Blockchain-Powered Authentication: Reinforcing Identification for Misinformation in AI-Generated Videos

Kasun Moolikagedara, Wei Qi Yan, Minh Nguyen

Department of Computer and Information Sciences

Auckland University of Technology, 1010 New Zealand

Abstract— The advent of artificial intelligence (AI) has revolutionized multiple aspects of digital content creation and manipulation, enabling sophisticated techniques for generating and editing videos. While AI-driven advancements offer numerous benefits, they also introduce challenges related to the authenticity and trustworthiness of digital media. The rapid dissemination of manipulated videos, deepfakes, and other forms of synthetic media has fuelled concerns about the spread of misinformation and its detrimental effects on society. This book chapter explores the integration of blockchains with AI video authentication to combat misinformation. Our approach leverages blockchain's immutable ledger and AI algorithms to verify content integrity, providing a robust solution against synthetic media threats. Our results show enhanced detection accuracy compared to the existing methods.

Key Words: Video Blockchain, AI-Generated Video, Intelligent Surveillance, Authentication, Misinformation

1. INTRODUCTION

In this chapter, we explore the potential synergy between video blockchain (Gedara, Nguyen, & Yan, 2023), (Moolikagedara et al., 2023), and AI-generated video identification systems to address the proliferation of misinformation in digital media. By leveraging blockchain’s decentralized and immutable ledger, coupled with cryptographic algorithms for video analysis. we enhance the authenticity and trustworthiness of digital videos. Our objective is to develop a robust framework that can accurately verify the origin and integrity of video content, thereby enabling users to make informed decisions and combat the spread of misinformation effectively.

One of the most pressing concerns in the digital age is the proliferation of misinformation and disinformation, especially through manipulated videos. The ability to create highly realistic but fabricated videos, known as deepfakes has made it increasingly difficult to discern truth from falsehood in digital content. The rapid spread of such misleading videos has serious implications for society, ranging from political manipulation to reputational damage and privacy breaches (Hu & Yan, 2020), integrity of digital content in the face of evolving technological capabilities. By exploring the integration of blockchain and AI, this project seeks to contribute to the efforts to combat misinformation and ensure the reliability of digital media in the digital age. There is a critical need to address the escalating issue of misinformation propagated through AI-generated videos. As AI advances, it becomes increasingly challenging to distinguish between authentic and fabricated content.

In response to these challenges, there is a growing need for innovative solutions for effectively authenticating video content and mitigate the risks associated with misinformation. Traditional methods of verifying video authenticity, such as digital watermarking or digital signatures are often insufficient in the face of AI-generated manipulation. This has led to explore new approaches that leverage cutting-edge technologies to address this complex issue (Wang & Liao, 2021).

One of the modern approaches is the integration of video blockchains with AI-generated video identification. Blockchain, a decentralized and immutable ledger technology, offers a promising solution for establishing the provenance and integrity of digital assets including videos. By combining blockchain's capabilities with cryptographic algorithms for video analysis, we aim to create a robust framework for authenticating video content and combating misinformation effectively. Through empirical testing and validation, the outcomes demonstrate the efficacy of this approach in mitigating the risks posed by manipulated videos and deep fakes (Mi et al., 2022). By exploring the integration of blockchain and cryptographic techniques, this project seeks to contribute to the efforts to combat misinformation and ensure the reliability of digital media in the digital age.

Overall, this research project is motivated by the urgent needs to develop innovative solutions that can safeguard the by leveraging video blockchain-powered authentication, we reinforce the identification of AI-generated videos, thereby enhancing the credibility and reliability of visual content. This approach offers a promising solution to combat misinformation, safeguarding the integrity of information shared online. Through this chapter, we seek to explore the potential of blockchain in mitigating harmful effects of misinformation, contributing to a more trustworthy digital landscape. With the introduction of AI-generated videos, text-to-video by using OpenAI. This problem has again become the main topic for most of the parties. Additionally, most of the influencers looked at the effect that will be going to happen with this new implementation. Our video blockchain has addressed this kind of problems by combining multiple methods and cryptographic functions (Moolikagedara et al., 2023). It aims to address this problem that is affecting globally. At the moment, we are finding out how to deal with AI-generated content differently. Video blockchain has been employed to secure video files generated in digital surveillance for the smart cities. In addition, it can be implemented by combining appropriate methods to address this misinformation problem.

In this book chapter, we take use of our proposed algorithm to resize and compare the process in the Video Blockchain(VB) that ensures the integrity and authenticity of the AI-generated video content by creating a tamper-proof record by using video blocks. It makes use of a Merkle tree to construct the blockchain for each frame and computes the root hash as a unique identifier for the frame. Using the Block Matrix (BM) algorithm we proposed in this chapter, we compress the blocks of each frame for efficient storage and retrieval.

The Bucketisation algorithms hash the frames of the video into buckets, create a hashed representation of the video content, and combine the results from the Super-bit LSH (Super-bit Locality-Sensitive Hashing) hashing, blockchain verification, and block matrix analysis to determine if the video is likely to be AI-generated misinformation.

The results of this book chapter demonstrate the efficacy of the proposed blockchain-powered authentication framework in enhancing AI-generated video identification against misinformation. By leveraging blockchain's immutable ledger, we effectively track the provenance of video content and verify its authenticity, thereby mitigating the risks posed by manipulated videos and deepfakes. Furthermore, the integration of cryptographic algorithms enables automated detection and analysis of suspicious content and enhances the overall security and reliability of the system. Consequently, this research work has the following contributions:

- We introduce the new algorithms by combining video blockchain with a block matrix.
- LSH Bucketisation connects with the improved video blockchain and block matrix.
- We present the analysis of theoretical effectiveness comparing this method with preset research works.
- We present the experimental testing results based on our selected datasets.

