

**TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING**



**LALITPUR ENGINEERING COLLEGE
KHOLKA POKHARI, LALITPUR**

**A PROPOSAL OF MAJOR PROJECT
ITHUB**

SUBMITTED BY

ABHISHEK NEUPANE (LEC-076-BCT-02)
RABINDRA ADHIKARI (LEC-076-BCT-025)
SANJISH MAHARJAN (LEC-076-BCT-032)
SUSHIL KAFLE (LEC-076-BCT-045)

SUBMITTED TO

DEPARTMENT OF COMPUTER ENGINEERING

Date

ABSTRACT

Deepfakes are realistic-looking fake media generated by deep-learning algorithms that iterate through large datasets until they have learned how to solve the given problem (i.e., swap faces or objects in video and digital content). The massive generation of such content and modification technologies is rapidly affecting the quality of public discourse and the safeguarding of human rights. Deepfakes are being widely used as a malicious source of misinformation in court that seek to sway a court's decision. Because digital evidence is critical to the outcome of many legal cases, detecting deepfake media is extremely important and in high demand in digital forensics. As such, it is important to identify and build a classifier that can accurately distinguish between authentic and disguised media, especially in facial-recognition systems as it can be used in identity protection too. In this work, we compare the most common, state-of-the-art face-detection classifiers such as Custom CNN, VGG19, and DenseNet-121 using an augmented real and fake face-detection dataset. Data augmentation is used to boost performance and reduce computational resources. Our preliminary results indicate that VGG19 has the best performance and highest accuracy of 95% when compared with other analyzed models.

Keywords: deepfake detection; digital forensics; media forensics; deep learning; VGG19; face-image manipulation

Contents

1	INTRODUCTION	1
1.1	Background	2
1.2	Problem Statement	3
1.3	Objectives	4
1.4	Scope	4
2	LITERATURE REVIEW	5
2.1	Existing Systems	6
2.2	Proposed Systems	7
3	FEASIBILITY STUDY	8
3.1	Economic feasibility	8
3.2	Operational feasibility	8
3.3	Technical feasibility	8
4	METHODOLOGY	9
4.1	Software Development Life Cycle	9
4.2	System Development Tools	10
4.3	Functional Requirement	10
4.4	Non Functional Requirement	10
5	BLOCK DIAGRAMS	11
5.1	System Architecture	11
5.2	System Design	12
5.3	Use Case Diagram	13
5.4	Sequence Diagram	14
5.5	Class Diagram	15
5.6	Dataflow Diagram	16
5.7	Activity Diagram	17
6	EXPECTED OUTCOMES	18
7	LIMITATIONS	19
8	FUTURE ENHANCEMENTS	20

List of Figures

1	Agile Model	9
2	Agile Model	13
3	Level 0 DFD	16
4	Level 1 DFD	16

1 INTRODUCTION

Deepfake technology has revolutionized the world of digital media manipulation. By combining artificial intelligence and image/video processing, deepfakes have garnered widespread attention. Deepfakes involve the creation of realistic media portraying individuals in situations they never experienced or saying things they never said. As this technology becomes more sophisticated, concerns arise regarding its impact on politics, entertainment, and personal privacy. This report provides an overview of deepfakes, including their underlying processes, societal implications, ethical challenges and the way of detection. In navigating this landscape, it is crucial to find a balance between innovation and responsible use in our increasingly digitized society.

In the last few years, cybercrime, which accounts for a 67% increase in the incidents of security breaches, has been one of the most challenging problems that national security systems have had to deal with worldwide.[1] Deepfakes, at present time, are being widely used to swap faces or objects in video and digital content. This artificial intelligence-synthesized content can have a significant impact on the determination of legitimacy due to its wide variety of applications and formats that deepfakes present online (i.e., audio, image and video). Considering the quickness, ease of use, and impacts of social media, persuasive deepfakes can rapidly influence millions of people, destroy the lives of its victims and have a negative impact on society in general [1].

