National College of Ireland

DeepFake Videos Detection Using Machine Learning

MSc Research Project
Data Analytics

Nikhil Reddy Byreddy
Student ID: 17136563

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Nikhil Reddy Byreddy
Student ID:	17136563
Programme:	Data Analytics
Year:	2019
Module:	MSc Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	12/12/2019
Project Title:	DeepFake Videos Detection Using Machine Learning
Word Count:	7090
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	12th December 2019

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

DeepFake Videos Detection Using Machine Learning

Nikhil Reddy Byreddy
17136563

Abstract

With the advancement of technology and machine learning the feasibility for people to produce fake videos and spread fake news is becoming more and more easy especially with the creation of GAN models which are capable of producing accurate enough fake videos. The main agenda of this project is to help detect these fake videos from real using machine learning techniques. In this project the data from faceforencis was considered and frames are generated from videos followed by application of novel technique in classification by using face detection over videos. These images are transformed using various parameters and models are compared against them. Highest accuracy of 95% with AUC value of 0.95 is obtained using laplace transformed images. Quantum machine learning is also applied and was able to train the model with 50% less time than classical model.

1 Introduction

Creating a fake video and images is nothing new in today's era. It has become one of most discussed topics in digital society. No doubt a simple image or photo can speaks a lot about a person but an audio or a video recording are more persuasive than images (chesney and citron; 2019). Imagine a video which shows us the some political discussion of some renowned person and this video could be manipulated using some tools which is so much persuasive to make us believe that its real. The increasing rate of the digital technology can convert the worst nightmare into reality. According to the recent research, in the last 9 months there is a rise of the fake videos to double the number. Researchers of a cyber-security company have found out that this year 14,698 deepfake videos was published online out of which 96% was replacing the faces of the celebrity for sexual activities(Jones; 2019). Due to the increasing rate of the fake content in digital media leads to a questionable form whether the image or video we are watching or the audio we are listening is fake or real.

1.1 Motivation and Project Background

On the year 2017 there was a face switching of a celebrity to a porn actor. When videos are created with false speeches of the world leaders then the security of the country is in risk like a video of former U.S president was released which clearly explains about the video manipulation (Bloomberg; 2018).Face swap also known as deep fakes is one of the big steps in AI development. DeepFake are one of the visualization techniques which changes the face to one person in other person face (*Deepfakes github*; 2018). It is very difficult to recognize which is fake and which is real videos. So the intelligent systems i.e.

deep learning techniques comes in picture to distinguish between them (pothabattula; 2019). There are already many cases seen where the fake videos and images of celebrities and political figures have been released by people which creates wrong impression about them. For creating the realistic images and video, deepfake algorithm requires a huge number of images and video data to train a model. Videos and images of celebrities and politicians are easily available in large amount, so they are the main predators of the deepfake techniques (Nguyen et al.; 2019). According to one of the report by tucker (2019), even satellite images of earth can be faked to misguide the troops by creating fake bridges and road. To overcome such unexpected circumstances GANs or Generative adversarial nets are used to identify the objects and confirm whether its real or fake.

1.2 Research Question

RQ: "Detection of deepfakes videos using machine learning can be obtained how? If so by how much accuracy?"

Sub RQ: "How much training time does quantum machine learning saves when compared to classical neural network model?"

1.3 Sub Research Question

Following are the research objectives of the project to solve the research questions

1. Literature review on classification of fake/real videos
2. Generate frames from videos
3. Detect face in image and generate new images
4. Transformation on images
5. Extraction of arrays from images
6. Implementation of machine learning models
7. Evaluation of models

1.4 Research Objective

Obj1: Literature review of the deepfakes, quantum machine learning and different machine learning techniques for video classification(2010-2019)

The rest of the paper is organised in the sequential order described below.The next section closely examines the previous work done on deepfakes and quantum machine learning. In the segment 3 of the paper the methodology and the process which gives a details study about the dataset extraction, pre-processing steps and the data mining techniques used is discussed. In the section 4 of the paper overall discuss about the implementation, evaluation process and the final outcome of the project is being shown. The last chapter 5 completes the paper by giving a overall conclusion and future work related to it.

2 Related Work

This discussion in the this section is related to the prior literature review of the video classification.The sections are divided into many other subsections that are 1)Details study about deep fakes 2)Quantum machine learning and its uses 3)Literature survey on video classification using different machine learning techniques 4) Comparison and summary

2.1 Details study about deep fakes

Deepfake is one of the trending topics which attracts many researches recently. According to (Nguyen et al.; 2019), deepfake is coined from deep learning and fake, which means it is one of the algorithms which is used for swapping the faces from one from to the other by deep learning techniques which is unidentifiable by human eyes. Many difficult problems like big data analytical problems to a small visualization is being solved by deep learning. This deep learning techniques is sometimes hazardous for the society questioning the security of the country. So, it becomes necessary to review the history of the deepfake and techniques to detect the deepfake. It is said that whatever is broken by AI is fixed by AI itself.The deep learning techniques helps in identifying the selective and notable features of the videos and images which distinguishes the deepfake i.e. the original video and the changed video. The exposure of the deepfake is divided into two sections which are namely fake image detection and fake video detection. The fake video detection is again divided into visual artifacts within frame which is uses deep classifiers and temporal features beyond the frame where deep learning recurrent classification models are used. To reduce the impact, several big companies are using their own filtering techniques to remove deepfakes.

According to an article published by (hui; 2018) explains clearly about the mechanism which is involved in deepfakes. Deep learning is mainly used for high-dimensional data which is mostly used for compressing the images and for dimensionality reduction. To create a fake image there should be almost thousand images of both the source and the target. An autoencoder-decoder structure is been created which helps in the process of making false images. Using deep learning technique i.e. CNN an autoencoder is used which extract the latent features like facial expression, skin-tone, eye color, face angle etc of the face and then the decoder is used to reconstruct the image. As it is very difficult for the encoder to remember all the features, so it only obtains those features which are most important to reconstruct the real image. So, when we try to recreate the images the decoder takes the features from the real images and just generate the newly generated tampered image. During swapping of the faces we use two encoder and decoder pairs where the features extracted from the encoder is being divided between two network pairs. In this process the work of the common encoder is to search and train itself with the similar features of the both the faces which is not very hard as all faces have common features like nose, mouth position, eyes etc. It is very important for the encoder to get proper features else it may create noise or distortion in the result.

Rossler, Cozzolino, Nießner, Verdoliva, Riess and Thies (2019) have explained in the paper that the fake face creation can be segregated into two parts, one is facial expression manipulation and other is facial identity manipulation. In former it displaces the facial expression of one face to another face. The most popular techniques which is used is Face2Face which is a real time facial recreation of a monocular target video order.

The expressions are being collected by an RGB sensor which is transferred online to the target. The target video are manipulated in a photo-realistic fashion which is visually not possible to notice the changes (Thies et al.; 2018). The other false face creation is identity manipulation is changing the identity which replaces the face of one person with other. Also known as face swapping. Deepfake does these swapping with the help of deep learning techniques. The graphic based techniques used for the manipulations are Face2Face and FaceSwap which are popular and the deep learning approaches are the Deepfakes and the neuralTextures. In this paper it even discusses how to automatically detect the changes made in a video or images. The advancement in the deep learning techniques helps to learn the image features with CNN and overcome the detection problem by the training the network in a supervised manner. With random dimensions and compressions, a benchmark has been proposed from this four-manipulation process which was never done before.

Many researchers have experimented different methods for detections of deepfake videos. In one of the researches done by Parkhi et al. (2015), where the main purpose is to verify and identify different face alignment and metric learning from the large dataset which has around two million faces collected from various sources. This is done by CNN architecture which even filter the unnecessary details. A similar type of research is done by Schroff et al. (2015), where it speaks about the FaceNet where using deep convolutional network learns the Euclidean embedding per image. The convolutional network is trained in such a manner that the square L2 distances in the embedded space is directly proportional to the face similarity i.e. the real face has small distances whether the different faces have large distance. The face is verified by checking the distance between two different faces which in turn becomes a KNN classification problem. The approach is verified using different dataset like labelled faces in the wild(LFW) which achieved 99.63% accuracy and youtube faces DB obtained 95.12% accuracy.

One of the research studies done by Korshunov and Marcel (2019) the replacement of faces is done by pre-trained algorithm known as GAN (Generative adversarial network) which is popularly known as Faceswap-GAN(Nguyen et al.; 2019). GAN was first proposed by Warde-Farley et al. (2014), where a detailed study of the adversarial process was done in which two models are being trained. A generative model which holds all the distributed data and a discriminative model which calculates the probability of samples generated from the training data. The generative model is trained in such a way so that it can increase the probability of the discriminative model from committing any mistakes. So this model leads in generative the deepfake which ultimately effects the quality of the video.

