TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

A

Project Report

on

Convolutional Neural Networks for Multiclass Face Recognition

Submitted By:

Rajan Gyawali

073/MSIC/611

Submitted To:

Department of Electronics and Computer Engineering

Lalitpur, Nepal

November 2018

ACKNOWLEDGEMENT

I would like to express my deep gratitude to Department of Electronics and Computer Engineering for incorporating project work as a part of our syllabus. It makes us the realization of our knowledge in the real-world applications and provides a benchmark to proceed for the thesis and research.

I am very grateful to Dr. Surendra Shrestha, Head of Department of Electronics and Computer Engineering, Pulchowk Campus for his precious support.

I am deeply indebted to Dr. Dibakar Raj Pant for his guidance and supervision for the selection and completion of the project.

I am also very grateful to Prof. Dr. Dinesh Kumar Sharma, Prof. Dr. Shashidhar Ram Joshi, Prof. Dr. Subarna Shakya, Dr. Sanjeeb Prasad Panday, Dr. Nanda Bikram Adhikari, Dr. Aman Shakya, Daya Sagar Baral and other faculties for their encouragement and precious guidance.

I am thankful to our program coordinator Dr. Basanta Joshi for providing suitable platform to complete the project.

ABSTRACT

A face recognition system based on Convolutional Neural Network is developed. The network contains five convolutional layers, each layer followed by the Rectified linear Unit and Max Pooling layer. Besides, it contains two fully connected layers and a Softmax classifier. The convolutional network extracts successively larger features in a hierarchical set of layers. A database of 13000 images of 5749 individuals is used which contains quite a higher degree of variability in expression, pose and facial details. The system can classify the images with an accuracy of 96.44%

Keywords: convolutional neural network, face recognition, classification

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
FAR	False Acceptance Rate
FRGC	Face Recognition Grand Challenge
GPU	Graphical Processing Unit
ML	Machine Learning
NLP	Natural Language Processing
PCA	Principle Component Analysis
ReLU	Rectified Linear Unit
SVM	Support Vector Machine

1. Introduction

1.1 Background

Face recognition is becoming an important tool in human- computer interaction. Its usage includes security systems, video surveillance, commercial areas and even in the social networking sites like Facebook as well. It is the process of recognizing the face of a concerned person by a computer vision system. Face recognition system has gained its popularity due to its nonintrusive nature and has been the main method of person identification compared to other biometric techniques like fingerprint.

The techniques used in the best face recognition systems may depend on the application of the system. There are two broad categories of face recognition systems.

1) Finding a person within a large database of faces (e.g., in a police database). These systems typically return a list of the most likely people in the database. Often only one image is available per person. It is usually not necessary for recognition to be done in real-time.
2) Identifying particular people in real-time (e.g., in a security monitoring system, location tracking system, etc.), or we want to allow access to a group of people and deny access to all others (e.g., access to a building, computer, etc.). Multiple images per person are often available for training and real-time recognition is required.

1.2 Problem Statement

Machine learning algorithms' performance relies on the choice of data representation or features on which they are applied. In the face recognition process using traditional methods like PCA, SVM, etc. based on shallow learning have been facing challenges like pose variation, facial disguises, lighting of the scene, the complexity of the image background and changes in facial expressions. They only utilize from some basic features of images and depend on artificial experience to extract sample features.

This has made the necessity of representation learning or feature learning. Feature engineering is important but labor-intensive and highlights the weakness of above

mentioned methods: their inability to extract and organize the discriminative information from the data. Feature engineering is a way to take advantage of human ingenuity and prior knowledge to compensate for that weakness.

Representation learning or deep learning can be excellent at revealing complex structures in high dimensional data and is therefore applicable to lots of domains of science, business and government sectors. Deep learning methods like Convolutional Neural Network (CNN) has been gaining popularity in the field of image recognition.

1.3 Objectives

The objectives of this project are:

i. To detect faces.
ii. To recognize correct face using CNN and validate.

1.4 Scope and Application

With robust face alignment and recognition, the use of such systems can be used in various sectors like military camps, banks, school, colleges and university, mines. These face recognition systems can be used in highly sensitive areas which require access control and authentication.

Based on the assessment of the applications in the field today, a majority of facial recognition use-cases appear to fall into three major categories:

- **Security**: Companies are training deep learning algorithms to recognize fraud detection, reduce the need for traditional passwords, and to improve the ability to distinguish between a human face and a photograph. Similarly, the US FBI has been using face recognition for criminal identification.
- **Healthcare**: Machine learning is being combined with computer vision to more accurately track patient medication consumption and support pain management procedures.

- **Marketing**: Fraught with ethical considerations, marketing is a burgeoning domain of facial recognition innovation, and it’s one we can expect to see more of as facial recognition becomes ubiquitous.

2. Literature Review

Representation learning has become a field in itself in the machine learning community, sometimes under the header of Deep Learning or Feature Learning. The rapid increase in scientific activity on representation learning has been accompanied and nourished by a remarkable success both in academia and in industry [1].

CNN was proposed firstly by LeCun and applied it on handwriting recognition [2]. From his contributions, many scientists got true inspiration to work in this field. Krizhevsky, Sutskever and Hinton achieved best results when they published their work in ImageNet Competition. In 2012, AlexNet significantly outperformed all the prior competitors and won the challenge by reducing the top-5 error from 26% to 15.3%. The second-place top-5 error rate, which was not a CNN variation, was around 26.2%. The network was deeper, with more filters per layer, and with stacked convolutional layers. It consisted 11x11, 5x5,3x3, convolutions, max pooling, dropout, data augmentation, ReLU activations, SGD with momentum. It attached ReLU activations after every convolutional and fully-connected layer [3].