Deepfake technology has been driven by various motivations, including individual attacks, political manipulation, and the spread of false information. Its impact extends beyond personal attacks to manipulating satellite images and using stock images for identity protection. Cyber attackers continuously adapt their strategies, making it challenging to identify deepfake media and stay ahead of evolving threats.

The societal implications of deepfake technology are profound. Misinformation and disinformation fueled by deepfakes erode public trust, damage reputations, and violate privacy at personal and professional levels. Deepfakes can also disrupt democratic processes and contribute to societal polarization. Addressing the legal and ethical concerns surrounding deepfakes requires technological advancements, policy development, media literacy, and careful consideration of privacy rights and the manipulation of visual evidence.

To tackle these implications, it is essential to advance deepfake detection methods, bolster cybersecurity measures, promote media literacy for individuals to discern manipulated content, and establish clear legal frameworks governing the responsible use of

deepfake technology. By taking a comprehensive approach, we can effectively navigate the ethical challenges and societal impacts posed by deepfakes.

1.1 Background

At present context of time, the rapid advancements in mobile camera technology and the widespread use of social media platforms have made it easier than ever to create and share digital pictures. Deep learning has played a crucial role in developing technologies that were previously unimaginable. One notable example is modern generative models, which can produce highly realistic images, speech, music, and video. These models have been applied in various fields, such as enhancing accessibility through text-to-speech technology and generating training data for medical imaging.

There will always be drawbacks to any technological breakthrough. Since deepfakes are still relatively new and expanding quickly, their excessive use as a result of rising human interest has resulted in misuse of this technology. It is simple for widespread false information to proliferate among the populace when there is no controlling element and a weak mechanism in place to identify deep fakes. Since their initial emergence in late 2017, a variety of open-source deep fake generation techniques and tools have appeared, resulting in an increase in the amount of synthetic media clips. Others may be destructive to people and society, even though many are probably intended to be amusing. Due to the accessibility of editing tools and the strong demand for topic expertise, false digital contents have been growing in number and in realism up until recently.

Deep fakes are now widely disseminated on social media platforms, which encourages spamming and the spread of false information. Just picture a deep fake image of Donald Trump getting arrested which was trending on twitter or a deep fake of a well-known celebrity assaulting their supporters. These types of misinformation can brainwash the audience and are awful and endanger and mislead the general public.

Deep fake detection plays a crucial part in overcoming such a circumstance. Therefore, we provide a novel deep learning-based method that can successfully separate artificial intelligence-generated fake photos (Deep fake images) from authentic digital materials. In order to identify deep fakes and stop them from spreading across the internet, it is crucial to develop technology that can detect fakes.

Although with all these downsides mentioned, the deepfake technology holds impor-

tance in several areas. It offers creative expression and entertainment possibilities, enhances research and development in fields like computer vision, and aids forensic analysis in legal investigations. Deepfakes also emphasize the need for media literacy and critical thinking skills, promoting education and awareness. Ethical considerations and policy development are crucial in addressing the responsible use of deepfakes and protecting individuals' rights. Understanding the significance of deepfake technology enables us to navigate its implications effectively and harness its potential while mitigating potential harm.

1.2 Problem Statement

With the help of visual effects, convincing modifications of digital photographs and videos have been proven for many years. However, new developments in deep learning have dramatically increased the realism of fake content and made it more widely available. These purportedly artificial intelligence-generated works of media are also known as "deepfakes". It is easy to create deep fakes utilizing artificial intelligence techniques. However, it is extremely difficult to identify these Deep Fakes. In the past, there have been numerous instances of deep fakes being used to effectively incite political unrest, stage terrorist attacks, make revenge porn, blackmail individuals, etc. Therefore, it becomes crucial to identify these deep fakes and stop their spread through social media. Therefore, with the growing curiosity we have taken a step forward in detecting the deep fakes using vggface2 based artificial Neural network.

