INTELLIGENCE FOR DEEPFAKE DETECTION: AN EXPERIMENTAL MEASURE TO COMBAT RISING CYBER THREAT

BY

MD. RAISUL ISLAM
ID: 202-15-3813
AND

MD. AMINUL ISLAM RAKIB
ID: 202-15-3832

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. Afjal Hossan Sarower
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised By

Ms. Tapasy Rabeya
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

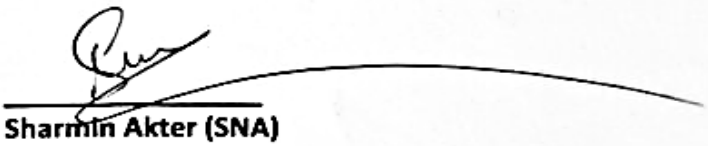
APPROVAL

This Project titled " **Intelligence for Deepfake Detection: An Experimental Measure to Combat Rising Cyber Threat** ", submitted by **Md. Raisul Islam** and **Md. Aminul Islam Rakib** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 15-07-2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque (SMAH)
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Board Chairman

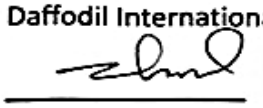


Sharmin Akter (SNA)
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1

Tapasy Rabeya (TRA)
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2



Dr. Md. Zulfiker Mahmud (ZM)
Professor
Department of CSE
Jagannath University

External Examiner

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Afjal Hossan Sarower, Lecturer (Senior Scale), Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

30.06.24,

Mr. Afjal Hossan Sarower
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

30-6-24

Ms. Tapasy Rabeya
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

01-07-2024

Md. Raisul Islam
ID: 202-15-3813
Department of CSE
Daffodil International University

01.07.2024

Md. Aminul Islam Rakib
ID: 202-15-3832
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Mr. Afjal Hossan Sarower, Lecturer (Senior Scale),** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Cyber Security*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Deepfake detection technology primarily combines artificial intelligence and machine learning, including deep learning techniques such as Multi-layer Perceptrons (MLP), Convolutional Neural Networks (CNN), and EfficientNet etc. to detect deep fake videos or images using this model. An overview of deepfake detection systems is given in this abstract, along with a focus on important methods, difficulties, and potential future developments. There are several methods for detecting deep fakes, such as analyzing facial features and artifacts using images, analyzing voice characteristics and lip-syncing using audio, and utilizing hybrid approaches that combine several modalities. Difficulties in detecting deep fakes include the quick development of methods for generating deep fakes, the scarcity of labeled training data, and the susceptibility of detection systems to hostile attacks. Despite these obstacles, continuous research endeavors seek to augment the efficiency and expandability of deepfake detection systems via developments in digital forensics, machine learning, and interdisciplinary cooperation. In the future, deepfake detection research will focus on creating multimodal detection systems, integrating blockchain technology for tamper-proof authentication, and investigating explainable AI methods for analyzing detection data.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES **PAGE NO**

LIST OF TABLES

TABLES **PAGE NO**

CHAPTER 1
INTRODUCTION

1.1 Overview

Regarding digital media, deep fakes represent an emerging threat in cyberspace [1] around the world nowadays. Deepfake is mainly fake video using people's images or changing the abstracts of faces like nose, mouth, eye, ear, backgrounds, etc. A hybrid of the terms "deep learning" and "fake" [2], deep fakes are incredibly lifelike films that have been digitally altered to show individuals saying and acting in ways that have never actually occurred [3]. Stated differently, deep fakes employ artificial intelligence (AI) and facial mapping technologies to replace a person's face in a video with that of another person [4]. Deepfakes are computer-generated images that are so lifelike that it is nearly impossible to tell them apart from actual images or movies [5].

Combining the terms deep learning and fake results in the phrase deepfake. Using deep neural networks, deepfakes are produced that show events that never occurred in an attempt to amuse, discredit people, propagate false information, and other purposes [6]. Artificial intelligence (AI) technology creates deepfake, synthetic media, which includes photos, movies, and audio. These types of media depict things that are not genuine or have never happened [7]. By using deep fake videos, social, economic, and other activities are done the wrong way. The capacity of deepfakes to disseminate misleading information that seems to originate from reliable sources is the biggest threat they provide. For instance, a deepfake film purporting to show Ukrainian President Volodymyr Zelenskyy pleading with his soldiers to surrender was made public in 2022 [8]. A deepfake featuring Mark Zuckerberg bragging about how Facebook "owns" its users fooled the founder of the company [9]. An edited film of President Obama was created using video editing in this Buzzfeed piece featuring actor/director Jordan Peele, highlighting the risks associated with manipulating video technology [10]. Deepfakes threaten International security, hold Potential for anti-terrorismt [11], the percentages of this occur are increasing day by day continuously.

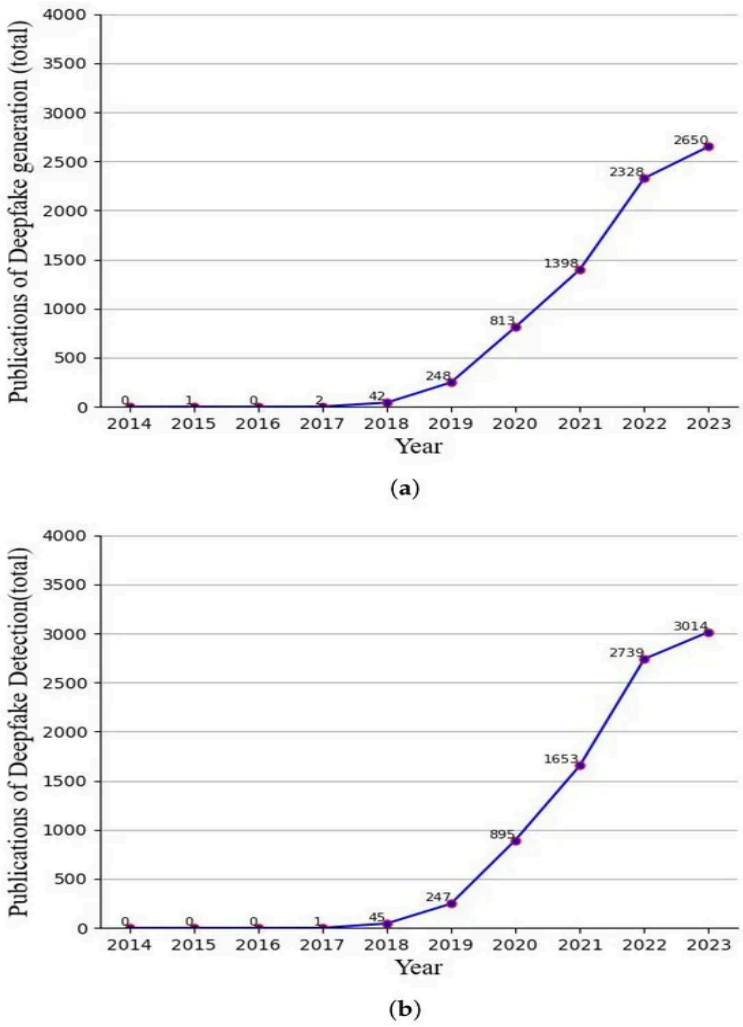


Fig-1.1.1: Relative publication data obtained from Dimensions database [12] at the end of 2024 by searching “deepfake generation” and “deepfake detection” as keywords: (**a**) number of deepfake-generation-related scholarly papers from 2014 to 2024; (**b**) number of deepfake-detection-related papers from 2014 to 2024.

EfficientNet

A class of convolutional neural networks (CNNs) called EfficientNet was created [13] with the goal of achieving cutting-edge performance while using less computing power. EfficientNet, which was unveiled by Google AI in 2019, maximizes accuracy and efficiency by methodically increasing the network's depth, breadth, and resolution by the use of compound scaling [14]. By ensuring balanced scaling, this method produces models that, when compared to standard architectures, perform very well on picture classification tasks while using less computing power [15]. EfficientNet models have achieved impressive

results on benchmarks like ImageNet, making them a popular choice for various deep learning applications.. An enhanced target identification method based on EfficientNet is suggested to increase the detection accuracy for small-scale targets in complicated scenarios [16]. Artificial intelligence (AI)-generated phony face videos are now easier to create than ever because to the rapid advancements in media generating methods. Facebook launched the Deep Fake Detection Challenge (DFDC) to hasten the creation of innovative methods to reveal forged movies. This challenge showcased many techniques to address this issue. All of the winners, according to an analysis of the best-performing systems, used EfficientNet networks that had previously been trained using films of people manipulating their faces [17].

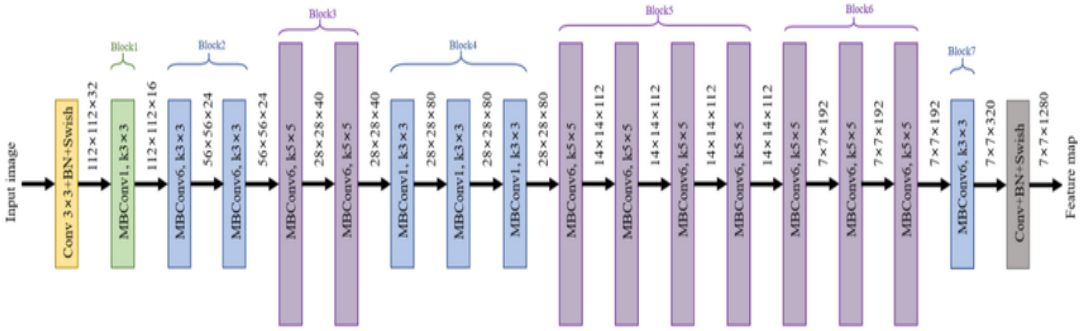


Figure 1.1.2: Schematic representation of EfficientNet [18]

Original Convolutional Neural Networks (CNNs)

A particular kind of deep learning technique called a Convolutional Neural Network (CNN), or ConvNet, is primarily designed for applications requiring object identification, such as picture categorization, detection, and segmentation [19]. Ths seminal paper marks a significant milestone in the development of Convolutional Neural Networks (CNNs). Published in *Neural Computation* in 1989, this research introduces a multi-layer neural network architecture designed for visual pattern recognition tasks. The paper [20] demonstrates the effectiveness of backpropagation in training deep neural networks for the specific task of handwritten digit recognition, showcasing remarkable accuracy and efficiency. This article [21] discusses methods for identifying deepfakes.This preprocessing technique makes use of convolutional neural networks (CNNs) as well as other neural networks. The result is ultimately what distinguishes the real image from the fake. After the model has been trained on a dataset, this is achieved by using an appropriate CNN to classify the pictures as true or false.

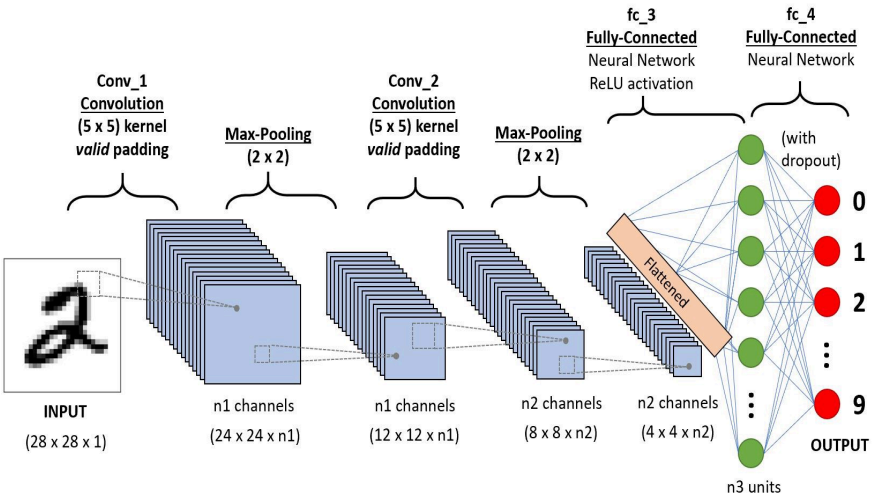


Figure 1.1.3: Schematic representation of CNN [22]

Xception

The Xception (Extreme Inception) model, an extension of the Inception architecture, is highly effective for deepfake detection due to its superior performance in capturing intricate patterns in image data. Introduced by François Chollet, Depthwise separable convolutions, which greatly decrease computational cost while retaining good accuracy, are used in Xception in lieu of the traditional Inception modules. The architecture comprises entry, middle, and exit flow, each consisting of convolutional layers and residual connections, enabling the model to learn complex features efficiently. For deepfake detection, Xception's ability to extract and analyze subtle facial details makes it a robust choice, often outperforming traditional convolutional neural networks in identifying manipulated images and videos. This research [23] provides a precise and extremely effective approach for identifying deepfakes, which are a serious danger to society since they are becoming widely available via mobile apps. By using the cutting-edge Xception network, a convolutional neural network (CNN), we tackle the problem of recognizing and classifying deepfakes in both pictures and movies.

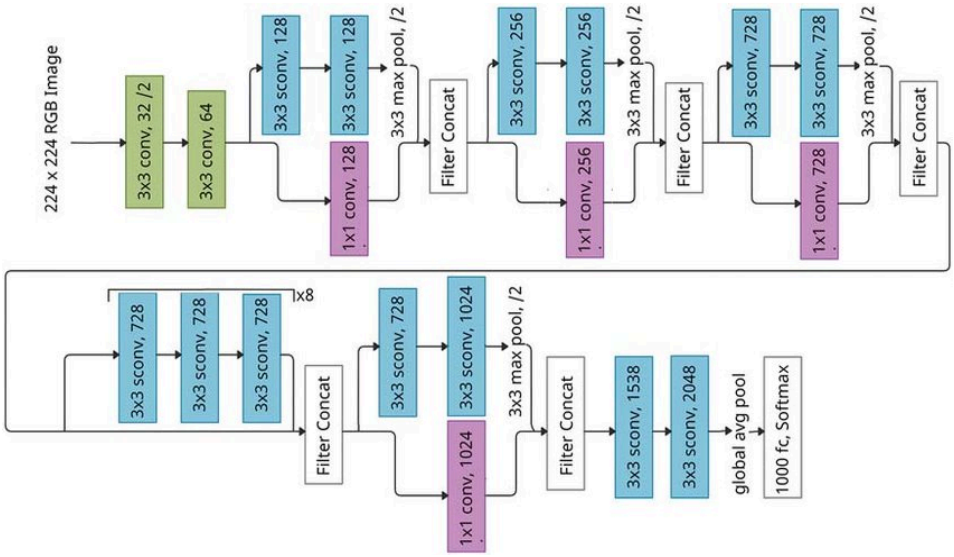


Figure 1.1.4: Schematic representation of Xception [24]

Multilayer Perceptron (MLP)

One kind of feedforward artificial neural network that is often utilized for applications like deepfake detection is the Multilayer Perceptron (MLP). An input layer, one or more hidden layers, and an output layer make up the MLP architecture. Numerous neurons that are totally linked to the neurons in the layers above and below may be found in each layer. Typically, elements collected from visual data, such as pixel values or more abstract features produced from convolutional layers in a previous CNN, are fed into the input layer in the context of deepfake detection. By executing nonlinear transformations on the inputs, the hidden layers enable the model to acquire intricate patterns and representations that are essential for differentiating between authentic and altered pictures. The output layer typically consists of a single neuron with a sigmoid activation function that determines whether or not the input picture is a deepfake based on a likelihood score. The MLP is a potent technique for spotting minute abnormalities and discrepancies typical of deepfakes because of its capacity to capture complex connections within the data.