According to Korshunov and Marcel (2019), the face recognition algorithms like VGG and Facenet neural network does not work properly on deepfake videos and fails to distinguish between original and tampered video with a 95% of error rate. The lip-sync based algorithms even failed in detetcting the mismatch between speech and the lip movement. But, the image based approach with the SVM classifier is figured out to give 8.97% error rate on deepfake videos. It is therefore concluded that image-based approaches have higher percentage of accuracy in recognizing the deepfake videos than rest of the methods. Still it is challenging for advanced face swapping techniques.

DeepFake are even generated in the field of virtual reality. Both the end i.e. the target and the source are fed into the generator network as two different input which remake the original image. An additional discriminator network is used in a GAN to separate the real input from the forged output.Source and target are trained separately via gen-

erator network which includes both the encoder-decoder module. After this process the virtual fakes can be easily be generated by using source images as input in the target generator.(Bose and Aarabi; 2019)

2.2 Quantum Machine Learning and its uses

Quantum machine learning is the association of quantum computing and the machine learning concepts, which means the machine learning algorithms are implemented on quantum based circuits of a quantum computer(Ying; 2010). Quantum computer were first conceptualised by Nobel Laureate physicist Feynman in 1982(Feynman; 1982). He researched that certain quantum phenomena can not be simulated by classical computers without an exponential decrease in the performance.

In 1985 on the basis of the Feynman's ideas Deutsh(Deutsch; 1985) proposed his Quantum Turing Machine. Basically Deutsch explained the technique of quantum parallelism which in turn used the principle of superposition from quantum mechanics. Using this superposition principle the Turing machine was able to encode a huge number of inputs in less memory and also can perform the calculations simultaneously on all the inputs. After Deutsch in 1994 Shor(Shor; 1994) made a advanced effort in quantum computing which explored the power of quantum parallelism and created a polynomial time algorithm for prime factorisation which took exponential less time compared to classical computers. Next in 1996 Grover (Grover; 1996) found the quantum algorithm to find a single stored element from unsorted database in a square root of time of that taken by classical computers. From here the quantum computing became the most exciting field of science.

In the following section lets see the basic operation of Quantum computing which makes it faster and efficient. To understand this will consider the Deutsch–Jozsa algorithm (James; 2001)he basic unit in Quantum computing is called as Quantum bits which is denoted as 'qubit'. Qubit has the state of both 0 and 1 at the same time for e.g. the horizontal and vertical polarisation of photon[]. Mathematically the qubits are represented by a unit vector , in the Dirac notation as follows,

$$|\Psi| = \alpha_0|0\rangle + \alpha_1|1\rangle \tag{1}$$

where $|0\rangle$ and $|1\rangle$ two basis states, and α_0 and α_1 are complex numbers with

$$|\alpha_0|^2 + |\alpha_1|^2 = 1 \tag{2}$$

The states $|0\rangle$ and $|1\rangle$ are called computational basis states of qubits and they correspond to the two states 0 and 1 of classical bits. The number α_0 and α_1 are called probability amplitudes of the state $|\Psi|$. The main and advantageous difference between classical and quantum is that, the qubits can be superposition of both $|0\rangle and|1\rangle$as in Eq 1.To perform mathematical operations we need to know the state at any point of time just like classical computer to enable this α and β are subjected to normalization that is the square and sum of these amplitudes at any point of time is equal to 1 i.e, probability of 1(Debnath et al.; 2016).An example state of qubit is:

$$|\Psi| = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \tag{3}$$

As the qbits can have both a 0 and 1 at a same time two qbits can represent 4 states and three qbits can represent 8 states. So by this nature the qbits can perform faster and calculate more efficiently.
The two concepts which makes the quantum computing faster are Superposition and entanglement.
Superposition is the concept where qubits can represent two states at same time that is state 0 and 1. Both states are superposed which helps in representing more values with the less variables , which makes the calculations faster than normal theory of 0/1(Ying; 2010).
Entanglement is the major and crucial concept of quantum computing. This is the state where the two qubits can not be represented by the vector product of two qubits. This state gives another information or the value which can also represent a value rather than normal multiplied product, like this the physical resources in quantum computing are used extensively and information processing is made faster.(Ying; 2010)

In 2019 the authors Fastovets et al. (2019) conducted the experiments to investing on machine learning methods in quantum computing theory. They used the IBM's quantum processor. According the authors quantum machine learning is the approach which is hybrid approach including the classical and quantum algorithms. Quantum approaches are used to analyse the quantum state while, the quantum algorithms will improve the efficiency of the classical data science algorithms exponentially. In this paper they implemented the classical K means algorithm using the Quantum minimisation algorithm and SWAP test (Mengoni and Di Pierro; 2019) Also implemented the tree tensor networks which help in implementing quantum machine learning algorithms. After the implementation they found that the algorithms with quantum computing worked exponentially faster than the normal ones. Hence the usage of quantum machine learning is the future of machine learning to solve the extensively large data set and predict the answers

2.3 Literature survey on video classification using different machine learning techniques

In the past few years many scholars and researchers have worked on the image and video classification using different machine learning techniques. One of the most powerful which is used by many researchers is CNN which is helpful in detecting the forgery image. In one of the article,Singh (2019) mentioned that IEEE information forensics and security technical committee conducted a challenge by providing an open dataset to detect the forged and pristine image. CNN are used to extract the useful features with minimal loss function. The parameters like gradient descent are given to the network which finally performs the classification functions. CNN does the similar work as visual cortex. The same is even used by Ramprasath et al. (2018) where they performed CNN considering Digit of MNIST data set as a benchmark for classification, obtaining an accuracy of 98%.
According to (Guera and Delp; 2019),machine learning free software tool helps in creating manipulated videos.Here the two staged analysis in done on the collected data.CNN is used to obtain the frame-level features which in return helps to train RNN(Recurrent neural network) to segregate whether the video is tampered or not. RNN picks up the temporal divergence between the frames in the process of swapping face. Here LSTM which is one type of RNN is used for processing the order of the frame. A softmax layer is used which calculates the probabilities of the frame sequence whether it is fake or real.

The LSTM unit acts as an intermediate layer which is being trained without any auxiliary loss function. A suspected video is seen to give a 94% more accuracy in detecting the manipulation of videos than other detector.
The firmness of the pattern in any kind of data can be checked by machine learning techniques which helps us to know whether the data have been manipulated or not. But for determining it a good and large number of datasets is required so the video tampering is not seen properly in small dataset (Johnston et al.; 2019).A self-supervised method which uses Siamese neural network which detect whether the pixel patches have same image metadata, or same image pipeline or only have a part of the original image. An auto recorded EXIF meta-data of the real images was trained to check whether the image has been tampered or its still the original one i.e the content of the image is from single imaging pipeline or its different. Three processing techniques has been used i.e. re-scaling, Gaussian blur and JPEG compression and 8- features from EXIP metadata was generated. This was a new research done in the field of image tampering localisation(Liu et al.; 2018).
Due to the advancement of the AI, the fake digital data have rose to great heights. It has the capability to change the actual content of the video, audio or image and give a false impression. So, it is very necessary to trace back and find out the provenance of digital media. If the proof of authenticity (PoA) of the digital content is found out then it will be easier to remove the suspicious content. According to the research done by (Hasan and Salah; 2019), a novel method (blockchain based solution)was introduced which provide a framework using Ethereum smart contracts to track and find out the original source even if the content was tampered multiple times. It uses the hashes of interplanetary file system(IPFS) which stores the digital content. If the origin of the source is trustworthy and reputable then the content can be proved to be real and legitimate.

2.4 Comparison and summary

From the above discussion about the state-of-art related to the deepfakes and quantum machine learning we can come into the conclusion that there are very few methods through which we can perfectly find out whether it the digital content is genuine or if it is tampered one. In most of the papers, deep learning techniques i.e. CNN is used for finding out the difference between fake and real images. In Rossler, Cozzolino, Nießner, Verdoliva, Riess and Thies (2019) a benchmark was set up with 1000 additional videos which was evaluated using low quality versions of selected trained models and a result was found out in different detection methods. One of the benchmark dataset for understanding the video is YouTube-8M. The videos are filters out by manual or automated curation strategies.Deep CNN is used to extract hidden representation. The features are being compressed and available for the download.(Abu-El-Haija et al.; 2016)In the later section a detailed study about the methodology and the ways in order to get a model which helps in finding out the fake and real videos is discussed.

3 Methodology

3.1 Introduction

As this project in a basic sense falls under the section of data mining and data science, one of the mostly used methodology in this category KDD is chosen for this project.