This chapter has been organized as follows: We present the review of the related work in a similar capacity, also covering blockchain and block matrix to identify the current stage of the related work. In Section 3, we describe the proposed method of resolving the problem related to misinformation in AI- generated video content. In Section 4, we exhibit our experimental results and analysis.

2. Related work

In this section, we conduct a comparative review of the state-of-the-art approaches to AI-generated video identification and visual misinformation detection. In addition, we highlight the problems these solutions addressed to date. This ensures the best solution for our research problems.

The State-of-the-art blockchain for video identification

The decentralized nature of blockchain is a key attribute that ensures the verification and re-verification of each transaction. This intrinsic feature is not confined solely to currency-based implementations but extends to a diverse array of applications. The paramount advantage of this decentralized ledger lies in its ability to maintain the immutability of data once it has been appended to the blockchain.

The particularly advantageous in the context of combating misinformation propagated through AI-generated videos. Despite the clear advantages, it is important to note that a standardized mechanism for effectively addressing misinformation using blockchain has yet to be universally established. However, in comparison to other available solutions, blockchain emerges as one of the most viable and effective options available at present.

To ensure the integrity of the recorded videos, the unique features of the distributed and tamper-proofing characteristics in blockchain have been employed. In blockchain, timestamping features are applied to verify and transfer unaltered data to a distributed repository. Similarly, the captured data from a closed- circuit television (CCTV) camera in smart cities have been further explored.

The blockchain-based system can guarantee that the recorded data is stored without being altered or tampered with. It avoids manipulating the data integrity of original videos. Because a distributed ledger of blockchain records the data with metadata, the CCTV system will assist law enforcement and clients to secure data from surveillance.

BleddSPS is a public safety system with a decentralized secure architecture, which supports immutability, audibility, and traceability to ensure smart city safety. SD-IoD is the software-defined Internet architecture that makes use of smart contractor and blockchain to secure real-time monitoring systems by using drones. Maintaining the correct order of the recording data can ensure reliability and integrity.

The processing time for visual data, as well as resistance, is evaluated in computational methods. According to the analysis, video blockchains enable smart cities to continue the operations without a single point of failure. In this book chapter, computational methods of visual blockchains are taken into consideration, a new prototype is created for visual data storage by using blockchains, and all visual data is linked together by using a blockchain with a decentralizing or flattened method.

Securing surveillance is a crucial for face detection, human behaviour analysis, and traffic rule violation detection. These tools have shown significant contributions in preventing crimes, anomalous incidents, and privacy policy violations. Our previous work related to blockchain and computer vision leads to more robust method to address malicious attackers and hackers can illegally manipulate video repositories and surveillance cameras, thereby rendering the recorded footage unusable in criminal cases.

Attackers may manipulate or tamper with video footage, which leading to compromised integrity. Tamper-resistant and immutable blockchain features are employed to protect stored data and ensure data integrity.

Hashing is a reliable method for creating confidentiality between two blocks in the chain. Cryptographic hash functions convert confidential data into a random string of fixed size. Security requirements such as one-witness and collision-resistance are necessary.

Misinformation related to AI-generated videos

With the implementation of OpenAI Sora system, the topic of AI-generated content has rapidly spread as a means to address the growing problem of misinformation, especially with the advent of text-to-video AI technology. One of the primary challenges identified in this context is the difficulty in distinguishing between real videos and AI-generated videos. Our naked eyes alone often struggle to make this distinction, requiring considerable efforts to discern the genuine from the fabricated.

However, a number of cues, such as anomalies or inconsistencies, can tip off viewers to the presence of AI- generated content. The familiar items or elements that appear slightly off or unusual may be easier for the human eye to detect. To combat this challenge, classification techniques have been employed, utilizing improved algorithms.

The impact of misinformation spread through videos can be severe, leading to negative consequences for a wide range of stakeholders (Wang & Liao, 2021). In the work on combating online misinformation videos, researchers have characterized the problem, identified detection methods, and proposed future directions for addressing this issue. The research categorized misinformation into three levels based on characterization: Signal, semantic, and intent.

However, it is important to note that while this research work which represents a significant advancement in the field of misinformation detection may be outdated due to the continuous improvement of the AI technologies. These advancements highlight the need for ongoing research and development in this area to stay ahead of evolving challenges posed by AI-generated content.

Revolution of the super-bit LSH bucketization

LSH is a method in data mining and machine learning to efficiently calculate approximate nearest neighbours in high-dimensional spaces. This concept combines super-bit hashing and LSH, aiming to map similar points to the same buckets with high probability. This bucketization method enables a fast and accurate search for the nearest neighbours.

$$R = \frac{2 \times \log(\frac{1}{a})}{\log(\frac{1}{w})}, \ w>0, \ a>0 \qquad (1)$$

where R represents the number of revolutions needed to achieve the desired recall. The parameter b corresponds to the number of bits in the hash, while α is the approximation factor, and w represents the width parameter. This shows a way to estimate the number of revolutions required for the Super-bit LSH Bucketization to achieve a specific level of recall.

The Super-bit LSH has been widely employed in various research fields, including distributed frameworks, automatic processing, image and video processing, blockchain, biological sciences, and geological sciences. This innovative approach, combining super-bit hashing and LSH, has significantly impacted these fields.

