Deep Fake: A Study of DL based Tools

Submitted in the partial fulfillment of the requirements

for the degree of B.Tech in Computer Engineering

by

Piyush Pal 20CE1180

Aakash Mishra 20CE1151

Harsh Jaiswal 20CE1264

Supervisor

Dr. Vanita Mane

Department of Computer Engineering

Ramrao Adik Institute of Technology

Sector 7, Nerul, Navi Mumbai

(Under the ambit of D. Y. Patil Deemed to be University)

May 2024

Ramrao Adik Institute of Technology

(Under the ambit of D. Y. Patil Deemed to be University)

Dr. D. Y. Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai 400 706

CERTIFICATE

This is to certify that, the Major Project report entitled

Deep Fake: A Study of DL based Tools

is a bonafide work done by

Piyush Pal 20CE1180

Aakash Mishra 20CE1151

Harsh Jaiswal 20CE1264

and is submitted in the partial fulfillment of the requirement for the degree of

B.Tech in Computer Engineering

to the

D. Y. Patil Deemed to be University

Supervisor	Project Co-ordinator
(Dr. Vanita Mane)	**(Dr. Aditi Chhabria)**
Head of Department	Principal
(Dr. Amarsinh V. Vidhate)	**(Dr. Mukesh D. Patil)**

Major Project Report Approval

This is to certify that the Major Project entitled ***"Deep Fake: A Study of DL based Tools"*** is a bonafide work done by ***Piyush Pal 20CE1180***, ***Aakash Mishra 20CE1151***, and ***Harsh Jaiswal 20CE1264*** under the supervision of ***Dr. Vanita Mane***. This Major Project is approved in the partial fulfillment of the requirement for the degree of ***B.Tech in Computer Engineering***

Internal Examiner:

1. ……

2. ……

External Examiners:

1. ……

2. ……..

Date: ../.../……

Place: ……

DECLARATION

I declare that this written submission represents my ideas and does not involve plagiarism. I have adequately cited and referenced the original sources wherever others' ideas or words have been included. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action against me by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: _____

Piyush Pal 20CE1180

Aakash Mishra 20CE1151

Harsh Jaiswal 20CE1264

Abstract

Our digital environment is becoming increasingly threatened by deepfakes, which are synthetic media in which one an individual's likeness is substituted for another. Media trust can be undermined, false information can be disseminated, and reputations can be harmed by these edited recordings. In order to tackle this issue, this study investigates the creation of a deep learning-based deepfake detection system using an LSTM (long short-term memory) network for classification and a Res-Next Convolution Network for feature extraction. This method makes use of ResNext50's strong picture recognition capabilities as well as RNNs' (Recurrent Neural Networks) LSTM (Long Short Term Memory) capacity to identify temporal correlations in video sequences. The system that was put into place was able to distinguish between actual and deepfake videos with an accuracy of 95.76%. This study advances the continuous endeavour to create strong deepfake detection techniques, encouraging a more dependable and safe online environment. In order to improve the model's performance and replicate real-world circumstances, we test our model using extensive datasets such as Face-Forensics++[13], Celeb-DF V2[12], and Deepfake Detection Challenge 2011. This initiative aims to increase the capacity to detect deepfakes, which could protect people from being manipulated online, preserve public debate, and increase confidence in digital information.

Keywords:

Residual Network- 50 (32 x 4) (Res-Next)

Deep Learning (DL)

Artificial Intelligence (AI)

Contents

Abstract **i**

List of Tables **iv**

List of Figures **v**

1 Introduction **6**

1.1 Overview 6

1.2 Motivation 7

1.3 Problem Statement 7

1.4 Objectives 7

1.5 Organization of the report 8

2 Literature Survey **9**

2.1 Survey of Existing System 9

2.2 Limitations of Existing System or Research Gap 12

3 Proposed System **14**

3.1 Problem Statement 14

3.2 Proposed Methodology/Techniques 14

3.3 System Design 15

3.4 Details of Hardware/Software Requirement 17

4 Results and Discussion **20**

4.1 Implementation Details 20

4.2 Result Analysis 24

4.3 Comparison Analysis 24

5 Conclusion and Further Work **26**

References **27**

A Weekly Progress Report **30**

B Plagiarism Report **31**

C Project Competitions **32**

Acknowledgement **34**

List of Tables

3.1 Hardware Requirements of the Model……..17
4.1 Testing Scenarios of the model…….…….23
4.2 Model Development……..……..24
4.3 Comparison Analysis…….24

List of Figures

3.1 Dataset division of the Model 15
3.2 System Design of the Model 15
3.3 System Architectural Diagram of the model 16
4.1 Test Case 1 21
4.2 Test Case 2 22
4.3 Test Case 3 22
A.1 Weekly Progress Report 28

Chapter 1

Introduction

1.1 Overview

The burgeoning ubiquity of deepfakes, hyper-realistic manipulated videos meticulously crafted through Generative Adversarial Networks (GANs), has become a preeminent threat to online trust and the integrity of information. The model is meticulously architected to prioritize paramount accuracy, dexterity in handling a diverse spectrum of manipulation techniques, and interpretability to engender user confidence. To achieve these ambitious objectives, we leverage a novel approach that capitalizes on the inherent limitations of deepfake creation tools. Our model takes advantage of this vulnerability by employing a pre-conditioned ResNeXt-50 as the main building block. ResNeXt, which is well-known for its skill in feature extraction, can detect these anomalies that are caused by deepfakes and inserted into video frames. Following feature extraction, a Long Short-Term Memory (LSTM) network is meticulously fed the features. LSTMs are lauded for their ability to scrutinize temporal nuances, a critical aspect of deepfake detection as inconsistencies across video frames frequently expose manipulation attempts. This expansive corpus, encompassing 1676 videos, amalgamates datasets from Kaggle's Deepfake Detection Challenge, CELEB-DF v2, and Face Forensics ++. The dataset undergoes meticulous preprocessing, encompassing tasks such as frame extraction, facial detection, and cropping, to ensure standardization and optimize training efficacy. Evaluation of the model's efficacy is facilitated through a comprehensive confusion matrix, in accurately discerning between genuine and counterfeit videos (true/false positives/negatives). By furnishing a robust deep learning-based apparatus for deepfake detection, the endeavor charts a course towards bolstering digital integrity and fostering trust within a media landscape increasingly susceptible to manipulation.