The reason for choosing KDD over CRISP DM is because usually CRISP DM ends with deployment of the project as it is designed to suit the business applications and in KDD this is not a mandate step to finish which really suits for this project. In the following sections you can find the details about methodology used and also about the design process used to implement this project.

3.2 KDD methodology of Deefake video detection

The modified KDD methodology for this propject is inspirted from Azevedo and Santos (2008). Here The data of 1) Real and manipulated (fake) videos data of 100 videos each is sourced from Faceforencis data set. 2) In the next stage of feature extraction images are extracted from videos 3) From the extracted images the faces are detected and extracted as new images and applied few transformation (laplacian, canny edge and adaptive mean) techniques 4) In the next stage the arrays are extracted based on image pixel values using computer vision libraries. 5) From the extracted arrays neural network models are implemented. 6) In the last stage the results of implemented values are evaluated based on ROC curve and the best model is chosen.

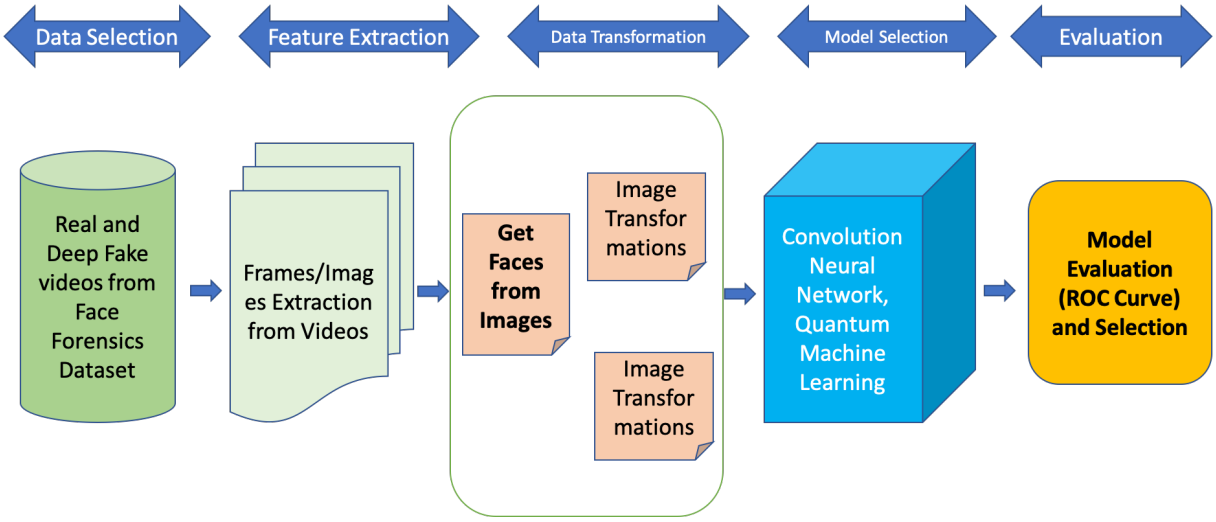

Figure 1: KDD Methodology Of DeepFake Videos Detection

4 Design Specification

For this project a 2 tier design process is implemented. From 2 the 2 tiers are named as 1) Presentation Layer and 2) Business Logic Layer.

4.1 Project Design Process Flow

In the presentation layer the images, all the transformations on images and model implementations and evaluations are visualized using matplotlib and CV libraries in python. In the business logic layer the process of implementation of project is outlined step wise and it is implemented using GPU provided by Google in its colab environment and to implement quantum machine learning model "strawberry fields" library is used.

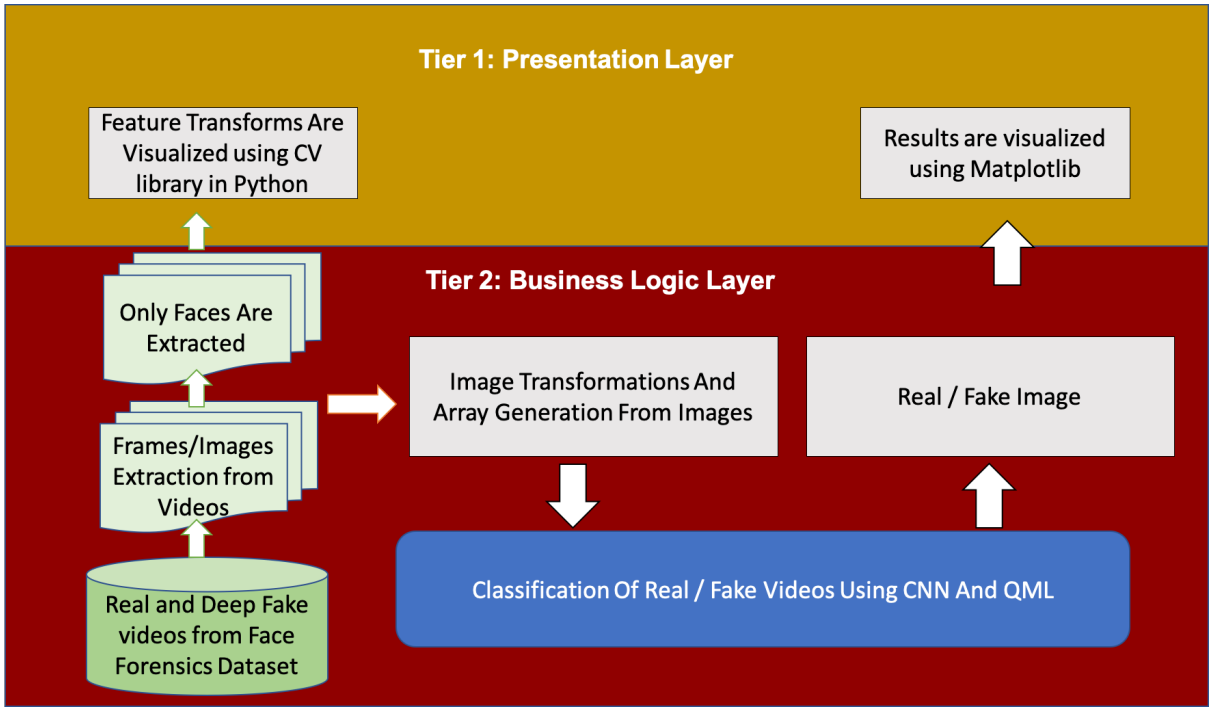

Figure 2: Project Design process Of DeepFake Videos Detection

5 Dataset

The dataset chosen for this project is FaceForensics++ [1] FaceForensics++ is a forensics dataset contains over 1000 videos sourced from youtube and few videos are made by hiring actors. But for the purpose of the project and due to the computational expenses only 100 videos each from original and deep fake category is used.

The reason for choosing only 100 videos is mainly because of computational problems with the limited usage access in Colab.

Figure 3: Deepfake Generation

[1](https://github.com/ondyari/FaceForensics/tree/master/dataset)

6 Implementation

In this section a thorough explanation on the extraction of frames, face detection, transformation on images feature extraction is discussed. All models are evaluated and compared using AUC score / ROC curve data and also compared with the models from the literature review. Quantum machine learning model is also compared against the best model and results are outlined in this section.

6.1 Extraction Of Frames From Videos

Using computer vision library from each video a image is extracted for every 1 sec. There by the research objective (2) is successfully implemented. The generation of frames from videos is necessary for this project as to train machine learning models it is absolute to feed it with numbers i.e to generate structured format of data from unstructured format of data (videos). Also the images are generated for every second rather than every micro second or so for two reasons one being to save computational expenses and second to avoid any duplicates of images. For example if from a video for every micro second a image is generated since there won't be much movement in the video it ends up with loads of duplicated images. Since the videos in the dataset are 200 in total and each one is about 15 sec - 25 sec length after extraction of images there are 4072 images.

6.2 Face Detection And Extraction

Using computer vision library a face is detected in the image and the face is extracted out.The image below looks blurred because of the noise clearance done and used low quality videos due to computational problems.

This is the most crucial and novel step implemented in this project. It is the critical step as it removed the unwanted background from the images and there by reduces the size of image as well. The detection of faces from images is done using transfer learning from CV2 library in python. Transfer learning refers to a pretrained model which can be used in new implementations. After the implementation of this resulted in 742 images with only faces. It must be noted that though there are less than 742 real/fake faces the direction difference of face in a image is considered as a new image. There by it can be concluded here that research objective (3) is successfully resolved.

