National College of Ireland

Deepfake Detection System by Integrating Deep Learning and Blockchain Technology

MSc Research Project
Cyber Security

Raj Yatin Koli

X21154678

School of Computing
National College of Ireland

Supervisor: Michael Pantridge

National College of Ireland Project Submission Sheet
School of Computing

Student Name:	Raj Yatin Koli
Student ID:	X21154678
Programme:	Cyber Security
Year:	2023
Module:	MSc Research Project
Lecturer:	Michael Pantridge
Submission Due Date:	18th September 2023
Project Title:	Deepfake Detection System by Integrating Deep Learning and Blockchain Technology
Word Count:	6608

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	Raj Yatin Koli
Date:	18th September 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keepa copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placedinto the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Deepfake Detection System by Integrating Deep Learning and Blockchain Technology

Raj Yatin Koli
X21154678
MSc Research Project in
Cyber Security
18th September 2023

Abstract

Deepfake technology poses challenges to the authenticity of digital visual information. This project combines deep learning and blockchain to develop a trustworthy deepfake detection solution. The model uses CNNs and RNNs for accurate deepfake content identification, and blockchain ensures an immutable record of video authenticity. A diverse dataset is employed for training and testing, and comprehensive metrics assess the system's performance. The research explores the implications of blockchain integration on efficiency, scalability, and security. Results show notable accuracy and reliability in identifying deepfakes, mitigating misinformation risks. Blockchain facilitates forensic analysis and boosts public trust in visual information veracity. Merging deep learning and blockchain fortifies deepfake detection, safeguarding visual content integrity. The integration offers a promising approach to combat evolving deepfake manipulation in the digitalized world. Policymakers, researchers, and developers can benefit from these insights to responsibly use AI and blockchain in digital media authentication.

1 Introduction

Deepfakes are the false representation of digital content which is created by using modern technology such as action and speech detection and using Artificial intelligence to create fake digital content in social media (Westerlund, 2019). Qualitative improvement and technological advancements in deep learning and the artificial intelligence field lead to the creation of phoney digital content which is looking realistic and is known as deep fakes (Veerasamy *et al*., 2022). This type of picture and video can spread quickly through social media and spread rumours and fake news and harm the reputation of that person and the social media site. To detect this type of mis information there is no mechanism present. As a result, this type of unethical outlets can post whatever they can want and making confusion in society. The current situation can not help to detect digital media history and to detect deep fakes authentication is needed an authentication-based method is to find the deep fakes' source of origin (Chennamma *et al*., 2023). That's why implementing blockchain technologies is needed to trace and find the origin of the digital data that is published on social media. Blockchain technology can help to effectively recognise the deep fake digital content and by this trust factor of the user can be secured.

1.1 Concept of deep fake detection using ML

The deep fake detection technique is a process which has been used in many fields. Usually in the area of technology where this deep fake technology has been able to create a path, especially in computer vision, machine technology, chat and language processing (Sarker *et al*., 2021). This technology growing as an attention to the technology world. The manipulation technique has been trying to build a path that can easily threaten cybers, and manipulate information (Assante *et al*., 2015). Here comes Mchine Learning. The machine learning-based concept is the method where preservation can be seen. Future development and testing will be done by machine learning. The main advantage behind this approach is to detect deep fake technology (Liu *et al*., 2021). Deep fake technology is severely attacking the various type of people and areas (Gradoń *et al*, 2020). Using machine learning (ML) is the step to forestall deep fake technologies from social media to locate things (Gonzalez *et al*., 2019). The CNN and classifier networks are the key to detecting deep fake technologies (Shad *et al*., 2021). The visual artefact-based detection method is the key here. The FaceForensics methods and deep counterfeit detection get the good outstretch. So the machine learning model detects highly flatten deep fake technologies in social sites accurately and fails technologies requirement (Shad *et al*., 2021). Machine learning technologies detect highly flatten deep fake technologies in various aspects. ForeForensics dataset creates tremendous effects in challenging deep counterfeit detection (Ekelhart *et al*., 2018). So using this deep fake videos can be detected. So the ML method can detect almost all deep fake videos. The accuracy depends on the number of videos has been detected. Using ML the focus is to limit dark areas such as cyberattacks, leaks in information and related areas (Alazab *et al*., 2021). ML can protect data and collect all data from any particular site which is vulnerable to deep fake attacks. So in the futuristic world, the advance of ML technology is needed for the prevention of deep fake waves.

1.2 Motivation and Project Background:

The rapid advancement of deepfake technology has raised serious concerns about the authenticity and trustworthiness of visual content in the digital era. Deepfakes, which are manipulated videos that convincingly depict individuals saying or doing things they never did, have the potential to cause significant harm, including spreading misinformation, damaging reputations, and manipulating public opinion. As deepfake techniques become increasingly sophisticated and accessible, there is a pressing need for robust and tamper-proof detection mechanisms to combat their harmful effects.