1.2 Motivation

Understanding the significance of developing a Deep Fake detection system can be driven by looking at a few real-world scenarios where such a solution would be highly beneficial:

1. Societal Threat: The rampant spread of deepfakes poses a multi-faceted danger, fueling exploitation, vengeance, and online bullying. The pandemic's influence on social media use has amplified this concern.
2. Combating Deception: This project is driven by the urgent need to counter this insidious threat. We aim to be pioneers in deepfake detection using cutting-edge deep learning methods.
3. Truth in the Digital Age: The erosion of trust in online content due to deepfakes is a critical motivator. We strive to build robust tools that safeguard against the dissemination of fabricated videos.

1.3 Problem Statement

The capacity of deepfake films to deceive users and alter information has led to widespread is concerned about their prominence. For the purpose of countering misleading information and protecting individuals and institutions, correctly categorising these videos is vital. 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.

1.4 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.
3. Bolster digital defenses against deepfake manipulation. This will ensure the authenticity and reliability of digital media in the modern age.

1.5 Organization of the Report

The report is organized as follows:

Chapter 1 gives an introduction about the topic, the motivation and the objective of the project.

Chapter 2 reviews the literature related to the project area.

Chapter 3 focuses on defining the system's problem statement. That includes problem formulation, proposed techniques, model architecture, and hardware/software requirements.

Chapter 4 focuses on the results obtained from the implementation including details of the implementation and analysis of the results.

Chapter 5 describes the conclusions drawn from the project and future work to enhance the techniques for an improved model.

Chapter 2

Literature Survey

3.1 Survey of Existing System

[1] Nicola et al. proposed a model. They used a strong GPU with a DFDC Kaggle dataset & FF++ 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. This research focuses on detecting manipulated faces in videos, particularly those created using modern techniques. The proposed method involves combining different Convolutional Neural Network (CNN) models built upon a base network (EfficientNetB4). These models incorporate two distinct approaches: attention layers and Siamese training. The authors use these networks to detect modified faces in over 119,000 videos culled from two openly accessible video databases, with promising results.

[2] Shoel et al. proposed "Deepfake Detection: A Systematic Literature Review". Deepfakes, hyper-realistic AI-powered videos where faces are seamlessly swapped, pose a serious threat. These faked movies tend to be used to target politicians and celebrities in an attempt to disseminate false information and sway public opinion. Fortunately, scientists are retaliating. Whilst deepfake detection has progressed recently thanks to advances in machine learning and deep neural network approaches, detection on its own is inadequate. We require multiple approaches in order to effectively counter this threat. Developing regulations for deepfake creation and use is essential, alongside public education programs to teach critical online evaluation skills. Finally, continued technological advancements are crucial to refine deepfake detection and content authentication. By combining these efforts, we can mitigate the dangers of deepfakes

[3] N. Yu et al. proposed a model. They developed a simple but effective way to detect deepfakes. These markers let them easily identify deepfakes and even who made them. Their method works across different models and is hard to trick. It makes deepfake detection simple and outperforms

current methods. This approach holds creators accountable, promoting responsible use of this technology. There are a total of four phases in their solution procedure. First, train a decoding algorithm and transmitter for steganography of images. Afterwards, integrate synthetic biometrics into the learning set using the encoder. Next, use the original method to train an algorithm that generates models. Lastly, deciphering fingerprints gathered from the deepfakes that were generated. For this, they employed the CelebA, LSUN, & CIFAR-10 datasets. They obtained a 100% accuracy percentage.

[4] Darius et al. proposed another model. This research proposes a method to automatically detect manipulated faces in videos, especially those created with techniques like Deepfakes and Face2Face. Traditional methods struggle with video compression, so this approach utilizes deep learning with lightweight networks to analyze the "medium-scale" features of images. Tests on existing and custom video datasets show impressive success rates, exceeding 98% accuracy for Deepfakes and 95% for Face2Face. 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.

[5] D. Coccomini et al. proposed a model. This paper suggests a novel feature extraction method that combines convolutional neural network EfficientNet B0 with Vision Transformers. The aforementioned method does not require sophisticated techniques such as distillation or ensemble learning, yet it produces outcomes on par with state-of-the-art approaches. The authors also provide a straightforward voting system for managing several faces in a movie. On the Deep Fake Detection Challenge (DFDC) dataset, the recommended method achieves a staggering Area Under the Curve (AUC) of 0.951 and an F1 score of 88.0%, providing its efficacy with regard to identifying deepfakes.

[6] Shahroz et al. proposed a model. The present investigation presents an innovative method for detecting deepfakes using a Convolutional neural network LSTM based residual neural network (CLRNet), with an emphasis on the minute disparities between frames in altered videos. CLRNet analyzes sequences of images to learn these temporal patterns. Additionally, the authors leverage transfer learning to make the method adaptable to various deepfake creation techniques. Experiments on a large dataset demonstrate that CLRNet outperforms existing methods by achieving better generalization across different deepfake generation approaches. They achieved

an accuracy of 98.05%.

[7] Yuzen Li and Siwei Lyu proposed an unique approach. The first attempt to identify these "artefacts" with a Convolutional Neural Network (CNN) is presented in this study. This method emphasises on the distinct aberrations in deepfakes, as opposed to previous approaches that require a large amount of actual and fake movies for training. It achieves this by simulating these artifacts for negative training examples, saving time and resources. Additionally, this method is more robust because these artifacts are common across various deepfake creation tools. Evaluations on benchmark datasets confirm the effectiveness of this approach in identifying deepfakes. They achieved an accuracy of 93%.

[8] Ming – Ching et al. proposed a model. They present a new method to create high-quality deepfakes that can fool existing detection systems. They created 90 lifelike deepfakes by using a special autoencoder for face switching and a sophisticated blending method. It's interesting to note that the detector performs better when it is retrained using our superior fakes. This suggests that current detectors lack exposure to diverse and realistic manipulations, and their training data should be expanded to include high-quality deepfakes like those we created. An eye blinking trimmed frame was subjected to temporal analysis using the Long-term Recurrent Convolution Network (LRCN). They achieved an overall accuracy of 98.05%.

[9] Huy H. et al proposed a model. The development of deepfake generation and detection is examined in this chapter along with its difficulties. Additionally, it talks about potential approaches to strengthen the deepfake detection system's resilience across a range of media types (such as movies and photos taken in the wild). Lastly, it offers a direction for further study on bogus media. Capsule networking to detect modified photographs and vids: This technique makes use of a capsule network to identify manipulated images and videos in many contexts, such as computer-generated video recognition and replay attack detection. Their rate of precision was 97.13%.

