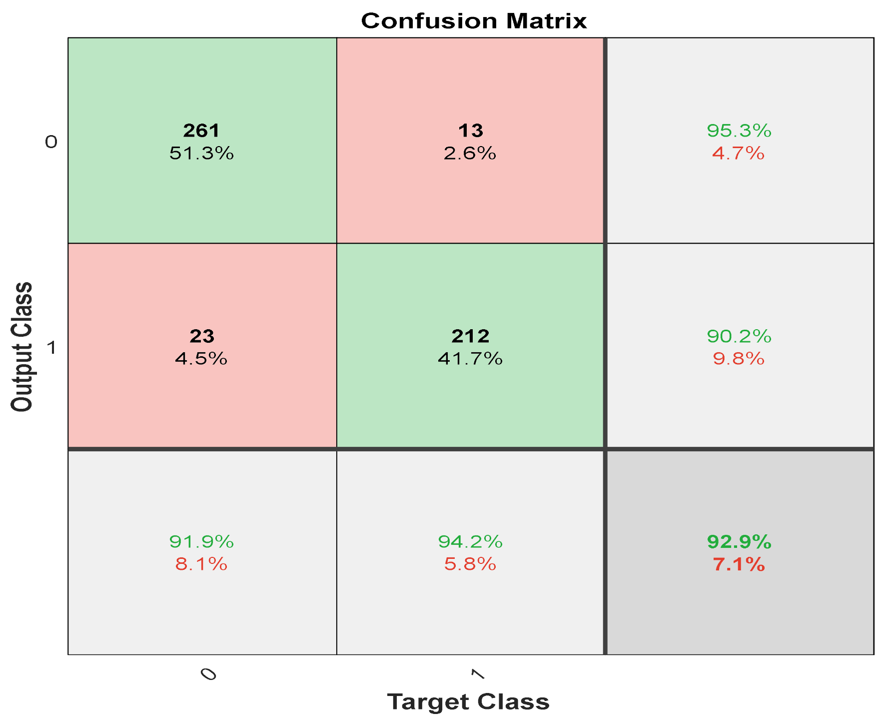


Figure 4.4: True Positive Rate and False Positive Rate on Smile

Table 4.7: Confusion Matrix for Smile Classification



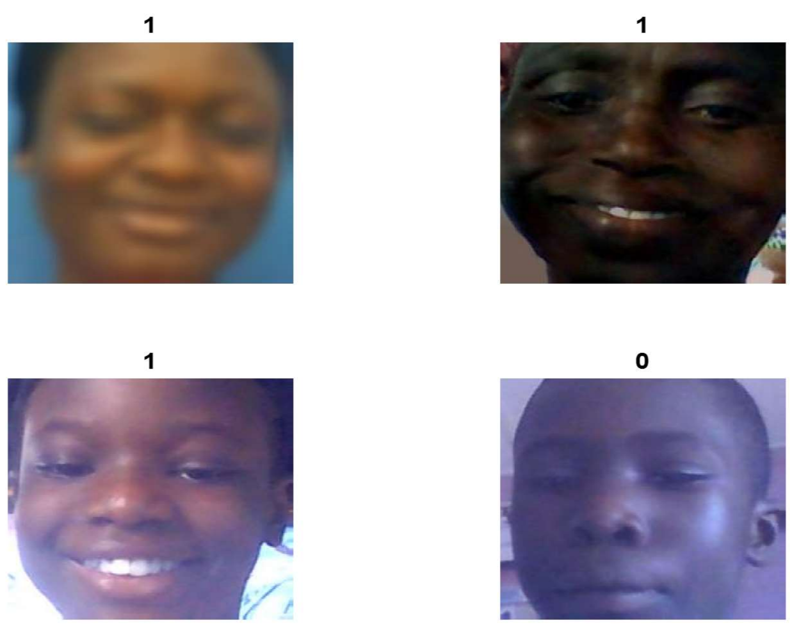


Figure 4.5 Smiling" is denoted by digit 1 while "Not smiling" is denoted by digit 0

Computing, Information Systems & Development Informatics Journal
Vol. 11 No. 2, 2020 - www.isteams.net/cisdijournal

