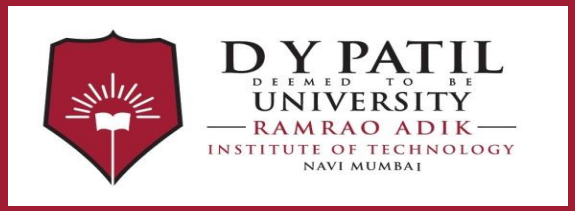


Ramrao Adik Institute of Technology
Department of Computer Engineering

MP-01 MAJOR PROJECT FINAL PRESENTATION

On

"Deepfake: A study of DL based tools"

By

Roll No.	Name of Student
20CE1140	Piyush Pal
20CE1151	Aakash Mishra
20CE1264	Harsh Jaiswal

Guided By: Dr. Vanita Mane Ma'am

Outline

1. Introduction
2. Literature Survey of the existing systems & Limitations
3. Problem statement
4. Proposed Methodology/ Techniques
5. Implementation Details
6. Design & Architecture
7. Results & Discussions
8. Conclusion & Future Scope
9. References

Introduction

Definition:

- It is a synthetic technique that can replace the person in an existing image or video with someone else's characteristic or likeliness.
- Deep fakes, a blend of "deep learning" and "fake," are synthetic media, typically videos or images, manipulated using AI to depict people saying or doing things they never did.

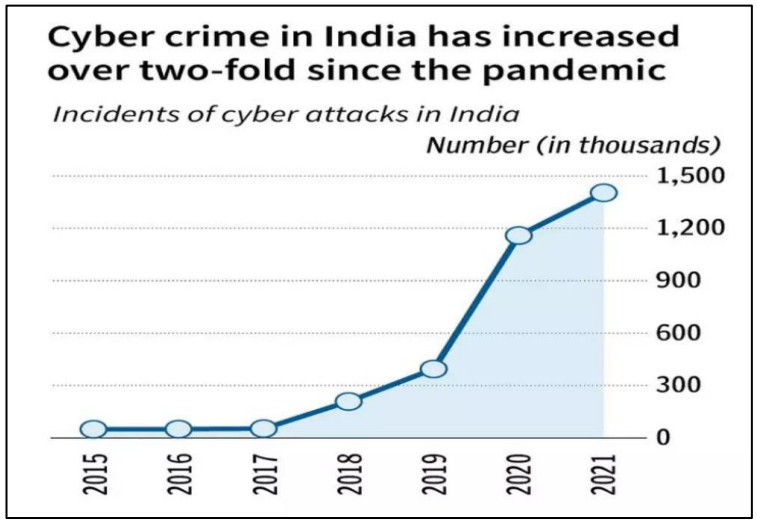
Why Deep fakes are a Growing Concern? [11][12]

- Used in social engineering attacks to trick users into revealing sensitive information.
- Can damage reputations and spread misinformation.
- Difficult to detect due to increasing sophistication.

Introduction

Motivation:

The surge of deepfakes poses a significant societal threat, leading to risks such as exploitation for revenge or bullying, exacerbated by the increased social media usage induced by COVID-19.



Img src: thehindubusinessline.com [13]

Objectives:

1. Develop cutting-edge deep fake detection algorithms addressing challenges such as adversarial attacks and imbalanced data, utilizing diverse datasets for training.
2. Enhance trust with interpretable outcomes, fortifying defenses to safeguard digital integrity against deep fake manipulation.

Literature Survey & Limitations of the existing systems

Sr. No.	Authors	Description	Accuracy	Limitation
1	*Nicolo et al.*	*They used a strong GPU with a DFDC Kaggle dataset & FF++ of 120000 videos and efficiently driven it into the EfficientNetB4 model. The model is also topped with an attention mechanism which will let analyst to interfere in the classification. On one hand they proposed a Siamese training strategy*	*88.13%*	*Implementing and training multiple CNN-based classifiers, each potentially with its own attention mechanism and siamese training strategy, can be computationally intensive and require significant resources.*
2	MD Shoel et al.	This paper covers the systematic literature surveys done for 12 different articles and mechanism and answered several questions like preferred dataset for deepfakes, possible techniques, evaluation metrics to be set etc.	Null	NULL (Review Paper)

Literature Survey & Limitations of the existing systems

3	Ning Yu et al.	Their solution pipeline consists of four stages. First train an image steganography encoder and decoder. Then use the encoder to embed artificial fingerprints into the training data. Next, train a generative model with its original protocol. Finally, decoding the fingerprints from the generated deepfakes. They have used CIFAR-10, LSUN & CelebA dataset for the purpose.	100%	The proposed solution's effectiveness may be limited to current generative models and might not generalize well to future models with different architectures or training methodologies.
4	Darius et al.	They designed their model using MesoNet. They generated their own dataset using 175 forged videos and compressed it using H.264 Codec. Meso-4 and MesoInception-4 networks are used in the process which consists of 4 ConV layers, two dropout layers, and a sigmoid function.	91.7%	The dataset used for training is collected from publicly available videos on the internet. This introduces potential biases, as the dataset might not be representative of all possible variations and techniques used in creating forged videos.