While considering the connection between blockchain technology and LSH, this method has shown promise in enhancing blockchain processes. It is applied to improve data storage, retrieval, and verification in blockchain systems, contributing to the advancement of blockchain technology. Furthermore, it was designed and employed for audio signal processing, digital image / video processing, text / documentation processing (Wang & Liao, 2021), and biological sciences.

In blockchain, this has been employed as image copyright protection and provided the copyright over the network under distribution constraints. Thus, it was employed with blockchain to find out the copyright ownership by following image metadata.

In this section, we provide a comparative review of the current state of AI-generated video identification to combat misinformation. It contrasts the problem identification and solution breakdown to date, ensuring the selection of the best solution for research problems. In this section, we highlight its decentralized nature and ability to maintain data immutability, particularly beneficial in combating misinformation through AI-generated videos.

3. Methodology

In this section, we will elucidate into the proposed method for enhancing the future of Video Blockchain by adding Block Matrix to enrich the AI-generated video misinformation detections.

Overview of the video blockchain against misinformation

Previously, while discussing video blockchain, it is initially introduced as a method to secure surveillance videos. However, its application has been expanded to include securing IoT device data, and autonomous vehicle data and ensuring the privacy protection of data added to the blockchain-distributed network. These advancements highlight the versatility and adaptability of blockchain technology in addressing various security and privacy concerns across different domains.

- The execution time for each iteration is calculated: $t_{exec,i} = t_{end,i} - t_{start,i}$.

  The execution times are grouped into bins: $\{B_1, B_2, \ldots, B_k\}$.
- The frequency of the execution times is counted in each bin, $f_j = count\{t_{exec,i} \in B_j\}\ for\ j = 1,2 \ldots, k$.
- The histogram is represented by the set of frequency corresponding to each bin.

In our video blockchain, we selected the SHA256 by following the same process we used to select the Merkle three for video blockchain. The selection of cryptographic function was carried out by using Multi-Criteria Decision Making (MCDM). Future more, we have shown the Merkle tree is the most suitable data structure.

We combine the Schnorr signature, Merkle tree, and Block Matrix for detecting AI-generated videos, the key steps are proposed. Let V be a video file with n frames, each divided into m fixed-size blocks B_{ij}, where $1 \leq i \leq n$ and $1 \leq i \leq m$.

Firstly, the Schnorr signature scheme is employed for video authentication. This scheme involves a private key K and a public key P. Each block is signed using K to produce signature σ_{ij}, which is verified by using P.

Next, a Merkle tree is constructed for each frame to ensure data integrity. A cryptographic hash function H (e.g., SHA-256) is applied to each block to calculate its hash These hashes are employed to build the Merkle tree for frame , where leaf nodes represent block hashes and internal nodes represent the hash of their children. The root hash of serves as a unique identifier for .

Finally, Block Matrix is employed to organize and compress the blocks of each frame for efficient storage and retrieval. By combining this blockchain method, a secure and efficient method is established for detecting AI-generated videos. The Block Matrix can be harnessed to analyse the video frames for patterns or anomalies indicative of AI-generated content, enhancing the overall security and integrity of the video data

Empowering Video Blockchain for Misinformation

As video blockchain and AIGV are described in the previous section, this method is combined with Block Matrix to enhance the feature of the video blockchain method. In our published papers (Gedara, Nguyen, & Yan, 2023)(Moolikagedara et al., 2023), we have adopted Block Matrix to secure privacy. and implement a secure and suitable autonomous vehicle video network In the Block matrix, Data Block Matrix, denoted as M, is a $n \times m$ matrix where each element represents a block   of the video frame .

$$M = (\mathrm{B}_{ij})_{nxm} \tag{2}$$

where it comes to the video blockchain method, let H be a cryptographic hash function (SHA-256) to calculate the hash of each block. The Video Blockchain, denoted as B, is a chain of hashes where each block contains the hash of the previous block. Let be the initial hash value (e.g., the hash of the first block of the first frame). The Data Block Matrix M organizes the blocks of each frame into a structured matrix for efficient storage and access.

***Algorithm**: AI-Generated Video Misinformation Detection*

***Input**: A video file V consisting of frames.*

***Output**: A binary value indicating whether the video is likely to be AI-generated misinformation.*

***Block Matrix Algorithm**:*

- *Divide each frame of the video into fixed-size blocks (16x16 pixels).*
- *Store the blocks of each frame in a matrix, where each row represents a block, and each column represents a frame.*
- *Apply compression algorithms (JPEG) to each block to reduce data size.*
- *Store the compressed block matrix as a binary file.*

***Super-bit LSH Bucketisation Algorithm**:*

- *Generate a Super-bit structure S using the Generate Super-bit Structure function.*
- *Loop through the dataset records:*
    - *Select the next record π.*
    - *Compute hashes for the record using the Super-bit structure: h_i*

$ComputeHashesRecord$ (π, S)

- *Map the computed hashes into buckets $H = MapSignatures$ ($h_1, h_2, \ldots, h_p, S$).*

***Detection**:*

- *Analyse the compressed block matrices and the mapped hashes to detect patterns or anomalies indicative of AI-generated content.*

***Output**: Return a binary value indicating whether the video is likely to be AI-generated misinformation (1) or authentic content (0).*

In eq. (2), The Video Blockchain B ensures the integrity and authenticity of the video content by creating a tamper-proof record of the video blocks. Together, the Data Block Matrix data structure and the Video Blockchain method provide a comprehensive approach to organizing, storing, and verifying video content, enabling the detection of AI-generated video misinformation.