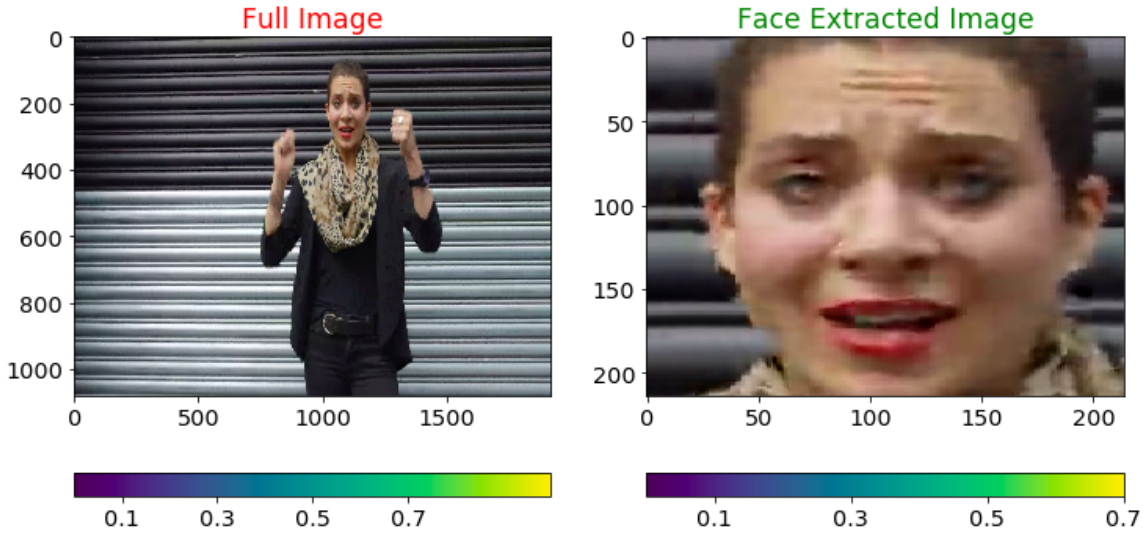

Figure 4: Face Detection And Extraction

6.3 Image Transformation Using Laplace

Images are transformed using laplace function because in this project there are minute differences between the fake and real faces and in real world advanced cameras the images are fine tuned automatically by reduced the noise in the edges to give more appealing images to human eyes. The main purpose of Laplace transform is to highlight these discontinuity of edges. This can be visualised in Figure 5 as the edges of faces are highlighted.

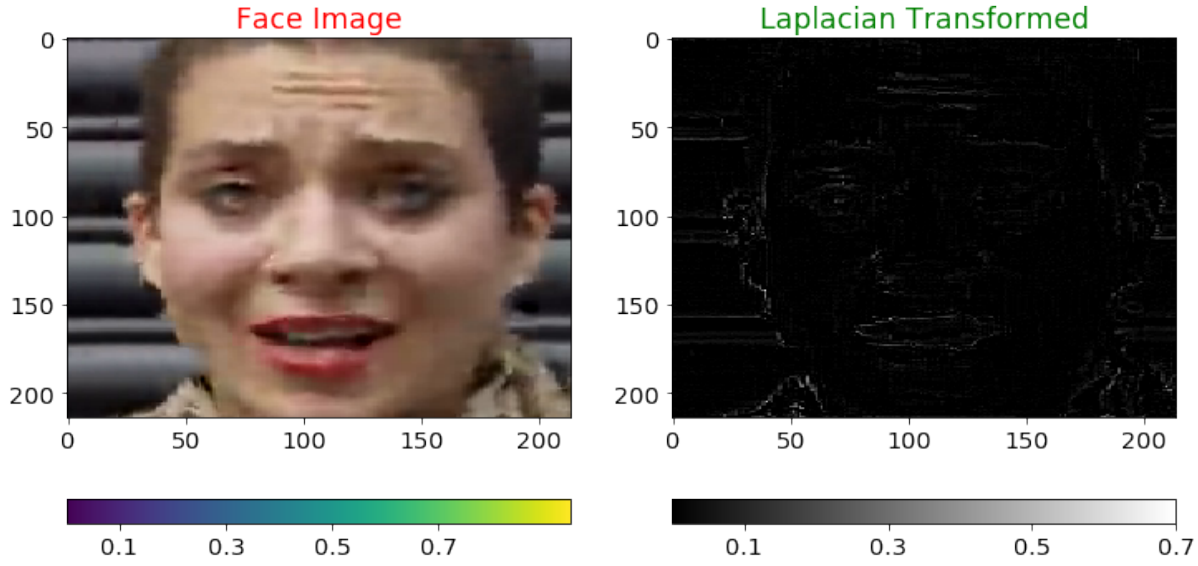

Figure 5: Laplace Transform on Images

6.4 Image Transformation Using Canny Edge

Although the canny edge detection serves almost the same purpose as Laplace transform based on literature it is found that one of these gives has higher advantage to models

in different scenarios. This is because mathematically Laplace transform is obtained by second derivative of the image where as canny edge is obtained on first derivative of an image. From figure 5 and figure 6 it can be seen that there are few differences in the images obtained if closely observed.

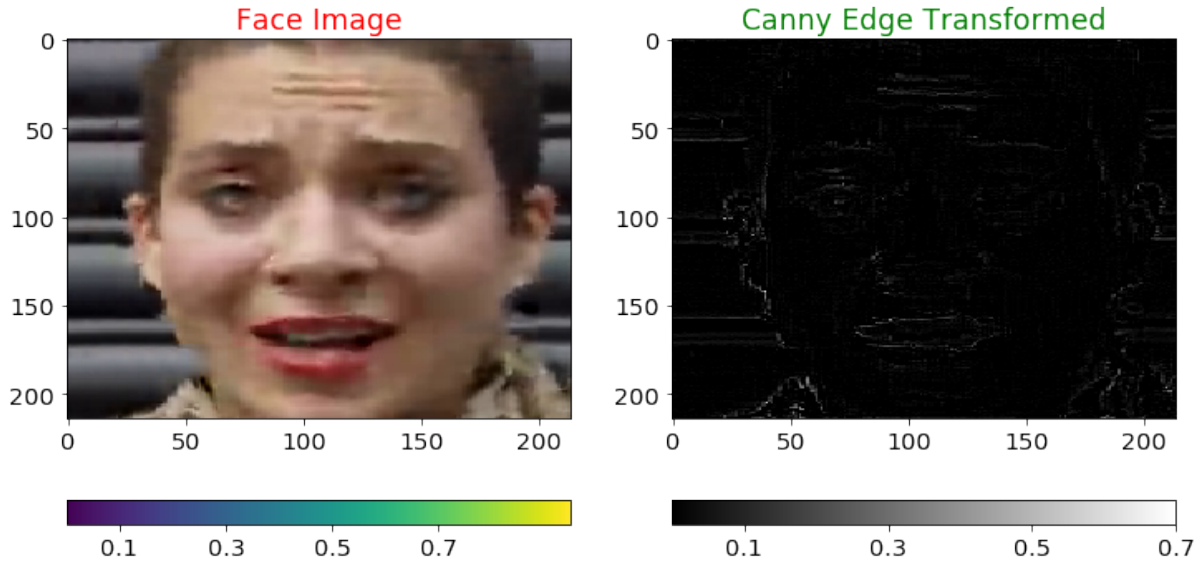

Figure 6: Canny Edge Transform on Images

6.5 Image Transformation Using Adaptive Mean

The purpose of using adaptive mean transformation on images is to highlight the hidden or less intensified pixels in a image. It can be seen clearly in the image 7.

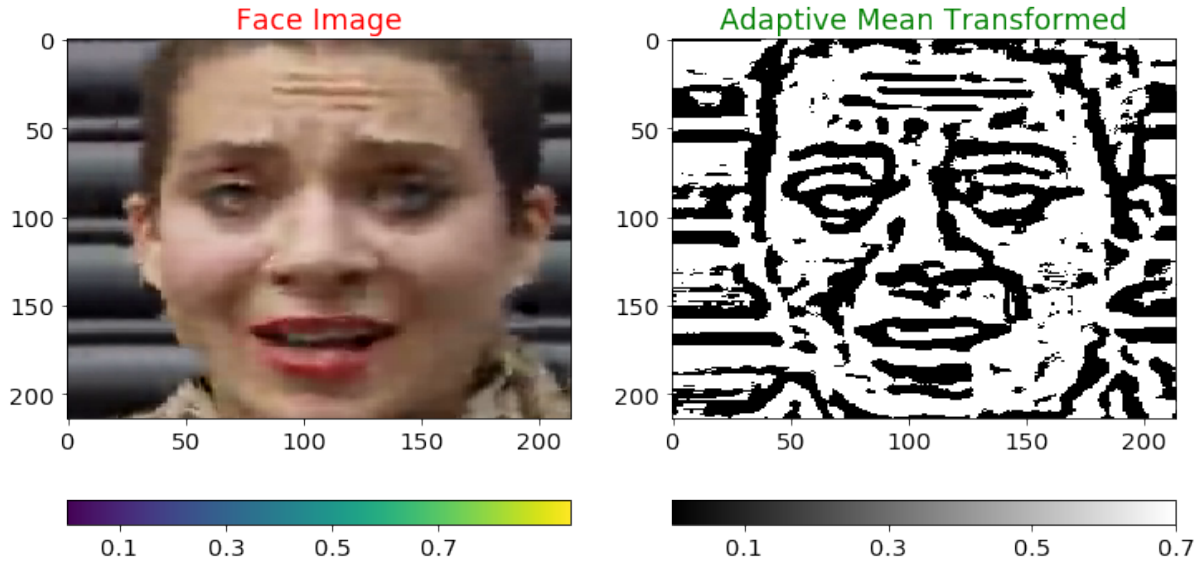

Figure 7: Adaptive Mean Transform on Images

6.6 Extraction of arrays from images

This step is mandatory because the machine learning models/algorithms won't be able to process unstructured data like images with out any transformation. To enable this the image processing technique of extraction of arrays from digital image is done by using computer vision library called CV2 in python.