Literature Survey & Limitations of the existing systems

5	*D. Coccomini et al.*	*They used DFDC third generation dataset (5000 videos) on their model which is based on EfficientNet. The proposed methods analyze the faces extracted from the source video using MTCNN to determine whenever they have been manipulated. They proposed two mixed convolutional-transformer architectures i.e. Efficient ViT & Convolutional Cross ViT*	*93%*	*The inference strategy aggregates the inferred output both in time and across multiple faces in a video shot. While this approach aims to make the detection decision on a video basis, it may not effectively handle cases where manipulation occurs intermittently or at varying intensities within a video.*
6	Shahrouz et al.	They proposed a novel architecture based on Convolutional LSTM and Residual Network for deepfake detection. To compare their method with different baselines, they used DeepFake (DF), FaceSwap (FS), Face2Face (FS), NeuralTextures(NT), and DeepFakeDetection (DFD) datasets. There are two ConvLSTM2D layers, each followed by dropout, BatchNorm, and ReLU	98.05 %	NULL(Review Paper)

Literature Survey & Limitations of the existing systems

7	Yuezun Li, Siwei Lyu	Face Warping Artifacts used the approach to detect artifacts by comparing the generated face areas and their surrounding regions with a dedicated Convolutional Neural Network model. In this work there were two-fold of Face Artifacts.	93%	Their method is based on the observations that current deepfake algorithm can only generate images of limited resolutions, which are then needed to be further transformed to match the faces to be replaced in the source video. Their method has not considered the temporal analysis of the frames.
8	*Ming-Ching et al.*	*Detection by Eye Blinking describes a new method for detecting the deepfakes by the eye blinking as a crucial parameter leading to classification of the videos as deepfake or pristine. The Long-term Recurrent Convolution Network (LRCN) was used for temporal analysis of the cropped frames of eye blinking.*	*98.05%*	*As today the deepfake generation algorithms have become so powerful that lack of eye blinking can not be the only clue for detection of the deepfakes. There must be certain other parameters must be considered for the detection of deepfakes like teeth enchantment, wrinkles on faces, wrong placement of eyebrows etc.*

Literature Survey & Limitations of the existing systems

9	Huy H. et al.	Capsule networks to detect forged images and videos uses a method that uses a capsule network to detect forged, manipulated images and videos in different scenarios, like replay attack detection and computer-generated video detection.	97.13%	In their method, they have used random noise in the training phase which is not a good option. Still the model performed beneficial in their dataset but may fail on real time data due to noise in training. Our method is proposed to be trained on noiseless and real time datasets.
10	*David et al.*	*Recurrent Neural Network (RNN) for deepfake detection used the approach of using RNN for sequential processing of the frames along with ImageNet pre-trained model. Their process used the HOHO dataset consisting of just 600 videos.*	*99.3%*	*Their dataset consists small number of videos and same type of videos, which may not perform very well on the real time data. We will be training out model on large number of Realtime data.*

Literature Survey & Limitations of the existing systems

Analysis of **issues** present in surveyed solutions:

1. **Generalization:** Existing deep fake detection models struggle to generalize across different types of manipulations, impacting their performance on new variations.
2. **Data Imbalance:** Imbalanced labeled data leads to biased models with less accuracy in detecting deepfakes compared to real videos.
3. **Resource Intensiveness:** Many techniques demand substantial computational resources, hindering real-time deployment, especially on mobile or social media platforms.
4. **Transparency:** Lack of interpretability makes it difficult for users to understand and trust the model's decisions.
5. **Statistical models:** simplistic features, limited adaptability, lack of deep learning capabilities, and lower efficiency compared to CNN-based DNN models. Deep Learning excels due to its ability to learn complex patterns and adaptability, continuously improving with more data and emerging techniques.

Problem statement

The prevalence of deepfake videos has raised serious concerns due to their ability to delude viewers and manipulate information. Correctly distinguishing these videos is critical for vying erroneous data and safeguarding people and organizations. Thus, the project's goal is to use cutting-edge methods from Deep Learning to create a deepfake detection model. To improve the media authenticity and reliability, this model must demonstrate strong efficacy in differentiating between authentic and deceptive videos.