LSH Bucketisation is a method for hashing high-dimensional data into buckets for efficient approximate nearest neighbor search. The algorithm takes as input a Dataset D, the number of stages n stages, the number of buckets n buckets, and the data dimensionality data dimensionality. It generates a super-bit structure S using the Generate Super-bit Structure-function π and computing hashes for the record using the super-bit structure S. These hashes are then mapped into buckets using the map-signatures function, forming the mapped set H, which is returned as the output. The algorithm utilizes the super-bit LSH to efficiently hash high-dimensional data and enable fast nearest neighbor search operations.

This algorithm leverages the Block Matrix algorithm to efficiently store and access video data and the Super-bit LSH Bucketisation algorithm to hash high-dimensional data, enabling efficient detection of AI-generated content. The final step involves using advanced analysis to classify the videos based on the detected patterns or anomalies.

Empowering Video Blockchain for Misinformation

In this section, we explore the improvement of our final findings. By referring to Algorithm 1, we demonstrate the individual methods included in Algorithm 1's overall process for AIGV (AI-generated video) misinformation detection using video blockchain. Firstly, we utilize video frames where resizing may occur. We define *V* as the input video composed of frames Fi was ranges from 1 to, with being the total number of frames in the video. Each frame F_i is divided into blocks of size pixels. In this paper, we choose an 8x8 block.

$$B_{ij} = Block\ (F_{i,\,j})\ ,\ \ i \in [1, n], j \in [1, k] \tag{3}$$

We store these blocks in a matrix **M**,

$$M_{jk} = B_{kj}\ ,\ \ j \in [1, k], k \in [1, n] \tag{4}$$

In eq. (3), the process starts with the input video V, which consists of frames, where $i \in [1, n]$ and n is the total number of frames. This process leads to dividing each frame into given pixel block sizes. The process of eq. (5), leads to compression a compression algorithm (JPEG) to each block. It initiates the same value to the next process to enhance the process of video detection.

$$C_{ij} = Compress\ (B_{ij})\ , i \in [1, n], j \in [1, k] \tag{5}$$

$$C = BinaryFile\ (C_{ji}) \tag{6}$$

In the process of compression, each block B_{ij}, and then C_{ij} represents the compressed version of block B_{ij}. Furthermore, eq.(6) processes the C_{ij} to Binary file to enhance the space efficacy and speed of access, it allows faster operation when it's handling a large number of compressed blocks during processing.

$$T_{ij} = ExtractVisualImages\ (B_{ij}\ )\ , i \in [1, n], j \in [1, k] \tag{7}$$

$$BCL = BlockChainLink\ (T_{ij}\ ) \tag{8}$$

This is the main stage of the AI-generated video detection by using Video Blockchain Method with improved function. We see that eq. (7) follows the creation of the token set and blockchain linking process. Also, its extracts visual image from blocks B_{ij} to create a token set, in eq. (8), T_{ij} Video Blockchain is applied to link and identify the tokens.

$$S = GenerateSuperBitStructure \tag{9}$$

For each compressed block matrix record π:

$$h_i = ComputeHashesRecord\ (\pi, S), for\ i \in\ [1, p] \tag{10}$$

It sees that eq. (9) and eq. (10) lead to the generating super bit structure to involves creating a set of hyperplanes in the feature space.

In eq. (10), the hash function is applied to the feature vectors of the compressed block matrix M. These hash functions map high-dimensional data points into a lower-dimensional space, preserving the locality of similar items. The main requirement of this process is to ensure that similar blocks are mapped to the same or nearby buckets.

$$H = MapSignatures\ (h1, h2, h3, \ldots., hp) \tag{11}$$

The final step of eq. (11), we achieve the requirement of this process by computing hash codes are computed for each compressed block matrix record, which need to be mapped to specific buckets. Buckets H serve as containers for storing similar items together.

Moreover, we summaries the process of all eq. (3) to eq. (11) for better understating S by using a predefined function. Then, it iterates through the dataset records, selecting each record π in turn. For each record, it computes hashes hi using the Super-bit structure.

$$result = Classify\ (C, T, H) \tag{12}$$

The stage of detection and classification has been the major part of this whole process. The compressed block matrices C, token sets T, and mapped hashes H are analysed for patterns or anomalies indicating AI-generated content.

Moreover, in eq. (11), the algorithm analyses the compressed block matrices and the mapped hashes to identify patterns or anomalies that may indicate AI-generated content. This analysis is crucial for distinguishing between authentic and AI-generated videos. Finally, the algorithm is use of statistical.
We know that eq. (12) is a binary value indicating the likelihood of the video being AI-generated misinformation (1) or authentic content (0). The key equations in the algorithm involve computing hashes for each record and mapping these hashes into buckets for analysis, providing a structured approach to detecting AI-generated video misinformation.

In this study, we outlined a robust method for detecting AI-generated misinformation in videos. Our approach encompasses a multi-step process, starting with the division and compression of video frames, followed by the extraction and linking of visual tokens using video blockchain. Additionally, we employ Super-bit Locality Sensitive Hashing (LSH) for efficient Bucketisation of computed hashes, methods to classify the video as either likely to be AI-generated misinformation or authentic content based on the analysis results.

Hence, we understand that eq. (12) is a binary value indicating the likelihood of the video being AI-generated misinformation (1) or authentic content (0). The key equations in the algorithm involve computing hashes for each record and mapping these hashes into buckets for analysis, providing a structured approach to detecting AI-generated video misinformation.

4. Our experiments

In this section, we present our experimental results and provide the comparative experiment analysis to better understand the performance of our implementation. We are use of metrics to evaluate the proposed video Blockchain algorithm for AI-generated video (AIGV) detection. Our results are divided into three main sections: Experimental settings, detection performance, and comparisons in detecting misleading AI-generated videos.