[10] David et al. proposed an unique model. This work introduces a temporal-aware process for proactively identifying deepfake videos. Their technique captures frame-level characteristics using convolutional neural networks (CNNs). Then, using these features, an RNN (recurrent neural network) is trained to identify whether or not a video has been manipulated. They contrast our method with a huge library of deepfake movies that they assembled from multiple internet

video portals. They showed how their system may achieve competitive performance in this challenge even with a simple architecture. Recurrent Neural Network (RNN) for deepfake detection employed a method that combined an ImageNet initially trained model with RNN for sequential frame processing. The approach they used made use of the 600-video-strong HOHO dataset. They acquired a 97.10% proficiency percentage.

3.2 Limitations of Existing System

1. Implementing and training multiple CNN-based classifiers[1], each potentially with its own attention mechanism and siamese training strategy, can be computationally intensive and require significant resources.
2. The proposed solution's [3] effectiveness may be limited to current generative models and might not generalize well to future models with different architectures or training methodologies.
3. The dataset used in this model[4] 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.
4. In a video shot, the inference technique aggregates the estimated outcome over numerous faces and in time. 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[5].
5. The research[7] notes The technique they use is based on the observation that the existing deepfake algorithm is capable of producing images with an appropriate resolution, which implies that extra processing is required in order to ensure the images match the individuals that need to be replaced it in the video in question. The chronological examination of the frames has not been taken into account by their method.
6. In relation to the study[8], since deepfake emergence algorithms are currently so strong, it is not impossible to detect deepfakes just by looking for indicators such as eye blinking. For the process of detection, several additional characteristics need to be considered into account.

7. They incorporated noise at random in the training period of their procedure[9], which is an undesirable choice. nevertheless, the model did adequately on their dataset; despite this, noise in the training set might have caused it to perform poorly on real-time data. We advocate utilising real-time and noiseless datasets to train our technique.
8. Their database[10] is composed up of just a handful of the same kind of motion pictures, which wouldn't interact effectively using real-time data. The algorithm we use will be developed utilising an enormous quantity of real-time data.

Chapter 3

Proposed System

3.1 Problem Statement

Recognizing that deepfake films can mislead viewers and distort information, their proliferation has given rise to major issues. Accurately differentiating these movies is essential for combating false information and protecting individuals and institutions. Consequently, the goal for this study is to use leading-edge Deep Learning techniques to formulate a deepfake detection model. To further enhance media authenticity and validity, this model has to demonstrate that it is particularly good at differentiating between authentic and fraudulent videos.

3.2 Proposed Methodology/Techniques

The project's objective is to determine the difference among real and fake videos from the real-world input scenario. The model is being trained on three different kinds of datasets [Refer Figure 3.1] consisting of real and fake videos in the same ratio. These videos were further preprocessed and later trained in the model. The entire working is done with the help of google colab as it becomes easier to use with the pytorch library framework namely the CUDA technologies that has been developed by NVIDIA. The accuracy compared on the basis of cross entropy loss function and a confusion metrics has been drawn for the same. The model when given an input, identifies basic 16 types of parameters which are blinking of eyes, teeth enhancement, bigger distance of eyes, iris segmentation, face angle, hairstyle, skin colour, background colour, etc. The dataset division is given below

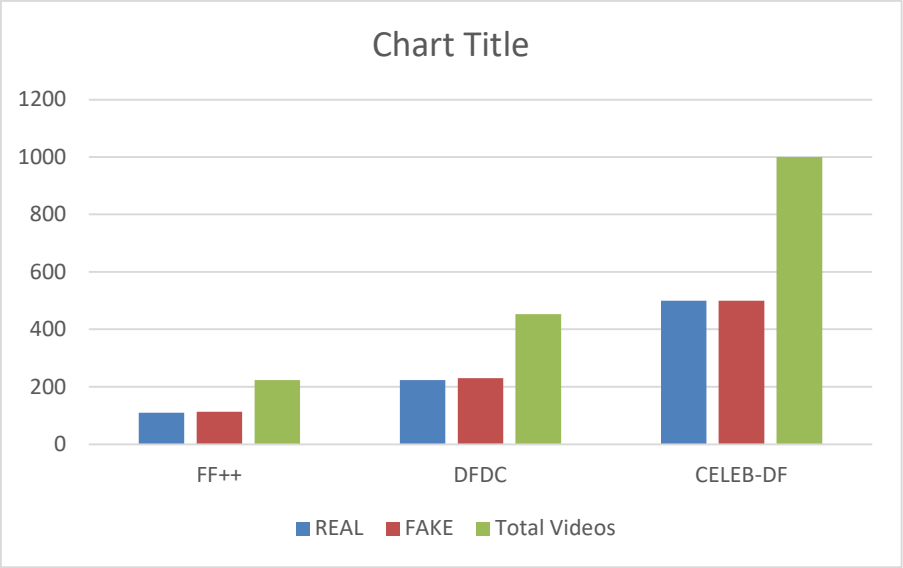

Figure 3.1: Dataset division of the Model

3.3 System Design

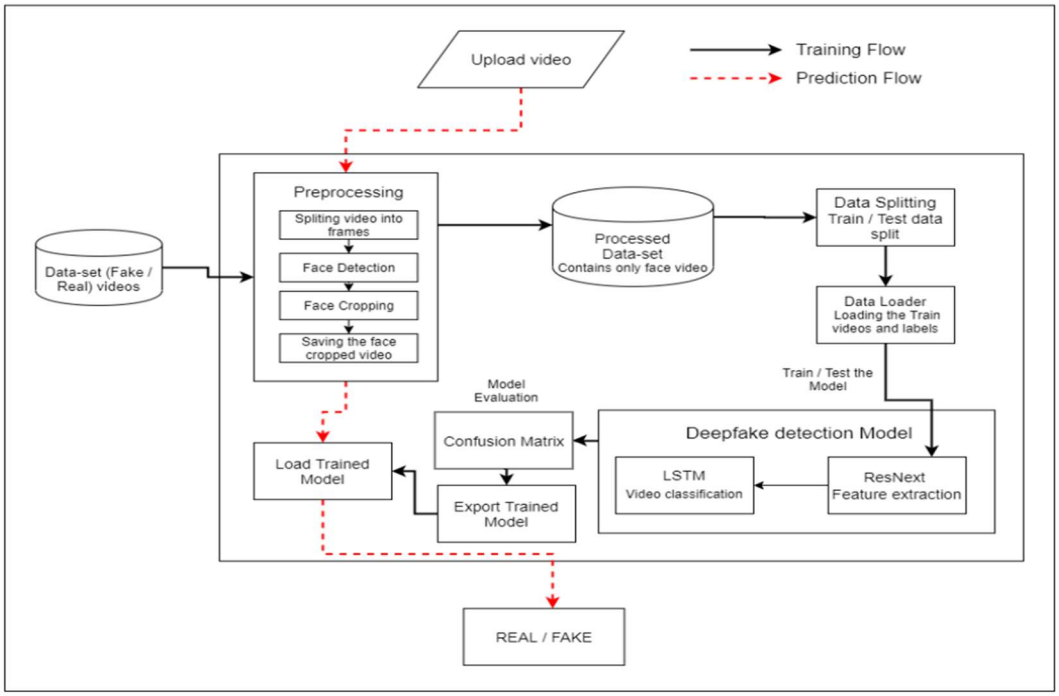

Figure 3.2: System Design of the Model

We first plan to collect dataset from 3 main sources that's Kaggle, CELEB-DF and FF++ which results in approximately 850 videos per class and 1700 videos being fed to the model in total [Refer Figure 3.2]. 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 cv2.VideoCapture is used to read the videos and get the mean number of frames in each video. The videos were stored in different folders uploaded in the google drive. The videos after preprocessing were sent for model training, confusion matrix was drawn and the performance was evaluated. Lastly prediction was made in the phase 3 of the model development. The system architecture can be clearly seen in the Figure 3.3