The runner-up at the ILSVRC 2014 competition is dubbed VGGNet by the community and was developed by Simonyan and et.al. VGGNet consists of 16 convolutional layers and is very appealing because of its very uniform architecture. Similar to AlexNet, only 3x3 convolutions, but lots of filters. Trained on 4 GPUs for 2–3 weeks. It is currently the most preferred choice in the community for extracting features from images. The weight configuration of the VGGNet is publicly available and has been used in many other applications and challenges as a baseline feature extractor [4].

Musab Coskun and et. al [5] have proposed a convolutional neural network for face recognition with number of convolutional layers. They have used Georgia Tech Database and showed that the approach has improved the face recognition performance with better recognition results.

Sharma S and et. al [6] have published the CNN based efficient face recognition technique using Dlib. They have emphasized the importance of the face alignment, thus the accuracy and False Acceptance Rate (FAR) is observed. Their computational analysis has showed

the better performance than other state-of-art approaches. The work had been done on Face Recognition Grand challenge (FRGC) dataset and giving accuracy of 96% with FAR of 0.1.

3. Methodology

The deep learning technique has been used for the purpose of face recognition. One of the best proven method for this is the Convolutional Neural Network. CNN is a kind of ANN that employs convolution methodology to extract features from the input data to increase the number of features. With increase in the computational power of Graphical Processing Units (GPU), CNN has achieved remarkable cutting-edge results over a number of areas including image recognition, scene recognition, edge detection and semantic segmentation [5].

The general structure of face recognition process in this project is divided into these stages:

i. Data Collection
ii. Pre-processing
iii. Features Extraction
iv. Classification
v. Recognition

3.1 System Block Diagram

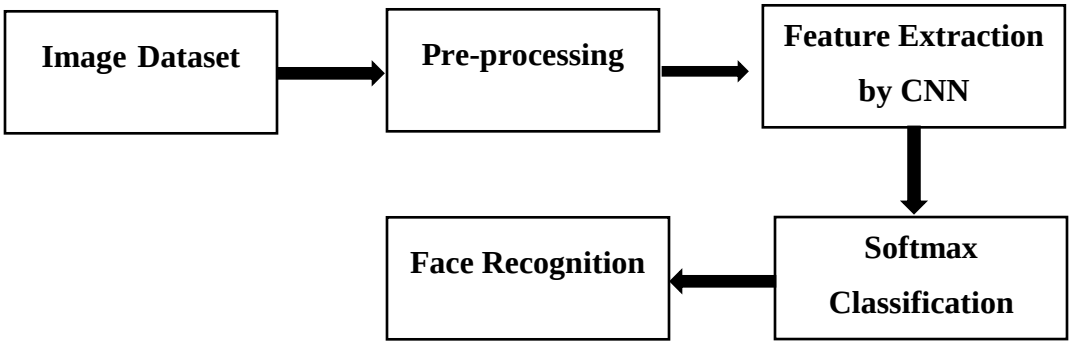


Figure 3. 1: System Block Diagram

3.1.1 Data Collection

The image dataset used is Labelled Faces in the Wild. It contains 13,000 images of 5749 individuals collected over the internet. Using all of the individuals would result a useless model because many of the individuals have only a single image. The project has used

images of only seven individuals for the purpose of face recognition. The individuals with the following number of images are used to build a training model:

- *Colin Powell – 224 images*
- *Donald Rumsfeld – 118 images*
- *George W. Bush – 493 images*
- *Gerhard Schroeder – 107 images*
- *Tony Blair – 141 images*
- *Ariel Sharon – 74 images*
- *Hugo Chavez – 69 images*

Altogether 106 images of these individuals are used to test the model.

Colin Powell

Donald Rumsfeld

George W. Bush

Gerhard Schroder

Tony Blair

Ariel Sharon

Hugo Chavez

Figure 3. 2: Training Samples Images

3.1.2 Image preprocessing

Building an effective neural network model requires careful consideration of the network architecture as well as the input data format. The most common parameters are the number of images, image height, image width and number of channels.

All the RGB images have been converted to gray scale images with size of 150 X 150. Before resizing the image, the face part has been segmented using a pre-built model.

Colin Powell

Donald Rumsfeld

George W. Bush

Gerhard Schroder

Tony Blair

Ariel Sharon

Hugo Chavez

Figure 3. 3: Preprocessed Training Samples

3.1.3 Feature Extraction by CNN

CNNs are the feed forward neural networks made up of many hidden layers. CNNs consist of filters or kernels or neurons that have learnable weights or parameters and biases. Each filter takes some inputs and does convolution. The components of CNN consist of following layers:

i. Convolutional Layer
ii. Rectified Linear Unit (ReLU) Layer
iii. Pooling Layer
iv. Fully Connected Layer

Convolutional Layer

This layer is the core building block of a convolutional network that performs most of the computational heavy lifting. Its primary purpose is to extract features from the input data which is an image. Convolution preserves the spatial relationship between pixels by learning features using small squares of input image. This produces a feature map or activation map in the output image and after then feature maps are fed as input data to the next convolutional layer. For the purpose of face recognition five convolutional layers have been used followed by ReLU and Max Pooling.