1.3 Objectives

- Our project aims at discovering the distorted truth of the deep fakes.
- Our project will reduce the Abuses' and misleading of the common people on the world wide web.
- Our project will distinguish and classify the video as deepfake or pristine.
- Provide a easy to use system for used to upload the video and distinguish whether the video is real or fakepSS

1.4 Scope

At present time there are numerous tools available for creating false videos in the current deepfake technology landscape, but there are few trustworthy tools available for spotting them. The idea creation of a deepfake detection software to solve this discrepancy and stop the widespread dissemination of deepfakes is what our project is based upon. Users will be able to post images through our platform and segregate them as authentic or deepfake. This project can be developed to include the development of a plugin for browsers that will automatically detect deep fakes. Notably, our idea can me implemented on different social sites as well as in various various governmental organizations. A synopsis of the program with the size of the input, bounds on the input, input validation, input dependency, the i/o state diagram, and the major inputs and outputs are explained in this report.

2 LITERATURE REVIEW

2.1 Existing Systems

2.2 Proposed Systems

Our proposed Deepfake Detection System is an online application created to aid users in identifying and mitigating the existence of deepfake content. It prioritizes offering robust deepfake detection capabilities to users. Users can take advantage of the system's deepfake detection tools, which enable them to submit images for analysis and detection of potential deepfake content. The system employs advanced algorithms and machine learning methods to accurately identify manipulated or artificially generated media. This assists users in recognizing and addressing the risks associated with deepfake content, such as spreading misinformation, perpetrating fraud, or violating privacy. The deepfake detection system strives to empower users by equipping them with the necessary resources to actively safeguard themselves and others against the potential negative consequences of encountering deepfakes. By leveraging cutting-edge algorithms, the system aids users in detecting and raising awareness about the presence of deepfakes, fostering a safer and more knowledgeable digital environment.

3 FEASIBILITY STUDY

3.1 Economic feasibility

This is a low-budget project with no development costs. The total expenditure of the project is just computational power. The dataset and computational power required for the project are easily available. The computational power is easily provided by google collab. So, the project is economically feasible. The system will be simple to comprehend and use. As a result, there will be no need of trained personnel to use the system. This system will have the capacity to expand by adding more components.

3.2 Operational feasibility

The project is operationally feasible since after the completion of the project, it can be operated as intended by the user to solve the problems for what it has been developed.

3.3 Technical feasibility

The purpose of technical feasibility is to establish whether the project is possible in terms of software, hardware, manpower, and knowledge to complete. It will take into account determining resources in support of the suggested scheme. The system is platform independent because it is written in Python. Advanced machine learning libraries are available and the technology is cutting-edge. As a result, the system is technically possible.

4 METHODOLOGY

4.1 Software Development Life Cycle

Agile method of Software Development uses iterative approach. Agile method cycles among Planning, Requirement Analysis, Designing, Development and Testing stages. These cycle is called sprints. Each sprints are considered as a miniature project on itself. Using this method allowed us to update various parts of project at any point of project development. In this model an iterative approach was taken where working software was delivered after each iteration some new features is added to main system. It works in incremental and iterative approach. Agile model mainly focuses on customer collaborations, on individuals and iterations and welcomes changes at anytime in SDLC process. We prefer to use agile model in this system as it helps in developing realistic systems and promotes teamwork during software development. Also system is easy to manage and it can accommodate new changes at any stages of software development phase.

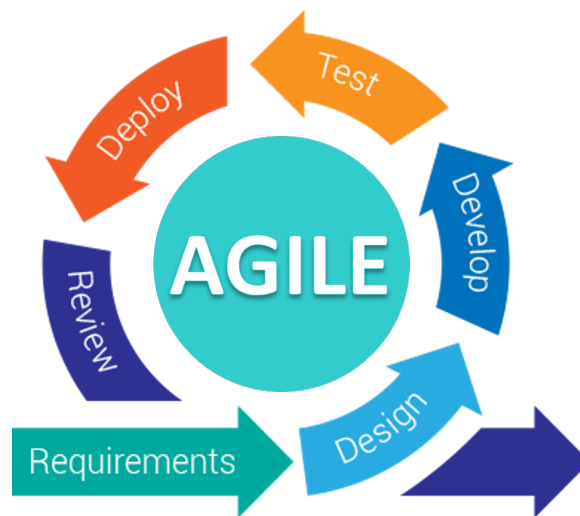


Figure 1: Agile Model

4.2 System Development Tools

Our static Deepfake detection System requires Python, Tensorflow, OpenCV, Machine Learning which are listed below:

