TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING

Khwopa College Of Engineering
Libali, Bhaktapur
Department of Computer Engineering

A PROPOSAL ON

DEEPFAKE DETECTION

Submitted in partial fulfillment of the requirements for the degree

BACHELOR OF COMPUTER ENGINEERING

Submitted by

Manish Pyakurel	KCE077BCT020
Rupak Neupane	KCE077BCT028
Sarjyant Shrestha	KCE077BCT033
Srijan Gyawali	KCE077BCT036

Khwopa College Of Engineering
Libali, Bhaktapur
2023-12-10

Chapter 1

Introduction

1.1 Background Introduction

In recent years, the landscape of digital image manipulation has undergone a transformative shift with the emergence of deepfake techniques. This innovative approach, rooted in deep learning methodologies, has gained significant traction as a means of fabricating images by seamlessly replacing facial features from one individual with those of another. Coined as "deepfakes" by a Reddit user in 2017, these manipulations often leverage advanced adversarial models, such as Generative Adversarial Networks (GANs). Notably, this technology has been controversially utilized to superimpose celebrity faces onto explicit content, raising concerns related to fake pornography, misinformation, financial fraud, and hoaxes. Despite the ethical challenges associated with deep fakes, it is essential to acknowledge the positive applications within fields such as virtual reality, film editing, and production. The core working principles behind deep fakes involve intricate processes of merging, replacing, combining, and superimposing images. Leveraging deep learning and machine learning techniques, these manipulations give rise to convincingly altered digital images and videos, demonstrating both the potential benefits and ethical considerations associated with this rapidly advancing technology.

1.2 Problem Statement

In the rapidly evolving landscape of computer and automation technologies, the realm of possibilities continues to expand. Artificial Intelligence (AI) stands as a pivotal force, driving unprecedented advancements in areas such as predictive analytics, weather forecasting, automation, and the creation of sophisticated entities like

deep fakes, which encompass AI-generated videos, audios, and images. While these technological strides are undeniably transformative, the misuse and exploitation of such capabilities pose significant concerns.

In recent times, there's been a surge in the creation of deep fakes, where the faces of celebrities or ordinary people are manipulated using just a single image and advanced deep learning algorithms. This issue is becoming more significant, as it circulates potentially harmful and illegal images of the victims to the public.

The rise of these deceptive practices not only threatens individual privacy but also has broader implications for public trust and safety. As deep fakes become increasingly convincing, the potential for malicious use, misinformation, and damage to reputations grows. It is crucial to address this issue head-on by developing sophisticated detection mechanisms to safeguard against the harmful consequences of manipulated images. This proposal seeks to contribute to the ongoing efforts in mitigating the risks associated with deep fakes, reinforcing the integrity of visual content in the age of advanced AI technologies.

1.3 Objective

The main aim of this project is:

- To identify manipulated digital media content, particularly facial features and images.
- To implement cutting-edge deep learning and machine learning techniques.

Chapter 2

Literature Review

Deepfakes, which involve the unauthorized swapping of face images, are frequently carried out without the knowledge or consent of individuals, including celebrities and politicians. Notably, historical instances, such as the facial image swapping in a photograph of Abraham Lincoln (Badale et al., 2018), underscore the longstanding nature of this challenge. Addressing these concerns, Yang, Li & Lyu (2019) proposed a model leveraging head pose inconsistency to detect deepfakes, enabling the creation of synthetic faces for various individuals while preserving the original facial expressions. Jagdale & Shah (2019) introduced the NA-VSR algorithm for super resolution, involving video conversion into frames, median filtering to remove noise, and bicubic interpolation for pixel density enhancement. Additionally, Yadav & Salmani (2019) elucidated the working principles of deep fake techniques, emphasizing the high precision value in face image swapping. Generative Adversarial Neural Networks (GANs) play a pivotal role in deepfake generation, comprising a generator and a discriminator. The generator synthesizes fake images from a given dataset, while the discriminator evaluates the authenticity of the generated images. The inherent risks of deepfakes, including character defamation, potential harm to individuals, and the dissemination of fake news in society, highlight the importance of addressing these challenges. Existing approaches encounter issues such as inefficiency in detecting deepfake images, high error rates, prolonged computation times, and data access inaccuracies. This work, FF-LBPH-DBN, focuses on minimizing computational complexity while efficiently applying various metrological parameters. Figure 1 presents a survey-based overview of existing approaches for detecting fake images (Vivek et al., 2018).