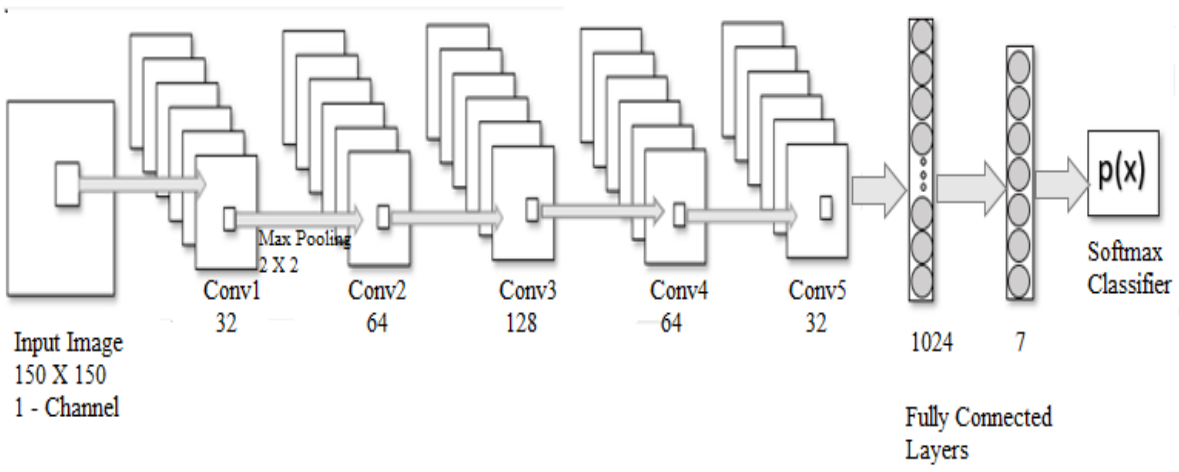


Figure 3. 4: Architecture of CNN for Face Recognition

The 1st convolutional layer has 32 filters, while 2nd has 64, 3rd has 128, 4th has 64 and finally 5th has 32 filters for feature extraction.

ReLU Layer

It is a non-linear operation similar to the rectification. It is an element wise operation that reconstitutes all negative values in the feature map by zero. In order to know how the ReLU operates, we assume there is a neuron input given as x and from the rectifier it is defined as f(x) = max (0, x).

Pooling Layer

This layer reduces the dimensionality of each activation map and continues to have the most important information. The input images are divided into a set of non-overlapping rectangles. Each region is down-sampled by a mon linear operation like average or maxima. This layer gains better generalization, faster convergence, robust to translation and distortion and usually placed between convolutional layers.

Here, Max Pooling has been used for pooling operation with kernel size 2.

Fully Connected Layer

This indicates that every filter in the previous layer is connected to every filter in the next layer. The output from the convolutional, pooling and ReLU layers are embodiments of high level features of the input image. Using fully connected layer employs these features for classifying the input image into various classes based on training set.

Fully connected layer is regarded as final pooling layer feeding the features to a classifier that uses Softmax activation function. The sum of output probabilities from the fully connected layer should be 1. The Softmax function takes a vector of arbitrarily real valued scores between zero and one that sum to 1.

Two fully connected layers have been used. The final fully connected layer has seven nodes for classification of seven different individuals.

3.1.5 Softmax Classification

The Softmax function squashes the outputs of each unit to be between 0 and 1. It also divides each output such that the total sum of the outputs is equal to 1. The output of the Softmax function is equivalent to a categorical probability distribution, it tells the probability that any of the classes are true.

Mathematically, the Softmax function is shown below, where z is a vector of the inputs to the output layer (if you have 10 output units, then there are 10 elements in z). And again, j indexes the output units, so j = 1, 2, ..., K.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$ ---------- (i)

The output of Softmax gives the name of the seven different individuals i.e. recognized face.

3.2 Algorithm

The algorithm for face recognition is as follows:

1. Resize the input images as 150 X 150 X 1.

2. Build a CNN structure with five convolutional layers, each layer followed by ReLU activation and Max Pooling.
3. After extracting all the features, use Softmax classifier for classification.

3.3 Tools Required

All the programming has been done in Python using the following tools and libraries.

- Jupyter Notebook
- Numpy
- Pandas
- Tensorflow

4. Result Analysis

A fully functional face recognition system has been developed. The system is able to classify the 5 different leaders of world with 96.44% accuracy. When the system is extended to make classification of seven different people the accuracy has been decreased. It is due to the low number of images for the added people. The result analysis has been done in two parts:

- 5 – Class Classification
- 7 – Class Classification

4.1 Five – Class Classification

- Total Samples of images: 1103
- Number of training samples: 1003
- Number of Validation Samples: 100
- Number of test samples: 96
- Number of Epochs per training: 5
- Learning Rate: 0.001
- Optimizer: Adam Optimizer
- Accuracy: 0.9644
- Loss: 0.40123