This is implemented on the following sets of images as listed below and there by resulting in 5 different datasets

1. Raw images
2. Only face images
3. Laplace transform on face images
4. Canny edge transform on face images
5. Adaptive mean transform on face images

There by the research objectives of (4) and (5) are resolved.

6.7 Implementation of classical neural network on all datasets

A CNN of 8 layers are used with first layer to get the input data, second later to apply activation function called 'relu', third layer to apply maxpooling2d to sample down the input by reducing the dimensionality to 2 from 3 and try to find the hidden feature patterns if any by doing so, fourth layer is to reduce the overfitting condition by using a dropout layer,in the next step of 5th layer flattening of the dimensions are done to feed a vector to output of neural network being used and 6th,7th and 8th layers are used to get the learning from all the layers consolidated and get the classification as an output.

In the next stage " model is compiled using loss function of categorical entropy as the project main work is to classify categories of real or fake videos. metrics used to measure and evaluate model training is accuracy. This can be seen here model.compile (loss= 'categorical_crossentropy',optimizer='Adam', metrics=['accuracy'])".

All the data sets are run with 50 epochs and batch size of 20 to maintain consistency of the results and training time.

6.8 Evaluation and Results of CNN on raw images

The data set with array of raw images with out transformation is used here. There are 4712 images and train and test datasets are formed in such a way that 80% of training and 20% of testing sets respectively i.e 3769 of train and 943 of test data. CNN is used to classify images and it resulted in overfitting of data.

It can seen in figures 8 with 100%accuracy and 9 1 AUC,that this is the case of over fitting. It resulted in over fitting because images are generated every second from videos and background is same in each video. The images generated from it have many duplicates in it with minimum differences. To avoid this face detection of images are done and the results can be seen in the next section.

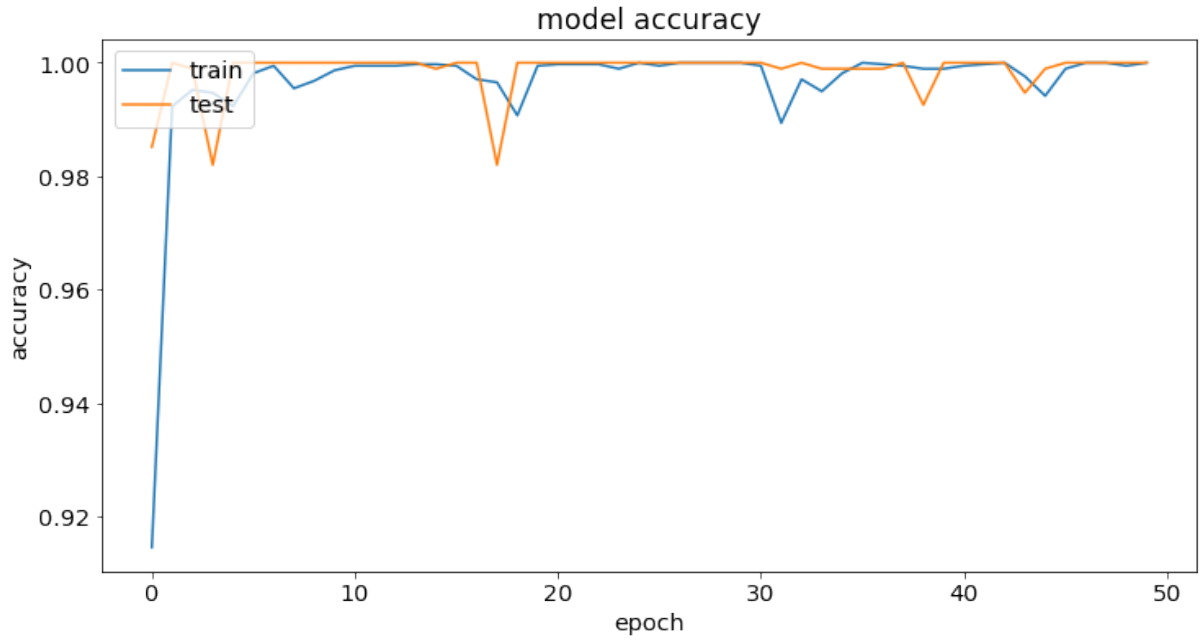

Figure 8: Accuracy With Raw Images - Overfit

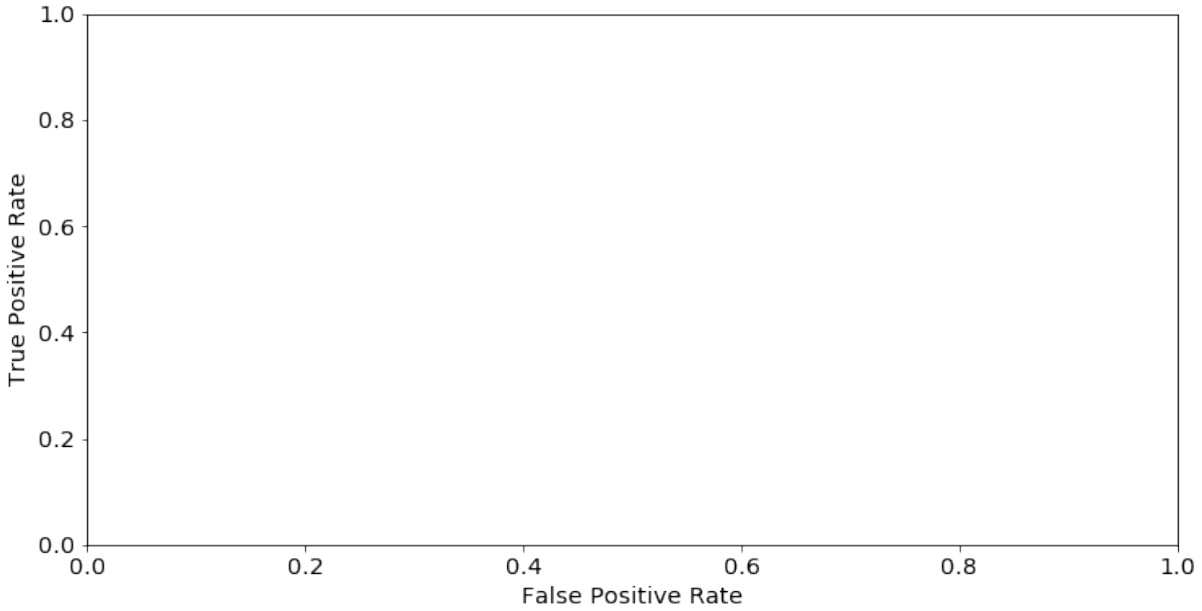

Figure 9: ROC With Raw Images - Overfit

6.9 Evaluation and Results of CNN on face images

The data set with array of face images with out transformation is used here. There are 742 images and train and test datasets are formed in such a way that 80% of training and 20% of testing sets respectively i.e 593 of train and 149 of test data. CNN is used to classify images and it resulted in good accuracy of 92% figure 10 and AUC 11 of 0.95.