1. Python
2. Pytorch
3. NumPy
4. OpenCV
5. Tensorflow

4.3 Functional Requirement

The functional requirements of the system are:

1. Detecting the Faces from Images and Videos.
2. Testing for realism of image.

4.4 Non Functional Requirement

These requirements are not needed by the system but are essential for the better performance of software. The points below focus on the non-functional requirement of the system.

- Reliability
- Usability
- Security
- Portability
- Speed and responsiveness
- Performance

5 BLOCK DIAGRAMS

5.1 System Architecture

5.2 System Design

5.3 Use Case Diagram

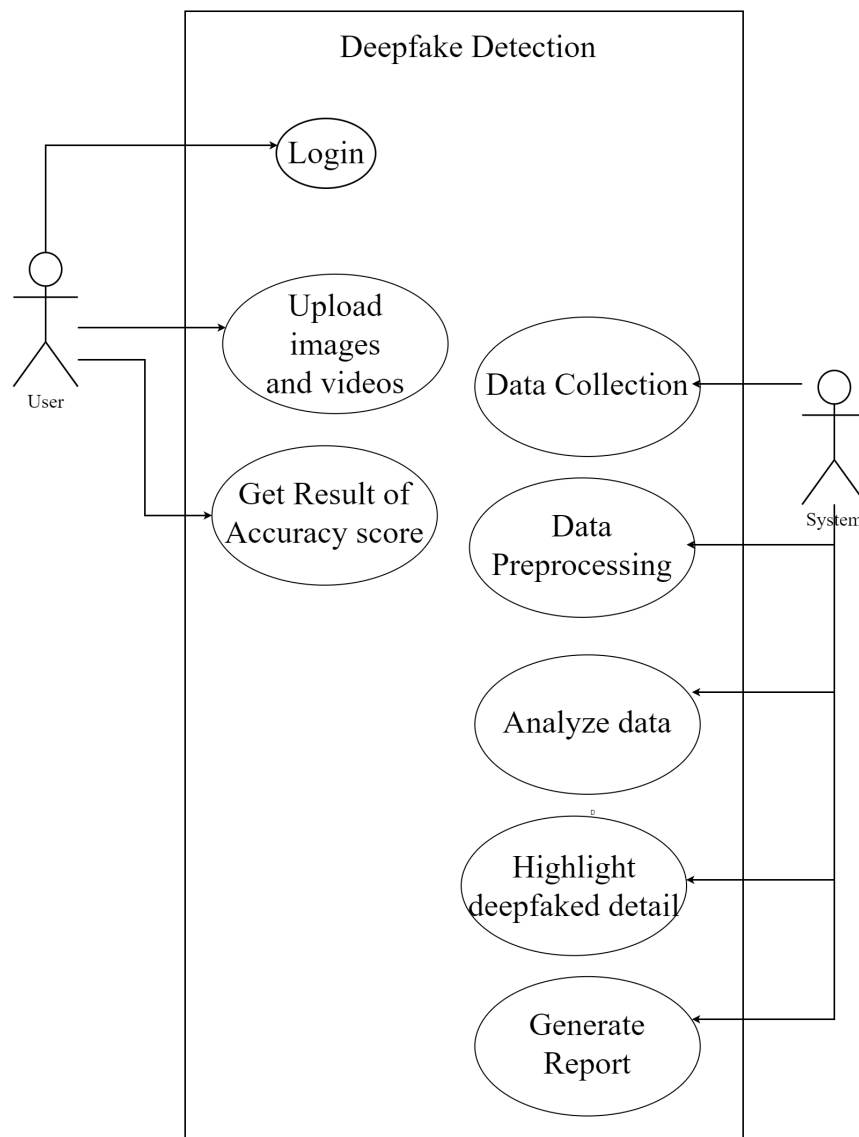


Figure 2: Agile Model

5.4 Sequence Diagram

5.5 Class Diagram

5.6 Dataflow Diagram

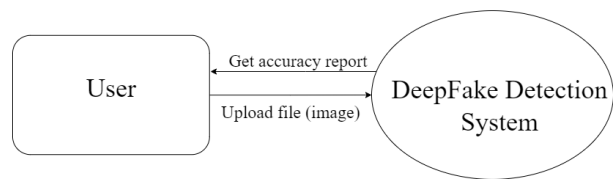


Figure 3: Level 0 DFD

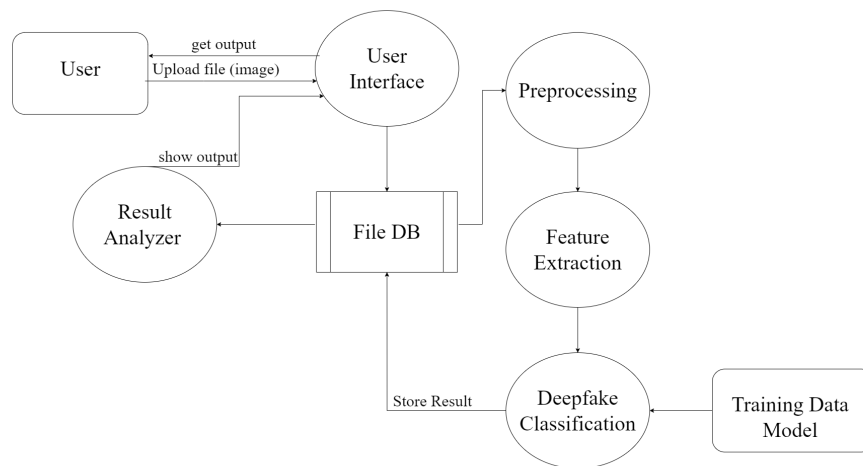


Figure 4: Level 1 DFD

5.7 Activity Diagram

6 EXPECTED OUTCOMES

- User-friendly interface for easy upload and clear result presentation.
- Accurate identification of manipulated media content.
- Robust performance against different deepfake techniques and adversarial attacks.

7 LIMITATIONS

- Deepfake detection projects face challenges due to rapidly evolving techniques and the need for diverse training data.
- Adversarial attacks can exploit weaknesses in detection algorithms, making deep-fakes harder to identify accurately.
- Deepfake detection algorithms often require significant computational resources, limiting their applicability on resource-constrained devices.