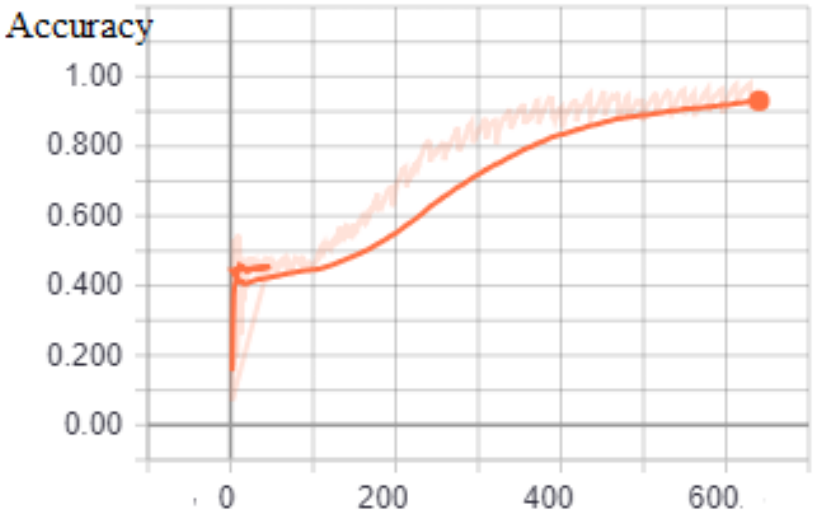


Figure 4. 1: Accuracy Vs Training Steps

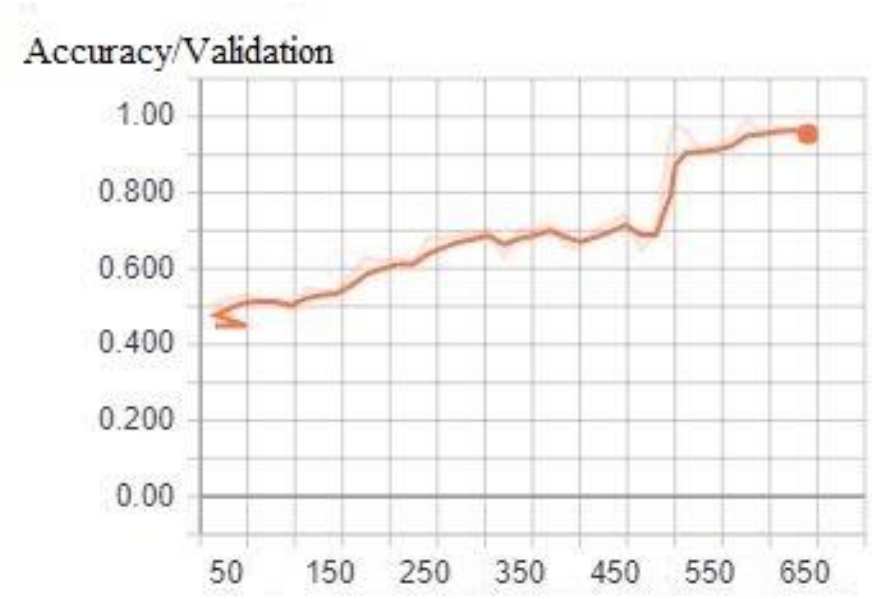


Figure 4. 2: Accuracy per Validation Vs Training Steps

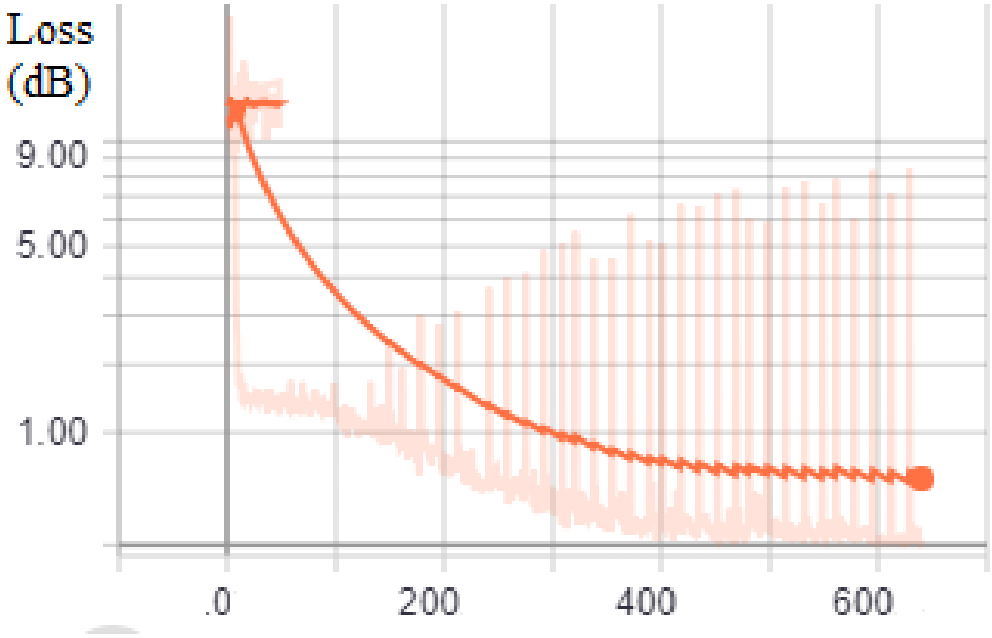
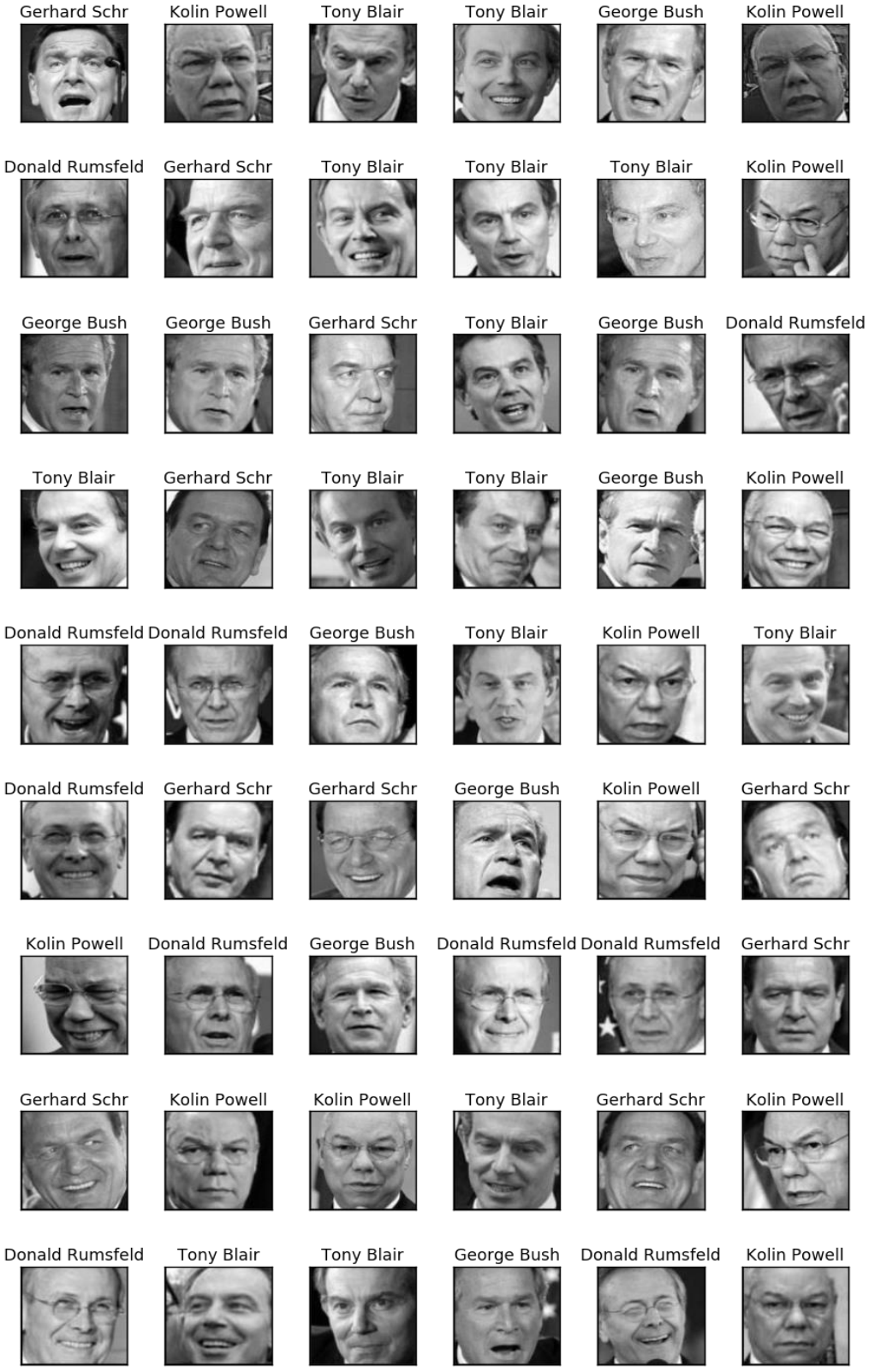