1.2 Background and Present State

Deepfake technology has rapidly advanced over the years, posing significant challenges in detecting and preventing the spread of manipulated content online. As of right now, anomalies in facial expressions, auditory signals, and visual clues are detected using advanced algorithms and machine learning approaches in deep face identification. On the other hand, deepfake generation has become increasingly accessible to the general public, making it crucial for researchers and tech companies to stay ahead in developing robust detection methods. With the rise of social media and digital platforms, the need for effective deep face detection tools is more pressing than ever to combat the potential misuse of this technology for malicious purposes. Deep fake creation and detection will need constant improvement as technology develops in order to preserve digital content integrity and shield people from the damaging impacts of false information.

Deepfake detection is a fast developing subject with substantial technology advancements and persistent problems. The dynamic nature of deepfake technology requires ongoing research and collaboration between engineers, ethicists, politicians, and legal experts, even though existing detection approaches show promise. Creating resilient and flexible methods to shield people from the possibly dangerous effects of deepfakes is the aim.

Figure 1.1.5: Trend in Google searches for deep fake queries [25]. (adapted from google.com) Retrieved from Google Trends: https://trends.google.com/trends/explore?date=2016-01-01%202023-01-12&q=deepfake

1.3 Problem Statement

Deepfakes are synthetic media that have been altered to look real, and they are becoming more and more realistic due to the rapid growth of deep learning algorithms. The authenticity and confidence of information in the digital sphere are seriously threatened by this technology. It is possible to utilize deep fakes with evil intent, like;

- ➢ Spreading false information and propaganda: fake news videos that can sway public opinion and affect elections can be made using deep fakes.

- ➢ Reputational harm: Produced videos using deep lakes have the potential to harm people and companies' reputations.

- ➢ Financial fraud: Deepfakes are useful for financial fraud since they allow people to imitate each other.

- ➢ Becoming harder to tell real news from fake: The increasing usage of fake news can cause people to lose faith in established media outlets.

The development of strong and trustworthy deep face detection algorithms is therefore imperative. It should be possible for these systems to:

- ➢ Declare with high confidence which videos are real and which are deep fake.

- ➢ Extrapolate to previously undiscovered deep fakes made using various methodologies.

- ➢ Possess scalability and processing efficiency for real-time use.

- ➢ Resist hostile attacks, such as deep fakes, that target detection systems explicitly.

Due to the following reasons, creating such a system is a difficult task: Deepfake-generating techniques are always evolving. It becomes harder and harder to identify deepfake creation processes as they advance in sophistication.
In spite of these obstacles, creating deep fake detection systems that work is essential to counteracting the harmful effects of this technology and preserving information integrity in the digital era.

1.4 Objectives

The main driving force behind this endeavor is to create a trusted platform where users can upload a video to check if it is fake or real. That's why people can easily authenticate the news and others. The digital age's ability to trust information is seriously threatened by the quickening growth of hyper-realistic synthetic media, or "deepfakes," produced using deep learning techniques. Democracies, institutions, and people themselves may be severely damaged by their capacity to modify audio and video material with ease.

The development of deep fake detection systems is primarily motivated by the following reasons:

1. Countering Misinformation and Disinformation: By influencing public opinion and reducing trust in traditional media sources and other internet platforms, deep fakes can be used to control misinformation and disinformation.
2. Safeguarding Individuals and Institutions: Deepfake can be used to damage the individual reputations of people and injure the reputations of institutions, which can be harmful to all. A deep fake detection system would shield people and organizations from nefarious manipulation for safeguarding.

The main objective of this paper is to develop a secure and transparent web platform where users can upload videos and images to detect fake or real shows as output.

This research makes the following key contributions:

1. Explore and select benchmark datasets for the deepfake detection system.
2. Achieving high accuracy with less FPR results in significantly lower power consumption on the existing dataset.
3. Creating a dataset focusing on the local celebrity.
4. To develop a hybrid model for detecting with audio and video analysis.
5. To develop a website in which people can upload a video to check whether this video is deepfake generated or not.

In conclusion, detecting deep fakes is not just a technical challenge but a social and ethical imperative. By building robust and accessible detection systems, we can safeguard the integrity of information, protect individuals and institutions, and promote responsible use of technology in our increasingly interconnected world.

1.5 Scope and Limitations

The research will offer a thorough overview of the deep fake detection landscape, covering both image and video-based analytic techniques. Important facets of the project's scope consist of:

- **Comprehending Deepfake Technology**: This basic knowledge will be the starting point for delving into the nuances of deep fake identification.

- **Methods for Deep fake Detection:** These will comprise image-based analysis techniques like digital forensic analysis, artifact detection, and face feature analysis. Hybrid methods that combine audio and picture analysis will also be investigated.

- **Challenges with Deep fake Detection:** The study will address the challenges, which include the rapid advancement of deepfake generating techniques, the scarcity of labeled training data, the vulnerability to adversarial attacks, and the computational intricacy.

- **Evaluation Metrics and Performance**: The research will examine the metrics for evaluating deepfake detection systems' performance, such as accuracy, precision, recall, and F1 score.

- **Current Deepfake Detection Systems:** A survey and comparison of current open-source and commercial deepfake detection systems will be carried out.

By drawing this line in the project scope, we hope to give stakeholders the information they need to navigate this intricate and quickly changing field and to gain a thorough grasp of deepfake detection systems.

1.6 Report Organization

The suggested report has a thorough format that walks readers through the methodology, conclusions, and ramifications of the study. Every chapter has a distinct function and adds to the document's overall cohesion and depth.

Chapter 1: Introduction

The quick spread of deepfakes produced by deep learning algorithms puts information trustworthiness in danger and might be detrimental to institutions, individuals, and democracy. Deepfake detection systems are designed to defend against malicious manipulation, refute misinformation and deception, and preserve reputations. This study describes the creation of a transparent and safe online platform for identifying deep fakes. Achieving high accuracy with low false positive rates, building a user-friendly website for deepfake verification, choosing benchmark datasets, and producing a local celebrity dataset are some of the major achievements. The initiative aims to maintain information integrity and encourage responsible technology usage by developing strong detection mechanisms.

Chapter 2: Literature Review

A thorough analysis of pertinent literature and earlier research is provided in the background chapter. This section provides a thorough overview of the deep-learning techniques to deep fake detection and production, as well as the theoretical underpinnings, technical developments, and methodology associated with deep fake detection. It lays forth the background for the present investigation and identifies knowledge gaps.

Chapter 3: Research Methodology

The study's methodology is described in the research methodology chapter. It offers understanding of the model architecture, data gathering techniques, study design, and the justification for selecting particular methodology. The chapter also covers any restrictions that could affect the study as well as ethical issues.

Chapter 4: Implementation

This chapter presents the experimental results obtained from the application of the result of learning and web platform of deepfake detection. Detailed about train model, model architecture, system testing and the evaluation of the whole model.

Chapter 5: Result and Analysis

Deepfake movies have the potential to damage public trust in the media and other public sources in addition to swaying public opinion. The ability to produce realistic recordings of public people, politicians, or celebrities behaving or speaking in ways they have never done might have a significant impact on society and democracy. How to avoid deepfakes and their drawbacks are covered in this chapter.

Chapter 6: Impact on Society. Environment and Sustainability

Aside from influencing public opinion, deepfake films can undermine confidence in the media and other public sources. The capacity to create lifelike recordings of politicians, public figures, or celebrities talking or acting in ways they never did might have profound effects on democracy and society. This chapter explains how to prevent deepfakes and their negative effects.

Chapter 7: Conclusion and Future Research

The whole report is summarized in this last chapter. It provides suggestions for real-world applications and further research in addition to summarizing the main results and restating the study's conclusions. Here is a description of the implications for future study.

1.7 Summary

An overview of the growing issues surrounding deepfake technology, which uses sophisticated machine learning algorithms to create extremely convincing fake media, is given in the introduction. Deepfakes were first exciting because of their creative potential, but because of their ability to trick and manipulate, they have come to be associated with anxiety in society. The prevalence of deepfakes undermines confidence and feeds disinformation campaigns, endangering the integrity of visual media. As a result, creating reliable deepfake detection systems has become crucial. Systems for facial and speech recognition can easily be tricked by fake content. Additionally, not every kind of deep fake has the ability to compromise biometric systems. In this paper, we propose a method for detecting deep fakes using deep learning algorithms to get the highest accuracy.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

To enable a comprehensive comprehension of the fundamental ideas and techniques used in our investigation on "Intelligence for Deepfake Detection: An Experimental Measure to Combat Rising Cyber Threat," By utilizing intelligence tools such as machine learning algorithms, facial recognition technology, and digital forensics, we can develop robust systems to detect deep fakes accurately. These tools can analyze patterns, anomalies, and inconsistencies in images or videos to identify potential deep fakes. Additionally, by gathering intelligence on emerging deepfake trends and tactics used by threat actors, we can stay ahead of the curve and proactively defend against new threats. Deepfake detection techniques are a collection of advanced approaches designed to recognize material that has been altered.

This approach relies heavily on image analysis, where computers examine pixel-by-pixel minutiae to find discrepancies that point to the existence of a deep fake. For example, distortions or artifacts in the picture may indicate manipulation. Moreover, the identification of deepfakes involving human beings depends on the detection of irregularities in face features. Algorithms may detect digital alteration in videos by closely examining non-verbal signs like eye movements and facial expressions.

2.2 Related Works

For the last couple years, there has been a concerted effort by various authors to develop methods for identifying deep fakes. This study [26] is about the morphing attack detection is a crucial research area, categorized into four main categories. The study uses digital images from Ferrara by 2014, extracted from 70 videos, and uses CNN, VGG16, and customized models. The proposed model achieves 89% accuracy, 90% FNR, 40 precision, and 79% VGG accuracy, highlighting the security risk posed by these attacks.

This paper's primary goal is to build a deep hybrid neural network model to detect deep fake videos. proposes a deep learning approach using multilayer perceptron (MLP) and convolutional neural networks (CNN) to detect fake videos. [38] The proposed model provides good classification results, with an accuracy of 84% and an AUC score of 0.87.

First of all, load all video data, and then do two different frame extractions (1) convert the image to numerical form by using CNN; and (2) identify facial landmarks. By using MLP, the coordinates of the eyes, nose, and lips are extracted. Using this data, the number of eye blinks and the shape of the eyes, nose, and lips are also extracted. Then combine the two methods and check the connotation layer to classify the deep fake as real or fake

2.3 Comparative Analysis and Summary

An essential part of this study on "Intelligence for Deepfake Detection: An Experimental Measure to Combat Rising Cyber Threat" is the comparative analysis that illuminates the subtle distinctions and complementary aspects of different approaches. By comparing the performance metrics and methodologies of the studies mentioned, a clearer understanding emerges regarding the effectiveness and limitations of each approach. The study underscored the importance of precision and the challenges associated with high false-negative rates (FNR), which can impact the reliability of detection systems in real-world applications.

On the other hand, the deep hybrid neural network model utilizing MLP and CNNs provided promising results [38], with an accuracy of 84% and an AUC score of 0.87. This approach emphasized the significance of combining image-based analysis with facial landmark extraction to enhance detection accuracy. The integration of eye blinks and facial feature shape analysis further refined the detection process, showcasing the advantages of a hybrid methodology.

Comparing these approaches reveals that while both models demonstrate substantial accuracy, the integration of multiple techniques, such as combining CNN with MLP, can enhance detection capabilities by leveraging the strengths of different algorithms. The analysis of non-verbal cues, such as eye movements and facial expressions, emerged as a critical factor in improving the robustness of deepfake detection systems.

Overall, by offering a thorough grasp of the present state-of-the-art in deepfake detection, this comparative study sets the scene for the next chapters. It highlights the importance of continued research and innovation in this field to address the evolving threats posed by deepfakes. In order to counter the growing cyber danger of deepfakes [53], it is important to have detection systems that are reliable, flexible, and multifarious. The findings gathered from this study are used to enrich the debate on the effect on society, suggestions, and implications for future research.

TABLE 2.3.1: Comparative Analysis Table

Author Name	Contribution	Research Gap/Limitation	Dataset Name/Source	Dataset Size/details	Model/Algorithms	Proposed Model/Performance
K. Raja *et al* [26]	Morphing Attack Detection - Database, Evaluation Platform and Benchmarking	Classified into 4 main categories and morphing attacks pose a threat to FRS systems and leave a major security risk to any nation where the malicious actor enters.	Digital images Ferrara by 2014, (printed and scanned) images,	34000 images extracted from 70 videos	CNN, VGG16, Customized Model (X+Y)	**CNN:** Accuracy: 89% **VGG:** Accuracy: 79%
V. Blanz et al. [27]	Face Recognition Vendor Test with Computer graphics and mapping from 2D images to a 3D representation	Estimates a textured 3D face model from a single image, along with all relevant scene parameters, such as 3D orientation, position, focal length of the camera	Morphable Model of 3D faces	200 textured Cyberware (TM)	Morphing Attack Detection (MAD)	**Accuracy: 73%**
CAVE et al. [28]	Enables automatic replacement of faces across different poses, lighting, facial expression, image resolution, image blur, and skin tone	This uses a 2D image base, but sometimes 2D images are not detached from light conditions.	Own created Dataset	33,000 faces	Face De-Identification, Switching Faces	**Accuracy: 68%**
IEEE Conference Publication et al. [29]	contribution is the semi-supervised labelling of the face segmentation system and face swapping pipeline.	Need to improve the detection of 3D expressions and sensitive face landmark detection.	IARPA Janus CS2	1275 videos	Face segmentation, 3D face shape fitting, Face swapping and blending	**Accuracy 77.2%**
N. Damer et al. [30]	first synthetic-based dataset SMDD, 30,000 attacks and 50,000 bona fide samples	The limitation is the SMDD data set only creates MAs using one morphing	SMDD, LMA-DRD (D), MorGAN-LMA, MorGAN-GAN, and FRLL-Morph.	500k synthetic images	Inception-MAD, MixFace Net-MAD, PW-MAD	**Accuracy: 88%**
E. Sarkar et al.	provide an open-source morphing tool and publicly available	GAN-based morphing techniques are not well-studied	FERET [12], FRLL [13]	FERET: References: 418,439,	OpenCV, FaceMorpher, StyleGA	**OpenCV: Accuracy: 59.8 - 97.8%**