Datasets and Evaluation Metrics

To evaluate the AI-generated misinformation detection algorithm, we utilize four diverse and challenging datasets to assess the performance and generalization capabilities of our deepfake detection methods: WildDeepfake, Celeb-DF, and DeeperForensics-1.0.

WildDeepfake (Heidari et al., 2024) is a dataset is designed for real-world deepfake detection. It contains sequences extracted from 707 deepfake videos collected from verified internet sources. Although it is smaller compared to other two datasets, WildDeepfake is harnessed to initiate our evaluation process due to its real-world relevance.

DeeperForensics-1.0 (Korshunov & Marcel, 2022) is a dataset which contains over 10,000 deepfake videos generated using a state-of-the-art face- swapping algorithm. The comprehensive and meticulously- designed perturbations make it an excellent benchmark for testing the robustness of deepfake detection models under diverse conditions.

Celeb-DF (Ramachandran, Nadimpalli, & Rattani, 2021) dataset is a includes 5,639 deepfake videos, fulfilling various research evaluation requirements. It is particularly valuable for evaluating detection methods due to its high visual fidelity and challenging nature. The Celeb-DF dataset addresses the limitations of other datasets by providing high-quality deepfake videos that closely resemble real-world scenarios.

FaceForensics++ [14] is a newly-built dataset designed to facilitate research work in detecting manipulated facial images and videos. These datasets take use of face manipulation methods providing resource for training and evaluation detection algorithms. This data sets contain 1,000 deepfake videos.

To evaluate the algorithms for AI-generated video misinformation detection, we employ standard metrics including accuracy, precision, recall, and F1 score. These metrics provide a comprehensive evaluation of the detection capabilities, ensuring that the models are not only accurate but also reliable and robust in various scenarios.

Experimental Settings

The video blockchain was developed by using a private blockchain implemented with Python and JavaScript. Initially, we employed our tailored protocol, which incorporates cryptographic methods to ensure data integrity and immutability. To enhance AI-generated video detection, we integrated the Locality-Sensitive Hashing (LSH) method. For our experimental datasets, we took use of Visual Studio Code IDE to run, debug, and deploy the blockchain functions. Additionally, for data analysis and plot generation, we made use of Matplotlib. All software was run on Microsoft Windows 1164-bit with an Intel Core i7-8550U processor and 16GB of RAM. We adopted Python version 3.12.0 with the pip module installed.

Performance

In this section, we demonstrate our implemented algorithm using different elements. All of these performances have been completed within our experimental setup to ensure accurate results for better understanding. Face Forensics++ datasets have been using for these detection performance evaluation process.