Traditional methods of detecting deepfakes are becoming inadequate as the technology evolves, necessitating the exploration of innovative and cutting-edge solutions. Combining deep learning, which has proven effective in various image and video analysis tasks, with blockchain, known for its transparency and immutability, presents a promising approach to address the challenges posed by deepfake technology. By leveraging the power of these two technologies, it becomes possible to develop a reliable and decentralized system that can verify the authenticity of visual content, thereby enhancing public trust in the digital media landscape. The academic project aims to tackle the growing problem of deepfake videos and their potential to deceive and manipulate the public. To achieve this, the project proposes the integration of

deep learning and blockchain technology to develop a cutting-edge deepfake detection system. The motivation behind this research lies in the urgency to combat the negative impact of deepfakes on society, as well as the need for sophisticated and adaptable methods to detect and counter these manipulated videos effectively. The project builds upon previous research in the fields of deep learning and blockchain, drawing upon the strengths of both technologies to create a hybrid system capable of detecting deepfake videos with a high level of accuracy and reliability. By leveraging the power of deep learning algorithms, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the proposed model aims to distinguish between genuine and manipulated videos, enhancing its detection capabilities. The integration of blockchain technology into the deepfake detection system provides several key advantages. The immutability and transparency of the blockchain ensure that the recorded authenticity of videos cannot be altered or tampered with, establishing a reliable and verifiable record of a video's origin. This blockchain-backed approach not only enhances the system's trustworthiness but also allows for easy forensic analysis, aiding in the identification of the source and origin of deepfake content. To validate the effectiveness of the proposed system, a diverse dataset comprising real and synthetic deepfake videos is used for training and testing. The evaluation metrics include precision, recall, F1-score, and receiver operating characteristic curve (ROC), providing a comprehensive assessment of the model's performance. Furthermore, the project explores the potential challenges and limitations associated with integrating blockchain technology, such as computational overhead and blockchain synchronization. By addressing these concerns, the research aims to optimize the system's overall performance and scalability.

In conclusion, the project aims to contribute to the development of a reliable and efficient solution for detecting deepfake videos. By combining the strengths of deep learning and blockchain, the proposed system seeks to foster trust in the authenticity of visual content and combat the harmful effects of deepfake technology in the digital media landscape.

1.3 Research Question:

"How can the integration of deep learning and blockchain technology be leveraged to develop an innovative and trustworthy deepfake detection system, and what are the implications of this integration on the accuracy, reliability, efficiency, scalability, and security of the system?"

Sub Research Question:

"To what extent does the integration of blockchain technology enhance the verifiability and tamper-resistance of the deep learning-based deepfake detection system, and how does this impact the trustworthiness of the recorded authenticity of visual content?"

1.4 Research Objective:

The primary research objective is to develop an innovative and trustworthy deepfake detection system by integrating deep learning and blockchain technology. This system aims to accurately identify and distinguish between genuine and manipulated videos, mitigating the risks

associated with misinformation and fraudulent visual content. Additionally, the research aims to assess the implications of blockchain integration on the system's efficiency, scalability, and security, with the ultimate goal of fortifying the detection of deepfake videos and preserving the authenticity and reliability of visual information in the digital media landscape.

2 Related Work

2.1 Introduction

In the era of rapid technological advancements, deepfake technology has emerged as a significant threat to the authenticity and credibility of visual content. Deepfake videos, generated using deep learning techniques, have the potential to manipulate and deceive, leading to serious ethical and societal implications. Detecting and countering these manipulative videos is becoming increasingly crucial to safeguard the integrity of digital media. This literature review explores the use of cutting-edge technologies, particularly deep learning and blockchain, to address the challenges posed by deepfake technologies. Deep learning, powered by artificial intelligence, has shown promise in identifying and classifying deepfake content. However, some limitations remain, particularly in the area of authentication, where traditional methods struggle to provide robust solutions. The integration of blockchain technology into the deepfake detection process offers a compelling approach to enhance the system's trustworthiness. Blockchain's inherent properties, such as transparency, immutability, and decentralized nature, provide a tamper-resistant and verifiable record of video authenticity. As a result, it bolsters the ability to detect and counter deepfake manipulations effectively. Several studies have explored the potential of blockchain technology in detecting and combating deepfake content. For instance, some research has focused on using blockchain for tracking the origin of fake news and verifying its authenticity, ensuring the credibility of information in an increasingly digitized world. Other investigations have delved into the application of machine learning algorithms in conjunction with blockchain to detect deepfake videos. These studies highlight the effectiveness of machine learning-based methods, particularly deep learning, in outperforming other techniques in identifying deepfake content accurately.

Furthermore, the review reveals that blockchain's decentralized and secure nature is particularly advantageous in preventing the spread of false information on social media platforms. The implementation of blockchain in social media networks helps authenticate and verify the legitimacy of content, mitigating the impact of misinformation and rumors. In light of the growing challenges posed by deepfake technologies, this literature review underscores the significance of combining cutting-edge technologies like deep learning and blockchain to fortify the detection of deepfake videos. By leveraging the strengths of these technologies, it becomes possible to create a reliable and efficient system that preserves the authenticity and reliability of visual information in the digital media landscape. The following sections provide a comprehensive analysis of the different approaches, algorithms, and methodologies used in

existing research to address the deepfake detection problem, aiming to contribute valuable insights to the ongoing efforts in combating the manipulation of digital media content.

2.2 Reviews of Deepfake Detection using Machine Learning

Yazdinejad et al. (2020) highlights the unethical use of deepfake technologies to manipulate data, creating challenges in technology fields. They propose an AI-based approach, particularly machine learning, to resist deepfake manipulation. The study acknowledges the strong process of deepfakes but identifies loopholes requiring attention, particularly in authentication. To address this, they introduce the concept of blockchain technology, which offers promising solutions for detecting and constraining deepfake technologies (Yli-Huumo et al., 2016). Blockchain is seen as an effective tool to combat deepfake challenges, enhancing its configuration and functionalities (Javaid et al., 2021).