[31]	datasets			Probes: 418,439 FRLL: Reference : 1,984, Probes: 4,153	N2, MORGAN-II	
CSDL et al. [32]	They propose a novel solution which is inspired by recent progress in artistic style transfer, and defines the concepts of content and style as functions in the feature space of convolutional neural networks trained for object recognition.	Used dlib, which is accurate only up to a certain degree of head rotation and gains may be possible when using the VGGFace network for the content and style loss as suggested by Li et al.	CelebA,KSTAR,VGG,VGG-Face,ImageNet	200,000 images	feed-forward convolutional neural network to perform face swapping using style transfer.	**Accuracy: 87%**
S. Waseem et al. [33]	Effective deepfake detection solutions for face and expression swap efforts to protect the authenticity and trustworthiness of visual media.	The paper lacks a comprehensive survey on deepfake face and expression swap techniques, necessitating this study.	1st Gen.: UADFV, Deepfake-TIMIT 2nd Gen.: (DFD), (DFDC-P), MNIST+, (FSh), Celeb-DF, 3rd Gen.: (DFDC), (WDF), DF-Platter, (FMFCC-V), (OF), (FN), (FFIW), (KoDF), (DF-1.0)	Combine approximately 8200 videos	Autoencoders and generative adversarial networks.	**Accuracy:** LSTM+CNN 96% CNN 81%
G. Gao et al. [34]	By forcing the produced identities to maintain appropriate distances from the targets, they employ the IB principle for disentangled representation in order to recover the minimally required identity and perceptual	They only with InfoSwap method but also space with the work part of face detection like nose, eye, lip etc.	FaceForensics++ (FF++) datasets	96000 images after pre-processing 4000 images t	InfoSwap	**Accuracy: 60.41%**

	information..					
A. Siarohin et al. [35]	A self-supervised framework for image animation, utilizing a first-order motion model with learned key points and local affine transformations, coupled with an occlusion-aware generator, outperforming existing methods on various benchmarks.	They work with motion video and describe two key points displacements and local affine transformations.	The BAIR robot pushing, The UvA-Nemo, The VoxCeleb, Following Tulyakov	dataset of 22496 videos	Warframe, VoxCeleb	**Accuracy 93.8%**
A. Mousa et al. [36]	the new approach to speech morphing that avoids the extraction of fundamental as well as the detection of phone or syllable boundaries	They work with very few articles and it is work only with higher degrees of unsupervised morphing between any two utterances.	CMU ARCTIC speech corpus.	7KHz to 1KHz audio	DTW−based Alignment	**LSP accuracy 30-40%**
A. Firc et al. [37]	providing a united taxonomy for facial and speech deep fakes and providing an overview of deep fake creation tools	The gap between a comprehensive survey on deep fake creation and detection methods and the lack of security analysis regarding the potential threats posed by deepfakes to biometric systems.	DFDC preview, Cele-DF v2, FaceForensics++, UADFFV	124,000 deepfake videos	Xception, Capsule Network and DSP-FWA.	**Accuracy: 83%**
U. Kosarkar et al. [38]	A method to extract facial features such as eye blinks, eye shape, lip shape, and nose shape from the facial landmarks. A deep hybrid neural network model to detect deep fake videos using both facial landmarks data and image frame data	It struggles with face detection in low-light conditions and does not handle videos with multiple people in the frame, also they don't work with larger datasets for completeness.	(DFDC), FaceForensics++	used 318 videos	(MLP) and (CNN) to detect deepfake videos using facial landmarks.	**Accuracy: 84%**
R. Tolosana	the selection of specific facial regions, and	Work with only Xception, Capsule Network, and	**1st Generation**:	Total videos **13,877**.	Xception, Capsule Network,	**UADFV accuracy: 79%**

et al. [39]	fusion approaches in order to develop more robust fake detectors. An in-depth comparison in terms of performance among Identity Swap databases of the 1st and 2nd generation.	DSP-FWA but they need more options such as physiological measures, temporal features, or 3D decomposition	UADFV (2018), FaceForenics++ (2019) **2nd Generation:** Celeb-DF v2 (2019), DFDC Preview (2019)	There are **3,070** real videos from YouTube and **10,807** Fack videos	DSP-FWA	
H. Dang et al. [40]	Detecting manipulated facial images and videos is an increasingly important topic in digital media forensics.	Anticipating the predicted attention maps for manipulated face images and videos would reveal hints about the type, magnitude, and even intention of the manipulation.	DFFD(2019)	Real Images:58703 , Deep Fake Images: 240336	CNN + Attention mechanism	**Accuracy: 91.3%**
T. Jung et al. [41]	To detect Deepfakes generated through the generative adversarial network (GAN) model via an algorithm called DeepVision to analyze a significant change in the pattern of blinking	The main limitation of integrity verification algorithms performed only on the basis of pixels.	OWN	8 videos and 8 images	GAN	**Accuracy: 87.5%**
U. A. Ciftci et al. [42]	To create deep fake detector	create only superior detectors based on facial expression	CelebDF	7000 images	CNN	**Accuracy 91.3%**
Hu et al. [43]	Investigate visual features of facial landmark manipulation	The quality of the produced landmark the image was unsuitable, requiring a massive amount of target images	CelebV, VoxCeleb1	250,000 images from more than 7000 celebrities	LE-GAN	**CelebV:** CSIM = 0.661 PRMSE = 2.68 AUCON = 0.672 **VoxCeleb1:** PRMSE = 1.58 AUCON = 0.793

Ke & Wang, [45]	Examine feature restoration using transfer learning	The model has limited performance on low-level artifact features	Celeb-DF	7000 images	DF-UD	**Accuracy 88.92%**
Alnaim et al. [46]	Analyze visual artifacts of manipulated masked faces	Experiments were performed on limited datasets of masked faces	DFFMD	299 kb of JSON data	ResNetV2 with VGG19	**Accuracy 79.88 %**
Abdul Qadir et al. [47]	Provide a hybrid method that feeds the ResNet-Swish-BiLSTM classification with video input consisting of consecutive targeted frames.	Experiments of training model acts slowly	DFDC and Face Forensics	1,08500 images	BiLSTM-based residual network	**Accuracy 96.23 %**

2.4 Open Issue

Deepfake photos and videos are produced on a daily basis in vast quantities because to the widespread availability of software and technologies that make them. Academic academics now face significant challenges when examining and deciphering deep fake photos and movies. The absence of high-quality datasets is one of the biggest issues the researchers are dealing with. The scalability of the deep learning techniques used today is a problem. Stated differently, face swapping is detected by the deepfake algorithms [33] by using fragmented data sets. Nevertheless, using their methods on extensive datasets might lead to unsatisfactory outcomes. The scalability problem is another obstacle that the researchers need to take into account. Secondly, future research has to concentrate on developing more scalable and reliable models that can be used with any size and quality of high-quality information. In addition to the aforementioned difficulties, it is well recognized that big datasets are necessary for the training phase of deep learning models in order to get satisfactory results. Unfortunately, these datasets are either restricted or need authorization from social media providers. Fourth, new forms of created pictures and videos that the existing deep learning models may not be able to detect may arise as a result of the quick growth of deep fake GAN algorithms. Therefore, it is evident from all of the aforementioned difficulties that there is a strong need to create scalable and reliable deep learning models in order to identify phony photos and videos.

2.5 Summary

Important developments and ongoing shortcomings in the field of morphing attack detection and deepfake prevention are highlighted by the thorough evaluation of several research contributions. First off, much progress has been achieved in developing databases, assessment platforms, and benchmarking techniques for morphing attack detection thanks to research like [26]. Classifying morphing attacks thoroughly and assessing the security concerns they may pose to facial recognition systems is still urgently necessary, though. Accurately recognizing 3D emotions and sensitive facial features remains a difficulty, despite research initiatives such as [27] focusing on estimating textured 3D face models from single photos. Furthermore, while utilizing 2D picture bases and ignoring differences in lighting conditions, [28] presents automated face replacement capabilities.

In addition, although [29] presents synthetic-based datasets for the purpose of morphing attack detection, the use of a single morphing approach creates restrictions that impede thorough assessment. In a similar vein, [31] notes the gaps in knowledge on new risks to biometric systems and emphasizes the need for an extensive study on deepfake development and detection methods. Further evidence of ongoing difficulties with detection accuracy comes from research like [38], which has trouble with face identification in low light and trouble processing movies, including many subjects.

Studies such as [34] employ the information bottleneck principle for disentangled representation in their methodology, but the analysis's breadth is constrained by the use of a single approach. Furthermore, even if [35] suggests a self-supervised framework for picture animation, certain drawbacks in practical applications are overlooked due to the dependence on motion video input. Analogously, [37] offers a unified taxonomy for deep fakes including the face and speech, but it does not do a thorough security study of all possible dangers, which leaves important gaps in our knowledge of the hazards associated with deepfake technology.

Finally, a comparison study across identity swap databases from several generations is presented by [26], which emphasizes the necessity of using a variety of detection methods in addition to Xception, Capsule Network, and DSP-FWA. All things considered, filling up these holes and overcoming these constraints together can result in stronger and more successful tactics to counter morphing assaults and slow down the spread of deep fake technology.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Overview

The research investigates advanced deep learning approaches for detecting deepfakes, a rising cybersecurity threat, by evaluating the performance of a hybrid model with EfficientNet+Xception & Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN), and a hybrid CNN+Multi-Layer Perceptron (MLP) model. The methodology begins with the collection of a comprehensive dataset from publicly available sources like FaceForensics++ [54] and DFDC [49], ensuring a balanced composition of real and fake samples. Pre-processing steps include frame extraction, normalization, data augmentation, and face alignment to enhance model robustness. Each model architecture is meticulously designed: EfficientNet leverages efficient scaling for resource optimization, the CNN is custom-built for feature extraction, and the CNN+MLP hybrid combines the strengths of CNN for feature extraction with MLP for classification. The models undergo rigorous training with hyperparameter optimization, early stopping, and learning rate scheduling to prevent overfitting and ensure efficient convergence. Performance evaluation employs metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, along with confusion matrix analysis. Comparative analysis of performance and computational efficiency highlights the most effective model, providing insights into their practical applications in cybersecurity. The implementation utilizes Python with libraries like TensorFlow, Keras, PyTorch, OpenCV, and Scikit-learn, ensuring a robust and flexible development environment.

In this paper, we try to develop a custom website to detect deep fake videos, using the classification algorithms in deep learning. We use a public dataset like the latest databases of the 2nd generation, such as Celeb-DF [48] and DFDC [49], and create our own dataset by using our local celebrity, and then combine both of the datasets for training purposes. After successfully training our model by using deep learning algorithms, we then dynamically change our single dataset and show the result.

3.2 Data Collection Procedure

A. Datasets

The dataset used for this study comprises a total of 106,948 frames, meticulously curated to balance real and deepfake images, ensuring robust model training and evaluation.

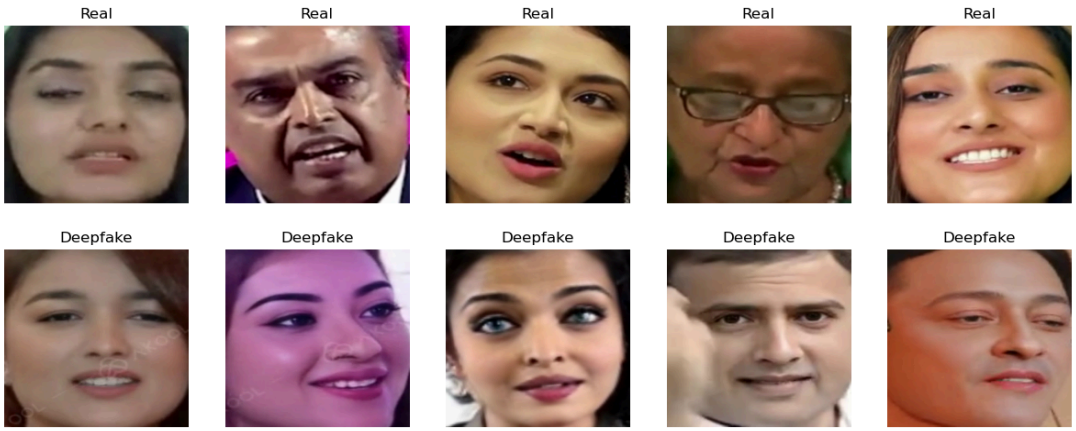


Fig- 3.2.1: combine dataset

The dataset is categorized into real frames and deepfake frames, with the following detailed breakdown:

Real Frames

Total Real Frames: 3,746. The dataset includes 3,746 real frames, with a breakdown of gender representation as follows:

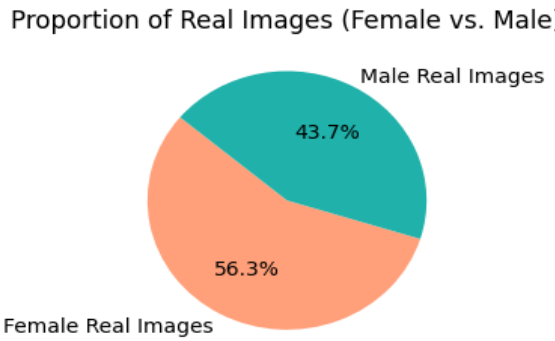


Figure- 3.2.2: comparison between female and male among the real videos

This distribution ensures that both male and female subjects are adequately represented in the real image subset, contributing to a balanced dataset for training and evaluating deepfake detection models.

Female Real Images: 2,109 and Male Real Images: 1,637

TABLE 3.2.1: Real Frame Collection Details.

Gender	No of videos	Avg duration of each video	Total frames
Female	15	28.12 second	2,109
Male	15	21.83 second	1,637

Deepfake Frames

The deepfake frames were generated using two different deepfake creation technologies, ensuring a diverse representation of synthetic content:

TABLE 3.2.2: Deepfake Frames Distribution Table

Technology	Male Deepfake Frames	Female Deepfake Frames	Total Frames
Technology-1	31,611	28,028	59,639
Technology-2	21,592	21,971	43,563
Total	53,203	49,999	103,202

TABLE 3.2.3: Deepfake Frame Collection Details

Gender	Roop-face swapper	Akool AI	The average duration of each video	No of videos	Total frame
Female	15* 15 = 225	15*15 = 225	26.51 sec	450	59,639
Male	15 videos* 15 face = 225	15* 15 = 225	19.36 sec	450	43,563

The deepfake frames were created by manipulating real videos to generate realistic fake content. Both technologies applied sophisticated face-swapping and generative techniques to produce high-quality deepfakes. The dataset was further processed to ensure the faces in the frames were accurately detected and cropped, maintaining uniformity across all images.

Overall Dataset

Total Dataset Size: 106,948 frames

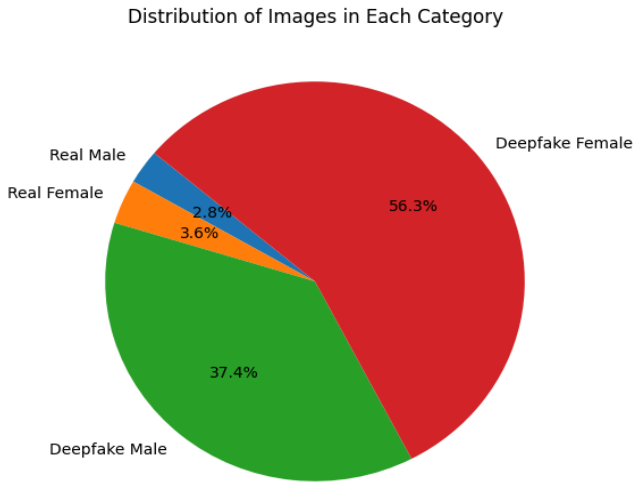


Figure- 3.2.4: Overall Distribution of whole dataset

The pie chart illustrates the distribution of images across different categories in the dataset. It shows a significant dominance of deepfake images compared to real images. Specifically:

- **Deepfake Female** images constitute the largest portion, representing 56.3% of the dataset.
- **Deepfake Male** images account for 37.4%.
- **Real Female** images make up 3.6%.
- **Real Male** images are the smallest group, comprising 2.8%.