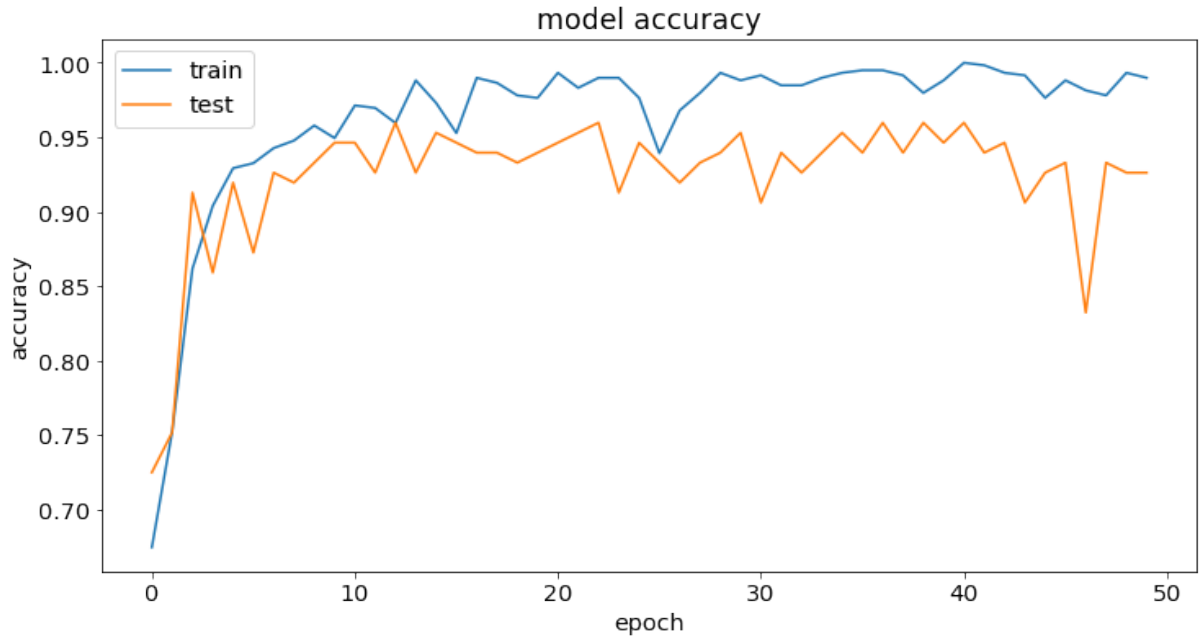

Figure 10: Accuracy With Face Images

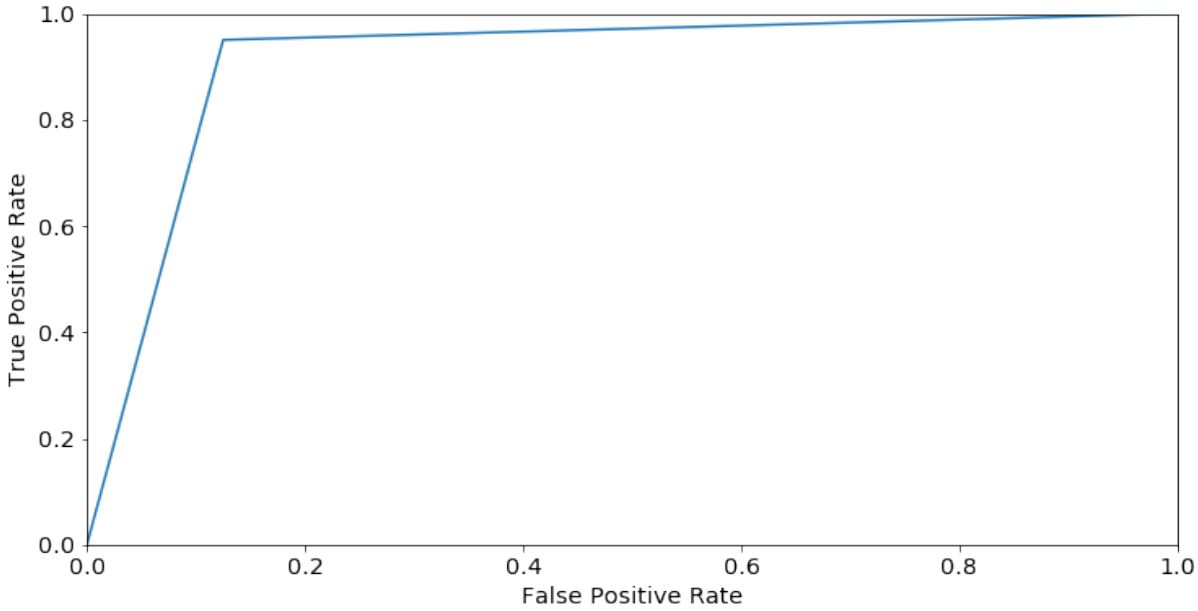

Figure 11: ROC With Face Images

From the above figures it can be concluded that it resulted in good results as the benchmark literature review paper Rössler, Cozzolino, Verdoliva, Riess, Thies and Nießner (2019) had accuracy of 85% and also the novel technique of face detection in classifying images overcame the problem of overfitting.

6.10 Evaluation and Results of CNN on face images with laplace transform

The data set with array of face images with laplace transformation is used here. There are 742 images and train and test datasets are formed in such a way that 80% of training and 20% of testing sets respectively i.e 593 of train and 149 of test data. CNN is used to classify images and it resulted in good accuracy of 91% figure 12 and AUC 13 of 0.96.

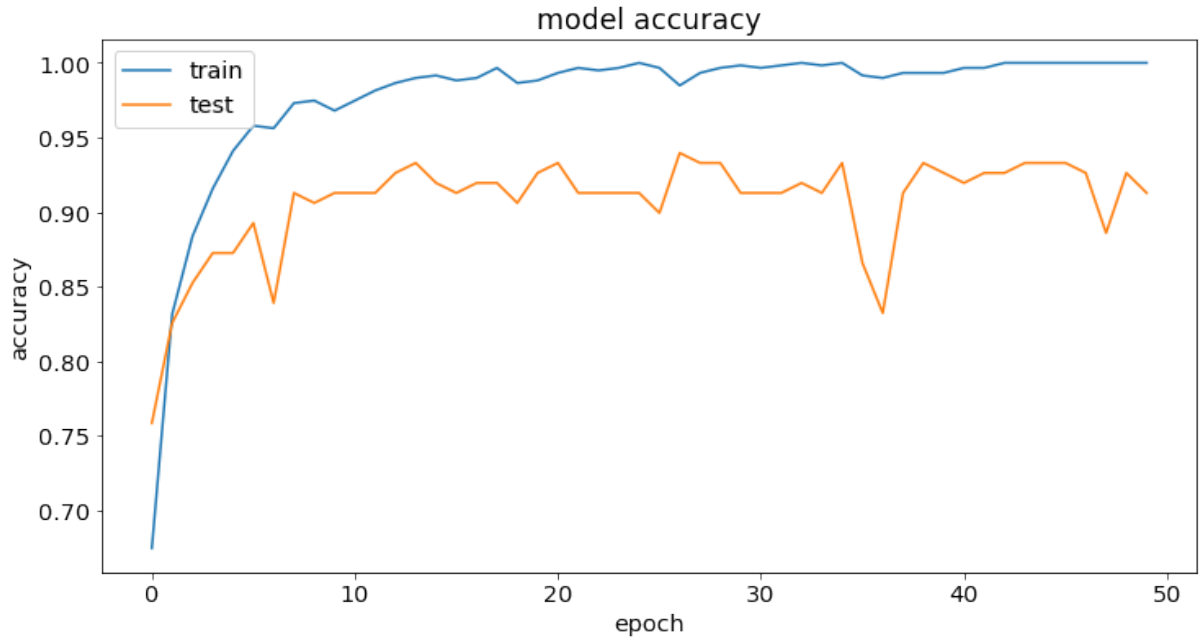

Figure 12: Accuracy With Face Images Laplace

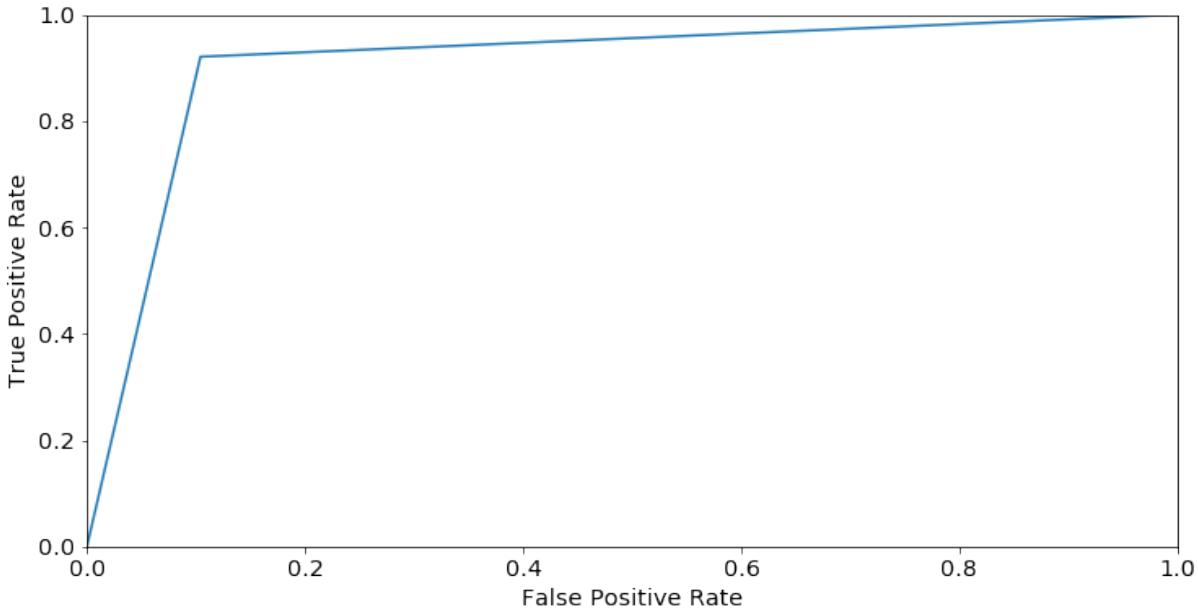

Figure 13: ROC With Face Images Laplace

From the above figures it can be concluded that it resulted in good results as the bench-

mark literature review paper Rössler, Cozzolino, Verdoliva, Riess, Thies and Nießner (2019) had accuracy of 85% and also it have beaten the AUC value of 0.95 using face images. Altough the accuracy is less here by 1% due to a point increase in AUC value this model is determined as best model.