Figure 4. 3: Loss Vs Training Steps

Figure 4. 4: Random Test Results of 5 - Class Classification

All the images have been recognized with full accuracy.

4.1.1 Total Test Results for 5 - Class Classification



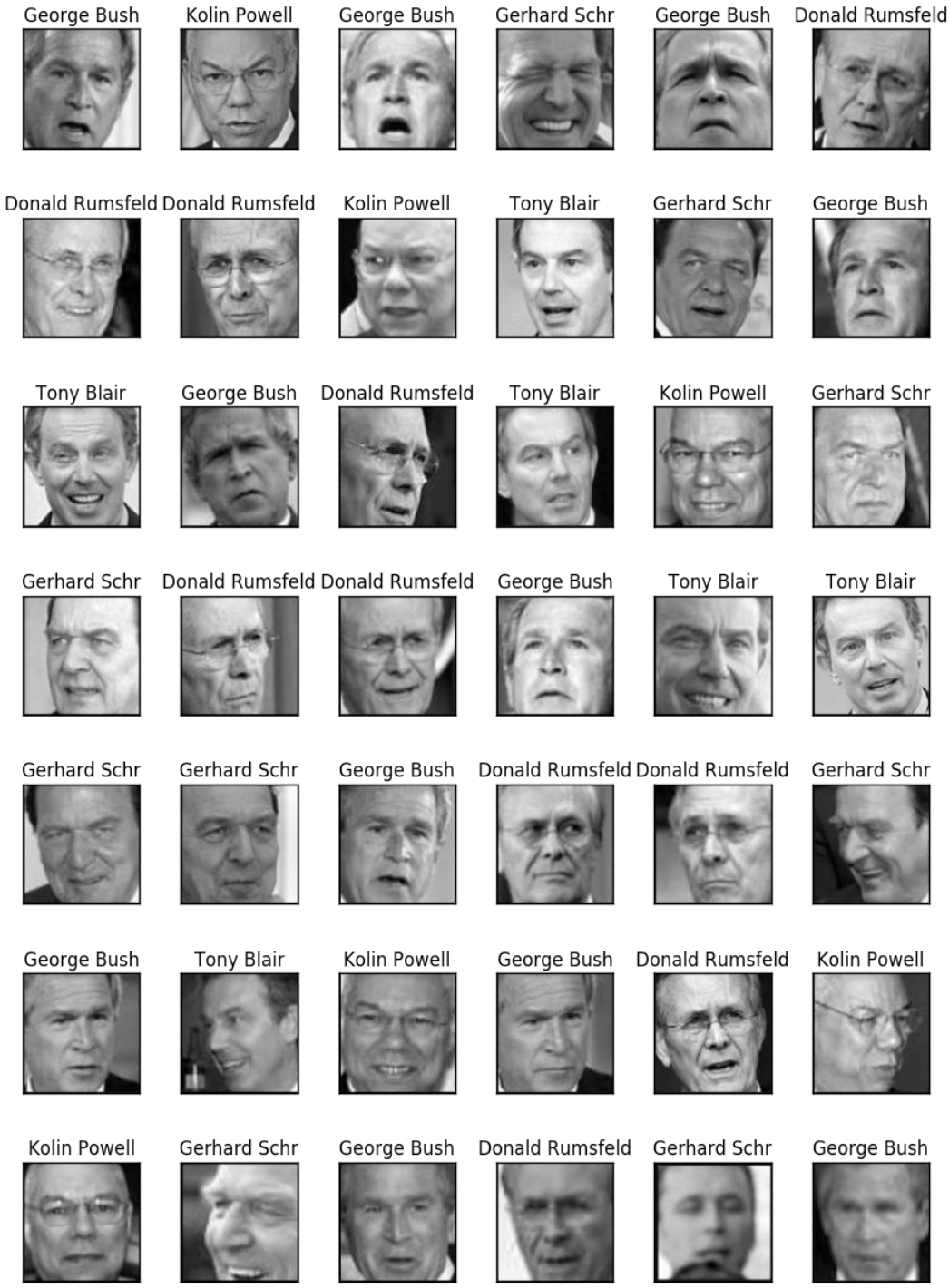


Figure 4. 5: Total Test Results of 5 - Class Classification

4.1.2 Confusion Matrix for Performance Evaluation of 5 – Class Classification

Table 4. 1: Confusion Matrix for Performance Evaluation of 5 – Class Classification