Details of Hardware & Software

Hardware Required:

- Intel i5 processor
- 8 GB RAM
- 7 GB Hard Disk Space
- NVIDIA GTX 3050 Graphic Chipset

Software Required:

- Google Colab / AWS Sagemaker
- Any Web Browser

Proposed Methodology

1. **Uploading the Dataset**
2. **Preprocessing by splitting, face recognition & Face cropping**
3. **Splitting into Training & Testing sets**
4. **Feature extraction using CNN ResNeXt & Classification using LSTM**
5. **Model Evaluation & Confusion Matrix**

Dataset used: 1676 Videos

1. Kaggle Deepfake Detection Challenge (453 Videos) [14]
2. CELEB-DF V2 (1000 Videos) [15]
3. Face Forensics ++ Dataset (223 Videos) [16]

Proposed methodology /Techniques

Preprocessing Details:

1) Using glob we imported all the videos in the directory in a python list. We pre-processed the data by splitting the data into test frames, face detection, face cropping and will save the face cropped data.
2) cv2.VideoCapture is used to read the videos and get the mean number of frames in each video.
3) To maintain uniformity, based on mean a value 100 is selected as idea value for creating the new dataset.
4) The video is split into frames and the frames are cropped on face location.
5) The new video is written at 30 frames per second and with the resolution of 112 x 112 pixels in the mp4 format.

Proposed methodology /Techniques

Model Implementation Details:

1) The dataset is split into train and test dataset with a ratio of 80% train videos (1300) and 20% (331) test videos. The train and test split is a balanced split i.e 50% of the real and 50% of fake videos in each split.
2) The pre-trained model of Residual Convolution Neural Network is used. The model name is resnext50_32x4
3) LSTM is used for sequence processing and spot the temporal change between the frames. 2048-dimensional feature vectors is fitted as the input to the LSTM.
4) A Rectified Linear Unit is activation function that has output 0 if the input is less than 0, and raw output otherwise. That is, if the input is greater than 0, the output is equal to the input. The operation of ReLU is closer to the way our biological neurons work.
5) Dropout layer with the value of 0.4 is used to avoid overfitting in the model and it can help a model generalize by randomly setting the output for a given neuron to 0.

Proposed methodology /Techniques

6) The training is done for 20 epochs with a learning rate of 1e-4 (0.0001),weight decay of 1e-3 (0.001) using the Adam optimizer.
7) To calculate the loss function Cross Entropy approach is used because we are training a classification problem.
8) A Softmax function is a type of squashing function. Squashing functions limit the output of the function into the range 0 to 1. This allows the output to be interpreted directly as a probability.
9) The trained model performs the prediction and return if the video is a real or fake along with the confidence of the prediction.

