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: ronkujoyo@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

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 around 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)](image)

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).](image)

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 which are:

![Fig 2.3: Face Recognition Operation described by Barnouti. et al. (2016)](image)
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:

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

(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 step 1 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 discuss them.

Table 4.1: CNN training parameters for smiling

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

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

<table>
<thead>
<tr>
<th>Test Image ID</th>
<th>Smiling (Actual or ground truth)</th>
<th>Smiling (Predicted by CNN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>495</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>496</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>497</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>498</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>499</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>500</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>507</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>508</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>509</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 4.3: Results of the Created CNN Model for Smiling Detection

<table>
<thead>
<tr>
<th>Smiling</th>
<th>Number of images tested</th>
<th>Correct classification</th>
<th>Misclassification</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>Accuracy (%)</th>
<th>F1-score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes (1)</td>
<td>225</td>
<td>212</td>
<td>13</td>
<td>94.2222</td>
<td>95.2555</td>
<td>92.9273</td>
<td>94.7360</td>
</tr>
<tr>
<td>No (0)</td>
<td>284</td>
<td>261</td>
<td>23</td>
<td>95.2555</td>
<td>94.2222</td>
<td>92.9273</td>
<td>94.7360</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td>94.7388</td>
<td>94.7388</td>
<td>92.9273</td>
<td>94.7360</td>
</tr>
</tbody>
</table>

Table 4.4: Comparison of the Training Set and Testing Set

<table>
<thead>
<tr>
<th></th>
<th>Training set</th>
<th>Testing set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of images</td>
<td>2058</td>
<td>509</td>
</tr>
<tr>
<td>Testing time (s)</td>
<td>3786</td>
<td>33.8265</td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>Output Class</th>
<th>Target Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>261 (51.3%)</td>
<td>13 (2.6%)</td>
</tr>
<tr>
<td></td>
<td>23 (4.5%)</td>
<td>212 (41.7%)</td>
</tr>
<tr>
<td>1</td>
<td>91.9%</td>
<td>94.2%</td>
</tr>
<tr>
<td>0</td>
<td>8.1%</td>
<td>5.6%</td>
</tr>
</tbody>
</table>

Figure 4.5 Smiling” is denoted by digit 1 while “Not smiling” is denoted by digit 0
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
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
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
12. 'conv4' Convolution 384 3x3x192 convolutions with stride [1  1] and padding [1  1  1  1]
13. 'relu4' ReLU
14. 'conv5' Convolution 256 3x3x192 convolutions with stride [1  1] and padding [1  1  1  1]
15. 'relu5' 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
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 imageld 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 Imageld.

For example from the table 4.2, the first Test imageld (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 imageld 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