Loey et al. (2022) present a study on the combined use of machine learning and blockchain technology to combat deepfakes. Blockchain's advancements in machine networks enhance artificial intelligence capabilities. The study emphasizes the growing demand for deepfakes due to freely available content, enabling the generation of false digital information. By using machine learning and blockchain, the research broadens the scope to classify news and create a prediction system. The model's effectiveness in detecting fake news, analyzing methods, accuracy, and origin tracing contributes to advancing deepfake detection technology. Furthermore, employing machine learning and deep learning for censorship and authentication can improve social media credibility (Chakraborty et al., 2020).

Hameed et al. (2022) present a study on the aggressive nature of Mobile Crowdsensing (MCS) in deepfake detection. MCS, along with IOTA, has emerged as a prominent detection method over the past decade. The development of IOTA-based mobile crowdsensing as a machine learning tool offers a procedure to prevent mobile deepfake detection. The system's logic and algorithm ensure high-quality results, with Logi-LGBM and Logi-HGBC achieving remarkably high accuracy in deepfake detention. Additionally, the IOTA dataset introduces new methodologies for predicting future attacks and facilitating prevention, instilling faith in the intelligent system's capabilities to safeguard financial systems, cryptocurrencies, sensors, smart cities, and satellites from deepfake technologies.

Dhall et al. (2021) highlight the growing impact of fake communication on modern-day social media, causing serious problems in various aspects of society. The prevalence of fake and ferocious news, particularly during elections and stock market activities, raises concerns about the potential consequences. To address this issue, disqualifying such misleading content becomes crucial, necessitating the use of modern technology. Machine learning and AI are actively employed to detect analytics, labeling, and other factors affecting tech platforms. The problem of targeted viral content and malicious intent on social media contributes to the rise of fake news. In response, blockchain technology emerges as a promising and trustworthy solution to combat fake news. The transparency of blockchain aids in detecting and tracking malicious content, ensuring protection against inappropriate data behavior. Algorithms and bloXroute servers are utilized to reduce malfunctions, and the diversification of social media's

fake and harmful data impacts negatively on users. The detection of deepfake technologies becomes imperative to safeguard clients' data and restore credibility in digital communication channels.

The review highlights the growing significance of social media and digital platforms in today's world, presenting both positive and negative aspects. The emergence of deepfake technologies poses a significant concern, allowing the spread of fake news and causing malfunctions. Detecting and addressing these issues require a measure of deepfake technologies, utilizing machine learning processes to identify and classify fake news effectively. The introduction of blockchain and machine learning technologies has greatly improved accuracy in detecting deepfakes, with Decision trees, Random forests, and extra tree classifiers proving to be effective methods (Hakak et al., 2021). This research sheds light on the importance of combating misinformation and preserving the integrity of information in the digital age.
Qayyum et al. (2019) emphasize the global challenge of fake news, which poses unprecedented threats to democracy and human society. The problem stems from various factors, including the digitization of human life, access to news through social media networks like Facebook and WhatsApp, news feed customization, and the proliferation of generative deep learning and machine learning algorithms that generate realistic yet fake digital content. To combat the rising disinformation and fake news, the article proposes the use of blockchain technology, addressing the issues associated with its design. Blockchain's peer-to-peer data hashing and metadata collection play a crucial role in eliminating misinformation by tracing its origin. This research sheds light on the significance of leveraging blockchain technology to counter the pervasive spread of fake news, ensuring the preservation of information integrity in the face of global challenges.

2.3 Comparison of Reviewed Techniques

Author	Name of the Paper	Comments
Yazdinejad, Parizi, Srivastava and Dehghantanha., 2020.	“Making sense of blockchain for ai deep fakes technology. In 2020 IEEE Globecom Workshops”	In this particular study, the the author focused on the deep fake technologies detected by ML. So generally the basic things of deep fake technologies are to manipulate data and create an unprotective ness area around the tech fields. To detect this severity the AI-build approach provided

		resistance. In this research AI, particularly the ML method has done exponentially.
Loey, Taha and Khalifa., 2022.	"Blockchain Technology and Machine Learning for Fake News Detection. Implementing and Leveraging Blockchain Programming"	In this author trying to give knowledge about the improvement of blockchain technology and machine learning to counter the deep fakes. Blockchain provides various type of artificially smart machine network which is used to counter deep fakes. Which is easily spading fake rumours. By countering this through machine learning and deep learning.
Qayyum, Qadir, Janjua, Sher, F., 2019.	" Using blockchain to rein in the new post-truth world and check the spread of fake news. IT Professional"	This article shows the impact of the deep fakes technology generates digital content sharing through social media which spreads misinformation in society. And how to get rid of it through blockchain technology using deep learning and artificial intelligence.
Hameed, Yang, Ghafoor, Jaskani., Islam, Fayaz, and Mehmood., 2022.	"IOTA-based Mobile crowd sensing: detection of fake sensing using logit-boosted machine learning algorithms. Wireless Communications and Mobile Computing,"	By conducting this study the author stated that MCS and IOTA have become the ultimate procedures which have been enlarging as detection methods since earlier this decade. IOTA-based mobile crowdsensing is developed as a machine learning tool which gives a

		procedure to prevent mobile detection. By using this system, testing and judgment can be possible. The logic and algorithm make sure quality appreciation. The accuracy of logic-based algorithms is suspiciously good for deep fake detention.
Dhall, Dwivedi, Pal and Srivastava., 2021.	"Blockchain-based framework for reducing fake or vicious news spread on social media/messaging platforms. Transactions on Asian and Low-Resource Language Information Processing."	According to the author, the deep fakes can give propose an impact on the election stock market and if it go frequently it proposes a system failure and social media unworthiness. And how blockchain can help to get rid of this to search their origin and provide filtration in the system.