		Predicted Class				
Actual Class		Colin	Donald	George	Gerhard	Tony
	Colin	16	0	1	0	1
	Donald	0	20	0	0	0
	George	0	0	20	0	0
	Gerhard	0	0	0	18	0
	Tony	0	0	0	0	20

The Confusion matrix parameters are:

a) Accuracy = $\frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{True\ Positive(TP) + True\ Negative\ (TN) + False\ Positive(FP) + False\ Negative(FN)}$

b) Sensitivity = True Positive Recognition Rate = $\frac{True\ Positive\ (TP)}{True\ Positive(TP) + False\ Negative(FN)}$

c) Specificity = True Negative Recognition Rate = $\frac{True\ Negative\ (TN)}{False\ Positive(FP) + True\ Negative(TN)}$

d) Precision (Exactness) = $\frac{True\ Positive\ (TP)}{True\ Positive(TP) + False\ Positive(FP)}$

e) Recall (Completeness) = $\frac{True\ Positive\ (TP)}{True\ Positive(TP) + False\ Negative(FN)}$ = Sensitivity

For class Colin:

a) Accuracy = $\frac{16+78}{16+78+0+2} = 0.98$

b) Sensitivity = $\frac{16}{16+2} = 0.89$

c) Specificity = $\frac{78}{0+78} = 1$

d) Precision = $\frac{16}{16+0} = 1$

e) Recall = $\frac{16}{16+2} = 0.89$ and so on.

4.2 Seven – Class Classification

- Total Samples of images: 1225
- Number of training samples: 1125
- Number of Validation Samples: 100
- Number of test samples: 106
- Number of Epochs per training: 5
- Learning Rate: 0.001
- Optimizer: Adam Optimizer
- Accuracy: 0.8980
- Loss: 0.82535