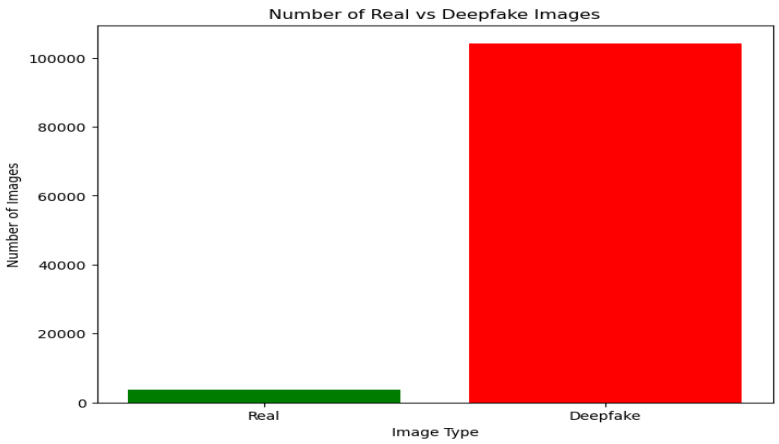


Figure- 3.2.5: Overall Distribution of Real and Deepfake images

The bar chart provides a clear comparison of the count of images in each category:

- **Deepfake Female** images are the most numerous, exceeding 30,000 frames.
- **Deepfake Male** images follow with over 20,000 frames.
- Both **Real Female** and **Real Male** images have significantly lower counts, each with fewer than 4,000 frames

B. Process of Data Collection

Another main contribution of this paper is to create a benchmark dataset. To create the dataset, inside the main data have 2 different folders male and female. In this dataset, We have selected 15 males and 15 females, a total of 30 people based on our Asian subcontinent. On the other hand, We have also selected 15 for male and 15 for female videos to make deepfake videos by using those faces. To make deepfake videos, We have chosen 2 technologies, (1) roop-face swapper technology and (2) Akool AI tool.

Roop-face swapper technology [50]: With this technology, we have created a total of 450 videos (15*15 =225 videos) for both male and female persons. This technology sets up and utilizes the `roop` repository, which facilitates a face-swapping technology designed to replace faces in videos with a specified photo, leveraging deep learning and GPU acceleration. After cloning the repository and installing necessary dependencies including `onnxruntime-gpu`, `opennsfw2`, and `keras`, it configures paths for a source photo and a folder of input videos. Each video undergoes processing where the faces are swapped with the specified photo using the `roop` tool. Parameters such as frame and output video quality are adjusted for optimal performance, utilizing CUDA for GPU acceleration. This technology is valuable for applications in video editing, deepfake detection, and creative media production, offering automated face replacement capabilities to enhance or alter video content seamlessly.

Akool AI tools [51]: Using these AI tools, we have created a total of 450 videos (15*15 =225 videos) for both male and female persons AKOOL's face-swapping app is an industry-leading tool designed to boost your marketing team's efficiency and generate eye-catching advertising campaigns in minutes. This Generative AI platform helps innovative companies transform stock images and easily create personalized visual marketing adverts for commercials, social media content, and e-commerce platforms--at a fraction of the cost of hiring an experienced graphic designer.

C. Frame Extraction and Face Cropping Procedure

We have used a Python script to process videos to extract and save faces from frames using the MTCNN (Multi-task Cascaded Convolutional Networks) face detector [52]. It imports necessary libraries such as `os` for file and directory operations, `cv2` (OpenCV) for video capture and image processing, `MTCNN` for face detection, and `matplotlib.pyplot` for displaying images. The script defines paths for the source folder containing videos and the destination folder where the extracted face images will be saved.

The `get_all_videos` function traverses the source folder and its subfolders to collect all video files with the `.mp4` extension, returning a list of video file paths. The `process_videos` function initializes the MTCNN face detector and iterates over each video file path. For each video, it captures the video and retrieves its frames per second (FPS), setting a frame interval to capture roughly 5 frames per second. For each frame at the specified interval, it detects faces using MTCNN. If faces are detected, it processes only the first detected face (`face_0`), cropping the face from the frame, resizing it to 500x500 pixels, and saving the cropped face image to the destination folder. Optionally, it displays the face image using `matplotlib` for verification.

In the main execution section, the script calls `get_all_videos` to get a list of all video files in the source directory and its subfolders, ensures the destination folder exists (creating it if necessary), and calls `process_videos` to process the videos and save the detected faces. This script effectively handles the task of face detection in videos, processes approximately 5 frames per second by adjusting the frame interval, and saves only the first detected face in each frame, making it suitable for applications requiring efficient face extraction and storage.

Preprocessing

Frame Extraction

Frames are extracted from videos at a consistent interval (e.g., every 1 second) using libraries like OpenCV. This step generates a sequence of images that capture different moments in the video, providing a comprehensive representation of the video content. We have chosen 5FPS (Frame Per Second) for frame extraction for both real and deepfake videos to collect frames.

Normalization

Pixel values of the extracted frames are normalized to a specific range, typically [0, 1] or [-1, 1]. We have chosen the labels '0' for real video image frames and '1' for deepfake video image frames. First of all, select the path of different types of dataset paths then define the real (0) or fake (1). Normalization standardizes the input data, ensuring that the model trains more efficiently and converges faster. It also helps in stabilizing the learning process.

Data Transformation and Augmentation

We have applied it to the images using the transforms.Compose function from the torchvision.transforms module. This function allows for the chaining of several transformations that will be applied sequentially to each image. The first transformation, transforms.Resize (224, 224), resizes the input image to 224x224 pixels. This is necessary because certain models, like ResNet, require input images to be of a fixed size. Following this, transforms.RandomHorizontalFlip() is used to randomly flip the image horizontally with a probability of 0.5. This technique is a form of data augmentation that helps improve the generalization ability of the model by increasing the diversity of the training dataset artificially.

The next transformation transforms.RandomRotation(10), randomly rotates the image by up to 10 degrees in either direction. This further enhances the robustness of the model to variations in the orientation of input images. After these augmentation steps transforms.ToTensor() converts the image from a PIL Image or numpy.ndarray format (with shape H x W x C) into a torch.FloatTensor (with shape C x H x W) and scales the pixel values to the range [0.0, 1.0].

Finally, the transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]) transformation normalizes the image using the specified mean and standard deviation values for each color channel (Red, Green, Blue). These values are typically used for pre-trained models in PyTorch and help to speed up the convergence of the model during training while also maintaining numerical stability.

Face Detection and Alignment

A face detection algorithm (e.g., MTCNN - Multi-Task Cascaded Convolutional Networks) is used to locate and extract facial regions from each frame. By focusing on the face, which is the region of interest for deepfake detection, this step helps in reducing noise and

irrelevant information. Aligning faces ensures consistency in the facial features across different frames, which is crucial for accurate model training.

These steps collectively prepare the video data for deepfake detection by ensuring that the input is standardized, varied, and focused on the critical facial regions where manipulation often occurs. Each step plays a vital role in improving the quality and effectiveness of the deepfake detection model.

Training:

The training loop iterates over a specified number of epochs (num_epochs = 10 in this case). For each epoch:
The model is set to training mode and an accumulator is initialized to track the cumulative loss for the epoch. Within each epoch, the data is iterated over in batches using train_loader, where inputs are the data and labels are the corresponding ground truth labels. Both are transferred to the appropriate computing device (e.g., GPU) using inputs.to(device) and labels.to(device). The optimizer's gradients are reset and a forward pass through the model computes the predicted outputs. The loss between predicted outputs and actual labels is calculated. Backpropagation is performed to compute gradients of the loss with respect to model parameters, followed by an optimizer step to update model parameters based on the gradients. The current batch's loss is added to running_loss, and every 100 batches, the current epoch, batch index, and batch loss are printed for monitoring.

After completing an epoch of training: The model is switched to evaluation mode and validation loss is computed similarly to training, but without backpropagation. The validation loss is accumulated and averaged over all batches in test_loader. At the end of each epoch, both training and validation losses are recorded, representing the average loss per batch for that epoch.

Finally, after completing all epochs, the training loop concludes with a message indicating that training is complete.

Results:

Three parts based on the original individual network topologies, transfer learning, and ensemble methodologies show the experiment outcomes.

3.3 Hardware/ Software Requirement

Hardware Requirement:

To run the detection of deepfakes, the instrumentation involves a combination of hardware and software components designed for optimal performance and accuracy. On the hardware side, a powerful GPU-enabled workstation (such as NVIDIA RTX 3090 or Tesla V100) is essential for efficient training and inference of deep learning models. The software stack includes Python as the primary programming language, with key libraries such as TensorFlow and PyTorch for model development, training, and deployment. TensorFlow/Keras or PyTorch will be used to implement the EfficientNet, CNN, and CNN+MLP hybrid models. For data preprocessing, OpenCV is crucial for tasks like frame extraction, face detection, and alignment, while NumPy and Pandas facilitate data manipulation and analysis. During the training phase, Jupyter Notebooks or integrated development environments (IDEs) like PyCharm or Visual Studio Code are used for scripting and debugging. The training and evaluation pipelines include hyperparameter optimization tools like Keras Tuner or Ray Tune, and performance metrics are calculated using Scikit-learn. Additionally, visualization tools such as Matplotlib and Seaborn will be employed to plot training progress and evaluation results. Finally, the trained models are deployed in a real-time detection system using TensorFlow Serving or Flask APIs, ensuring they can process live video feeds or batch image inputs effectively for deepfake detection in practical applications.

Software Requirement:

The deepfake detection web application will enable users to upload videos, which the system will process to detect deepfake content using advanced neural networks. The application will support user authentication and authorization, allowing registered users to upload video files in various formats. Using OpenCV, the system will extract video frames and employ MTCNN for face detection. Detected faces will be cropped, preprocessed, and fed into EfficientNetB0 and Xception models for feature extraction. These features will be concatenated and passed through a fully connected layer with a sigmoid activation function to make predictions. The application will display the results, indicating whether the video contains real or deepfake content with a confidence level, and will include visualizations of training and validation metrics. The frontend will be developed using React.js or Vue.js, and the backend will utilize Django or Node.js, with PostgreSQL or MongoDB for data storage. The system will ensure secure data handling, scalability, and reliable performance, leveraging cloud storage and GPU-based compute resources for efficient processing.

3.4 Proposed Methodology/ System Design

Deepfake technology, which uses artificial intelligence to create realistic but fake videos, has raised significant concerns regarding the authenticity of video content. This detailed workflow outlines the steps involved in detecting deepfake videos using advanced neural network models, specifically EfficientNetB0 and Xception. The process involves extracting frames from an input video, detecting and preprocessing faces, extracting features using two different models, and finally making a prediction about the authenticity of the faces. This comprehensive methodology ensures reliable detection of deepfake content.

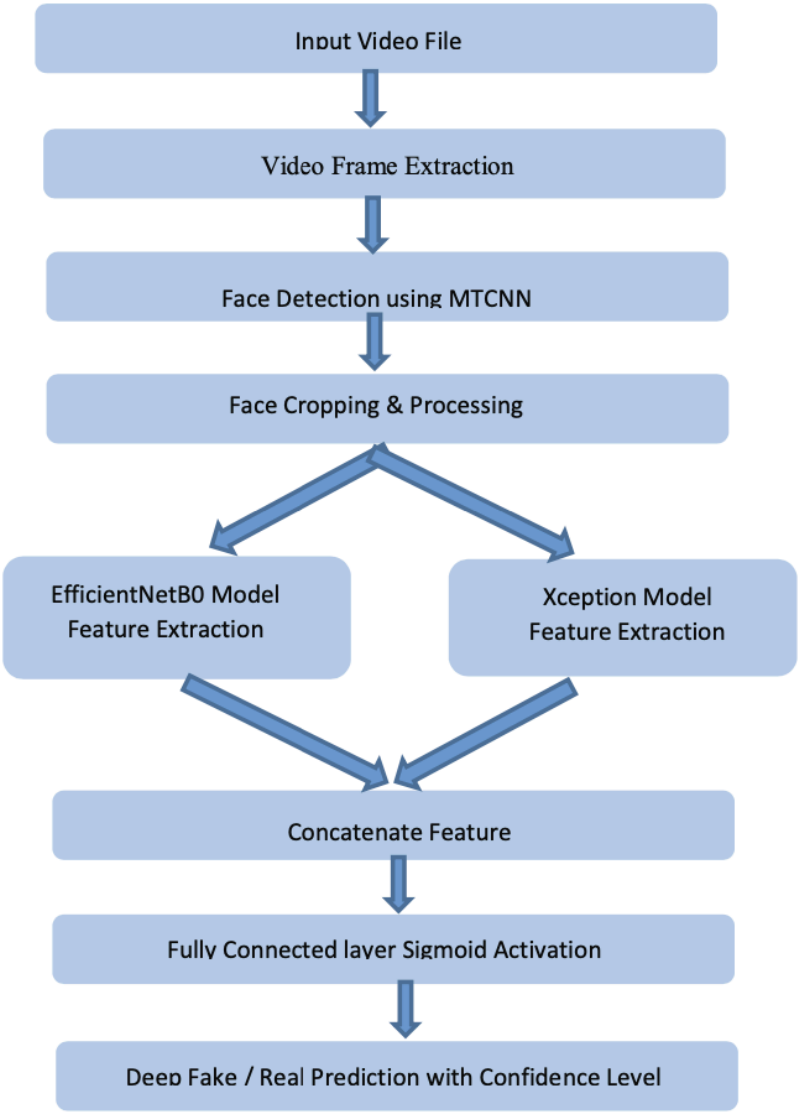


Figure-3.4.1: Deep fake Detection System Workflow

Input Video File:

The deepfake detection process starts with an input video file provided by the user. This video file, which can be in formats such as MP4, MOV, or AVI, contains sequences of frames that will be analyzed for the presence of deepfake content. The initial step involves loading the video file into the system, which serves as the raw data source for the subsequent stages. This step is crucial as it sets the foundation for the entire detection process.

Video Frame Extraction:

After receiving the input video file, the next step is to extract individual frames from the video. This is achieved using video processing tools like OpenCV, which allows for reading and manipulating video data frame by frame. The extraction process involves iterating through the video file and capturing each frame, converting the continuous video stream into discrete images. These frames are then stored for further analysis. Extracting frames is essential because it allows the system to analyze each frame separately for facial features and other relevant information.

Face Detection using MTCNN:

With the frames extracted, the system proceeds to detect faces within each frame using the Multi-task Cascaded Convolutional Networks (MTCNN). MTCNN is a robust and accurate face detection algorithm that identifies the presence and location of faces in images. This step involves scanning each frame to locate and mark regions where faces appear. The face detection process is critical as it ensures that the subsequent steps focus only on the relevant portions of the frames, i.e., the faces. Accurate face detection is essential for reliable deepfake detection.