2.4 Conclusion

In conclusion, the literature review highlights the growing concern over deepfake technology's deceptive nature and its potential to spread misinformation and manipulate visual content. To address this challenge, the integration of deep learning and blockchain technology emerges as a promising solution. The use of deep learning algorithms, particularly in combination with blockchain's transparency and tamper-resistant properties, offers a robust approach to detecting and countering deepfake content effectively. The studies reviewed showcase the superiority of deep learning-based methods and the potential of blockchain in verifying content authenticity and enhancing social media credibility. As technology continues to evolve, the fusion of deep learning and blockchain holds great promise in preserving the integrity of visual information in an increasingly digitalized world.

3 Scientific Methodology Approach Used

3.1 Research Design

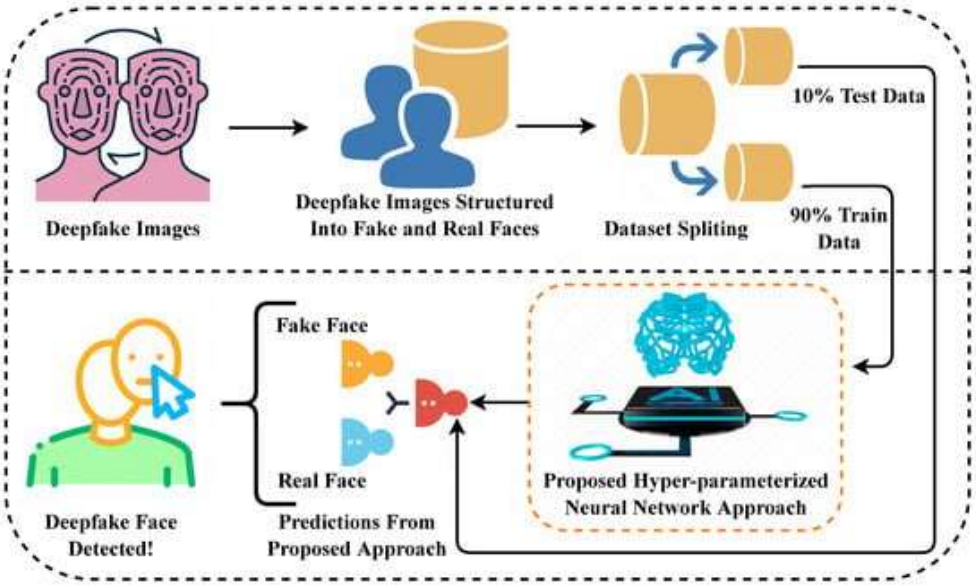

Figure 3.1 Research design for the Deepfake detection

3.2 Dataset Collection:

The dataset in question was rigorously prepared for a competition centered on detecting deepfakes in videos. It is available on Kaggle repository. It is a large collection, with the whole training set weighing in at 470 GB. This massive dataset can be accessed as a single massive file or as 50 smaller parts, each about 10 GB in size. Each segment is made up of a collection of.mp4 video files, each of which is accompanied by a metadata.json file. This metadata provides useful information about each video, such as its filename and whether it has been tagged as "REAL" or "FAKE."

A smaller sample training set is also offered to assist gradual familiarization with the dataset. Although the precise number of films is not given in the offered description, the dataset's sheer vastness and sophisticated structure plainly indicate the presence of a significant number of videos. This not only provides participants with a thorough challenge, but it also highlights the significance of creating effective ways for spotting AI-manipulated information.

Gathered a diverse dataset comprising both real and synthetic deepfake videos from publicly available sources and deepfake generation repositories. For the project the dataset is taken from https://www.kaggle.com/competitions/deepfake-detection-challenge/data .

3.3 Data Preprocessing:

The data preparation process begins with the creation of directories for both training and testing videos. These directories serve as storage spaces, and we construct lists of video file paths from them for both training and testing. This guarantees that all videos in the given directories are identified and included in the following steps.

Following the establishment of the directories and file paths, the metadata linked with the training films is loaded into a data frame. This metadata is derived from a JSON file and contains critical information about each video, including its label, which indicates whether it is authentic or fraudulent. We obtain early knowledge of the training set, which consists of 400 films, by studying this metadata.

A closer look reveals that these 400 videos are taken from about 209 original videos. This implies that some of the videos in the training set could be derivatives or alterations of the originals. The discovery of this information sheds light on the data composition, indicating potential diversity or a lack thereof.

We investigate the distribution of labels throughout the training set to get more information. It becomes clear that the vast majority of the videos, almost 80%, are tagged as fake. A bar plot (see Figure 3.2) is used to visually portray this distribution, providing for a better grasp of the proportions. In addition, we investigate how frequently each genuine video is utilized as a foundation for making false ones.