Design & Architecture

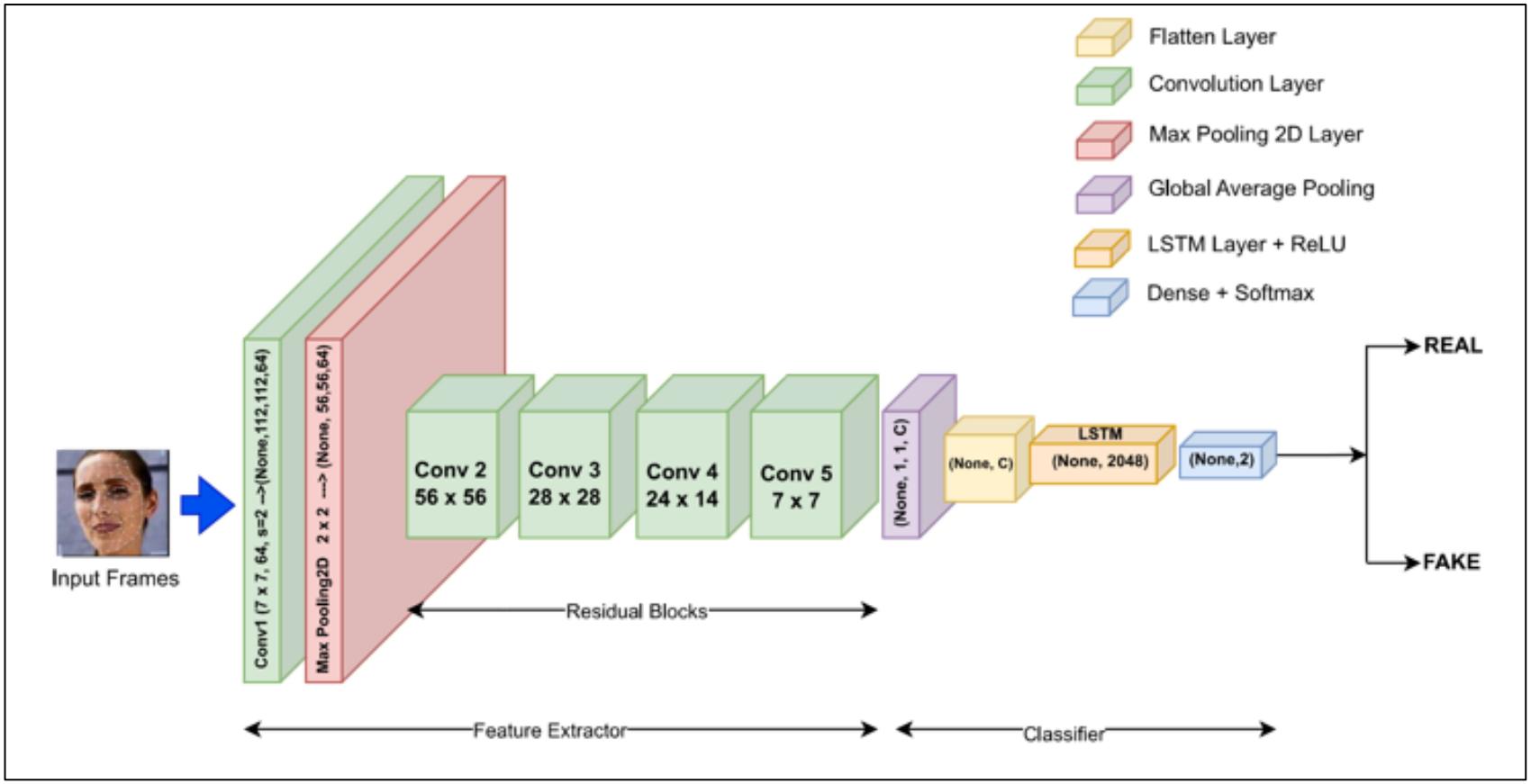


Figure: Deepfake Detection System Architecture