8 FUTURE ENHANCEMENTS

There is always a scope for enhancements in any developed system, especially when the project build using latest trending technology and has a good scope in future.

- Web based platform can be upscaled to a browser plugin for ease of access to the user.
- Currently only Face Deep Fakes are being detected by the algorithm, but the algorithm can be enhanced in detecting full body deep fakes.

References

- [1] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner, “FaceForensics++: Learning to Detect Manipulated Facial Images” in arXiv:1901.08971.
- [2] Deepfake detection challenge dataset : <https://www.kaggle.com/c/deepfake-detection-challenge/data> Accessed on 26 March, 2020
- [3] Yuezun Li , Xin Yang , Pu Sun , Honggang Qi and Siwei Lyu “Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics” in arXiv:1909.12962
- [4] 10 deepfake examples that terrified and amused the internet : <https://www.creativebloq.com/features/deepfake-examples> Accessed on 26 March, 2020
- [5] Keras: <https://keras.io/> (Accessed on 26 March, 2020)
- [6] PyTorch : <https://pytorch.org/> (Accessed on 26 March, 2020)
- [7] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. arXiv:1702.01983, Feb. 2017
- [8] TensorFlow: <https://www.tensorflow.org/> (Accessed on 26 March, 2020)
- [9] Face app: <https://www.faceapp.com/> (Accessed on 26 March, 2020)
- [10] Face Swap : <https://faceswaponline.com/> (Accessed on 26 March, 2020)