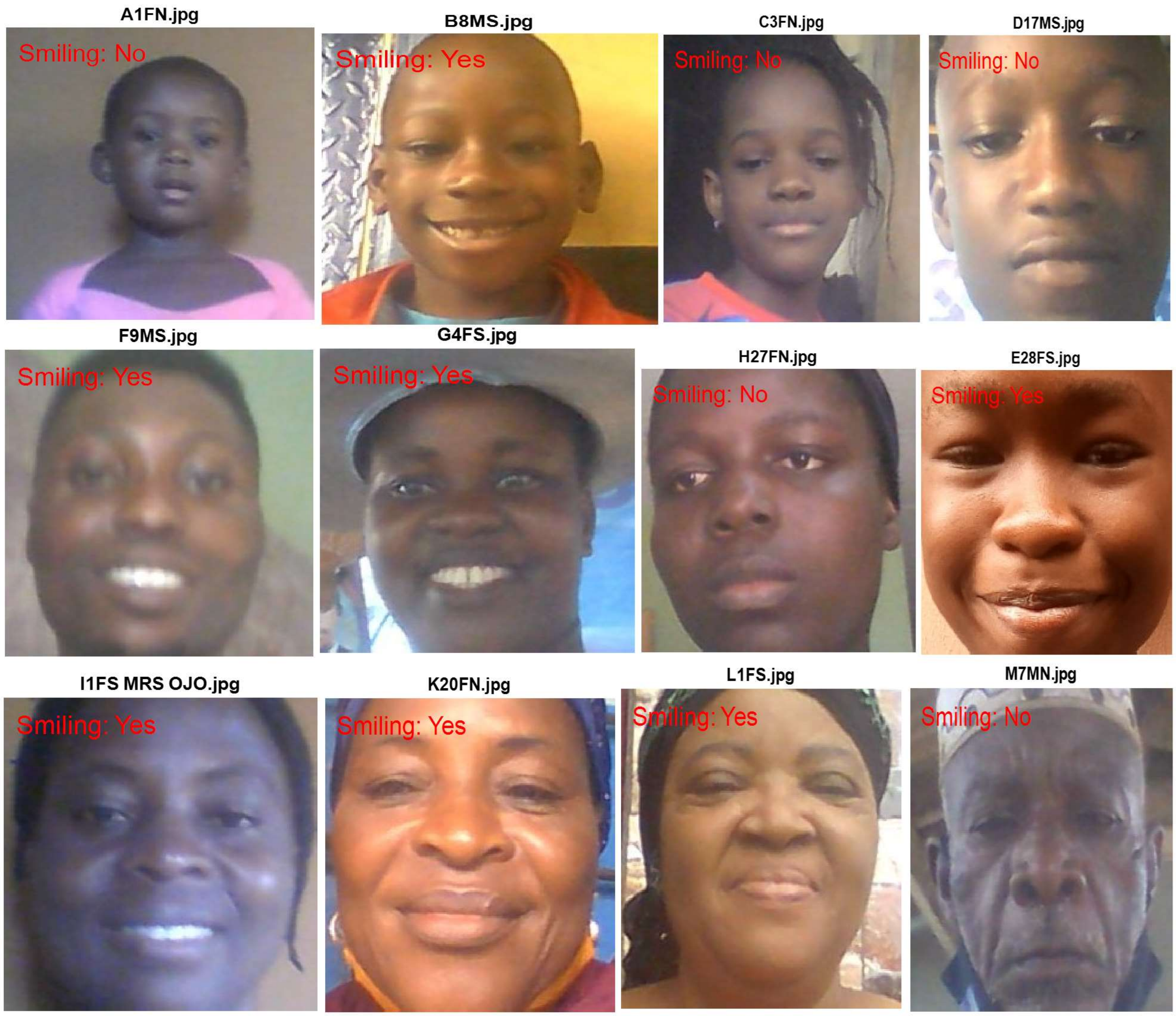


Figure 4.8: Tested samples for smile classification

4.1 Network architecture of CNN model for Smile detection

22x1 Layer array with layers:

1	'data'	Image Input	227x227x3 images with 'zerocenter' normalization
2	'conv1'	Convolution	96 11x11x3 convolutions with stride [4 4] and padding [0 0 0 0]
3	'relu1'	ReLU	ReLU
4	'norm1'	Cross Channel Normalization	cross channel normalization with 5 channels per element
5	'pool1'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
6	'conv2'	Convolution	256 5x5x48 convolutions with stride [1 1] and padding [2 2 2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
10	'conv3'	Convolution	384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu4'	ReLU	ReLU
14	'conv5'	Convolution	256 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
17	'fc_1'	Fully Connected	28 fully connected layer
18	'fc_2'	Fully Connected	512 fully connected layer
19	'fc_3'	Fully Connected	512 fully connected layer
20	'fc_4'	Fully Connected	2 fully connected layer
21	'softmax'	Softmax	softmax
22	'classoutput'	Classification Output	crossentropyex with classes '0' and '1'

4.2 Discussion on Smile Classification

The table 4.2 is the result of testing CNN model for smiling detection where 509 images were tested with brief discussion.