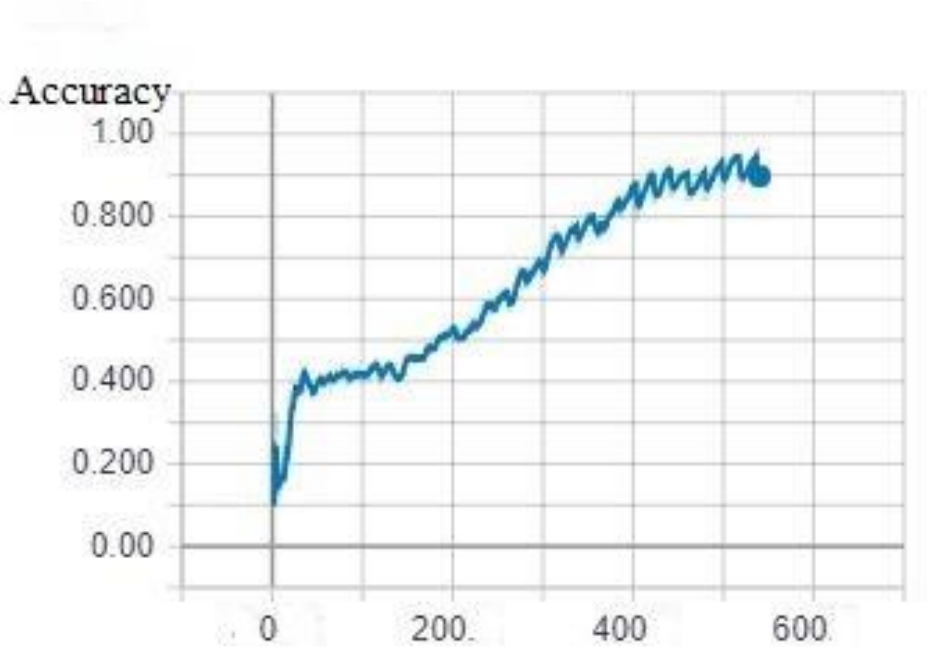


Figure 4. 6: Accuracy Vs Training Steps for 7 - Class Classification

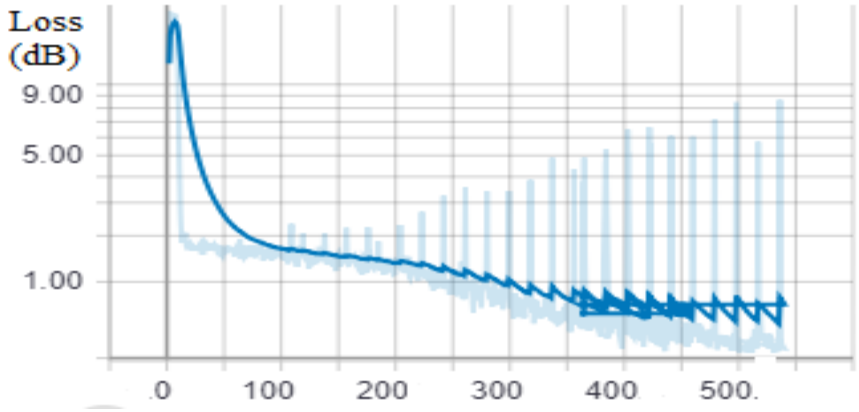
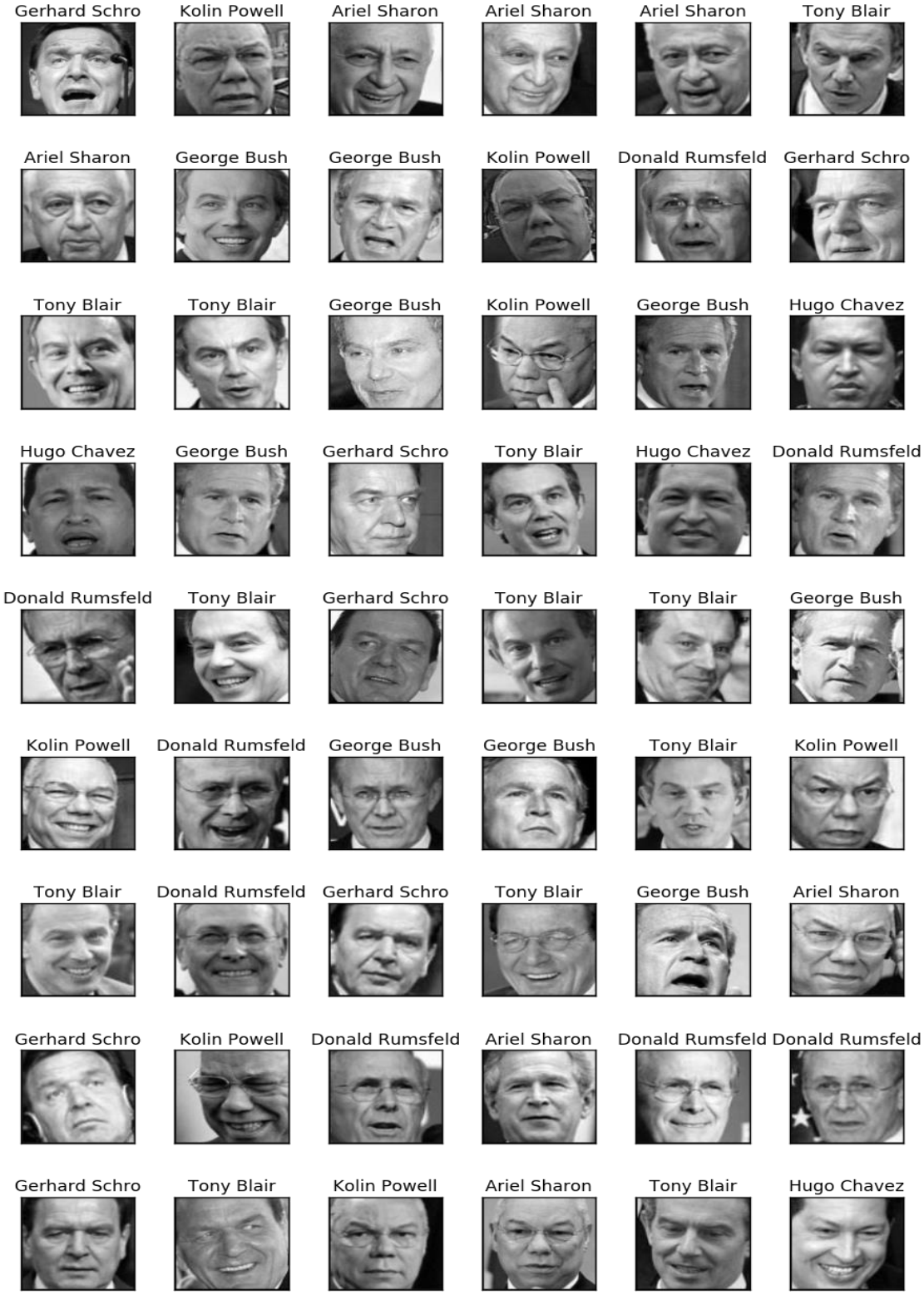