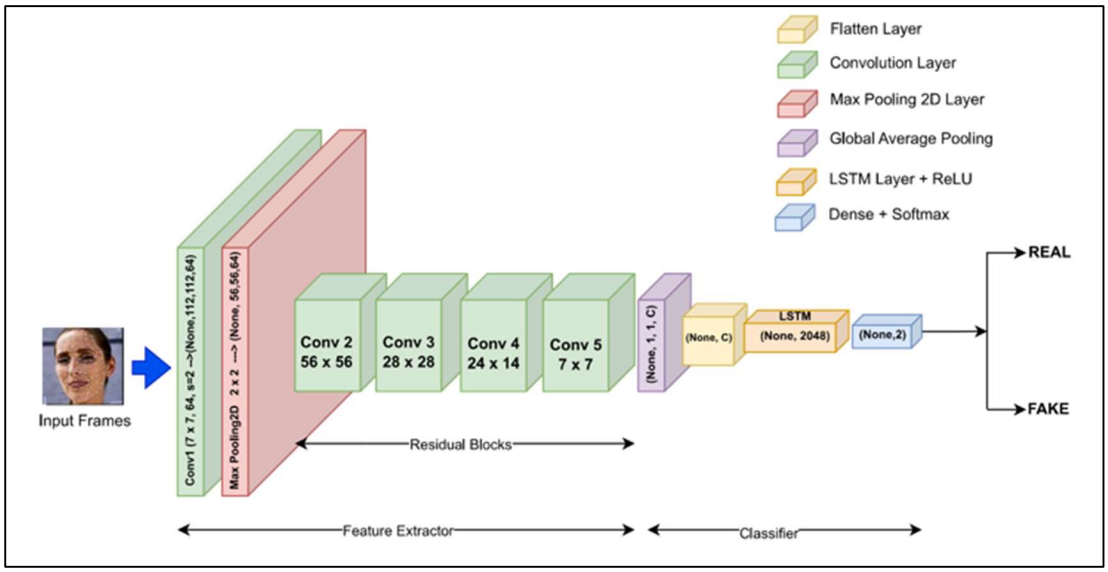

Figure 3.3: System Architectural Diagram of the model

3.4 Details of Hardware/Software Requirement

Sr. No.	Parameter	Minimum Requirement
1	Intel Xeon E5 2637	3.5 GHz
2	RAM	16 GB
3	Hard disk	100 GB
4	Graphic Card	NVIDIA GeForce GTX Titan

Table 3.1: Hardware Requirements of the Model

Libraries:

torch (PyTorch): A freely available framework for deep learning that is becoming quite well-liked. It presents a strong and adaptable neural network construction, training, and deployment platform. Torch offers functionalities for automatic differentiation, GPU acceleration, and a rich ecosystem of tools and libraries.

torchvision (PyTorch Vision): A companion library to PyTorch specifically designed for computer vision tasks. It simplifies common operations like image classification, object detection, image segmentation, and data transformations. Torchvision integrates seamlessly with PyTorch, allowing researchers and developers to focus on their models rather than low-level computer vision details.

os (Operating System): A core Python library providing functionalities to interact with the underlying operating system. This includes essential tasks like creating, deleting, and managing files and directories, as well as navigating the file system hierarchy. It also allows for process management and interaction with system resources.

numpy (Numerical Python): An elementary Python package for computational science. It offers powerful multi-dimensional array data structures (ndarrays) and a vast collection of mathematical functions. NumPy excels at numerical computations, linear algebra operations, and array manipulation, making it essential for data analysis and scientific simulations.

cv2 (OpenCV): A set of modules for real-time computer vision technology Intel first developed. It provides a comprehensive toolkit for image and video processing tasks. OpenCV is widely used

in applications like object detection, feature extraction, image manipulation, camera calibration, and augmented reality.

matplotlib: A mature and versatile Python library for creating static visualizations. It allows for generating various plots, charts, and histograms, often used for data exploration, presentation of results, and creating publication-quality figures. Matplotlib offers a wide range of customization options for tailoring visualizations to specific needs.

face_recognition: A specialized Python library for facial recognition tasks. It simplifies common operations like face detection, landmark localization (identifying key facial features like iris, nostrils, and jaws), and face identification (comparing faces to any known database). This library can be valuable for security applications, image organization, and user authentication systems.

json (JavaScript Object Notation): A library for working with JSON data format. JSON is a human-readable text format for data exchange, often used for configuring applications, storing data in web applications, and transferring data between systems. This library allows Python programs to parse, manipulate, and generate JSON data.

pandas: A powerful data analysis and manipulation library in Python. It offers high-performance, easy-to-use data structures like DataFrames (tabular data) and Series (one-dimensional data) for efficient data handling. Pandas excels at data cleaning, transformation, analysis, and visualization, making it a favorite tool for data scientists and analysts.

copy: A built-in Python library providing functions for creating shallow or deep copies of objects in memory. Copying objects can be useful in various scenarios, such as preserving the original state of an object while modifying its copy or creating multiple references to the same data with different functionalities.

glob: A built-in Python library for locating pathname patterns. It allows searching for files and directories that match specific naming conventions using wildcard characters like "*" (any characters) and "?" (single character). This functionality is helpful for automating tasks involving files with specific naming patterns.

random: A built-in Python library providing functionalities for generating pseudo-random numbers. It offers various functions for generating random integers, floating-point numbers, and selecting random elements from sequences. Randomness is crucial for many applications, including simulations, cryptography, and machine learning algorithms.

sklearn: A well-liked and approachable Python machine learning package. For preprocessing, model selection, dimensionality reduction, clustering, regression, and classification, it provides a large number of algorithms. Sklearn provides a consistent interface for these algorithms, making it easy to learn, apply, and compare different machine learning techniques.