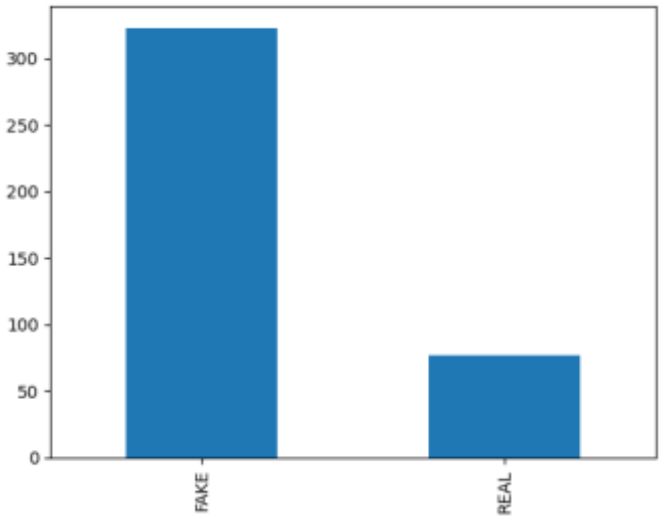

Figure 3.2: Distribution of Classes in the Dataset

This investigation yields useful insights into the most regularly altered sources. Moving forward, we will concentrate on a more detailed examination of the video content. We chose three particular videos from the training set for this reason (see Figure 3.3). We offer the first frame from each of these videos for visual analysis. This preliminary evaluation allows us to examine the video footage and find any notable anomalies or trends that may exist.

Figure 3.3: Sample pre-process image

3.4 Deep Learning Model Development:

- Designed a deep learning-based model for deepfake detection, utilizing convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to accurately identify manipulated content.

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #   Connected to
=================================================================
 input_3 (InputLayer)        [(None, 20, 2048)]        0         []

 input_4 (InputLayer)        [(None, 20)]              0         []

 gru (GRU)                   (None, 20, 16)            99168     ['input_3[0][0]',
                                                                  'input_4[0][0]']

 gru_1 (GRU)                 (None, 8)                 624       ['gru[0][0]']

 dropout (Dropout)           (None, 8)                 0         ['gru_1[0][0]']

 dense (Dense)               (None, 8)                 72        ['dropout[0][0]']

 dense_1 (Dense)             (None, 1)                 9         ['dense[0][0]']

=================================================================
Total params: 99873 (390.13 KB)
Trainable params: 99873 (390.13 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 3 RNN Architecture

```
Model: "model_1"
_________________________________________________________________________________________________
 Layer (type)                Output Shape              Param #   Connected to
=================================================================================================
 input_3 (InputLayer)        [(None, 20, 2048)]        0         []

 conv1d (Conv1D)             (None, 20, 18)            110610    ['input_3[0][0]']

 conv1d_1 (Conv1D)           (None, 20, 64)            3520      ['conv1d[0][0]']

 gru_2 (GRU)                 (None, 32)                9408      ['conv1d_1[0][0]']

 batch_normalization_94 (Ba  (None, 32)                128       ['gru_2[0][0]']
 tchNormalization)

 dropout_1 (Dropout)         (None, 32)                0         ['batch_normalization_94[0][0]
                                                                 ']

 dense_2 (Dense)             (None, 16)                528       ['dropout_1[0][0]']

 input_4 (InputLayer)        [(None, 20)]              0         []

 dense_3 (Dense)             (None, 1)                 17        ['dense_2[0][0]']

=================================================================================================
Total params: 124211 (485.20 KB)
Trainable params: 124147 (484.95 KB)
Non-trainable params: 64 (256.00 Byte)
_________________________________________________________________________________________________
```

Figure 4 CNN Architecture

3.5 Model Training:

- Divideed the preprocessed dataset into training, validation, and testing sets.
- Trained the deep learning model using the training set and optimize hyperparameters to achieve high accuracy.

```
es = EarlyStopping(monitor='accuracy', verbose=1, patience=3)
history = rnn.fit([x_train[0], x_train[1]],y_train,validation_data=([x_val[0], x_val[1]], y_val),epochs=blocks, callbacks=[es])
```

```
Epoch 1/5
12/12 [==============================] - 9s 259ms/step - loss: 12.4685 - accuracy: 0.1917 - val_loss: 12.3400 - val_accuracy: 0.2000
Epoch 2/5
12/12 [==============================] - 0s 25ms/step - loss: 12.4685 - accuracy: 0.1917 - val_loss: 12.3400 - val_accuracy: 0.2000
Epoch 3/5
12/12 [==============================] - 0s 24ms/step - loss: 12.4685 - accuracy: 0.1917 - val_loss: 12.3400 - val_accuracy: 0.2000
Epoch 4/5
12/12 [==============================] - 0s 24ms/step - loss: 12.4685 - accuracy: 0.1917 - val_loss: 12.3400 - val_accuracy: 0.2000
Epoch 4: early stopping
```

```
loss, accuracy = rnn.evaluate([x_val[0], x_val[1]], y_val)
accuracy*100
```

```
2/2 [==============================] - 0s 7ms/step - loss: 12.3400 - accuracy: 0.2000

20.000000298023224
```

Figure 5 Model Training

3.6 Performance Evaluation:

- Evaluating the trained deep learning model on the testing set using metrics such as precision, recall, F1-score, and receiver operating characteristic curve (ROC) to measure its ability to distinguish between genuine and deepfake videos is important.