Design & Architecture

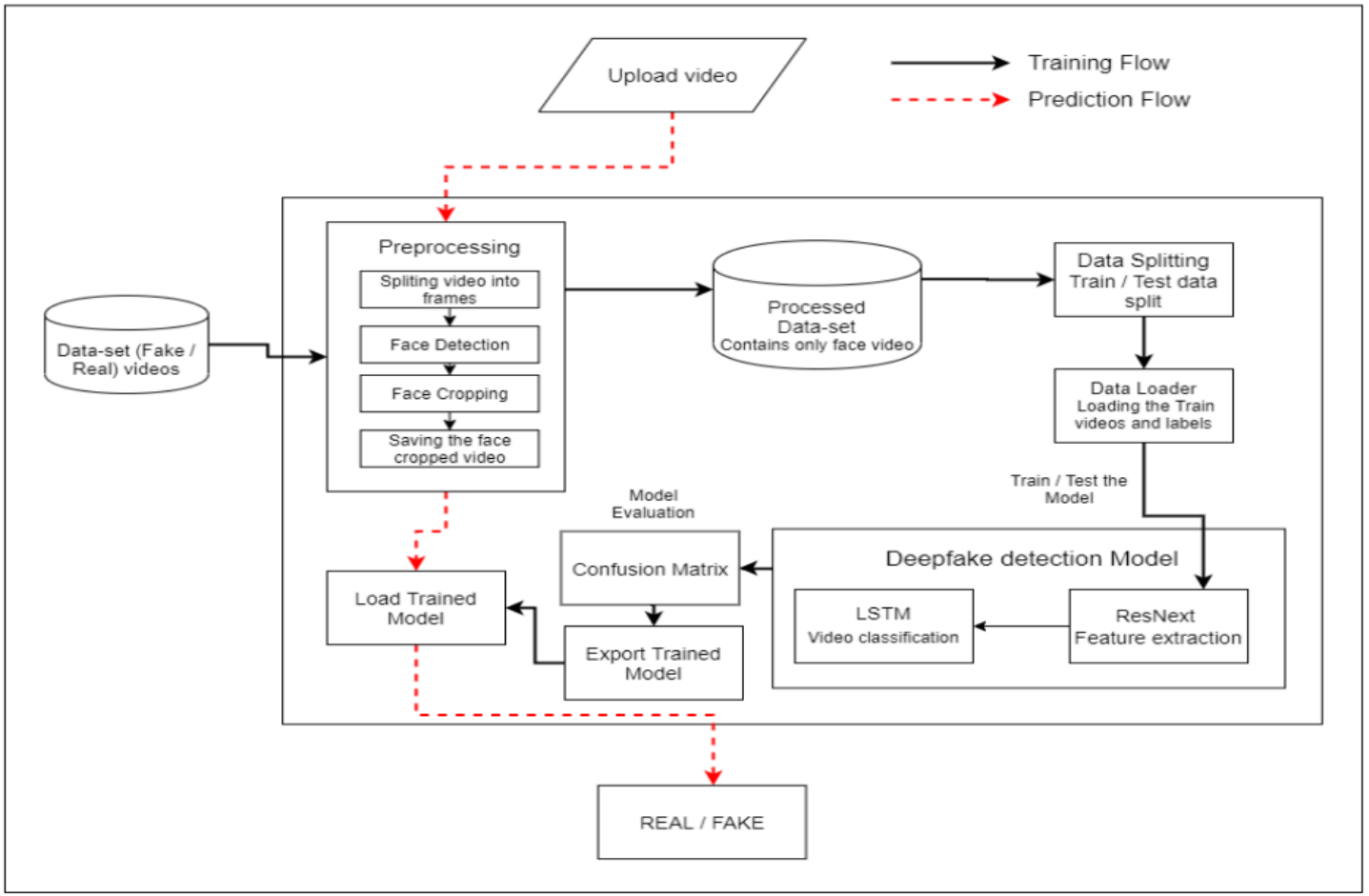
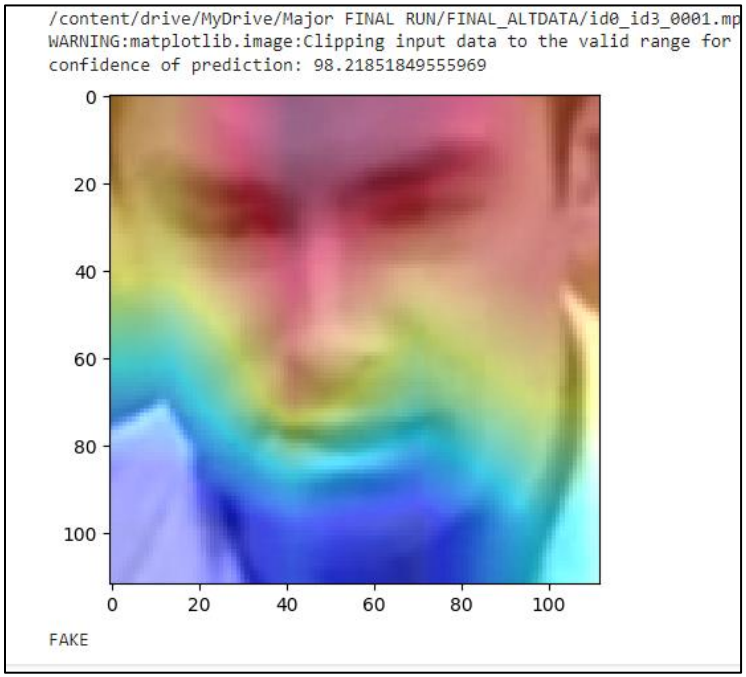