Table 2.1: Compilation of related work

Researcher	**Contributions**	**Scope**	**Advantage**	**Weakness**
L. Verdoliva (2020)	Presenting an overview of contemporary manipulation techniques	Fake media	Deepfake's backstory is presented. Issues and possible solutions are explored.	There has not been any in-depth review of the articles.
Tolosana et al. (2020)	Examining face-altering techniques	Image deepfake detection	Different criteria for evaluating articles are taken into account.	It is unclear how articles are chosen for review.
Mirsky and Lee (2021)	Providing deepfake creation and detection services	Deepfake in general	Challenges and potential guidance are discussed.	It is unclear how articles are chosen for review.
Castillo Camacho and Wang (2021)	Examining DL-based image forensic methods	Image forensic	Taking into account all aspects of the criteria for image forensics.	It is unclear how articles are chosen for review.
P. Yu et al. (2021)	Focusing on deepfake video detection, its history, current research, and plans	Deepfake video	An in-depth description of future work. In-depth examination of datasets.	There is no comparison between the articles.
Rana et al. (2022)	Demonstrating several cutting-edge deepfake algorithms	DL-ML and statistical models	There is a comparison between the articles.	There is no discussion of all kinds of deepfake applications.
Ours	Providing a comprehensive review of the literature on deepfake detection techniques based on DL-based algorithms	DL-ML methods in the video, image, audio, and hybrid multimedia detection	An in-depth description of future work. In-depth examination of datasets. Challenges and potential guidance are discussed.	Papers published before 2018 are not allowed.

Chapter 3

Requirement Analysis

3.1 SOFTWARE REQUIREMENT

Our Deepfake Detection project requires following softwares:

3.1.1 Python

Python is a general-purpose, high-level programming language. With a strong emphasis on indentation, its design philosophy prioritizes code readability. Python uses garbage collection and dynamic typing. It is compatible with various programming paradigms, such as object-oriented, functional, and structured (especially procedural). It has an extensive standard library.

3.1.2 React

React is a free and open-source front-end JavaScript toolkit for creating component-based user interfaces. It is also referred to as React.js or ReactJS. It is maintained by a group of independent developers and businesses as well as Meta (previously Facebook).

3.1.3 FastAPI

A contemporary web framework for creating RESTful Python APIs is called FastAPI. Since its initial release in 2018, its robustness, speed, and ease of use have helped it rapidly acquire favor among developers. Based on Pydantic, FastAPI serializes and deserializes data using type hints for validation. For APIs created with it, OpenAPI

documentation is also automatically generated.

3.1.4 TensorFlow

TensorFlow is a free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. TensorFlow was developed by the Google Brain team for internal Google use in research and production. TensorFlow can be used in a wide variety of programming languages, including Python, JavaScript, C++, and Java.

3.1.5 Keras

Keras is a high-level, deep learning API developed by Google for implementing neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation. Keras is relatively easy to learn and work with because it provides a python frontend with a high level of abstraction while having the option of multiple back-ends for computation purposes. It supports frameworks like tensorflow.

3.2 FUNCTIONAL REQUIREMENT

3.2.1 Dataset Labeler

Dataset labeler is the labelling system that is used to annotate the texts as “Real”, or “Fake”.

3.2.2 Nepali News Web Scrapper

It is the tool that extracts various Nepali text from twitter as well as various Nepali news portals to fed into Dataset Labeler as well as vector creation process.