Face Cropping & Preprocessing:

Once faces are detected in the frames, the next step is to crop these faces and preprocess them for further analysis. Cropping involves isolating the face region from the rest of the frame, creating a smaller image that contains only the face. Preprocessing typically includes resizing the cropped face images to a standard size (e.g., 224x224 pixels) and normalizing pixel values. These preprocessing steps are necessary to ensure that the face images are in a consistent format suitable for input into neural network models. This step enhances the quality of the input data, leading to more accurate feature extraction and classification.

EfficientNetB0 Model Feature Extraction:

The preprocessed face images are then fed into the EfficientNetB0 model for feature extraction. EfficientNetB0 is a powerful convolutional neural network architecture known for its efficiency and performance in image classification tasks. This model extracts

high-level features from the face images, capturing essential patterns and characteristics that can help distinguish between real and deepfake faces. The features extracted by EfficientNetB0 serve as critical inputs for the next stages of analysis.

Xception Model Feature Extraction:

In parallel with the EfficientNetB7 model, the Xception model is also used to extract features from the preprocessed face images. Xception, which stands for "Extreme Inception," is another advanced convolutional neural network known for its accuracy in image classification and recognition tasks. By leveraging two different models (EfficientNetB7 and Xception), the system can extract a diverse set of features, enhancing the robustness and reliability of the final predictions. The features from both models complement each other, providing a richer representation of the face images. This redundancy increases the system's ability to detect subtle differences between real and fake faces.

Concatenate Features:

Once the features are extracted from both EfficientNetB7 and Xception models, they are concatenated to form a single, unified feature vector. Concatenation involves combining the features from the two models into a larger vector that encompasses all the relevant information captured by both models. This step ensures that the subsequent classification process benefits from the comprehensive feature set derived from both models, potentially improving the accuracy of deepfake detection. The combined features provide a holistic view of the face, capturing various aspects that may indicate whether the face is real or fake.

Fully Connected Layer & Sigmoid Activation:

The concatenated feature vector is then passed through a fully connected (dense) layer, which serves as the final classification layer. This layer applies a series of weights to the input features and produces a single output value. The output value is then passed through a sigmoid activation function, which squashes the value to a range between 0 and 1. The sigmoid activation function is particularly suitable for binary classification tasks, as it produces a probability score that indicates the likelihood of the face being real or a deepfake. This step is crucial for transforming the extracted features into a meaningful prediction.

Deepfake / Real Prediction with Confidence Level:

The final step involves making a prediction based on the output of the fully connected layer with sigmoid activation. The system determines whether the face in the video is real or a deepfake by comparing the output probability score against a threshold (typically 0.5). If the

score is above the threshold, the face is classified as a deepfake; otherwise, it is classified as real. Additionally, the system provides a confidence level, representing the certainty of the prediction. This confidence level helps users understand how reliable the prediction is, enabling them to make informed decisions based on the results. This step is the culmination of the entire process, delivering the final verdict on the authenticity of the faces in the video.

By following these detailed steps, the deepfake detection system systematically processes the input video to accurately identify deepfake content, leveraging advanced neural network models and comprehensive feature extraction techniques. Each step builds on the previous one, ensuring a thorough and reliable analysis of the video data. This detailed workflow ensures that the deepfake detection system is robust, accurate, and reliable, providing users with confidence in the results produced. Deepfake or real related to the performance of machine learning classification models are used to assess efficient net base approaches. Evaluations include AC, precision, recall, F1-score, and confusion matrix (CM). In those measures, there are additional variables called TP, TN, FP, and FN (FN).

Accuracy:

Accuracy of the classification model is one metric. The anticipated percentage of the model is its accuracy. The proportion of correctly categorized images to total samples is known as accuracy. Equation of accuracy.

$$Accuracy=(TP+TN)/(TP+TN+FP+FN)$$

Precision:

Precision is the chance of a positive label and how many are positive. Precision shows how many accurately anticipated instances were positive. Precision is helpful when FP matters more than FN. Mathematically, it's this equation.

$$Precision=TP/(TP+FP)$$

Recall:

The number of correctly categorized positive predicted events is measured by recall or sensitivity. Recall is a useful statistic in cases when FN outperforms FP. An further statistic for classification accuracy that accounts for both recall and precision is the F1-score. Since recall and accuracy are harmonic means, a thorough comprehension of these two metrics may be obtained from the F1 score. When Precision and Recall are equal, it performs best. Utilize the following equation to calculate it:

$$Recall=TP/(TP+FN)$$

F-1 Score:

A recall-precision measure of classification accuracy is the F1 score. The F1-score provides a comprehensive view as it is the harmonic mean of Precision and Recall. When precision and recall are equal, they are ideal. formula.

F-1 Score=2*(Precision*Recall)/(Precision+Recall)

Specificity is defined as the proportion of individuals with a negative test result who do not have the illness. A very particular test works well for ruling out most people who don't have the illness. This is why a positive result on an extremely specialized test might rule out the disease for a particular individual. The following is the equation:

Specificity =TN／((TN+FP))

The model can only imitate the training set of data so closely before it starts to lose its capacity for generalization. A specific table format known as the confusion matrix (CM) may be used to evaluate the performance of an algorithm. Critical predictive criteria including memory, specificity, precision, and accuracy are shown using CM. Confusion matrices can be very handy since they provide straightforward evaluations of values such as TP, FP, TN, and FN. Based on the research questions, the following sections address the research questions of this investigation.

EfficientNet

EfficientNet is a family of convolutional neural networks (CNNs) that have demonstrated state-of-the-art performance on a variety of computer vision tasks. Developed by Google AI, EfficientNet models achieve high accuracy while being more computationally efficient than many other architectures. This efficiency is achieved through a novel compound scaling method that uniformly scales the depth, width, and resolution of the network.

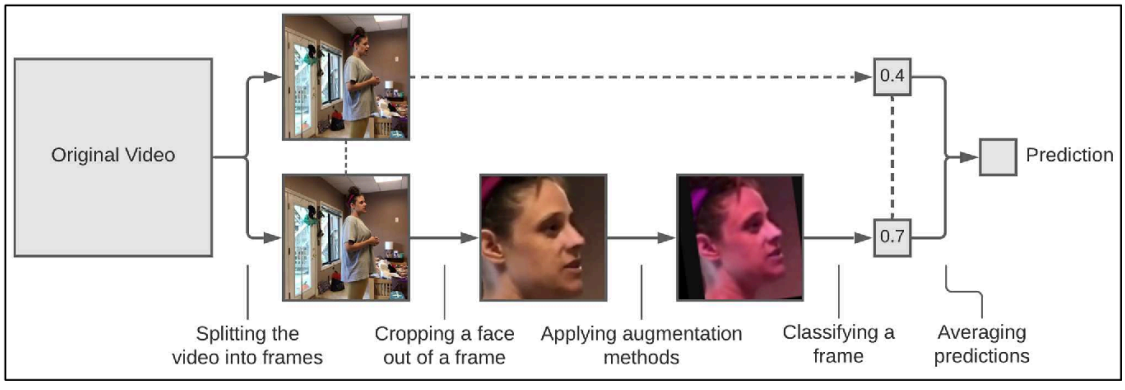


Figure-3.4.1: Working Flow with EfficientNet

3.5 Project Management and Financial Analysis

In this paper, we introduce a system like a web platform to detect whether a video is a deepfake or not. For developing this project, we need to manage scope analysis objectives and constraints on the dataset. To develop this project, we need to take on key responsibilities and risk management on a user-friendly and secure platform.
To develop this proposed project, we need to analyze the financial activities. We try to create our own dataset, which is made by local celebrities and needs to be costly. To develop this project, hosting on a web platform has to be costly. To maintain, to make it more secure, have some cost about that.

TABLE 3.5.1: Estimated Cost for Machine Learning based Project

SN	Components	Estimated Cost (BDT)
01.	Hardware/Infrastructure	40,000-45,000
02.	Visiting Stakeholders	1,500-3,000
03.	Software and Tools	7,000-10,000
04.	Data Collection and Processing	4,000-5,000
05.	Documentation and Report Writing	500-1,000
06.	Miscellaneous (e.g., licenses, tools)	500-1,000
07	Web Platform and Hosting	20,000 - 40,000
07.	Contingency (10% of total)	1,500-3,000
Total Estimated Cost		75,000 - 90,000

3.6 Summary

The suggested technique offers a methodical and exacting approach to creating a website interface and deepfake detection system. EfficientNet's efficient architecture, strong feature extraction capabilities, and adaptability make it a powerful tool for deepfake detection. By fine-tuning pretrained models and employing advanced training strategies, EfficientNet can effectively identify deepfake content, contributing to the fight against misinformation and digital manipulation. Using a variety of datasets, cutting-edge machine learning algorithms, and iterative development procedures, the methodology seeks to provide a dependable and approachable solution for preventing the dissemination of false media material in the digital age.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

In this research, we attempt to use deep learning classification algorithms to create a unique website for the purpose of detecting deep fake films. We make use of a public dataset similar to the most recent databases from the second generation, namely DFDC and Celeb-DF and create our own dataset by using our local celebrity, and then combine both of the datasets for training purposes. After successfully training our model by using deep learning algorithms, we then dynamically change our single dataset and show the result.To do thorough investigations, a thoroughly thought-out setup of hardware, software, and data resources is required. Powerful central processing units (CPUs) and graphics processing units (GPUs) are two examples of the high-performance hardware that makes up the computational core. PyTorch and TensorFlow are two deep learning frameworks that are used to create transfer learning models in a Python programming environment. The experimental dataset is derived from a variety of medical pictures and is carefully preprocessed, standardized, and normalized to provide a reliable and representative input for training and assessing the model. Part of the model setup that is crucial is the use of pre-trained deep EfficientNet topologies together with pneumonia detection fine-tuning.

To maintain ethical standards, ethical issues like informed consent and patient data protection are included into the setup. To promote openness, repeatability, and future cooperation in the field of advanced pneumonia diagnosis, thorough documentation including code details, model architecture, hyperparameters, and outcomes is kept up to date throughout the experimentation phase.

4.2 Train Model/ Prototype Design

Model Architecture and Process

Video Input: The process begins with input videos that need to be analyzed for real or fake content.

Face Detection (MTCNN): Multi-task Cascaded Convolutional Networks (MTCNN) is used for detecting faces in the video frames.

This step extracts the face regions from the video frames, producing a dataset of cropped face images.

Cropped Face Dataset: The extracted faces are organized into a dataset containing both real and fake face images.

Feature Extraction

Xception Network: One branch of the model uses the Xception network to extract features from the face images.

EfficientNet-B7: Another branch uses EfficientNet-B7 to extract a different set of features from the same images.

Feature Stacking: The features extracted by the Xception network and EfficientNet-B7 are combined (stacked) to form a unified feature set.

Feature Selection: In order to enhance the classification performance, the most relevant characteristics from the combined set are chosen in this stage.DC,

Classification (MLP): A Multilayer Perceptron (MLP) is used to classify the face images as either real or fake based on the selected features.

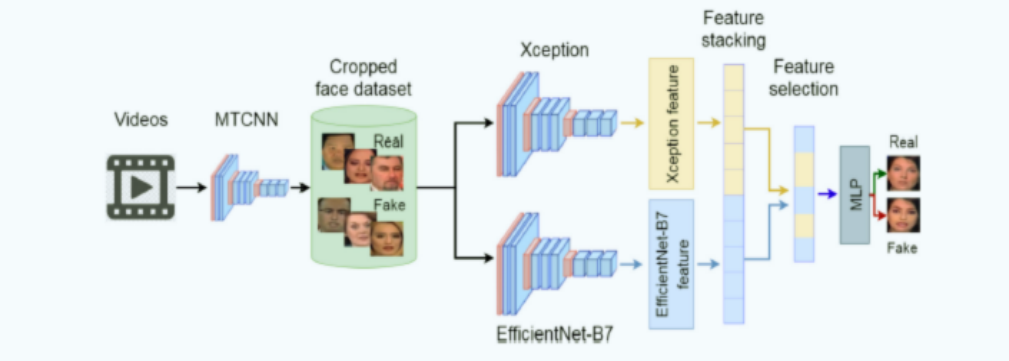


Figure-4.2.1: System Architecture

Prototype of Web Page

Our cutting-edge web application offers three main features to explore and create with advanced facial manipulation technologies:

Face Swap in Photo: Effortlessly swap faces in your photos with our state-of-the-art face swap technology. Upload a photo, select the faces you want to swap, and watch the transformation happen instantly. This feature uses advanced algorithms to seamlessly blend facial features, ensuring natural and realistic results.

Deepfake Video Creation: Create deepfake videos by applying face swap technology to video content. Upload your video and select the face you want to replace. Our system processes the video to replace the selected face with a new one, maintaining natural expressions and movements. This feature allows for creative video manipulations, producing high-quality deepfake videos that appear astonishingly real.

Deepfake or Real Video Detection: Ensure the authenticity of video content with our deepfake detection feature. Upload a video, and our system will analyze it to determine if it contains deepfake content. Leveraging advanced neural network models, we provide a detailed report with a confidence level indicating the likelihood of the video being real or a deepfake. This tool is essential for verifying video authenticity and combating misinformation.

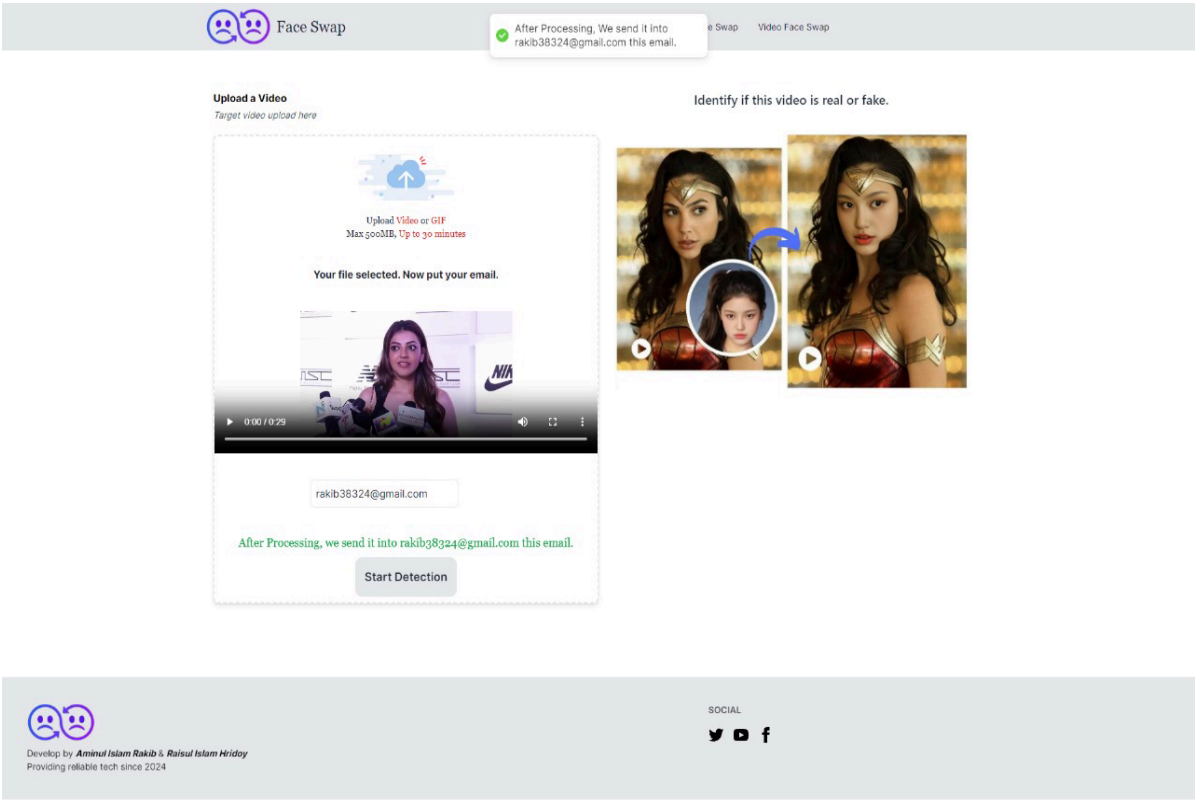


Figure-4.2.2: Web Page Prototype

These features harness the power of the latest AI and machine learning technologies, providing users with intuitive, powerful tools for facial manipulation and deepfake detection.