6.11 Evaluation and Results of CNN on face images with canny edge transform

The data set with array of face images with canny edge transformation is used here. There are 742 images and train and test datasets are formed in such a way that 80% of training and 20% of testing sets respectively i.e 593 of train and 149 of test data. CNN is used to classify images and it resulted in good accuracy of 85% figure 14 and AUC 15 of 0.93.

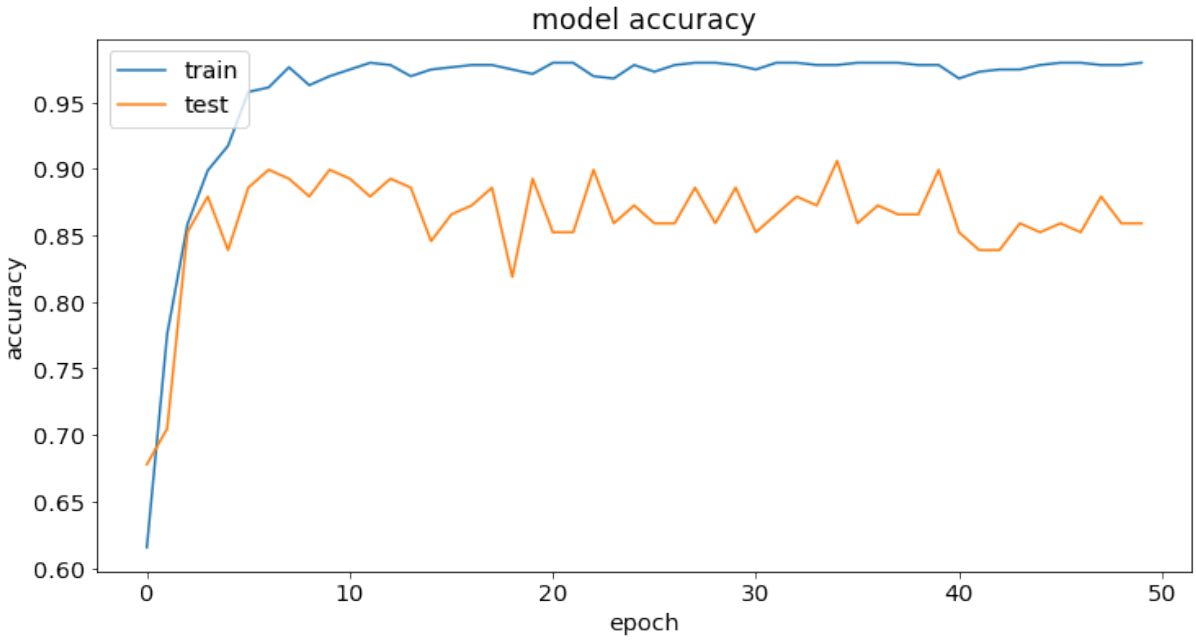

Figure 14: Accuracy With Face Images Canny Edge

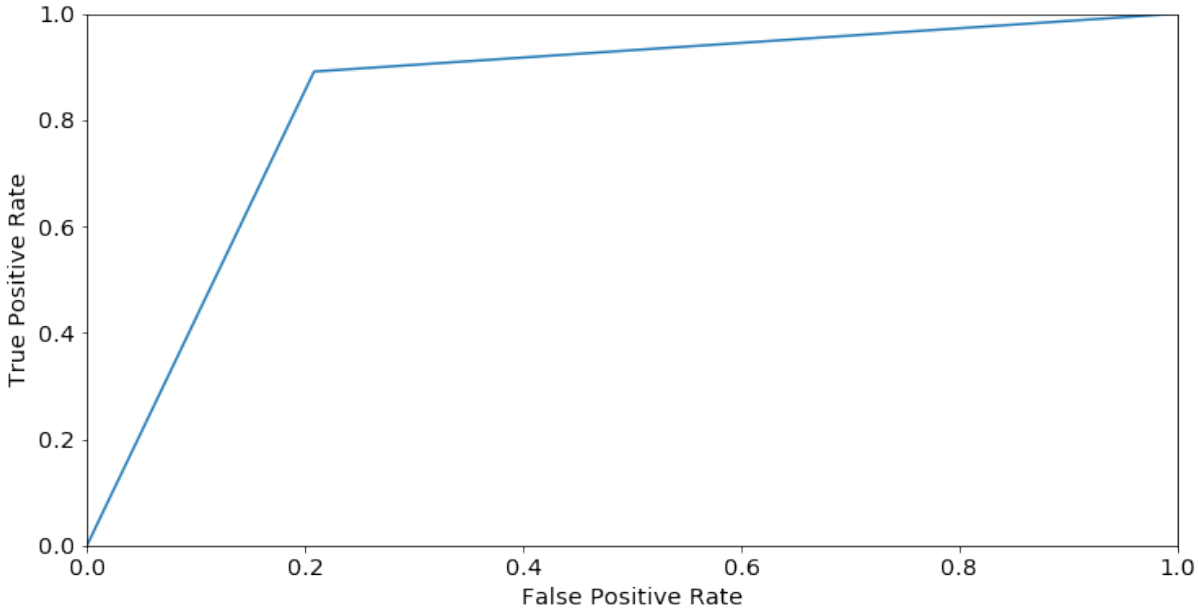

Figure 15: ROC With Face Images Canny Edge

These are interesting results when compared to laplace transformed dataset because in visualization there are not big differences between the images but the results are different this shows that there is clear logical and operational difference between both the transforms.

6.12 Evaluation and Results of CNN on face images with adaptive mean transformation

The data set with array of face images with adaptive mean transformation is used here. There are 742 images and train and test datasets are formed in such a way that 80% of training and 20% of testing sets respectively i.e 593 of train and 149 of test data. CNN is used to classify images and it resulted in good accuracy of 68% figure 16 and AUC 17 of 0.5.