Figure: System Design

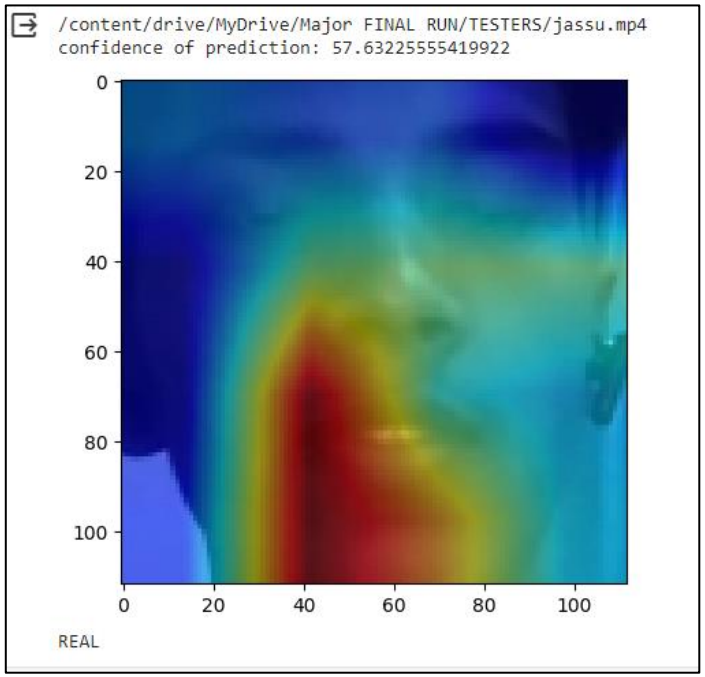
Results & Analysis

Input Image 1



Actual Result: FAKE
Predicted Result: FAKE

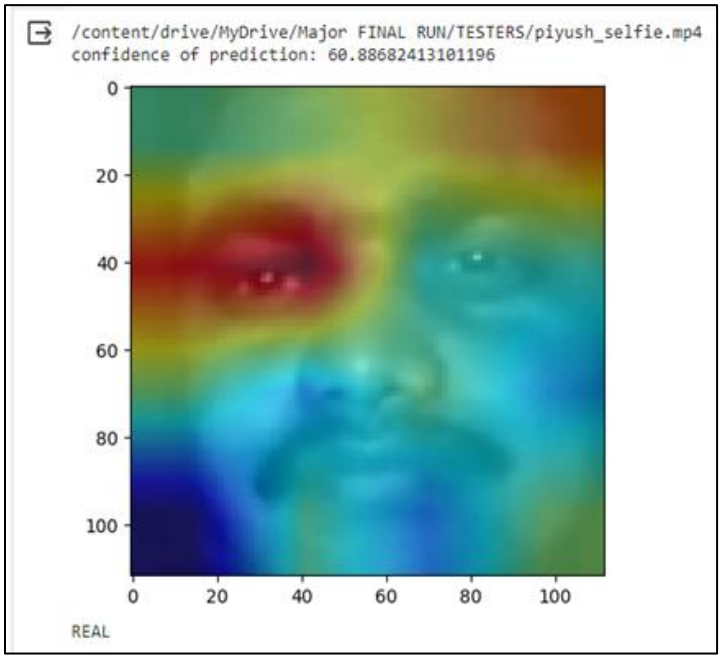
Input Image 2



Actual Result: REAL
Predicted Result: REAL

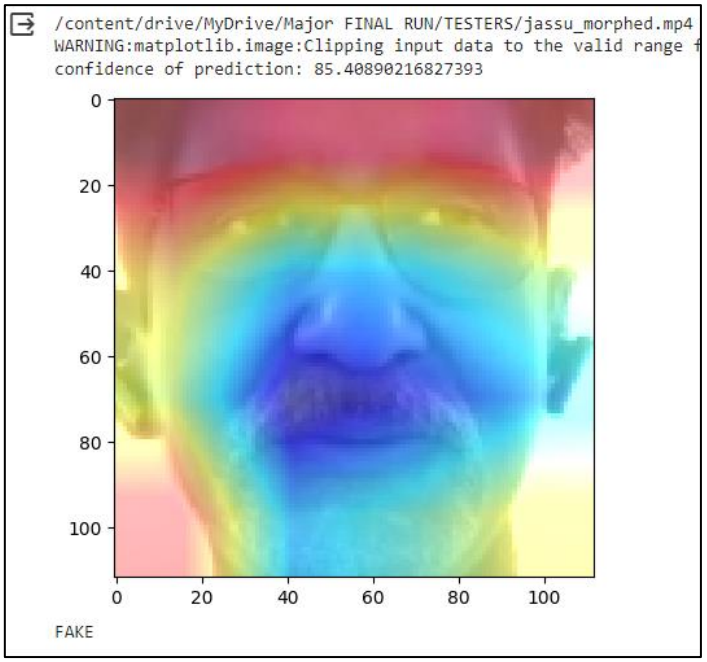
Results & Analysis

Input Image: 3



Actual Result: REAL
Predicted Result: REAL

Input Image: 4



Actual Result: FAKE
Predicted Result: FAKE

Result & Analysis

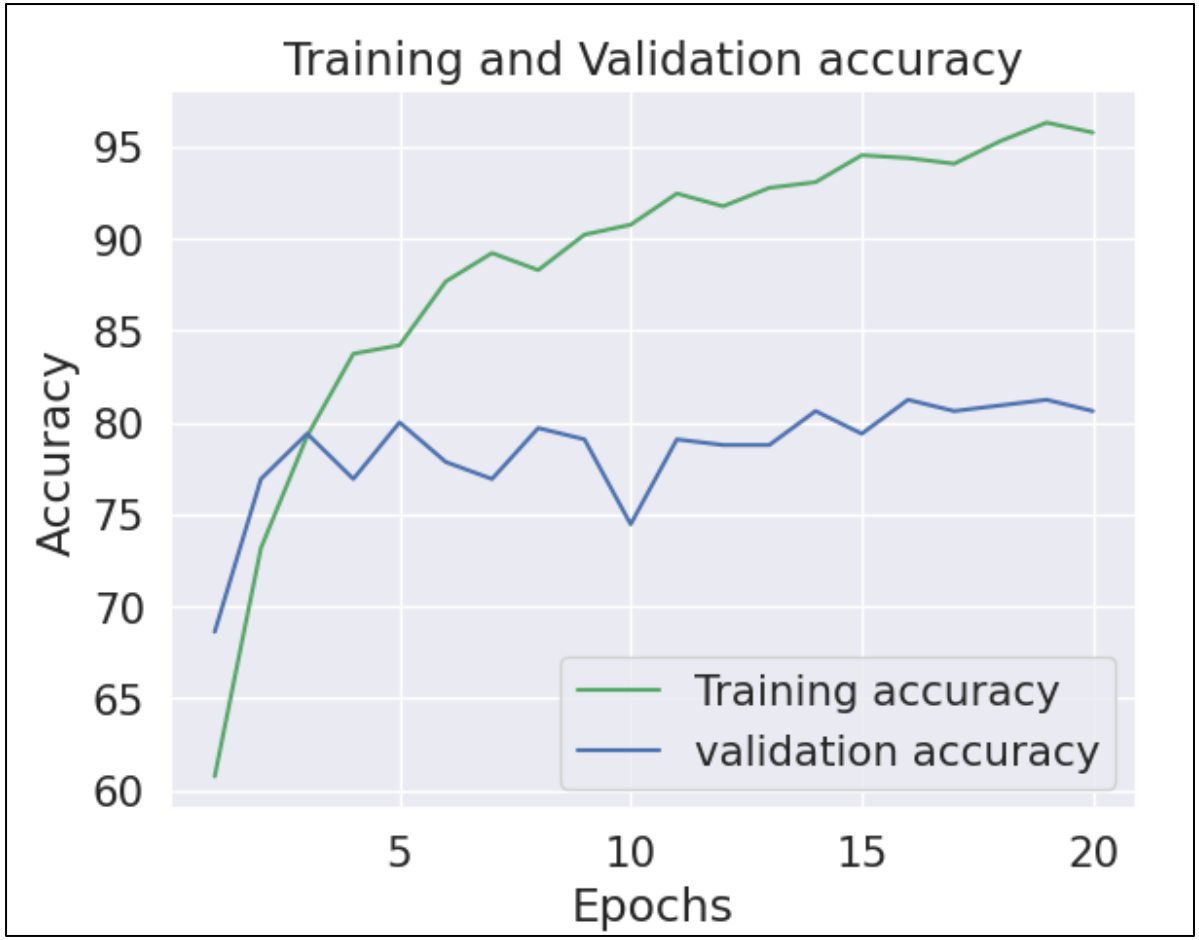


Figure: Training Accuracy v/s Validation Accuracy

Results & Analysis

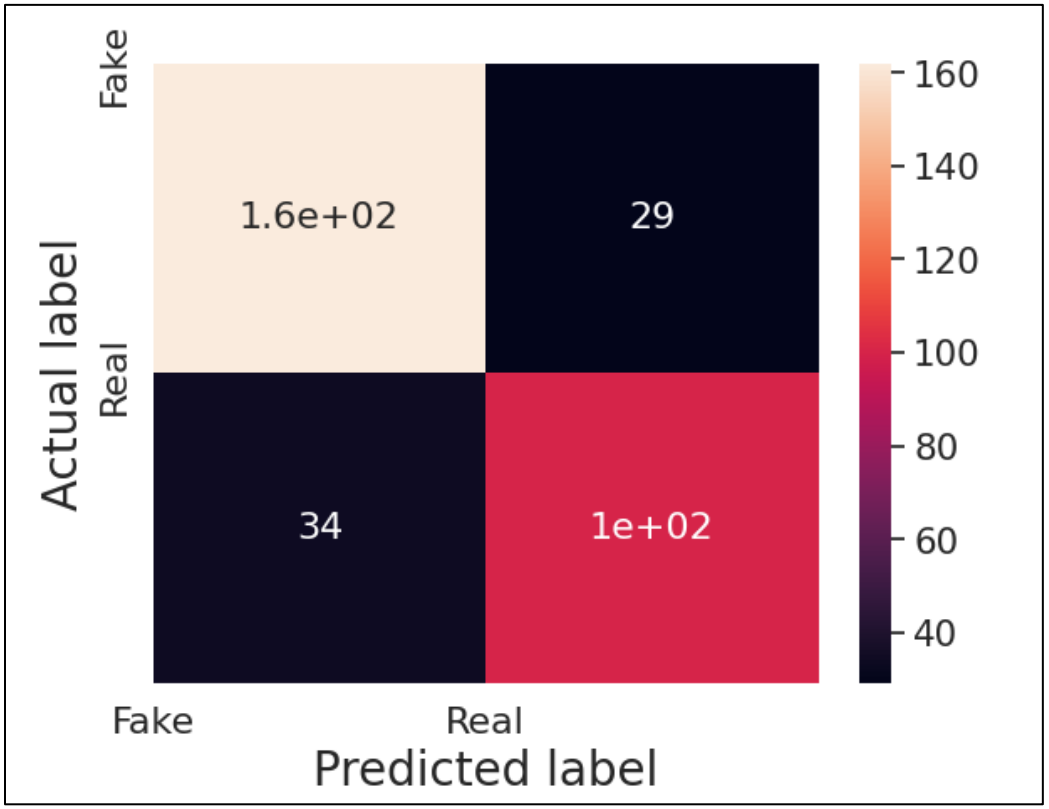


Figure: Confusion Matrix

[[162 29]
[34 100]]
True positive = 162
False positive = 29
False negative = 34
True negative = 100

Conclusion & Future Scope

Conclusion:

- The STAT (Statistical) models did not proved to be very accurate in compared to Deep Learning and Machine Learning models due to the increasing complexity of deepfake data.
- The Deep Learning and Machine Learning models show near or the same result but the later supremacy is achieved if several DL models are built together.
- The proposed model shows ample precision and rate of prediction towards detecting GAN developed deepfakes.

Future Scope:

- More datasets will be combined in order to check the model performance in complexity.