Chapter 4

Results and Discussion

4.1 Implementation Details

In this deep learning approach designed for video deepfake detection, a meticulously prepared dataset is split for training (80%) and testing (20%) [Refer Figure 3.2]. This split ensures an equal representation of real and fake videos in both sets, crucial for robust model performance. A initially-trained ResNext50_32x4 serves as the foundation [Refer Figure 3.3], tasked with extracting informative features from individual video frames. This pre-trained network offers a significant advantage by leveraging the knowledge gained from prior image recognition tasks, improving feature extraction efficiency. However, deepfakes often exhibit subtle inconsistencies in how faces change across frames. To capture these temporal anomalies, Long Short-Term Memory (LSTM), which is a typical RNN, is incorporated [Refer Figure 3.3]. LSTMs excel at processing sequential data like video frames, allowing the model to learn the temporal relationships between features extracted from each frame. This enables the model to identify inconsistencies that might be indicative of manipulation, such as unnatural head movements or flickering artifacts. A crucial step to prevent overfitting, a common challenge in machine learning, involves a dropout layer. During training, this layer randomly sets a certain percentage of neuron outputs to zero. This forces the system to not depend completely on any single feature or specific training examples, promoting better generalization to unseen data. The learning process itself is enhanced using the Adam optimizer, a popular algorithm that dynamically adjusts learning rates for different parameters within the model. Specific settings for the learning rate and weight decay are employed to guide the optimization process and achieve optimal convergence. The training iterates over the data for a specified number of epochs (typically 20 in this case). Considering the purpose of this classification problem is to discriminate between authentic and fraudulent films, a cross-entropy loss function is employed. This function calculates the discrepancy among the actual labelling (real or fake) for every video and the expected probabilities of the model.

Minimizing the cross-entropy loss during training drives the model towards making better classifications. Finally, a Softmax function transforms the model's output into a probability distribution. As a result, the final result can be directly understood as the probability that a video is authentic or fraudulent, coupled with a confidence level for the estimate. The confidence score [Refer Figures 4.1,4.2,4.3] provides valuable information about how certain the model is about its classification. In essence, this approach combines the feature extraction power of pre-trained CNNs with the sequential learning capabilities of LSTMs to effectively detect deepfakes in videos by analyzing both spatial features within each frame and the temporal relationships between frames.

Test Case 1: Author's real time video is morphed and given as input to the prediction model.

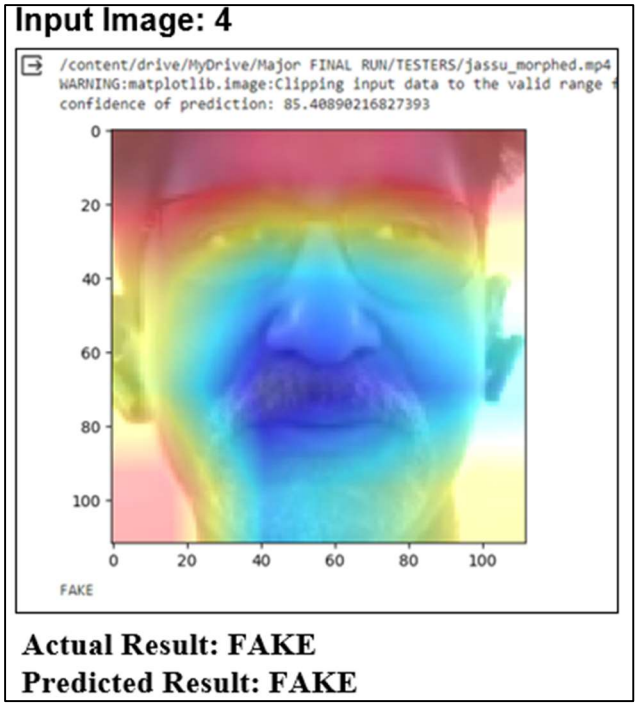

Figure 4.1: Test Case 1

Test Case 2: Author's real time video is given as input to the prediction model.

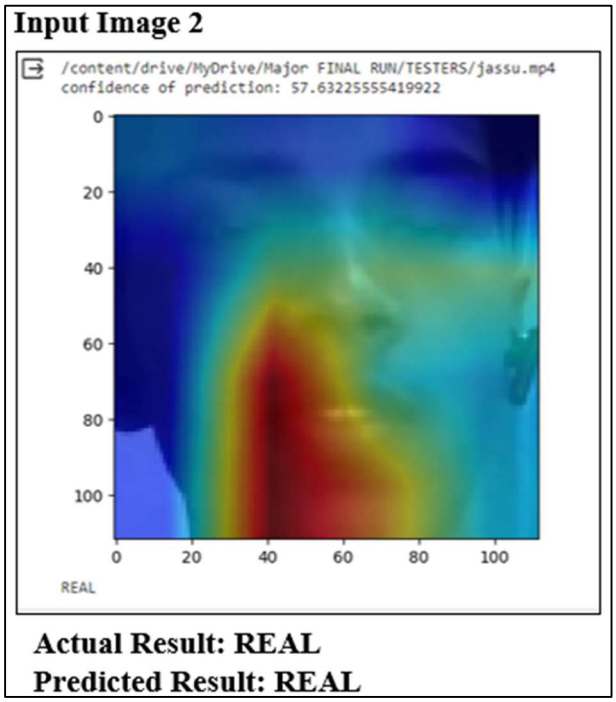

Figure 4.2: Test Case 2

Test Case 3: Author's real time video is given as input to the prediction model.