Table 1:Scores of the precision, F1 score, accuracy, and recall for Algorithm 1

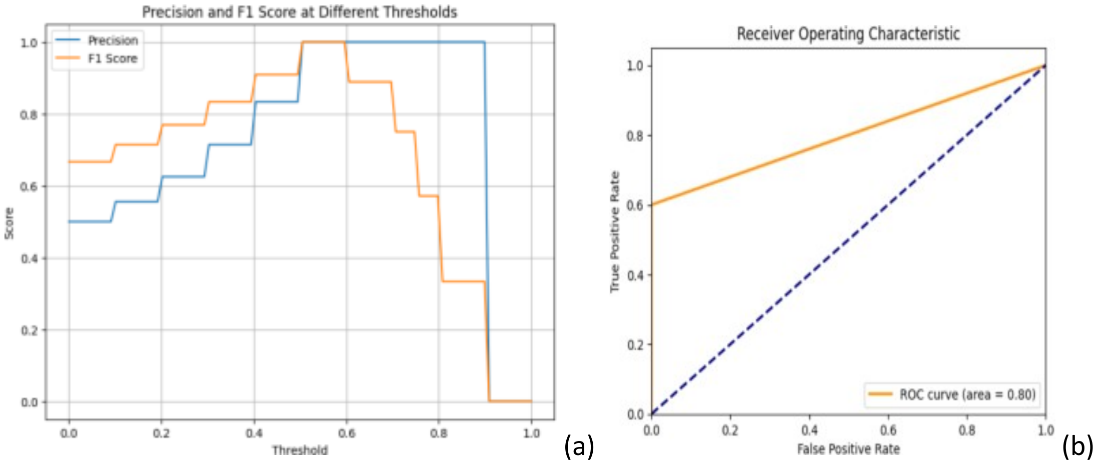
Metrics	Scores
Precision	1.0

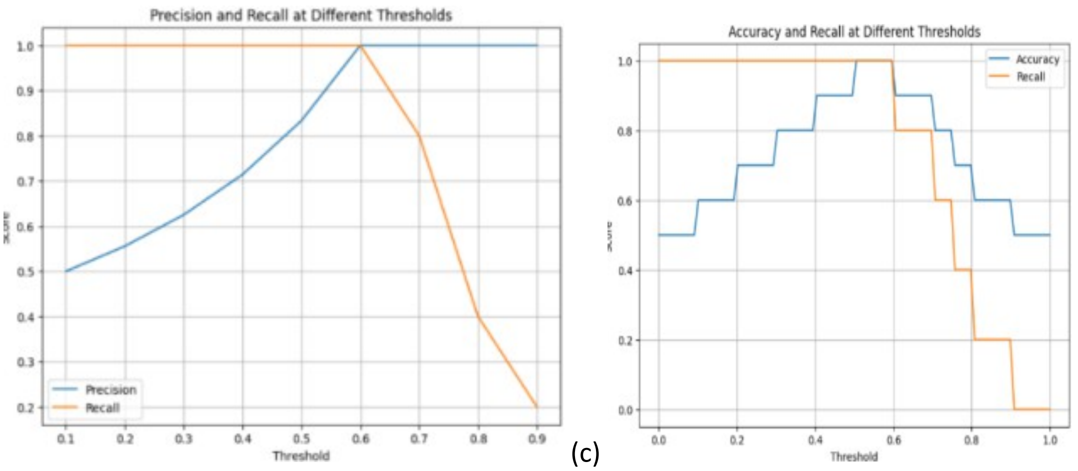
F1 score	0.75
Accuracy	0.8
Recall	0.6

Table 2:Detailed metrics breakdown by classes

	Precision	Recall	F1 score	Support
Class 0	0.714286	1.0	0.833333	5.0
Class 1	1.000000	0.6	0.750000	5.0
Accuracy	0.800000	0.8	0.800000	0.8
Macro Average	0.857143	0.8	0.791667	10.0
Weighted Average	0.857143	0.8	0.791667	10.0

In Table 1, its comprehensive evaluation of the proposed algorithm, it demonstrates its strengths to all four metrics. In Table 1 and Table 2, we extract the detailed metrics of the class 0, the precision is approximately 0.71, recall is 1.0, and F1 score is 0.83, indicating that while the model is perfect at identifying negatives, it includes false positives in its predictions. Moreover, regarding Class 1, the precision is perfect at 1.0, but recall drops to 0.6, resulting in an F1 score of 0.75, highlighting that the model is very precise but misses some positive instances. While comparing to the overall accuracy of the model is 80%, with both the macro and weighted averages showing high precision (0.857), but slightly lower recall and F1 score (both approximately 0.8 and 0.79, respectively), reflecting the model's strong but slightly imbalanced performance across classes.





(c) (d)

Figure 1:The precision, recall, Accuracy, F1 Score and ROC curve at different thresholds calculate using FaceForensics++ dataset to measure the performance of the algorithm 1 implementation.

In summary, Table 2 illustrates the model correctly classified all 5 instances of the negative class (Class 0) as negative, but misclassified 2 out of 5 positive class instances (Class 1) as negative, resulting in 3 true positives and 2 false negatives. After translated to percentages, the confusion matrix indicates that 100% of actual negatives were correctly predicted, while only 60% of actual positives were correctly predicted.

In Fig.1(a), *ground_truth* represents the true labels, and *predicted_probabilities* shows the predicted probabilities from our algorithm. The *precision_recall_curve* (•) function calculates precision and recall at different thresholds. The resulting plot shows how precision and recall change as the classification threshold varies.

In Fig.1(b), we calculate the accuracy and recall at different thresholds by iterating over a range of thresholds and converting predicted probabilities to binary predictions based on each threshold. It then calculates the accuracy and recall values for each threshold and plots them against the threshold values.

Fig.1(c) shows the calculation results of precision and F1 score at various thresholds by iterating over a range of thresholds and converting predicted probabilities to binary predictions based on each threshold. It calculates the precision and F1 score values for each threshold and plots them against the threshold values

Fig.1(d) indicates the ROC curve by using the *roc_curve*(•) function from scikit-learn. It calculates the area under the ROC curve (AUC) using the AUC function. Finally, it plots the ROC curve and displays the AUC value in the plot title, it shows the ROC curve by using the *roc_curve*(•) function from scikit-learn. It calculates the area under the ROC curve.

After evaluating its performance by using metrics like accuracy, precision, recall, and F1 score, the metrics can assist us to understand how well our algorithm performs in detecting AI-generated video misinformation without relying on machine learning models. In next, we evaluate the performance of the algorithm with different datasets and compare the result with the similar method that have used the same dataset to reevaluate their method performance.

Table 3:The evaluation of our proposed methods with similar research works by using WildDeepfake datasets

Methods	**Accuracy**	**Precision**	**Recall**	**F1 score**
DDM[08]	0.925	0.962	0.925	0.942

DVD[09]	0.968	0.975	0.932	0.953
Our Method	**0.978**	**0.981**	**0.965**	**0.968**

Table 4:The evaluation of our proposed method with similar research works by using DeeperForensics-1.0 data sets

Methods	**Accuracy**	**Precision**	**Recall**	**F1 Score**
DDDF[10]	0.94	0.96	0.88	0.92
IDD[11]	0.965	0.978	0.91	0.942
Our Method	**0.973**	**0.980**	**00952**	**0.972**

Table 5:The evaluation of our proposed method with the similar research works by using Celeb-DF datasets

Methods	**Accuracy**	**Precision**	**Recall**	**F1 Score**
DF[12]	0.95	0.97	0.90	0.63
MLDD[13]	094	096	0.88	0.92
Our Method	**0.982**	**0.932**	**0.912**	**0.949**

We benchmark our research outcomes against the existing research work to validate the effectiveness of our proposed method. Our studies employ machine learning models and methods similar to our research approach. This comparison provides valuable insights into the advancement and robustness of our method in addressing the research problem at hand.