1. Test Image ID: the Test imageId column represents the total number of Images tested. These images were randomly selected from the testing images.
2. The Smiling Actual or ground truth: is the column that shows the true emotion of the image selected. The Yes means the image is smiling while the No means the image is not smiling
3. Smiling Predicted by CNN: This column shows what the CNN classifier classified on the Test ImageId.

For example from the table 4.2, the first Test imageId (1) which represent the first image was Not smiling on the ground truth but, the CNN classifier classified it as not smiling which means the classifier is correct. But a cursory look at the Test ImageId 13 in the same table shows that while the ground truth is indicated that the image is not smiling, the classifier classified it as smiling which is incorrect etc. The table 4.3 shows the result of created CNN models for smiling detection where total number of 509 images was tested. 473 images were classified correctly and 36 images were misclassified. The Precision, Recall, Accuracy and F1-score which are parameters for evaluation of performance was calculated and their scores were highlighted in bold Precision **(94.7388), Recall (94.7388), Accuracy (92.9273) and F1 Score (94.7360)** with the Average performance of **94.7388 for both Precision and Recall.**

The table 4.4 shows the number of the training and testing set with testing time in seconds. The table 4.5 shows the number of epoch for both the training and testing for their accuracy versus loss with their graph representation as shown in figure 4.3 and 4.4. While, Table 4.6 shows the Precision versus Accuracy and True Positive Rate versus False Positive Rate scores and the figure 4.3 and figure 4.4 depict their plots. Finally, the Confusion Matrix for the Smiling Classification represents the result of the smiling classification as shown in table 4.7. Where digit 1 denotes smiling and digit 0 denotes not smiling. From table 4.3, it was observed that the total number of correctly classified images for smiling (1) was 212 with 13 images misclassified as not smiling resulting into 94.2% accuracy and 5.8% loss. While the total number of correctly not smiling images was 261 with 23 images misclassified as smiling resulting into 91.9% accuracy and 8.1 loss. The overall average accuracy was 92.9% and the overall average loss was 7.1%.

5. CONCLUSION

This paper presented a Multi-task CNN based model for face detection, verification, recognition and predict smile. The experimental setup was done on MATLAB 2018a whereby we experiment our local dataset to predict smile. Finally, the system was tested and performance evaluation such as Precision, Recall, Accuracy and F1-score were carried out and their scores were highlighted in bold. Precision **(94.7388), Recall (94.7388), Accuracy (92.9273) and F1 Score (94.7360)** with the Average performance of **94.7388 for both Precision and Recall.**

REFERENCES

1. Barnouti N. H., Mahmood S. S., and Matti W. E., (2016) et al. (2016). Face Recognition: A Literature Review. International Journal of Applied Information Systems (IJAIS) –Foundation of Computer Science FCS. Volume 11 No. 4, ISSN: 2249-0868
2. Dahghan A. Ortiz E. G, Shu G. and Masood G. S. (2017). DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks. Computer Vision Lab, Sighthound Inc., Winter Park, FL arXiv:1702.04280v2 [cs.CV].
3. Gan Y. (2018). Facial Expression Recognition Using Convolutional Neural Network Association for Computing Machinery. 978-1-4503-6529- (ACM).
4. Haider W., Bashir H., Sharif A., Sharif I. and Wahab A., (2014) (2014). A Survey on Face Detection and Recognition Approaches. International Science Congress Association Research Journal of Recent Sciences ISSN 2277-2502 Vol. 3(4), 56-62.
5. Harris J. H. and Yen D. C. (2002). Biometric Authentication: Assuring Access to Information. Information Management and Computer Security_x0019_
6. Jain A. K, Ross A and Prabhakar S. (2004). An Introduction to Biometric Recognition. Appeared in IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No. 1,
7. Lai M., Kumar K., Arain R. H., Maitlo A., Ruk A. and Shaikh H. (2018). Study of Face Recognition Techniques: A Survey. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6.
8. Mishra A., Swain M., and Dash B. (2012). An Approach to Face Recognition of 2-D Images using Eigen Faces and PCA. Signal & Image Processing: An International Journal (SIPIJ) Vol.3, No.2
9. Naeem M., Qureshi I., and Azam F., (2015). Face Recognition Techniques and Approaches: A Survey. Sci. Int. (Lahore), 27(1), 301-305.
10. Ranjan R, Sankaranarayanan S, Castillo D. & Chellappa R. (2016). An All-In-One Convolutional Neural Network for Face Analysis. Centre for Automation Research, UMIACS, University of Maryland, College Park, MD 20742. arXiv:1611.00851v1 [cs.CV] 3 Nov 2016