Figure 4. 7: Loss Vs Training Steps for 7 - Class Classification

Figure 4. 8: Random Test Results of 7 - Class Classification

4.2.1 Total Test Results for 7 - Class Classification



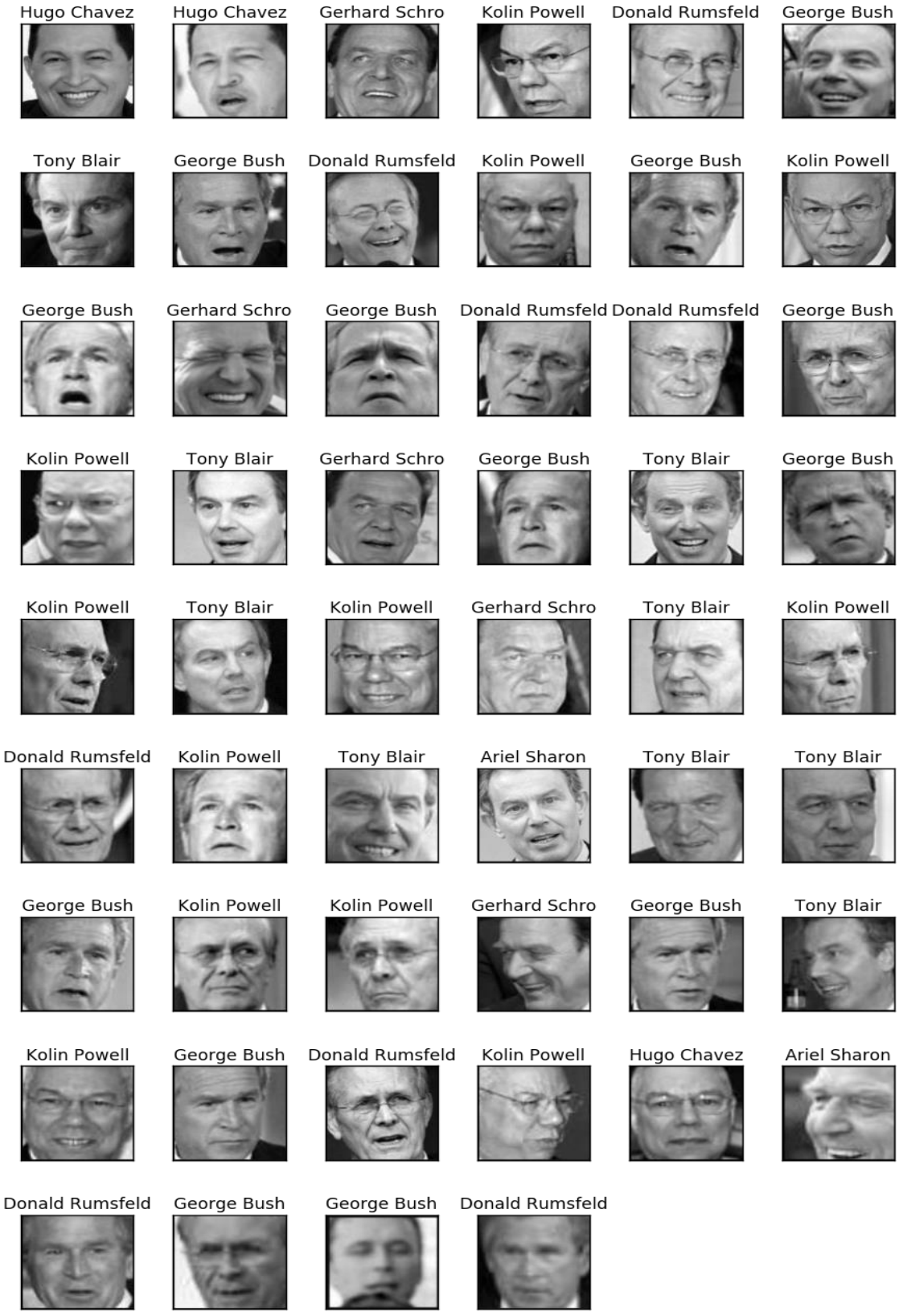


Figure 4. 9: Total Test Results of 7 - Class Classification

4.2.2 Confusion Matrix for Performance Evaluation of 7 – Class Classification

Table 4. 2: Confusion Matrix for Performance Evaluation of 7 – Class Classification

		Predicted Class						
		Colin	Donald	George	Gerhard	Tony	Ariel	Hugo
Actual Class	Colin	15	0	2	0	0	1	1
	Donald	4	13	3	0	0	0	0
	George	1	2	16	0	0	1	0
	Gerhard	0	0	0	13	4	1	0
	Tony	0	0	3	0	15	1	0
	Ariel	0	0	0	0	0	4	0
	Hugo	0	0	0	0	0	0	6

The Confusion matrix parameters are:

a) Accuracy = $\frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{True\ Positive(TP) + True\ Negative\ (TN) + \ False\ Positive(FP) + False\ Negative(FN)}$

b) Sensitivity = True Positive Recognition Rate = $\frac{True\ Positive\ (TP)}{True\ Positive(TP) + False\ Negative(FN)}$

c) Specificity = True Negative Recognition Rate = $\frac{True\ Negative\ (TN)}{False\ Positive(FP) + True\ Negative(TN)}$

d) Precision (Exactness) = $\frac{True\ Positive\ (TP)}{True\ Positive(TP) + False\ Positive(FP)}$

e) Recall (Completeness) = $\frac{True\ Positive\ (TP)}{True\ Positive(TP) + False\ Negative(FN)}$ = Sensitivity

For class Colin:

a) Accuracy = $\frac{15+87}{15+87+5+4} = 0.91$

b) Sensitivity = $\frac{15}{15+4} = 0.79$

c) Specificity = $\frac{82}{5+82} = 0.94$

d) Precision = $\frac{15}{15+5} = 0.75$

e) Recall = $\frac{15}{15+4} = 0.79$ and so on.

5. Work Schedule

Table 5. 1: Work Schedule

Activities	1st Week	2nd Week	3rd Week	4th Week	5th Week	6th Week	7th Week	8th Week	9th Week	10th Week
Material Collection										
Preprocessing										
Implementation										
Testing										
Output Analysis										
Documentation										

6. Discussion

Over the conventional machine learning algorithms like PCA, SVM, convolutional neural networks have shown higher accuracy in the face recognition problems. This system can classify the images with 96.44% accuracy for five individuals which have the significant number of training images. When the output classes have been extended from 5 to 7 with the number of training images being low for the extended classes, the accuracy of the system has been decreased. The convolutional neural networks are highly accurate when the number of training samples is very large. CNN is also not invariant to rotation and scaling. Despite having good performance compared to conventional learning algorithms, CNN has high computational cost and requires good GPUs for training complex datasets.

A good choice for face recognition system can be the Capsule Networks which require a smaller number of training samples. These networks are invariant to translation, rotation and scaling effects.

References

[1] A. C. P. V. Yoshua Bengio, "Representation Learning: A Review and New Perspectives".

[2] Y. LeCun, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Comput,* vol. 1, pp. 541-551, Dec. 1989.

[3] I. S. a. H. G. A. Krizhevsky, "ImageNet Classification with Deep Convolutional Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1-9, 2012.

[4] A. Z. Karen Simonyan, "Very Deep Convolutional Networks for Large Scale Image Recognition," in ICLR, 2015.

[5] A. U. O. Y. Y. D. Musab Coskun, "Face Recognition Based on Convolutional Neural Network," *IEEE*, 2017.

[6] K. S. S. K. R. Sharma S, "FAREC - CNN Based Efficient Face Recognition Technique using Dlib," in *Advanced Communication Control and Computing Technologies (ICACCCT)* , 2016.