Table 6:The evaluation of our proposed method with the similar research works by using FaceForensics++ datasets

Methods	**Accuracy**	**Precision**	**Recall**	**F1 score**
CNN-LSTM [14]	0.6178	0.5269	0.5131	0.5199
3D-ResNet[15]	0.6617	0.5624	0.5374	0.5496
Our Method	**0.7891**	**0.9852**	**0.5915**	**0.6932**

The comparison of our results with other methods that has been employed with different datasets presented in Table 3, Table 4, Table 5 and Table 6. These methods shown in the tables for experiments include:

- **Experiments with WildDeepfake datasets**: Deepfake Detection Method (DDM) (Ref 08), Deepfake Video Detection (DVD) (Ref 09).
- **Experiments with DeeperForensics-1.0, datasets**: Deepfake Detection with Data Farming (DDDF) (Korshunov & Marcel, 2022), Interpretability of Deepfake Detection (IDD) (Korshunov, Jain, & Marcel, 2022).
- **Deep Face Recognition** (DFR) (Ramachandran, Nadimpalli, & Rattani, 2021), Machine Learning Approach for Deepfake Detection (MLDD) (Lacerda & Vasconcelos, n.d.).

Our method shows the best performance among the all of compared method. Based on the two datasets of DeeperForensics-1.0 and Celeb-DF, we have the same best results compared with our proposed method. However, as the experimental results illustrate the state-of-the-art share-bit, video blockchain methods deliver superior performance on all of the four datasets. Moreover, there is considerable result with FaceForensics++ datasets. Compared to other average evaluation matrix, it shows high degree level of performance rate. But with this data, its comparatively getting low score with this dataset.

In addition, we identified that other methods also show similar low-level number with this dataset. We highlight the superior performance of our approach, particularly integrate the Video Blockchain method. Our method not only achieved higher accuracy and efficiency but also demonstrated robustness in handling AI-generated video detect.

By incorporating cryptographic methods within Video Blockchain, we ensure the integrity and immutability of the data, addressing key challenges in detecting misinformation. Furthermore, our method showcases a significant improvement in detection rates, outperforming traditional approaches and showcasing the potential for real-world applications in combating cryptographic methods within our Video Blockchain, we ensure the integrity and immutability of the data addressing key challenges in detecting misinformation. Furthermore, our method showcases a significant improvement in detection rates, outperforming traditional approaches and showcasing the potential for real-world applications in combating AI-generated video misinformation.

Ablation Studies Results with Different Block Sizes

In this section, Algorithm 1 shows the process of Block Division as a frame is divided into image blocks. To get this result, each frame Fi is divided into three different block sizes 8x8, 12 x 12 and 16 x 16.

- **Block Size 8x8**: The baseline accuracy is 0.4, which means the model correctly classified 40% of the frames with the relevant block size. Without image compression, the accuracy has been dropped to a considerable amount. This means that compression is crucial for performance. Finally, *no_hashes_record*, Ablating hash computation reduces accuracy to 0.2, showing its importance in maintaining performance.
- **Block Size 12x12**: In the baseline *no_extraction* accuracy remains 4.0. Both *no_compression* and *no_extraction* significantly increase the accuracy level, indicating high performance without each of these functions. However, *no_blockchain_link* results in lower accuracy compared to 8x8 block size frames. Removing super bit structure still provides a sensible accuracy of 0.6, indicating its moderate importance. Similarly, ablation of hash records does not significantly impact accuracy at this block size.
- **Block Size 16x16**: According to data illustrate in Fig. 2, the average accuracy is maintained without each function in Algorithm 1. The baseline accuracy is slightly better at 0.5. Removing compression increases accuracy to 0.6, indicating less dependence on compression at this larger block size. Moreover, accuracy drops to 0.4 without extraction, suggesting it's more necessary at this block size. Finally, we show the

blockchains linking improves accuracy to 0.7 and *no_super_bit_structure*, *no_hashes_record* take the accuracy remains consistent at 0.6.

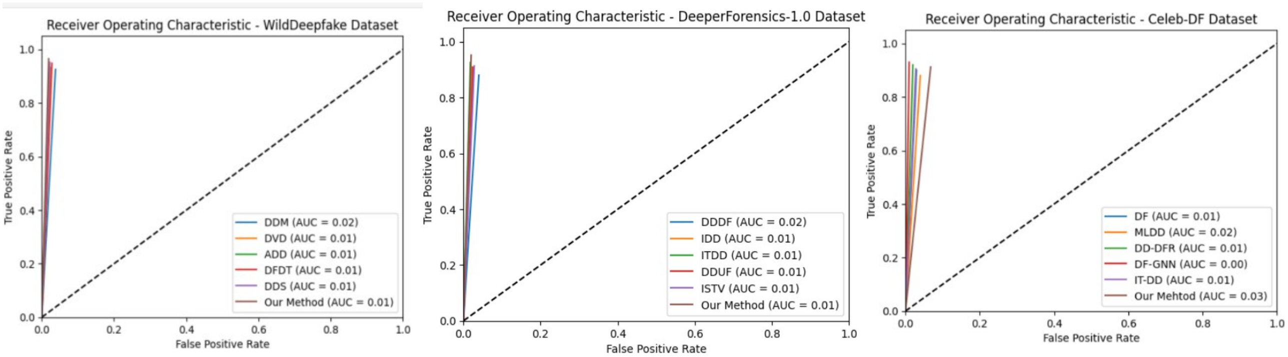


Figure 2:Receviver Operating Characteristic (ROC), analysis with WildDeepfake, DeeperForensics-1.0 and Celeb-DF datasets to understand the generating data.

By evaluating different block sizes, we summarize the conclusion. These results indicate in the TABLE VIII. The impact of different components in the pipeline varies with block size. In addition, smaller blocks (8x8) benefit from compression and has computation. On the other hand, mid-size blocks (12x12) do not need compression or extraction but rely heavily on blockchain linking. Finally, we see large blocks (16 x 16) do not require compression, may not need blockchain linking, but still require extraction. After compering all details, we identify the mid-size and large size blocks have more compatible with our algorithm.