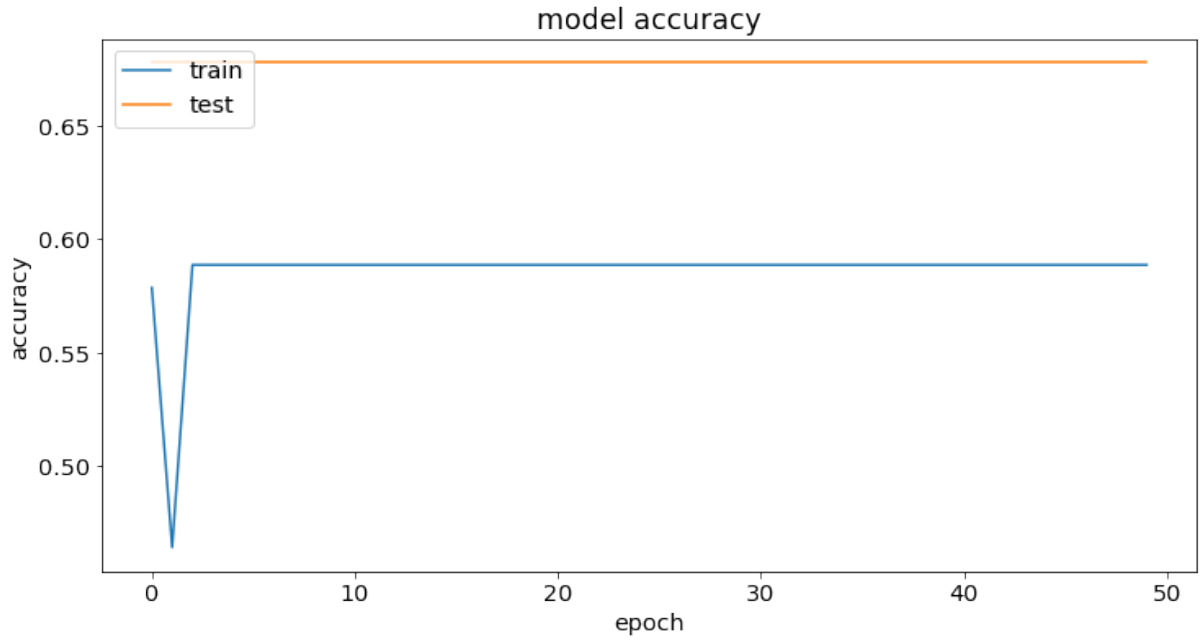

Figure 16: Accuracy With Face Images ADM

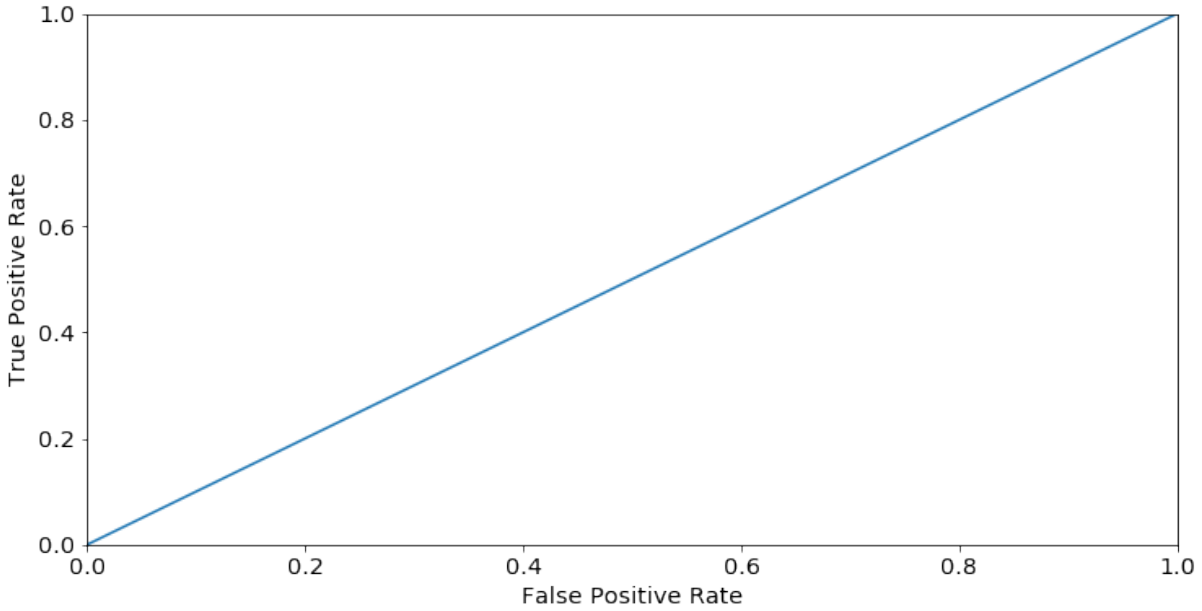

Figure 17: ROC With Face Images ADM

6.13 Implementation, Evaluation and Results of CNN on face images with Laplace transform

The main purpose of this implementation is that to see if there is any time saving can be achieved with quantum machine learning over classical cnn. The quantum machine learning is applied in CPU where are CNN is applied in GPU colab.

It resulted in only 50% of accuracy which is less that benchmark model. But the training time taken for this is 14 seconds while for CNN it is 33 seconds which is approximately

50% faster than the GPU implementation.

There by the research and sub research question has been answered.

7 Conclusion and Future Work

Better accuracy than literature review models by 10% with same number of videos and quality and 3% increase in accuracy overall. This answers research question. By implementation of quantum machine learning 50 % training time can be decreased but at the expense of low accuracy. This answers subresearch question.

The future work would be to develop better models in quantum machine learning so that computational time can be leveraged to get better accuracy as well. Audio files can also be used with videos to know about the semantics of the videos and there by achieve better results.

References

Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B. and Vijayanarasimhan, S. (2016). Youtube-8m: A large-scale video classification benchmark, *CoRR* **abs/1609.08675**.
URL: *http://arxiv.org/abs/1609.08675*

Azevedo, A. I. R. L. and Santos, M. F. (2008). Kdd, semma and crisp-dm: a parallel overview, *IADS-DM* .

Bloomberg (2018). how faking videos became easy — and why that's so scary$_2$018.
URL:*https://fortune.com/2018/09/11/deep-fakes-obama-video/*

Bose, A. J. and Aarabi, P. (2019). Virtual fakes: Deepfakes for virtual reality, *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, pp. 1–1.

chesney, r. and citron, d. (2019). Deepfakes and the new disinformation war.

Debnath, S., Linke, N. M., Figgatt, C., Landsman, K. A., Wright, K. and Monroe, C. (2016). Demonstration of a small programmable quantum computer with atomic qubits, *Nature* **536**(7614): 63–66.

Deepfakes github (2018).
URL: *https://github.com/deepfakes/faceswap*

Deutsch, D. (1985). Quantum theory, the church-turing principle and the universal quantum computer, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **400**(1818): 97–117.

Fastovets, D. V., Bogdanov, Y. I., Bantysh, B. I. and Lukichev, V. F. (2019). Machine learning methods in quantum computing theory, *Cornell University* .

Feynman, R. P. (1982). Simulating physics with computers, *International Journal of Theoretical Physics* **21**(6-7): 467–488.

Grover, L. (1996). L.k. grover, a fast quantum mechanical algorithm for database search,, *Proceedings of 28th Annual ACM Symposium on the Theory of Computing, 1996* .

Guera, D. and Delp, E. J. (2019). Deepfake video detection using recurrent neural networks, *Video and Image Processing Laboratory (VIPER)* .

Hasan, H. R. and Salah, K. (2019). Combating deepfake videos using blockchain and smart contracts, *IEEE Access* **7**: 41596–41606.

hui, j. (2018). How deep learning fakes videos (deepfake) and how to detect it?

James, D. F. V. (2001). Quantum computation and quantum information quantum computation and quantum information michael a, *Physics Today* **54**(11): 60–62.

Johnston, P., Elyan, E. and Jayne, C. (2019). Video tampering localisation using features learned from authentic content, *Neural Computing and Applications* .

Jones, R. C. (2019). Deepfake videos 'double in nine months'.
URL: *https://www.bbc.com/news/technology-49961089*

Korshunov, P. and Marcel, S. (2019). Vulnerability assessment and detection of deepfake videos.

Liu, A., Huh, M., Owens, A. and Efros, A. A. (2018). Fighting fake news: Image splice detection via learned self-consistency, *The European Conference on Computer Vision (ECCV)* pp. 101–117.

Mengoni, R. and Di Pierro, A. (2019). Kernel methods in quantum machine learning, *Quantum Machine Intelligence* .

Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T. and Nahavandi, S. (2019). Deep learning for deepfakes creation and detection.

Parkhi, O. M., Vedaldi, A. and Zisserman, A. (2015). Deep face recognition, *in* M. W. J. Xianghua Xie and G. K. L. Tam (eds), *Proceedings of the British Machine Vision Conference (BMVC)*, BMVA Press, pp. 41.1–41.12.

pothabattula, s. k. (2019). Detection of face manipulated videos using deep learning.

Ramprasath, M., Anand, M. and Hariharan, S. (2018). Image classification using convolutional neural networks, *International Journal of Pure and Applied Mathematics* **119**.

Rossler, A., Cozzolino, D., Nießner, M., Verdoliva, L., Riess, C. and Thies, J. (2019). Face forensics++: Learning to detect manipulated facial images.

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Nießner, M. (2019). Faceforensics++: Learning to detect manipulated facial images, *arXiv preprint arXiv:1901.08971* .

Schroff, F., Kalenichenko, D. and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823.

Shor, P. W. (1994). Algorithms for quantum computation: Discrete logarithms and factoring,, *Symposium on Foundations ofComputer Science, IEEE Press, Los Alamitos* .

Singh, V. (2019). Image forgery detection.
URL: *https://towardsdatascience.com/image-forgery-detection-2ee6f1a65442*

Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C. and Nießner, M. (2018). Face2face: Real-time face capture and reenactment of rgb videos.

tucker, p. (2019). The newest ai-enabled weapon: 'deep-faking' photos of the earth.
URL: *https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/*

Warde-Farley, D., Ozairt, S., Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Courville, A. and Bengio, Y. (2014). Generative adversarial nets, **1**.

Ying, M. (2010). Quantum computation, quantum theory and ai, *Artificial Intelligence* **174**(2): 162–176.