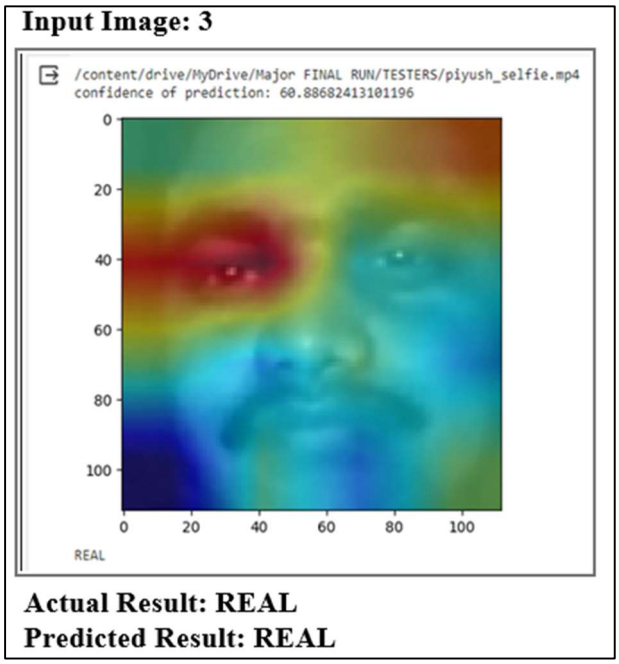

Figure 4.3: Test Case 3

Test Cases: Testing Scenarios

Table 4.1: Testing Scenarios of the model

Case_id	Testcase Description	Expected Results	Actual Result	Status
1	Upload an image file instead of video file	Error: Only video files allowed	Error: Only video files allowed	PASS
2	Upload a 200mb video file	Error: Max limit 100mb	Error: Max limit 100mb	PASS
3	Upload a file without any faces	Error: No faces detected. Cannot process the video	Error: No faces detected. Cannot process the video	PASS
4	Upload a Deep fake video	FAKE	FAKE	PASS
5	Upload a Real video	REAL	REAL	PASS
6	Upload a face cropped real video	REAL	REAL	PASS
7	Upload a face cropped deep faked video	FAKE	FAKE	PASS
8	Upload a video with many faces	FAKE/REAL	FAKE	PASS

4.2 Result Analysis

After the model was created and fitted, we trained it for 20 epochs. It has been noted that during the early training epochs, the accuracy is relatively poor also resulting higher training and validation loss. When the model undergoes training for 20 epochs, we see an improvement in model accuracy. The accuracy of the model at first epoch was 60.76% with a loss of 0.668 which was a poor result but as the epoch raised till 10 and 20 the accuracy reached to 90.75% and 95.76% respectively with a decrease in loss of 0.327 and 0.220 respectively.

We have compared our model with a number of other models. Based on the differences in the datasets and the methods applied, we have enumerated many elements of comparison between the models that are listed in Table 4.3: Comparison Analysis. It is not, however, a very good idea to compare the accuracy of several models to determine their efficacy. It really relies on how many movies are used and which generation the dataset is derived from.

Table 4.2: Model Development

No. of epochs	**Testing Loss**	**Training Loss**	**Testing Accuracy**	**Training Accuracy**
1	0.589	0.668	68.62%	60.76%
5	0.407	0.424	77.85%	87.66%
10	0.653	0.327	74.46	90.75%
15	0.559	0.268	79.38%	94.53%
20	0.536	0.220	**80.62%**	**95.76%**

4.3 Comparison Analysis

Table 4.3: Comparison Analysis

Authors	**Main Idea**	**Advantages**	**Dataset**	**TL**	**Method**	**Usage**
Darius et al.[4]	Meso-4 and MesoInception-4 networks are used in the process	-High Accuracy -Low Complexity	Face2Face, Deepfake *175 videos*	No	CNN	Video manipulation detection
Coccomini et al.[5]	Detecting deepfake using the Efficient Net technology on a mixed dataset	-Good AUC - F1 Score of 83.8%	DF-TIMIT, UADFC, FF++, GDDD, Celeb-DF, DFDC *5000 videos*	No	CNN – Efficient ViT	Video deepfake detection

Y. Li & S. Lyu[7]	Detection of warping leaves distinct artifacts by comparing the generated model with VGG-16	-Strong robustness -AUC 84.6%	UADFV *98 videos*	Yes	CNN – VGG16	Deepfake video and images detection
David et al.[10]	(RNN) for deepfake detection used the approach of using RNN for sequential processing of the frames along with ImageNet pre-trained model	-High accuracy -Training accuracy of 99.9%	HOHA *600 videos*	No	RNN-LSTM	Detection of manipulate media
OURS	**Presenting a CNN based model and a classifier**	**-Training Accuracy of 95.76%** **-Low response time** **-Low energy consumption** **-Trained on diversified 3^G Datasets**	**FF++, DFDC, Celeb-DF v2** ***1700 videos***	**Yes**	**CNN-ResNeXt 50 – RNN-LSTM**	**Detection of GAN generated synthesized videos**

Chapter 5

Conclusion and Further Work

We exhibited how reliable the proposed framework is and how to determine whether a video is real or deepfake using an artificial neural network-based approach. Our method is able to predict the outcome properly after analysing a single second of footage at a frame rate of 10 frames / second. We employed a previously trained ResNext CNN model to extract the frame level features and an LSTM for temporal chain processing to identify changes between the t and t-1 frame in order to implement the model. The video in frame sequences of 10, 20, 30, 40, 60, 80, and 100 can be processed by our model.

Our future work includes,

Any system that has been constructed has an opportunity for enhancements, especially if it has been developed with the newest, most prevalent technology and has an optimistic future.

• For convenience for users, web-based applications can be extended to browser plugins.

• The algorithm can be enhanced to detect complete body deep fakes, although at the moment it only detects facial deep fakes.

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).
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://www.kaggle.com/competitions/deepfake-detection-challenge/data
12. https://github.com/yuezunli/celeb-deepfakeforensics
13. https://github.com/ondyari/FaceForensics

Appendices

Appendix A

Weekly Progress Report

Figure A.1: Weekly Progress Report

D Y PATIL UNIVERSITY RAMRAO ADIK INSTITUTE OF TECHNOLOGY

RAMRAO ADIK INSTITUTE OF TECHNOLOGY
D. Y. PATIL VIDYANAGAR, SECTOR - 7, NERUL, NAVI MUMBAI - 400 706
WEBSITE: http://www.dypatil.edu/mumbai/rait/

Department of Computer Engineering
BE Major Project Weekly Project Performance Report
Even Semester 2023-2024

Title of the Project: **Deep Fake: A Study of DL based Tools.**	
Group No. **MP-01**	Roll No. & Name of Student 1: **Piyush Pal (20CE1180)**
	Roll No. & Name of Student 2: **Aakash Mishra (20CE1151)**
	Roll No. & Name of Student 3: **Harsh Jaiswal (20CE1264)**
Name of the Guide: **Dr. Vanita Mane**	

Expected Topics to be Covered Week-wise	Sr. No.	Date	Attendance			Progress/Suggestion
			1	2	3	
Week 1: Clear and Precise Objective	1	NIL	NIL	NIL	NIL	College not commenced
	2	NIL	NIL	NIL	NIL	College not commenced
	3	NIL	NIL	NIL	NIL	College not commenced
	4	18-01-2024	P	Akash	HS	Group Allotment Announcement
	5	19-01-2024	P	Akash	HS	First General meeting with Guide
Week 2: Abstract and Introduction	1	22-01-2024	P	Akash	HS	Topic Selection
	2	23-01-2024	P	Akash	A	Topic Selection revised
	3	24-01-2024	P	Akash	HS	Further research
	4	25-01-2024	P	Akash	HS	Abstract Presentation
	5	Republic Day	NIL	NIL	NIL	NIL
Week 3: Literature Survey	1	29-01-2024	P	Akash	HS	Papers on Face Verification.
	2	30-01-2024	P	A	HS	Papers on Deepfake Detection
	3	31-01-2024	A	Akash	HS	ll - - - - - - - ll - - - - - - - - -ll
	4	HORIZON	NIL	NIL	NIL	NIL
	5	HORIZON	NIL	NIL	NIL	NIL