4.3 System Testing and Model Evalution

Deepfake technology poses significant cybersecurity challenges. This project implements a deepfake detection system using MTCNN for face detection, Xception and EfficientNet-B7 for feature extraction, and an MLP for classification. Here we outline the system testing and model evaluation process.

System Testing

1. Unit Testing:

Face Detection (MTCNN): Verify accurate face detection and cropping in various conditions.

Feature Extraction: Ensure correct feature extraction by Xception and EfficientNet-B7.

Feature Stacking and Selection: Confirm proper feature combination and selection.

Classification (MLP): Test MLP accuracy with known real and fake images.

2. Integration Testing: Validate the end-to-end pipeline from video input to final classification. Test with real-world scenarios and various video formats.

3. Performance Testing: Measure processing time and system scalability.

Evaluate performance on different hardware configurations.

Model Evaluation

1. Data Preparation: Split dataset into training, validation, and test sets with balanced real and fake faces.

2. Training and Validation: Train MLP on training set, tune hyperparameters with validation set, and monitor loss/accuracy.

3. Performance Metrics: Accuracy, Precision, Recall, F1 Score, ROC-AUC: Measure overall and class-specific performance.

4. Testing: Evaluate final model on test set and generate confusion matrix.

Compare with baseline and state-of-the-art models.

5. Robustness Testing: Test model's performance under various perturbations and adversarial examples.

6. Interpretability: Use SHAP or LIME to understand feature importance and model decision-making.

4.4 Summary

Through comprehensive unit, integration, and performance testing, alongside detailed evaluation metrics and robustness checks, this approach ensures the deepfake detection system is reliable and effective in combating cyber threats.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

The performance of various models for deepfake detection was evaluated using test accuracy and test loss metrics. Here is a summary of the results:

1. EfficientNet:

- Test Accuracy: 0.9813
- Analysis: The hybrid model combining EfficientNet, Xception and Multilayer Perceptron (MLP) outperforms the CNN with a significantly higher test accuracy of 98.13%. This indicates that EfficientNet is more effective in feature extraction and classification, providing a more accurate deepfake detection capability.

2. CNN (Convolutional Neural Network):

- Test Accuracy: 0.9317
- Analysis: The CNN model demonstrates strong performance with a high accuracy of 93.17% and a comparatively minimal test loss, suggesting its efficacy in differentiating between authentic and counterfeit faces. But there's still opportunity for growth in terms of generality and accuracy.

3. MLP + CNN (Hybrid Model):

- Test Accuracy: 0.9104
- Analysis: The hybrid model combining MLP and CNN achieves an accuracy of 91.04%, which is slightly lower than the standalone CNN model. This suggests that the integration of MLP with CNN in this configuration may not provide additional benefits and could introduce complexity that impacts overall performance.

5.2 Experimental/ Simulation Result

In this work, we used a structured experimental technique to assess the effectiveness of several models for deepfake detection. Convolutional neural networks (CNNs), EfficientNet, and hybrid models that combine CNN and MLP are among the models that were examined. The following summarizes the main findings:

TABLE 5.2.1: Accuracy for EfficientNet+ Xception (Hybrid Model), CNN and Hybrid Model (CNN + MLP)

Architecture	Training Accuracy	Model Accuracy
EfficientNet, Xception and MLP	**98.13%**	**98%**
CNN	**93.17%**	**93%**
CNN + MLP	**91.04%**	**91%**

The performance of different models for deepfake detection, including EfficientNet, CNN, and a hybrid CNN+MLP model, was evaluated using precision, recall, F1-score, and support metrics.

EfficientNet, Xception, MLP (Hybrid Model):

EfficientNet demonstrated exceptional performance in detecting both real and deepfake images. For real images, it achieved a precision of 0.99, recall of 0.98, and an F1-score of 0.98, with a support of 526 images. For deepfake images, the precision was 0.97, recall 0.98, and F1-score 0.98, with a support of 382 images. Overall, EfficientNet attained an accuracy of 0.98 across 908 images. The macro and weighted averages for precision, recall, and F1-score were consistently 0.98, indicating balanced performance across both classes.

CNN:

The original CNN network also performed well but was slightly less accurate than EfficientNet. For normal images, the CNN achieved a precision of 0.92, recall of 0.97, and an F1-score of 0.94, with a support of 526 images. For deepfake images, the precision was 0.95, recall 0.88, and F1-score 0.92, with a support of 382 images. The overall accuracy for the CNN was 0.93. The macro and weighted averages for precision, recall, and F1-score were 0.93, indicating good but slightly less consistent performance compared to EfficientNet.

CNN+MLP (Hybrid Model):

The hybrid approach that blends MLP and CNN showed the lowest performance among the three models. For normal images, it achieved a precision of 0.90, recall of 0.96, and an F1-score of 0.93, with a support of 397 images. For deepfake images, the precision was 0.93, recall 0.85, and F1-score 0.89, with a support of 284 images. The overall accuracy of the hybrid model was 0.91 across 681 images. The macro averages for precision, recall, and F1-score were 0.92, 0.90, and 0.91, respectively, while the weighted averages were consistently 0.91, showing a slight drop in performance and consistency compared to the standalone CNN and EfficientNet.

TABLE 5.2.2: Precision, Recall, F1-Score, and Support (n) for EfficientNet, Xception and MLP (depending on the number of pictures, n=numbers)

EfficientNet, Xception and MLP				
	Precision	**Recall**	**F1-Score**	**Support**
Real	**0.99**	**0.98**	**0.98**	**526**
Deepfake	**0.97**	**0.98**	**0.98**	**382**
Accuracy			**0.98**	**908**
Macro Avg	**0.98**	**0.98**	**0.98**	**908**
Weighted Avg	**0.98**	**0.98**	**0.98**	**908**

TABLE 5.2.3: Precision, Recall, F1-Score, and Support (n) for CNN (depending on the number of pictures, n=numbers)

CNN

	Precision	Recall	F1-Score	Support
Real	**0.92**	**0.97**	**0.94**	**526**
Deepfake	**0.95**	**0.88**	**0.92**	**382**
Accuracy			**0.93**	**908**
Macro Avg	**0.93**	**0.93**	**0.93**	**908**
Weighted Avg	**0.93**	**0.92**	**0.93**	**908**

TABLE 5.2.4: Precision, Recall, F1-Score, and Support (n) for CNN and MLP (depending on the number of pictures, n=numbers)

CNN+MLP				
	Precision	**Recall**	**F1-Score**	**Support**
Real	**0.90**	**0.96**	**0.93**	**397**
Deepfake	**0.93**	**0.85**	**0.89**	**284**
Accuracy			**0.91**	**681**
Macro Avg	**0.92**	**0.90**	**0.91**	**681**
Weighted Avg	**0.91**	**0.91**	**0.91**	**681**

EfficientNet emerged as the best-performing model with the highest precision, recall, and F1-scores across both real and deepfake images, achieving an overall accuracy of 0.98. The original CNN also performed well but with slightly lower metrics and an overall accuracy of 0.93. The CNN+MLP hybrid model, while effective, showed the lowest performance with an accuracy of 0.91. These results highlight EfficientNet's superior capability in deepfake detection, making it the most reliable choice among the models tested.

5.3 Performance/ Comparative Analysis

The confusion matrix provides a more detailed understanding of model performance beyond simple accuracy. It helps in identifying specific types of errors (FP and FN), guiding improvements in model training and adjustment of decision thresholds to balance between detection accuracy and false alarms.

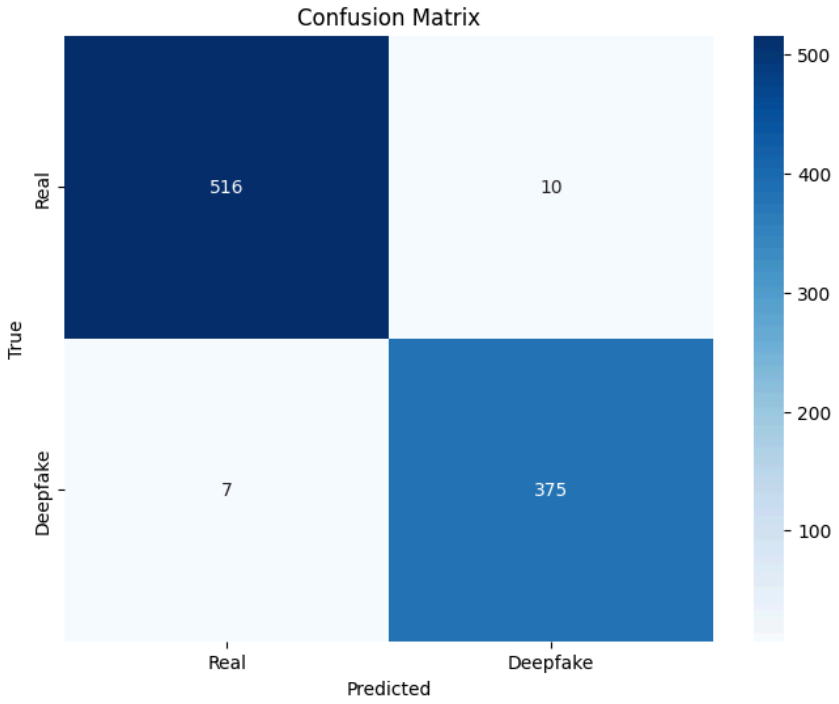


Figure 5.3.1: CM after EfficientNet, Xception and MLP

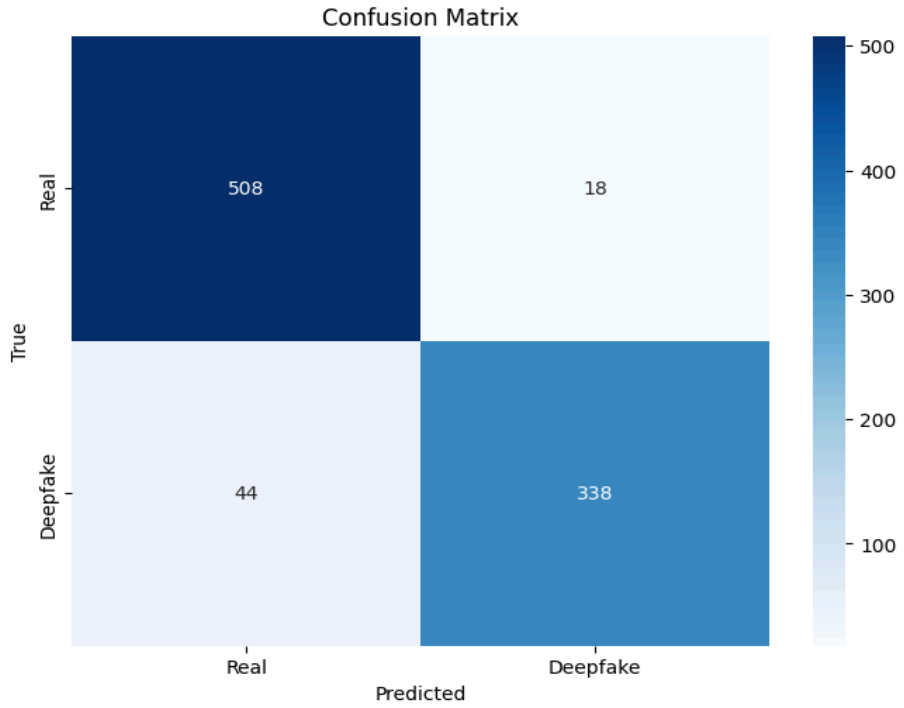


Figure 5.3.2: CM after CNN

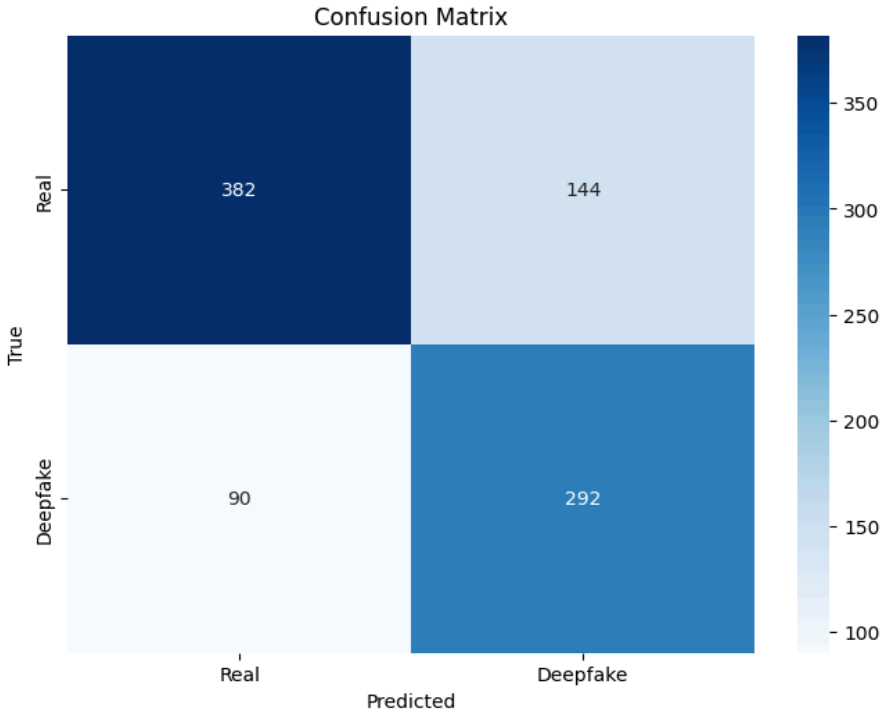


Figure 5.3.3: CM after CNN + MLP

In summary, for deepfake detection, the confusion matrix serves as a fundamental tool to assess the effectiveness of models in distinguishing between authentic and manipulated videos, guiding improvements in algorithm design and training strategies**.**

Training and Validation Accuracy and loss of Transfer Learning of models:

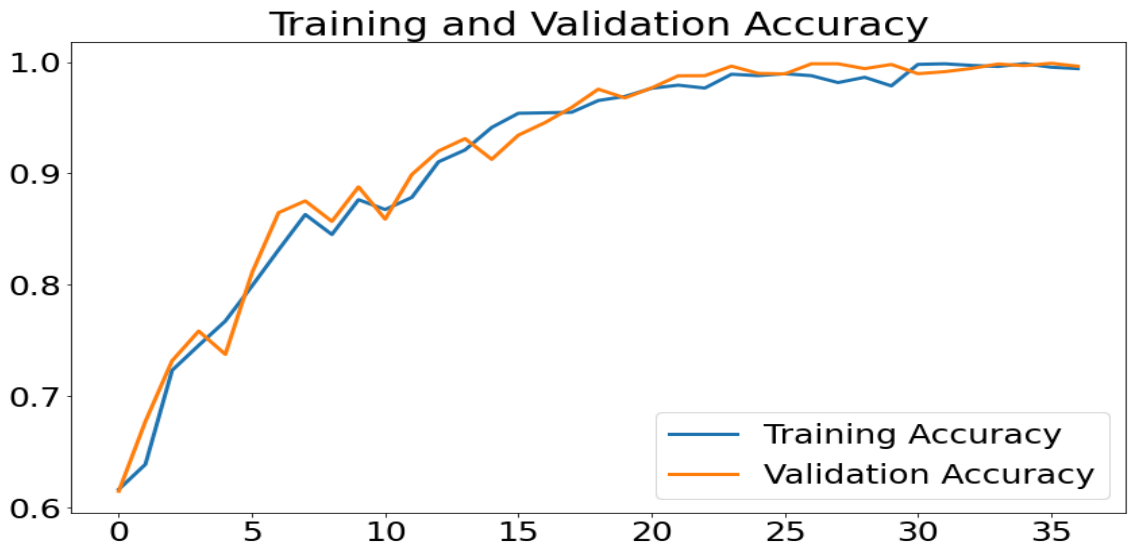


Figure 5.3.4: Training and validation accuracy and loss over the epochs (EfficientNet, Xception and MLP).