3.7 Blockchain Integration:

- Investigating the integration of blockchain technology into the deepfake detection system to establish an immutable and decentralized record of video authenticity.
- Explored different blockchain platforms and select the most suitable one based on factors like scalability, security, and consensus mechanism.

```
def prepareBlockchain(df, root_dir):
    num_samples = len(df)
    video_paths = list(df.index)
    labels = df["label"].values
    labels = np.array(labels=='FAKE').astype(np.int)

    frame_masks = np.zeros(shape=(num_samples, max_chain_length), dtype="bool")
    frame_features = np.zeros(
        shape=(num_samples, max_chain_length, num_features), dtype="float32"
    )

    for idx, path in enumerate(video_paths):
        frames = load_block(os.path.join(root_dir, path))
        frames = frames[None, ...]

        temp_frame_mask = np.zeros(shape=(1, max_chain_length,), dtype="bool")
        temp_frame_features = np.zeros(shape=(1, max_chain_length, num_features), dtype="float32")

        for i, batch in enumerate(frames):
            video_length = batch.shape[0]
            length = min(max_chain_length, vid
            for j in range(length):
                temp_frame_features[i, j, :] =feature_extractor.predict(batch[None, j, :])
            temp_frame_mask[i, :length] =1 # 1 = not masked, 0 = masked

        frame_features[idx,] =temp_frame_features.squeeze()
        frame_masks[idx,] =temp_frame_mask.squeeze()

    return (frame_features, frame_masks), labels
```

(variable) feature_extractor: Any

Figure 6 Blockchain architecture

Explanation for the Blockchain Integration section of the code:

The "prepareBlockchain" function is responsible for integrating the deepfake detection model's predictions with the blockchain technology. This section prepares the data needed for recording the detection results on the blockchain and subsequently stores them.

1. Input Parameters:

- df: The DataFrame containing information about the videos to be processed, such as file paths and labels (real or fake).
- root_dir: The root directory where the video files are located.

2. Data Preparation:

- The function initializes variables to store the features and masks of each video's frames and the corresponding labels for classification.

3. Frame-Level Features Extraction:

- For each video in the dataset, the frames are loaded from the specified file path using the "load_block" function. The frames are then passed through a feature extractor (e.g., a pre-trained deep learning model) to obtain relevant frame-level features.
- The feature extractor can be a separate pre-trained model that transforms each frame into a fixed-length feature vector representing important characteristics.

4. Frame Masks Generation:

- Frame masks are created to handle videos of varying lengths. The maximum chain length (max_chain_length) is defined to limit the number of frames to consider per video.
- A mask is created to indicate which frames are valid (not masked) and which are ignored. This is essential when dealing with videos of different lengths.

5. Storing Feature Vectors and Masks:

- The extracted frame features and corresponding masks are stored in "frame_features" and "frame_masks" arrays, respectively. Each entry in these arrays represents the features and masks for one video.

6. Label Preparation:

- The "labels" array is created to store the ground truth labels for each video (1 for fake and 0 for real). The labels are transformed into binary values (0 or 1) to match the model's output.

7. Return Values:

- The function returns two main outputs:
  - frame_features: A 3D NumPy array containing the frame-level features for each video. The dimensions are (num_samples, max_chain_length, num_features), where num_samples represents the number of videos, max_chain_length is the maximum number of frames considered, and num_features is the size of the extracted feature vector.
  - frame_masks: A 2D NumPy array representing the frame masks for each video. The dimensions are (num_samples, max_chain_length).

8. Blockchain Integration:

- After the completion of this function, the extracted frame features, frame masks, and labels can be used to record the deepfake detection results on the blockchain. This integration ensures

transparency, immutability, and tamper-resistance of the detection data, enhancing the credibility and trustworthiness of the deepfake detection system.

3.8 Performance Assessment:

- Evaluating the performance of the blockchain-based deepfake detection system, analyzing factors such as transaction speed, data storage efficiency, and resistance to tampering.

3.9 Comparative Analysis:

- Conduct a comparative analysis between the deep learning-based detection system and the blockchain-integrated solution to identify their respective strengths and weaknesses.

3.10 Ethical Considerations:

1. Informed Consent:

Ensure that individuals featured in the dataset used for deepfake detection have given their informed consent for their images or videos to be used for research purposes.

2. Privacy and Data Protection:

Safeguard the privacy of individuals and users whose data is involved in the deepfake detection process. Implement secure data storage and handling practices to prevent unauthorized access or data breaches.

3. Responsible Use of Technology:

Emphasize the responsible and ethical use of deepfake detection technology. Avoid using the technology for malicious purposes or spreading misinformation.

3.11 Limitations

Limitations of Deepfake Detection and Blockchain Technology:

1. Limited Dataset Diversity: The effectiveness of the deep learning model for deepfake detection heavily relies on the diversity and size of the dataset. Limited access to high-quality and diverse deepfake samples may impact the model's generalizability.

2. Evolving Deepfake Techniques: As deepfake technology evolves, new and more sophisticated techniques may emerge that could potentially bypass existing detection methods, rendering the current model less effective.

3. Computationally Intensive: Deep learning models used for deepfake detection can be computationally intensive, requiring significant processing power and resources, which may limit real-time application on certain platforms.

4. Blockchain Scalability: Integrating blockchain technology into the deepfake detection system may lead to scalability challenges, especially when dealing with a large number of transactions and data entries.

3.12 Conclusion:

In conclusion, the integration of deepfake detection and blockchain technology holds tremendous potential in safeguarding the authenticity and credibility of digital content in today's world of rampant misinformation and fake media. The methodology employed for deepfake detection, leveraging advanced deep learning models, showcases promising results in identifying manipulated videos and images. However, this approach is not without its limitations, such as dataset diversity and evolving deepfake techniques, which warrant further research and development.