Week 4: Limitations of Existing System	1	05-02-2024				Review papers on deepfake
	2	06-02-2024				Study of deepfake generation
	3	07-02-2024			A	Survey of deepfake crimes
	4	08-02-2024				Study of statistical methods
	5	09-02-2024				**MOCK PRESENTATION - I**
Week 5: Problem Definition Statement	1	12-02-2024		A		Additional Lit survey studied
	2	13-02-2024	A			Additional Lit survey studied
	3	14-02-2024				Identified Limitations
	4	15-02-2024				Problem Statement defined
	5	Tech HR				N/A
Week 6: Proposed Methodology	1	19-02-2024				Proposed Lstm -ResNeXt model
	2	20-02-2024				Study on proposed methodology
	3	21-02-2024				Other comparative solutions
	4	22-02-2024				Research on dataset selection
	5	23-02-2024		A		Research part
Week 7: Design & Architecture	1	26-02-2024				System Architecture defined
	2	27-02-2024				Model creation started
	3	28-02-2024				Further progress
	4	29-02-2024				**MOCK PRESENTATION - II**
	5	01-03-2024				Dry run conducted, result recorded
Week 8: Details of hardware & Software	1	02-03-2024				**CONCLAVE PARTICIPATION**
	2	04-03-2024	A			Hardware requirement purchased
	3	05-03-2024				from google colab worth 30$
	4	06-03-2024			A	Further progress--
	5	07-03-2023				Further progress--
Week 9: Implementation details	1	11-03-2024				Implemented with python language
	2	12-03-2024				Google Colab ENV.
	3	ABSENT	NIL	NIL	NIL	N/A
	4	14-03-2024				Dataset changed for further check
	5	15-03-2024				ll - - - - - - - ll - - - - - - - - -ll

Week 10: Result Analysis & Testing Strategies	1	18-03-2024				Result Analysis of various dataset
	2	19-03-2024				Study on accuracy improvement
	3	20-03-2024				Study on GUI development
	4	21-03-2024				II - - - - - - - II - - - - - - - - -II
	5	22-03-2024				**MOCK PRESENTATION - III**
Week 11: Conclusion and Future Work	1	HOLI	NIL	NIL	NIL	NIL
	2	26-03-2024				N/A
	3	27-03-2024				Preparation for Technovate'24
	4	28-03-2024				N/A
	5	Gudipadwa	NIL	NIL	NIL	NIL
Week 12: Report Submission	1	13-04-2024				**Technovate'24 Zonal Winner**
	2	18-04-2024				**Technovate'24 State Participation**
	3	25-04-2024				Report Plagiarism Check
	4	26-04-2024				Final Report Drafting
	5	29-04-2024				Final Report Submission

Dr. Vanita mane

Name and Signature of Guide

Appendix B

Plagiarism Report

Deep Fake: A Study of DL based Tools

ORIGINALITY REPORT

5%
SIMILARITY INDEX

PRIMARY SOURCES

1	openportal.isti.cnr.it Internet	19 words — < 1%
2	ijariie.com Internet	17 words — < 1%
3	"Information and Communications Security", Springer Science and Business Media LLC, 2020 Crossref	14 words — < 1%
4	Kalicharan Jalui, Aditya Jagtap, Saloni Sharma, Gilofer Mary, Reba Fernandes, Megha Kolhekar. "Synthetic Content Detection in Deepfake Video using Deep Learning", 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), 2022 Crossref	13 words — < 1%
5	scirp.org Internet	12 words — < 1%
6	Köse, Simge Yelkenci. "Hybrid Meta-Heuristic Approaches for Single and Multi-Objective Buffer Allocation Problems in Manufacturing Systems", Dokuz Eylul Universitesi (Turkey), 2024 ProQuest	10 words — < 1%
7	dash.harvard.edu Internet	10 words — < 1%

Appendix C

Project Competitions

Cyber-Security Conclave 2024

Nerul, March 02: We are thrilled to inform that we have taken active participation in the Cyber Security Conclave 2024 hosted by Ramrao Adik Institute of Technology, D Y Patil Deemed to be University. Through this competition we explored the intersection of cybersecurity and machine learning for cause of social well-being and fostering digital inclusivity. Dynamic panel discussions brought together experts from diverse backgrounds to tackle pressing cybersecurity challenges and safeguarding mechanisms. As a participating group from UG category, we have presented our latest finding w.r.t the topic of Deepfake Detection. The competition was promising in all ways and we were also awarded with "Certificate of Appreciation".

IEEE Technovation'24

Phase 1: Zonal Level

Nerul, April 13: IEEE Bombay Section under the leadership of Dr. Dattatray Sawant hosted the Technovation'24, a wide-spread event at Ramrao Adik Institute of Technology, DY Patil Deemed to be University. The competition featured hardware as well as software projects in the listings. RAIT was chosen to host the event featuring for Zone 1 (Mumbai, Thane, Palghar, Navi Mumbai, Silvasa, Diu & Daman). It was an opportunity to represent our innovation with the motivation of Dr. Vanita Mane Ma'am and win as the *First-place winner in the event from the entire Zone 1*.

Phase 2: State Level

Sangamner, April 18: The competition in the Phase 1 was conducted in three different zones of Mumbai, Pandharpur & Pune. The winners of all the zones were further instructed to participate for the final state level round which was hosted by Amrutvahini College of Engineering, Sangamner. The later competition was a bi-fold round. We were fortunate to reach the final round. This participation would have not been possible without the support of Principal, Dr. Mukesh Patil Sir, Head of the Department of Computer Engineering, Dr. Amarsinh Vidhate, and our Project Guide, Dr. Vanita Mane Ma'am. We were later felicitated by IEEE Bombay Section with Certificates and medals along with a cash prize of Five Thousand Rupees.