3.2.3 Inference System

It is the system formed after training the model using the dataset labeled from dataset labeler and run the prediction model in it.

3.3 NON-FUNCTIONAL REQUIREMENT

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

3.3.1 Reliability

The system is reliable. Sentiment prediction matches 80

3.3.2 Maintainability

A maintainable system is created and Sentiment Analyzer Engine is able to train on new input data and is scalable to millions of data points.

3.3.3 Performance

The forward pass from the neural network is a fast process. For the engine, fast matrix computation occurs.

3.3.4 Portability

Sentiment Analyzer engine is portable and it is easy to integrate into any web application or mobile application imaginable by the use of the REST API's made.

3.4 FEASIBILITY STUDY

The following points describes the feasibility of the project.

3.4.1 Economic Feasibility

The total expenditure of the project is just computational power. The dataset and computational power required for the project are easily available. Dataset is found from the internet and computational power using a powerful PC provided by the college. Therefore, the project is economically feasible.

3.4.2 Technical Feasibility

Although the datasets are easily available on the internet, it is estimated that it will take a large amount of time in order to train. Training huge news dataset takes a lot of computation power, with the help of the college provided machine the project is technically feasible.

3.4.3 Operational Feasibility

Chapter 4

Methodology

4.1 SOFTWARE DEVELOPMENT APPROACH

Prototyping model is a type of software develpoment model. It is an iterative approach where a basic prototype is constructed to gain a better understanding of the project. This prototype is typically incomplete or lacking many components. The model is then refined based on feedback and system is reconstructed iteratively until desired conditions are met.

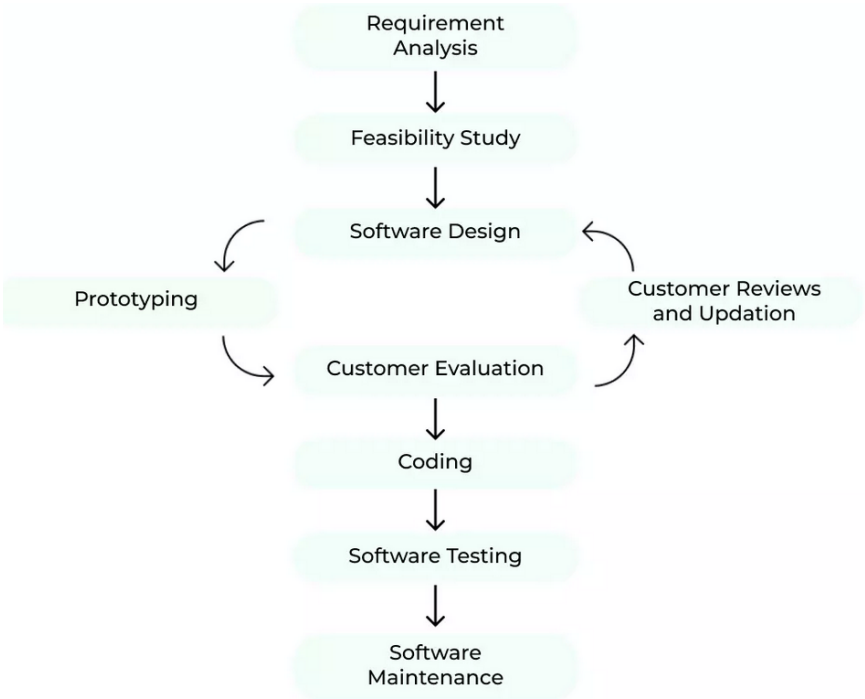

Figure 4.1: Prototype Model for Software Development

4.2 DATA COLLECTION

We have found many datasets on the internet from popular platforms like kaggle, github. For this project we will be using the datasets provided by ondyari/ FaceForensics `https://github.com/ondyari/FaceForensics`.
For POS tagging, the data from NELRALEC [10] The following table shows the amount of data collected from different sources and their usage in our project;