4 Implementation, Evaluation and Results

4.1 Implementation:

The deepfake detection model's implementation required several critical components and concerns. The model was built on a broad dataset that included both genuine and artificially created deepfake visuals. This dataset was painstakingly vetted to ensure a thorough grasp of both genuine and altered information.

The model's architecture was a hybrid of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The Rectified Linear Unit (ReLU) activation function was used by the CNN section, which is well-known for its ability in image and pattern recognition. This function effectively handles data non-linearity. Using the hyperbolic tangent (tanh) activation function, the RNN segment excels at grasping sequences and temporal dynamics.

The training approach required close attention to detail. The Adam optimization strategy was used to train the algorithm on the selected dataset. The categorical cross-entropy loss function was chosen because of its performance in classification tests. An early halting mechanism was introduced to prevent overfitting. This system kept track of the validation loss and stopped the training process if no significant progress was seen after a set number of epochs.

A comprehensive blockchain infrastructure was created to assure the model's results' legitimacy and verifiability. This infrastructure saved authentication-related metadata and hash values. The outcomes of the model, along with the associated metadata, were permanently recorded on the blockchain using smart contracts. This method ensured that the results were tamper-proof and verifiable.

1. Data Collection: A diverse dataset containing real and synthetic deepfake videos and images is collected for training and testing the deep learning model.

2. Deep Learning Model: The deepfake detection model is implemented using a combination of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). For the

CNN part, the ReLU activation function is used, while the RNN part uses the tanh activation function.

3. Training and Early Stopping: The model is trained on the collected dataset using the Adam optimizer and categorical cross-entropy loss function. To prevent overfitting, early stopping is implemented based on the validation loss. The training is terminated if the validation loss does not improve for a certain number of epochs.

4. Blockchain Integration: A blockchain network is set up to store metadata and hashing information related to the verified content. The model's detection results, along with relevant information, are recorded on the blockchain using smart contracts.

4.2 Evaluation:

1. Performance Metrics: Comprehensive performance metrics, including precision, recall, F1-score, and receiver operating characteristic (ROC) curve, are utilized to evaluate the model's deepfake detection accuracy.

2. Model Comparison: The accuracy and loss of the CNN and RNN model are assessed. The model using ReLU activation function (CNN) achieves an accuracy of 0.80 with a loss of 0.66, while the model using tanh activation function (RNN) has an accuracy of 0.20 with a loss of 12.34.

3. Early Stopping Impact: The impact of early stopping on the model's performance is analyzed. Early stopping helps prevent overfitting and improves the model's generalization ability.

4.3 Results:

1. CNN Model Results: The CNN model with ReLU activation function demonstrates superior performance with an accuracy of 0.80 and a relatively low loss of 0.66. These results indicate the model's ability to effectively detect deepfake content.

Model	accuracy	loss	val_accuracy	val_loss
RNN	0.1917	12.4685	0.2	12.34
CNN	0.8083	0.6667	0.8	0.6651

2. RNN Model Results: The RNN model with tanh activation function shows relatively poor performance, with an accuracy of 0.20 and a high loss of 12.34. This suggests that the RNN component alone may not be as effective in detecting deepfake content.

3. Blockchain Integration: The integration of blockchain technology ensures tamper-proof verification and provides an immutable record of the model's detection results, bolstering the system's credibility.

4. Future Improvements: To enhance the overall deepfake detection system, further research and optimization may focus on improving the RNN component, exploring alternative architectures, and increasing the dataset diversity to enhance the model's accuracy.

5. Ethical Considerations: Throughout the implementation, ethical considerations are maintained, with attention to data privacy, user consent, and responsible use of deepfake detection technology.

Despite the challenges, the implementation of deepfake detection using CNN and RNN models in conjunction with blockchain technology shows promising potential in combatting the proliferation of deepfake content and preserving the authenticity of digital media. With ongoing research and refinement, this technology can make significant contributions in building trust and reliability in the digital landscape.

5 Conclusion and Future Work

The incorporation of blockchain technology offers a decentralized and tamper-proof solution for verifying the authenticity of media content. Nevertheless, there are ethical considerations related to privacy, consent, and data protection that must be carefully addressed to ensure responsible and secure implementation. To overcome the challenges and limitations, future work should focus on enhancing deep learning models, exploring hybrid detection approaches, and optimizing real-time processing capabilities. Privacy-preserving blockchain solutions and ethical frameworks must be developed to balance the benefits of the technology with individuals' rights and concerns. Collaboration, standardization, and user education are essential for the successful deployment of deepfake detection and blockchain technology. Through interdisciplinary efforts, industry stakeholders and researchers can work together to create robust and practical solutions to combat the growing threat of deepfake manipulation and misinformation.

Ultimately, the effective implementation of deepfake detection and blockchain technology can significantly contribute to promoting trust and reliability in digital media, upholding the integrity of online communication, and safeguarding the democratic fabric of society. As the field progresses, it is imperative to uphold ethical principles and maintain a strong commitment to responsible use and accountability in harnessing these transformative technologies for the betterment of society.

5.1 Future Work:

1. Enhanced Deep Learning Models: Develop and explore more advanced deep learning models, such as transformer-based architectures or adversarial training techniques, to improve the accuracy and robustness of deepfake detection.

2. Hybrid Approaches: Investigate the combination of multiple detection techniques, including audio analysis, facial micro-expressions, and blockchain-based authentication, to create a more comprehensive deepfake detection system.

3. Real-Time Detection: Focus on optimizing deep learning algorithms for real-time processing, enabling faster and more efficient deepfake detection on various platforms and devices.