Figure 5.3.5: Training and validation accuracy and loss over the epochs (CNN).

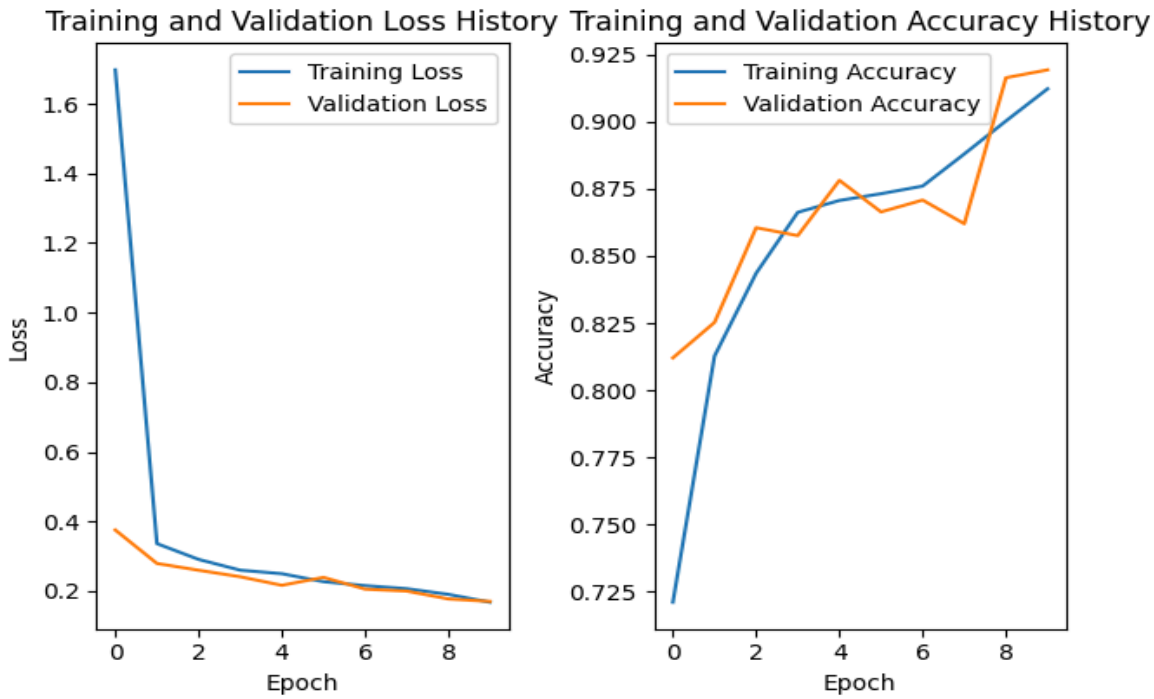


Figure 5.3.6: Training and validation accuracy and loss over the epochs (MlP + CNN).

5.4 Summary

EfficientNet emerges as the most effective model for deepfake detection, achieving an impressive test accuracy of 98.13%. This model's performance surpasses that of the traditional CNN model, which, while robust, does not reach the same high level of accuracy. The hybrid MLP + CNN model, though conceptually innovative and promising, currently underperforms compared to both the CNN and EfficientNet models. To unlock the full potential of the hybrid model, further optimization and extensive experimentation are required.

After the training phase, these models will be integrated into a comprehensive web platform. This platform will not only facilitate deepfake creation but also provide advanced deepfake detection and face-swapping functionalities. Users will be able to generate realistic deepfakes, verify the authenticity of multimedia content, and swap faces in images and videos, leveraging the cutting-edge capabilities of EfficientNet and other deep learning models. The platform aims to offer a user-friendly interface while maintaining high accuracy and reliability in its operations.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The rise of deepfake technology poses significant threats to individuals, institutions, and society. Using deep learning algorithms for deepfake detection is crucial to mitigate these threats and protect information integrity.

Personal Reputations and Privacy:
Deepfakes can damage personal reputations and invade privacy by creating explicit or damaging content. Detection systems help protect individuals by identifying and removing harmful media promptly.

Public Trust and Media Integrity:
Deepfakes undermine public trust in the media by spreading misinformation. Reliable detection systems maintain media credibility and ensure the public receives accurate information.

Institutional and Organizational Security:
Deepfakes can be used to impersonate executives and manipulate communications, causing financial and reputational harm. Detection systems safeguard institutions by verifying the authenticity of digital content.

Political Stability and National Security:
Deepfakes can influence elections and incite violence, threatening political stability and national security. Robust detection systems help governments counteract malicious deepfake campaigns.

Legal and Ethical Implications:
Deepfakes raise legal and ethical issues around consent, privacy, and intellectual property. Detection systems assist legal authorities in enforcing digital content laws and holding creators accountable. Deepfakes significantly impact personal privacy, public trust, institutional security, political stability, and legal frameworks. Intelligent deepfake detection systems are essential for combating these cyber threats and ensuring responsible AI use in the digital age.

6.2 Impact on Society & Environment

Societal Impact

Countering Misinformation and Disinformation:

The dissemination of false narratives that have the potential to sway public opinion, upend political situations, and erode confidence in conventional media is made possible by deepfakes, which represent a serious threat to the accuracy of information. By employing deep learning algorithms for deepfake detection, society can better identify and mitigate these fabricated media, ensuring that information remains trustworthy and credible.

Protecting Reputations:

Deepfakes can be used maliciously to damage individual reputations, harm professional careers, and defame public figures. A robust detection system can shield individuals from these attacks, maintaining personal and professional integrity and preventing the emotional and psychological distress caused by such violations.

Safeguarding Institutions:

Institutions, including governments, corporations, and educational bodies, are vulnerable to deepfake attacks that can tarnish their reputations and undermine their authority. Deep learning-based detection systems can help preserve institutional credibility, prevent fraud, and maintain public trust in these entities.

Enhancing Public Awareness:

Increasing public knowledge of deepfake presence and risks may also be greatly aided by the creation and use of deepfake detection tools. Educating the public on how to identify and report suspected deepfakes can foster a more informed and vigilant society, reducing the overall impact of these deceptive media.

Environmental Impact

Resource Utilization:

Developing and running deep learning algorithms for deepfake detection requires significant computational resources, which can lead to high energy consumption. This energy use contributes to the carbon footprint of data centers and computing facilities. However, optimizing algorithms to achieve high accuracy with lower power consumption can mitigate this environmental impact.

Promoting Sustainable Practices:

By focusing on efficient algorithm design and leveraging green computing technologies, the development of deepfake detection systems can promote sustainable practices within the tech industry. Researchers and developers can prioritize energy-efficient models and explore renewable energy sources for powering data centers.

Long-Term Environmental Benefits:

While the immediate environmental impact of deepfake detection involves energy consumption, the long-term benefits include reducing the spread of misinformation that can lead to social and political instability. Stable societies are better equipped to address environmental challenges collectively, making the indirect impact of combating deepfakes significant for environmental sustainability. The deployment of deep learning algorithms for deepfake detection has profound implications for both society and the environment. By combating the spread of misinformation, protecting reputations, and safeguarding institutions, these technologies contribute to a more trustworthy and stable societal framework. Additionally, by emphasizing energy-efficient practices and sustainable development, the environmental impact of these technologies can be managed effectively. Ultimately, deepfake detection systems not only address a pressing cyber threat but also promote responsible and ethical use of technology in our interconnected world.

6.3 Ethical Aspects

This report's objective is to investigate the use of deep learning algorithms for deepfake identification and the moral ramifications of doing so in the fight against the growing danger that modified or fake videos represent to the internet.

Deepfakes are artificial intelligence (AI)-generated pictures or movies that change pre-existing material or superimpose the face of one person onto another, so manipulating reality. As these technologies advance, there is a growing concern about the potential misuse deepfakes for malicious purposes, such as spreading misinformation, fake news, or even blackmail. Deep learning algorithms provide a viable method for identifying deepfakes. Large datasets of actual and altered movies may be used to train these algorithms to find patterns and traits that tell genuine videos apart from deepfakes.

While the use of deep learning algorithms for deepfake detection shows promise, it also raises several ethical considerations. One major concern is the potential violation of privacy

rights, as these algorithms may require access to vast amounts of personal data to effectively learn and detect deepfakes.

In addition, there's a chance that algorithms may unintentionally discriminate against certain people or groups on the basis of gender, color, or other characteristics. This is known as algorithmic bias. This might cause deepfake detection to produce false positives or false negatives, reinforcing societal prejudices and disparities already in place. In conclusion, the use of deep learning algorithms for deepfake detection presents both opportunities and challenges in combating the rising cyber threat posed by fake videos. It is essential to address the ethical implications of using these algorithms and ensure that they are deployed responsibly and with respect for privacy and fairness.

6.4 Sustainability Plan

In order to counter the danger posed by synthetic media, the deepfake detection project's sustainability strategy prioritizes long-term viability, efficacy, and flexibility. Key strategies include continuous improvement of detection algorithms, scalable cloud-based infrastructure, and robust data security measures. Financial sustainability will be achieved through grants, a freemium model, and partnerships with tech companies. Operational sustainability involves user education, regular performance and security audits, and adaptability to emerging threats. Additionally, the project emphasizes environmental sustainability through energy-efficient computing and partnerships with green data centers. This comprehensive plan aims to maintain the platform as a reliable and trusted tool against deepfake cyber threats.

6.5 Summary

Deepfake detection technology has a profound impact on society by enhancing the integrity of information, protecting individuals and organizations, and fostering trust in digital environments. While there are environmental considerations due to the computational demands of deepfake detection, ongoing efforts to optimize and improve efficiency are crucial. Ultimately, the sustainable development and deployment of deepfake detection rely on ethical practices, collaboration, and regulatory frameworks to ensure long-term benefits for society.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

The study "Intelligence for deepfake detection: an experimental measure to combat rising cyber threat" offers a thorough examination of cutting-edge AI methods in the field of cybersecurity. Using a dataset of 3746 authentic photographs and 104201 deepfake photos, the research delves into how well deep EfficientNetB-7 distinguishes and separates two categories of images or videos: actual and deepfake. In order to construct a deepfake video utilizing face swapping technology and distinguish between actual and deepfake films, this work explains three deep learning based models, generates a dataset using local celebrities, and creates a web platform. The emergence of deepfake technology presents serious problems for media integrity, cybersecurity, and public confidence. This research examined the effectiveness of EfficientNet, Xception, and Multi-Layer Perceptron (MLP) models in detecting deepfakes. EfficientNet and Xception showed high accuracy, with EfficientNet balancing performance and computational efficiency and Xception excelling in image classification tasks. The MLP model, while simpler, highlighted the importance of model architecture in deepfake detection.

7.2 Further Suggested Works

Future research should aim to expand the dataset to include a broader range of subjects and scenarios. By incorporating diverse faces, lighting conditions, and backgrounds, the models can be trained to detect deepfakes more accurately across a variety of contexts. Additionally, including more types of deepfake manipulations, such as audio deepfakes and more complex video alterations, will further strengthen the models' robustness.

Future research should investigate the integration of additional cutting-edge deep learning approaches, such as Transformer-based models and Generative Adversarial Networks (GANs), even though this study concentrated on EfficientNet, Xception, and MLP models. These models have shown promise in various image and video processing tasks and could enhance deepfake detection capabilities. Combining multiple detection techniques in a hybrid model could also improve overall accuracy and reliability.

Implementing real-time deepfake detection is crucial for practical applications, especially in social media and live streaming platforms. Future research should focus on optimizing the models for real-time processing and ensuring they can handle high volumes of data without significant performance degradation. Techniques such as model compression, pruning, and distributed computing should be investigated to enhance scalability and efficiency.

The proposed web platform should undergo continuous development to improve its user interface and user experience. Incorporating feedback from users and cybersecurity experts will help in making the platform more intuitive and effective. Additionally, providing comprehensive tutorials and educational resources within the platform will enhance user understanding and engagement.

The ethical implications of deepfake technology need to be thoroughly examined. Future work should focus on developing guidelines and policies to govern the use and detection of deepfakes. Collaborating with legal experts, policymakers, and ethicists will be essential in creating a framework that balances technological advancements with privacy and security concerns.

Exploring the application of deepfake detection techniques in other domains, such as fraud detection in financial institutions, authentication in secure communications, and integrity verification in digital forensics, can open new avenues for research. Adapting and testing the models in these varied contexts will demonstrate their versatility and practical utility.

Deepfake technology is rapidly evolving, and so should the detection models. Future research should focus on developing adaptive learning mechanisms that allow models to continuously update themselves based on new data and emerging deepfake techniques. Implementing automated retraining pipelines can ensure that detection models remain effective against the latest deepfake methods.

By addressing these areas in future work, researchers can further enhance the effectiveness and applicability of deepfake detection technologies, ultimately contributing to a more secure and trustworthy digital landscape.

7.3 Limitations/ Conflict of Interests

This research, while comprehensive, has several limitations. The dataset, although substantial, may lack sufficient diversity in terms of faces, lighting conditions, and backgrounds, potentially limiting the models' effectiveness in real-world scenarios. The focus on EfficientNet, Xception, and MLP models excludes other advanced techniques that could enhance detection accuracy. Additionally, the current implementation does not support real-time detection, which is crucial for practical applications. The proposed web platform, though promising, requires further development to improve user interface and experience. There are no reported conflicts of interest associated with this research. All findings and recommendations are based solely on the experimental results and the objective to advance deepfake detection technologies in the field of cybersecurity. The researchers have no financial or personal interests that could have influenced the outcomes of this study.