IEEE BOMBAY SECTION

STUDENT ACTIVITIES COMMITTE

IEEE BOMBAY SECTION

D Y PATIL
RAMRAO ADIK INSTITUTE OF TECHNOLOGY
NAVI MUMBAI

1st

Technovation'24

CERTIFICATE

of appreciation

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Aakash Mishra

FOR BEING FIRST WINNER IN TECHNOVATION 2024(~~HARDWARE~~ / SOFTWARE CATEGORY) DIVISION 1(Mumbai, Thane, Palghar, Navi Mumbai, Silvassa, Diu & Daman) HOSTED BY RAMRAO ADIK INSTITUTE OF TECHNOLOGY, NAVI MUMBAI ON 13 APRIL ,2024

Dr. Vishwesh Vyawahare
BRANCH COUNSELLOR
IEEE RAIT

Prof. Dattatray Sawant
CHAIR
STUDENT ACTIVITIES COMMITTE (SAC)
IEEE BOMBAY SECTION

Dr. Mukesh D. Patil
PRINCIPAL
RAIT

IEEE BOMBAY SECTION
STUDENT ACTIVITIES COMMITTEE

IEEE BOMBAY SECTION

D Y PATIL
RAMRAO ADIK INSTITUTE OF TECHNOLOGY
NAVI MUMBAI

1st

Technovation'24

CERTIFICATE

of appreciation

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Piyush Pal

FOR BEING FIRST WINNER IN TECHNOVATION 2024(~~HARDWARE~~ / SOFTWARE CATEGORY) DIVISION 1(Mumbai, Thane, Palghar, Navi Mumbai, Silvassa, Diu & Daman) HOSTED BY RAMRAO ADIK INSTITUTE OF TECHNOLOGY, NAVI MUMBAI ON 13 APRIL ,2024

Dr. Vishwesh Vyawahare
BRANCH COUNSELLOR
IEEE RAIT

Prof. Dattatray Sawant
CHAIR
STUDENT ACTIVITIES COMMITTE (SAC)
IEEE BOMBAY SECTION

Dr. Mukesh D. Patil
PRINCIPAL
RAIT

IEEE BOMBAY SECTION
STUDENT ACTIVITIES COMMITTEE

IEEE BOMBAY SECTION

D Y PATIL
RAMRAO ADIK INSTITUTE OF TECHNOLOGY
NAVI MUMBAI

1st

Technovation'24

CERTIFICATE

of appreciation

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Harsh Jaiswal

FOR BEING FIRST WINNER IN TECHNOVATION 2024(~~HARDWARE~~ / SOFTWARE CATEGORY) DIVISION 1(Mumbai, Thane, Palghar, Navi Mumbai, Silvassa, Diu & Daman) HOSTED BY RAMRAO ADIK INSTITUTE OF TECHNOLOGY, NAVI MUMBAI ON 13 APRIL ,2024

Dr. Vishwesh Vyawahare
BRANCH COUNSELLOR
IEEE RAIT

Prof. Dattatray Sawant
CHAIR
STUDENT ACTIVITIES COMMITTE (SAC)
IEEE BOMBAY SECTION

Dr. Mukesh D. Patil
PRINCIPAL
RAIT

IEEE BOMBAY SECTION

IEEE BOMBAY SECTION
STUDENT ACTIVITIES COMMITTEE

Finale
Technovation'24

CERTIFICATE

of appreciation

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Piyush Pal

FOR ACTIVELY PARTICIPATING IN FINALE OF TECHNOVATION 2024 (HARDWARE / SOFTWARE CATEGORY) HOSTED BY AMRUTVAHINI COLLEGE OF ENGINEERING SANGAMNER, AHMEDNAGAR ON 18 APRIL ,2024

D S Sawant

Prof. Dattatray Sawant
CHAIR
STUDENT ACTIVITIES COMMITTE (SAC)
IEEE BOMBAY SECTION

Anand Suhas Gharapure
CHAIR
IEEE BOMBAY SECTION

IEEE BOMBAY SECTION

IEEE BOMBAY SECTION
STUDENT ACTIVITIES COMMITTEE

★★★★★

Finale Technovation'24

CERTIFICATE

of appreciation

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Harsh Jaiswal

FOR ACTIVELY PARTICIPATING IN FINALE OF TECHNOVATION 2024 (HARDWARE / SOFTWARE CATEGORY) HOSTED BY AMRUTVAHINI COLLEGE OF ENGINEERING SANGAMNER, AHMEDNAGAR ON 18 APRIL ,2024

D S Sawant

Prof. Dattatray Sawant
CHAIR
STUDENT ACTIVITIES COMMITTE (SAC)
IEEE BOMBAY SECTION

Anand Suhas Gharapure
CHAIR
IEEE BOMBAY SECTION

IEEE BOMBAY SECTION

Finale

Technovation'24

CERTIFICATE

of appreciation

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Aakash Mishra

FOR ACTIVELY PARTICIPATING IN FINALE OF TECHNOVATION 2024 (HARDWARE / SOFTWARE CATEGORY) HOSTED BY AMRUTVAHINI COLLEGE OF ENGINEERING SANGAMNER, AHMEDNAGAR ON 18 APRIL ,2024

IEEE BOMBAY SECTION
STUDENT ACTIVITIES COMMITTEE

D S Sawant

Prof. Dattatray Sawant
CHAIR
STUDENT ACTIVITIES COMMITTE (SAC)
IEEE BOMBAY SECTION

Anand Suhas Gharapure
CHAIR
IEEE BOMBAY SECTION

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

CYBERSECURITY
CONCLAVE 2024

RAIT acm-w

RAIT acm Chapter

E-CELL
MOTIF
RAIT

Certificate of Appreciation

PRESENTED TO

Aakash Mishra, Piyush Pal, Harsh Jaiswal

in recognition for delivering a presentation on

Deepfake: A growing threat in Cybersecurity

in Security Research Symposium at Cybersecurity Conclave 2024
conducted on 1st-2nd March 2024.

Dr. Mukesh. D. Patil
Principal,
RAIT

INSTITUTION'S
INNOVATION
COUNCIL
(Ministry of Education Initiative)

Cyber Intelligence Global LLP
www.cyintglobal.com

Acknowledgment

We would like to take this opportunity to express our sincere gratitude to D. Y. Patil deemed to be University Ramrao Adik Institute of Technology, Nerul for giving us the opportunity to study and explore the topic "Deep Fake: A Study of DL based Tools". We are grateful to our respected Principal Sir, Dr. Mukesh D. Patil, for his support and guidance. We would like to express our gratitude to Dr. Amarsinh V. Vidhate, HOD Computer Engineering for his support. We would like to thank our Major Project coordinator, Dr. Aditi Chhabria for her constant support and guidance. We would like to sincerely thank Dr. Vanita Mane for her feedback and supervision throughout this study. This project would not been possible without his guidance. We would also like to thank our classmates for their constant encouragement and support.

Date: _____