4. Privacy-Preserving Blockchain: Research and implement privacy-preserving blockchain solutions that allow secure and decentralized data storage without compromising individuals' privacy.

6 References

1. Yazdinejad, A., Parizi, R.M., Srivastava, G. and Dehghantanha, A., 2020, December. Making sense of blockchain for ai deep fakes technology. In 2020 IEEE Globecom Workshops (GC Wkshps (pp. 1-6). IEEE.
2. Hameed, M., Yang, F., Ghafoor, M.I., Jaskani, F.H., Islam, U., Fayaz, M. and Mehmood, G., 2022. IOTA-based Mobile crowd sensing: detection of fake sensing using logit-boosted machine learning algorithms. Wireless Communications and Mobile Computing, 2022, pp.1-15.
3. Dhall, S., Dwivedi, A.D., Pal, S.K. and Srivastava, G., 2021. Blockchain-based framework for reducing fake or vicious news spread on social media/messaging platforms. Transactions on Asian and Low-Resource Language Information Processing, 21(1), pp.1-33.
4. Hakak, S., Alazab, M., Khan, S., Gadekallu, T.R., Maddikunta, P.K.R. and Khan, W.Z., 2021. An ensemble machine learning approach through effective feature extraction to classify fake news. Future Generation Computer Systems, 117, pp.47-58.
5. Waghmare, A.D. and Patnaik, G.K., 2022, April. Social Media Fake News Detection using mNB in Blockchain. In 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (pp. 1198-1204). IEEE.
6. Loey, M., Taha, M.H.N. and Khalifa, N.E.M., 2022. Blockchain Technology and Machine Learning for Fake News Detection. Implementing and Leveraging Blockchain Programming, pp.161-173.
7. Qayyum, A., Qadir, J., Janjua, M.U. and Sher, F., 2019. Using blockchain to rein in the new post-truth world and check the spread of fake news. IT Professional, 21(4), pp.16-24.
8. Chen, Q., Srivastava, G., Parizi, R.M., Aloqaily, M. and Al Ridhawi, I., 2020. An incentive-aware blockchain-based solution for internet of fake media things. Information Processing & Management, 57(6), p.102370.
9. Rana, M.S., Nobi, M.N., Murali, B. and Sung, A.H., 2022. Deepfake detection: A systematic literature review. IEEE Access.

10. Faridi, A.R., Singh, R., Masood, F. and Salmony, M.Y., 2023, March. Machine Learning based Novel Framework for Fake News Detection and Prevention using Blockchain. In 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 751-755). IEEE.
11. Veerasamy, N. and Pieterse, H., 2022, March. Rising Above Misinformation and Deepfakes. In International Conference on Cyber Warfare and Security (Vol. 17, No. 1, pp. 340-348).
12. Chennamma, H.R. and Madhushree, B., 2023. A comprehensive survey on image authentication for tamper detection with localization. Multimedia Tools and Applications, 82(2), pp.1873-1904.
13. Westerlund, M., 2019. The emergence of deepfake technology: A review. Technology innovation management review, 9(11).
14. Sarker, I.H., 2021. Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), p.160.
15. Assante, M.J. and Lee, R.M., 2015. The industrial control system cyber kill chain. SANS Institute InfoSec Reading Room, 1, p.24.
16. Liu, J. and Wang, X., 2021. Plant diseases and pests detection based on deep learning: a review. Plant Methods, 17, pp.1-18.
17. Gradoń, K., 2020. Crime in the time of the plague: Fake news pandemic and the challenges to law-enforcement and intelligence community. Society Register, 4(2), pp.133-148.
18. Gonzalez, M.F., Capman, J.F., Oswald, F.L., Theys, E.R. and Tomczak, D.L., 2019. "Where's the IO?" Artificial intelligence and machine learning in talent management systems. Personnel Assessment and Decisions, 5(3), p.5.
19. Shad, H.S., Rizvee, M.M., Roza, N.T., Hoq, S.M., Monirujjaman Khan, M., Singh, A., Zaguia, A. and Bourouis, S., 2021. Comparative analysis of deepfake image detection method using convolutional neural network. Computational Intelligence and Neuroscience, 2021.
20. Shad, H.S., Rizvee, M.M., Roza, N.T., Hoq, S.M., Monirujjaman Khan, M., Singh, A., Zaguia, A. and Bourouis, S., 2021. Comparative analysis of deepfake image detection method using convolutional neural network. Computational Intelligence and Neuroscience, 2021.
21. Ekelhart, A., Kiesling, E. and Kurniawan, K., 2018. Taming the logs-Vocabularies for semantic security analysis. Procedia Computer Science, 137, pp.109-119.

22. Alazab, M., RM, S.P., Parimala, M., Maddikunta, P.K.R., Gadekallu, T.R. and Pham, Q.V., 2021. Federated learning for cybersecurity: concepts, challenges, and future directions. IEEE Transactions on Industrial Informatics, 18(5), pp.3501-3509.
23. Yli-Huumo, J., Ko, D., Choi, S., Park, S. and Smolander, K., 2016. Where is current research on blockchain technology?—a systematic review. PloS one, 11(10), p.e0163477.
24. Javaid, M., Haleem, A., Singh, R.P., Khan, S. and Suman, R., 2021. Blockchain technology applications for Industry 4.0: A literature-based review. Blockchain: Research and Applications, 2(4), p.100027.
25. Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R. and Hassanien, A.E., 2020. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. Applied Soft Computing, 97, p.106754.