REFERENCES

[1] A. Firc, K. Malinka, and P. Hanáček, "Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors," *Heliyon*, vol. 9, no. 4, p. e15090, Apr. 2023, doi: 10.1016/j.heliyon.2023.e15090.

[2] Wikipedia contributors, "Deepfake," *Wikipedia*, Jun. 27, 2024. https://en.wikipedia.org/wiki/Deepfake

[3] M. Westerlund, "The Emergence of Deepfake Technology: A review," *Technology Innovation Management Review*, vol. 9, no. 11, pp. 39–52, Jan. 2019, doi: 10.22215/timreview/1282.

[4] T. T. Nguyen *et al.*, "Deep learning for deepfakes creation and detection: A survey," *Computer Vision and Image Understanding*, vol. 223, p. 103525, Oct. 2022, doi: 10.1016/j.cviu.2022.103525.

[5] Depositphotos, "8 Ideas on how Deepfakes will change photography and marketing in 2020," *Medium*, Dec. 15, 2021. [Online]. Available: https://depositphotos.medium.com/welcome-to-the-world-of-deepfakes-b3484f320acc

[6] "Front Matter from Deepfakes and Synthetic Media in the Financial System:: Assessing Threat Scenarios on JSTOR," *www.jstor.org*, [Online]. Available: https://www.jstor.org/stable/resrep25783.1

[7] L. Payne, "Deepfake | History & Facts," *Encyclopedia Britannica*, Jun. 09, 2024. https://www.britannica.com/technology/deepfake

[8] B. Allyn, "Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn," *NPR*, Mar. 17, 2022. [Online]. Available: https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

[9] A. Chiu, "Facebook wouldn't delete an altered video of Nancy Pelosi. What about one of Mark Zuckerberg?," *Washington Post*, Jun. 12, 2019. [Online]. Available: https://www.washingtonpost.com/nation/2019/06/12/mark-zuckerberg-deepfake-facebook-instagram-nancy-pelosi/

[10] "Obama deep fake – Ars Electronica Center." https://ars.electronica.art/center/en/obama-deep-fake

[11] News team, "Deepfakes and AI's new threat to security - Cyber Defense Magazine," *Cyber Defense Magazine*, Mar. 20, 2024. https://www.cyberdefensemagazine.com/deepfakes-and-ais-new-threat-to-security/

[12] "The largest linked research database | Dimensions Data," *Dimensions*. https://www.dimensions.ai/dimensions-data/

[13] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *arXiv (Cornell University)*, Jan. 2019, doi: 10.48550/arxiv.1905.11946.

[14] "EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling."
https://research.google/blog/efficientnet-improving-accuracy-and-efficiency-through-automl-and-model-scaling/
[15] M. Malik *et al.*, "Waste Classification for Sustainable Development Using Image Recognition with Deep Learning Neural Network Models," *Sustainability*, vol. 14, no. 12, p. 7222, Jun. 2022, doi: 10.3390/su14127222.
[16] T. Wu, H. Zhu, H. Fan, and H. Zhou, "An improved target detection algorithm based on EfficientNet," *Journal of Physics. Conference Series*, vol. 1983, no. 1, p. 012017, Jul. 2021, doi: 10.1088/1742-6596/1983/1/012017.
[17] A. A. Pokroy and A. D. Egorov, "EfficientNets for DeepFake Detection: Comparison of Pretrained Models," *https://ieeexplore.ieee.org/xpl/conhome/9396056/proceeding*, Jan. 2021, doi: 10.1109/elconrus51938.2021.9396092.
[18] P. Ratan, "What is the Convolutional Neural Network Architecture?," *Analytics Vidhya*, Nov. 06, 2023. https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/
[19] Z. Keita, "An Introduction to Convolutional Neural Networks (CNNs)," Nov. 14, 2023.
https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns
[20] Y. LeCun *et al.*, "Backpropagation applied to handwritten Zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, Dec. 1989, doi: 10.1162/neco.1989.1.4.541.
[21] A. Karandikar, "Deepfake video detection using convolutional neural network," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 2, pp. 1311–1315, Apr. 2020, doi: 10.30534/ijatcse/2020/62922020.
[22] P. Ratan, "What is the Convolutional Neural Network Architecture?," *Analytics Vidhya*, Nov. 06, 2023. https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/
[23] A. V and P. T. Joy, "Deepfake Detection Using XceptionNet," *https://ieeexplore.ieee.org/xpl/conhome/10363431/proceeding*, Nov. 2023, doi: 10.1109/rasse60029.2023.10363477.
[24] K. Srinivasan *et al.*, "Performance comparison of deep CNN models for detecting driver's distraction," *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 68, no. 3, pp. 4109–4124, Jan. 2021, doi: 10.32604/cmc.2021.016736.
[25] "Google Trends," *Google Trends*, Jan. 01, 2016. https://trends.google.com/trends/explore?date=2016-01-01%202023-01-12&q=deepfake
[26] K. Raja *et al.*, "Morphing Attack Detection-Database, evaluation Platform, and

benchmarking," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4336–4351, Jan. 2021, doi: 10.1109/tifs.2020.3035252.

[27] V. Blanz, K. Scherbaum, T. Vetter, and H. Seidel, "Exchanging faces in images," *Computer Graphics Forum*, vol. 23, no. 3, pp. 669–676, Aug. 2004, doi: 10.1111/j.1467-8659.2004.00799.x.

[28] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping," *ACM Transactions on Graphics, Vol. 27, No. 3, Article 39*, Aug. 2008, doi: 10.1145/1399504.1360638.

[29] Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni, "On Face Segmentation, Face Swapping, and Face Perception," *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition*, May 2018, doi: 10.1109/fg.2018.00024.

[30] N. Damer, C. A. F. Lopez, M. Fang, N. Spiller, M. V. Pham, and F. Boutros, "Privacy-friendly synthetic data for the development of face morphing attack detectors," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2022, doi: 10.1109/cvprw56347.2022.00167.

[31] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel, "Are GAN-based morphs threatening face recognition?," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, doi: 10.1109/icassp43922.2022.9746477.

[32] I. Korshunova, W. Shi, J. Dambre, and L. Theis, "Fast Face-Swap Using Convolutional Neural Networks," *IEEE Xplore*, Oct. 2017, doi: 10.1109/iccv.2017.397.

[33] S. Waseem, S. A. R. S. A. Bakar, B. A. Ahmed, Z. Omar, T. A. E. Eisa, and M. E. E. Dalam, "DeepFake on Face and Expression Swap: a review," *IEEE Access*, vol. 11, pp. 117865–117906, Jan. 2023, doi: 10.1109/access.2023.3324403.

[34] G. Gao, H. Huang, C. Fu, Z. Li, and R. He, "Information Bottleneck Disentanglement for Identity Swapping," *IEEE Xplore*, Jun. 2021, doi: 10.1109/cvpr46437.2021.00341.

[35] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "First order motion model for image animation," *Neural Information Processing Systems*, vol. 32, pp. 7135–7145, Jan. 2019, [Online]. Available: https://papers.nips.cc/paper/8935-first-order-motion-model-for-image-animation.pdf

[36] H. Mori and H. Kasuya, "Speaker conversion in ARX-based source-formant type speech synthesis," *8th European Conference on Speech Communication and Technology, EUROSPEECH 2003 - INTERSPEECH 2003, Geneva, Switzerland, September 1-4, 2003*, Sep. 2003, doi: 10.21437/eurospeech.2003-666.

[37] A. Firc, K. Malinka, and P. Hanáček, "Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors," *Heliyon*, vol. 9, no. 4, p. e15090, Apr. 2023, doi: 10.1016/j.heliyon.2023.e15090.

[38] S. Kolagati, T. Priyadharshini, and V. M. A. Rajam, "Exposing deepfakes using a deep multilayer perceptron – convolutional neural network model," *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100054, Apr. 2022, doi: 10.1016/j.jjimei.2021.100054.
[39] R. Tolosana, S. Romero-Tapiador, R. Vera-Rodriguez, E. Gonzalez-Sosa, and J. Fierrez, "DeepFakes detection across generations: Analysis of facial regions, fusion, and performance evaluation," *Engineering Applications of Artificial Intelligence*, vol. 110, p. 104673, Apr. 2022, doi: 10.1016/j.engappai.2022.104673.
[40] B. U. Mahmud and A. Sharmin, "Deep Insights of Deepfake Technology : A review," *arXiv (Cornell University)*, Jan. 2021, doi: 10.48550/arxiv.2105.00192.
[41] T. Jung, S. Kim, and K. Kim, "DeepVision: DeepFakes detection using human eye blinking pattern," *IEEE Access*, vol. 8, pp. 83144–83154, Jan. 2020, doi: 10.1109/access.2020.2988660.
[42] T. Jung, S. Kim, and K. Kim, "DeepVision: DeepFakes detection using human eye blinking pattern," *IEEE Access*, vol. 8, pp. 83144–83154, Jan. 2020, doi: 10.1109/access.2020.2988660.
[43] C. Hu, X. Xie, and L. Wu, "Face reenactment via generative landmark guidance," *Image and Vision Computing*, vol. 130, p. 104611, Feb. 2023, doi: 10.1016/j.imavis.2022.104611.
[44] J. Ke and L. Wang, "DF-UDetector: An effective method towards robust deepfake detection via feature restoration," *Neural Networks*, vol. 160, pp. 216–226, Mar. 2023, doi: 10.1016/j.neunet.2023.01.001.
[45] N. M. Alnaim, Z. M. Almutairi, M. S. Alsuwat, H. H. Alalawi, A. Alshobaili, and F. S. Alenezi, "DFFMD: A Deepfake Face mask dataset for infectious disease era with deepfake detection algorithms," *IEEE Access*, vol. 11, pp. 16711–16722, Jan. 2023, doi: 10.1109/access.2023.3246661.
[46] T. T. Nguyen *et al.*, "Deep learning for deepfakes creation and detection: A survey," *Computer Vision and Image Understanding*, vol. 223, p. 103525, Oct. 2022, doi: 10.1016/j.cviu.2022.103525.
[47] A. Qadir, R. Mahum, M. A. El-Meligy, A. E. Ragab, A. AlSalman, and H. Hassan, "An efficient deepfake video detection using robust deep learning," *Heliyon*, p. e25757, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25757.
[48] "CelebFaces Attributes (CeLEBA) Dataset," *Kaggle*, Jun. 01, 2018. https://www.kaggle.com/datasets/jessicali9530/celeba-dataset
[49] "Deepfake Detection Challenge | Kaggle." https://www.kaggle.com/c/deepfake-detection-challenge
[50] S0md3v, "GitHub - s0md3v/roop: one-click face swap," *GitHub*. https://github.com/s0md3v/roop

[51] “Premium AI video suite for business | AKOOL.” https://akool.com/apps/faceswap

[52] B. Jiang, Q. Ren, F. Dai, J. Xiong, J. Yang, and G. Gui, “Multi-task cascaded convolutional neural networks for Real-Time Dynamic Face Recognition Method,” in *Lecture notes in electrical engineering*, 2019, pp. 59–66. doi: 10.1007/978-981-13-6508-9_8.

[53] Liz, “The rising threat of deepfakes in Cybersecurity: How to protect your organization,” Cypfer, Jul. 10, 2024. https://cypfer.com/the-rising-threat-of-deepfakes-in-cybersecurity/#:~:text=Deepfakes%20pose%20several%20cybersecurity%20risks,sensitive%20information%20or%20transferring%20funds.

[54] “FaceForensics++,” *Kaggle*, Feb. 13, 2024. https://www.kaggle.com/datasets/hungle3401/faceforensics

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: INTELLIGENCE FOR DEEPFAKE DETECTION: AN EXPERIMENTAL MEASURE TO COMBAT RISING CYBER THREAT

Student ID: 202-15-3813 & 202-15-3832

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **deepfake detection** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project demonstrates **fundamental engineering (K3)** principles by employing deep neural networks, data augmentation techniques, and diverse CNN architectures for image processing and classification tasks. The project demonstrates **specialist knowledge (K4)** by conducting transfer learning with advanced CNN, EfficientNet B7, Xception, MLP architectures, enhancing deepfake detection accuracy in video images.	**Page no:** [1-5] **Section:** [1.1] **Page no:** [44-47] **Section:** [5.2]
				The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering practice & technology (K6)** by employing hybrid models.	**Page no:** [36-38] **Section:** [4.2] **Page**

					no: [29-32] **Section:** [3.4]
				This project ensures K8 **(Research Literature)** by synthesizing insights from recent studies, to advance deepfake detection using deep learning, showcasing a comprehensive understanding of current methodologies.	**Page no:** [12-18] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by recognizing the hurdles in deepfake detection, including limitations of traditional methods and the complexities of integrating transfer learning and deep CNNs. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies.	**Page no:** [18-19] **Section:** [2.4, 2.5]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting Transfer Learning as the chosen significant solution for enhancing deepfake detection amidst multiple potential approaches.	**Page no:** [40-44] **Section:** [5.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project using Deepfake detection involves understanding a variety of complex issues spanning technical, ethical, and societal domains. Technically, it requires knowledge of machine learning, particularly deep learning, to identify and combat sophisticated fake media generated by AI algorithms, in line with EP-4 standards.	**Page no:** [12-18] **Section:** [2.2, 2.4]

5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in deepfake detection which ensures **EP-7**.	**Page no:** [21-29] **Section:** [3.2, 3.3, 3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in pneumonia detection through transfer learning and deep learning algorithms.	**Page no:** [29-33] **Section:** [3.4]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	Yes		Web-based deepfake detection with primary data creation fosters EA3: Innovation by enabling scalable testing and continuous improvement. User-generated data enhances detection accuracy, addressing digital security needs through cutting-edge, adaptive solutions.	**Page no:** [44-47] **Section:** [5.3]
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving cyber awareness through advanced deepfake detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.	**Page no:** [48-51] **Section:** [6.1, 6.2, 6.4]

5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in deepfake detection through deep learning, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [12-18] **Section:** [2.1, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [34-35] **Section:** [3.5, 3.6]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing patient privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced AI technologies which comply with **CO9**.	**Page no:** [50-51] **Section:** [6.3]
3	CO10	Yes	Capable of operating proficiently both individually and as a team leader in developing a deepfake detection system, demonstrating technical expertise, problem-solving, and effective collaboration across diverse and interdisciplinary teams to integrate sophisticated models and web applications for this project with **CO10**.	**Page no:** [29-32, 34, 36-38] **Section:** [3.4, 3.5, 4.2]
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [21-33, 40-47] **Section:** [3.2, 3.3, 3.4, 3.5, 5.2, 5.3]

INTELLIGENCE FOR DEEPFAKE DETECTION: AN EXPERIMENTAL MEASURE TO COMBAT RISING CYBER THREAT

ORIGINALITY REPORT

18%	12%	9%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	4%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	www.researchgate.net Internet Source	1%
4	Fakhar Abbas, Araz Taeihagh. "Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence", Expert Systems with Applications, 2024 Publication	<1%
5	deepai.org Internet Source	<1%
6	www.mdpi.com Internet Source	<1%
7	Submitted to Liverpool John Moores University Student Paper	<1%