- There is a need to add more dataset in order to improve the confidence of prediction of videos varying in different lightings & environment.
- Improving the model to detect the videos with consisting specs and other cosmetical features covering the actual parameters.

References

1. N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini and S. Tubaro, "Video Face Manipulation Detection Through Ensemble of CNNs," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 5012-5019, doi: 10.1109/ICPR48806.2021.9412711.

2. M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 25494-25513, 2022, doi: 10.1109/ACCESS.2022.3154404.

3. N. Yu, V. Skripniuk, S. Abdelnabi, and M. Fritz, “Artificial fingerprinting for generative models: Rooting deepfake attribution in training data,” arXiv.org, https://arxiv.org/abs/2007.08457 (accessed Feb. 5, 2024).

4. D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “Mesonet: A compact facial video forgery detection network,” arXiv.org, https://arxiv.org/abs/1809.00888 (accessed Feb. 5, 2024).

5. D. A. Coccomini, N. Messina, C. Gennaro, and F. Falchi, “Combining EfficientNet and vision transformers for video deepfake detection,” SpringerLink, https://link.springer.com/chapter/10.1007/978-3-031-06433-3_19 (accessed Feb. 5, 2024).

References

6. S. Tariq, S. Lee, and S. S. Woo, “A convolutional LSTM based residual network for Deepfake video detection,” arXiv.org, https://arxiv.org/abs/2009.07480 (accessed Feb. 5, 2024).

7. Y. Li and S. Lyu, “Exposing deepfake videos by detecting face warping artifacts,” arXiv.org, https://arxiv.org/abs/1811.00656 (accessed Apr. 12, 2024).

8. A. Beckmann, A. Hilsmann, and P. Eisert, “Fooling State-of-the-art Deepfake Detection with High-quality Deepfakes,” arXiv (Cornell University), Jun. 2023, doi: https://doi.org/10.1145/3577163.3595106

9. T.-N. Le, H. Nguyen, J. Yamagishi, and I. Echizen, “Robust Deepfake On Unrestricted Media: Generation And Detection,” Feb. 2022. Available: https://arxiv.org/pdf/2202.06228v1.pdf. [Accessed: Apr. 12, 2024]

10. D. Guera and E. J. Delp, “Deepfake Video Detection Using Recurrent Neural Networks,” 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Nov. 2018, doi: https://doi.org/10.1109/avss.2018.8639163

11. https://rb.gy/6kge84

12. https://rb.gy/vcisul

13. https://rb.gy/stdmug

References

14. https://www.kaggle.com/competitions/deepfake-detection-challenge/data

15. https://github.com/yuezunli/celeb-deepfakeforensics

16. https://github.com/ondyari/FaceForensics

Thank You

D Y PATIL
DEEMED TO BE
UNIVERSITY
RAMRAO ADIK
INSTITUTE OF TECHNOLOGY
NAVI MUMBAI