In the ablation study, the mean, max, and min scores offer numerical insights into the performance variation across different experiments, specifically various block sizes tested. The mean score, calculated as the average of all performance metrics, provides a central measure, such as an average accuracy, precision, or F1 score, indicating the typical performance level across the tested configurations. For instance, in Table VII, the mean score across block sizes 8x8,12x12, and 16x16 is 0.4667, calculated as the average of the original accuracy scores (0.4, 0.4, and 0.5). Conversely, the max score, representing the highest achieved performance, highlights the best-performing configuration. In our example, the max score across block sizes is 0.8, achieved in the "*no_extraction*" scenario with a block size of 12x12.

Conversely, the min score reflects the lowest observed performance, indicating the worst-performing configuration. For instance, in our data, the min score is 0.1, observed in the "*no_blockchain_link*" scenario with a block size of 12x12. These numerical insights provide researchers with a comprehensive understanding of the performance landscape, guiding decisions regarding model configurations or interventions by illuminating the variability in performance across different experimental conditions. Essentially, WildDeepfake comprises a mixture of both real and fake videos.

Similarly, DeeperForensics-1.0 is similar to the WildDeepfake dataset, we assess the performance of the detection algorithm on DeeperForensics-1.0 by evaluating its accuracy in distinguishing real from fake sequences. Moving on to Celeb-DF dataset, it is represented as a collection of real and fake video pairs. Performance evaluation on Celeb-DF involves measuring the ability of our algorithm to differentiate between real and fake videos. When comparing our results using the datasets: **Wild Deepfake**, **Deeper Forensics-1.0**, and **Celeb-DF datasets**, our Algorithm 1 shows virtuous performance on all three due to the high resolution and realistic facial expressions, though challenges remain with more complex real-world scenarios. Regarding the FaceForensics++ performance, despite of its structured setup, this dataset does not always yield the best results in real-world due to its less diverse set of actors and settings compared to datasets like WildDeepfake. The controlled nature of the dataset means it might not capture the variability and complexity found in real-world deepfakes.

Conversely, while examining the performance of FaceForensics++, despite of its structured setup, it doesn't consistently yield optimal results in real-world scenarios as show in the Table VII. This is due to its less diverse pool of actors and settings compared to datasets like WildDeepfake. FaceForensics++ is employed for performance evaluation which has highlight, compared to other datasets, it shows lower performance. This has been noted in multiple studies. The controlled nature of the dataset may not capture the variability- -and complexity inherent in real-world deepfakes. After the analysis of similar research work, we have identified that the reasons for lower performance is the limited scene diversity in the FaceForensics++ (FF++) datasets, where Vi represents different scene variations, n is the number of videos for FF++, n limited, therefor V is too low place. In summary, the performance is due to controlled nature of FaceForensics++ which does not capture the full range of variability and complexity in real-world deepfakes. This leads to reduced generalization performance of detection algorithms when applied to more diverse datasets like Wild Deepfake.

5. Conclusion

In this book chapter, we have demonstrated the effectiveness of integrating Video Blockchain with AI-generated video identification to enhance the authenticity and trustworthiness of digital content. Our primary goal was to classify AI-generated videos and combat visual misinformation. The implementation of a robust framework for authenticating video content and comparative analysis with other research works highlights the superior performance and robustness of our proposed solution, particularly through the integration of cryptographic methods.

By establishing this framework, we provide a sustainable solution for stakeholders concerned with the integrity of digital content. Our approach not only achieved higher accuracy and efficiency but also demonstrated significant improvements in successful rates, showcasing its potential for real-world applications. Mitigating misinformation risks is crucial for empowering users to make informed decisions and safeguarding the integrity of digital media.

Our experimental results, which are evaluated by using multiple metrics such as accuracy, precision, recall, and F1 score, indicate that our method outperforms to the existing approaches.

References

Atrey, P., Yan, W., Chang, E., Kankanhalli, M. (2004) A hierarchical signature scheme for robust video authentication using secret sharing. International Multimedia Modelling Conference, 330-337.

Atrey, P., Yan, W., Kankanhalli, M. (2007) A scalable signature scheme for video authentication. Multimedia Tools and Applications 34 (1), 107-135.

Bansal, M., Yan, W., Kankanhalli, M. (2003) Dynamic watermarking of images. International Conference on Information, Communications and Signal Processing.

Ding, W., Yan, W., Qi, D. (1999) Digital watermark image embedding based on U-system. International Conference on Computer Aided Design and Computer Graphics, 893-899.

Ding, W., Yan, W., Qi, D. (1999) Digital image scrambling based on Gray code. International Conference on CAD/CG 3, 900-904.

Ding, W., Yan, W. (1999) Digital watermark image based on discrete cosine transform. Journal of North China University of Technology China.

Feng, H., Ling, H., Zou, F., Yan, W., Lu, Z. (2010) Optimal collusion attack for digital fingerprinting. ACM International Conference on Multimedia, 767-770.

Feng, H., Ling, H., Zou, F., Yan, W., Lu, Z. (2012) A collusion attack optimization strategy for digital fingerprinting. ACM Transactions on Multimedia Computing, Communications, and Applications.

Feng, H., Ling, H., Zou, F., Yan, W., Sarem, M., Lu, Z. (2013) A collusion attack optimization framework toward spread-spectrum fingerprinting. Applied Soft Computing 13 (8), 3482-3493.

Garhwal, A., Yan, W. (2018) BIIIA: A bioinformatics-inspired image identification approach, Multimedia Tools and Applications.

Garhwal, A., Yan, W. (2019) BIIGA: Bioinformatics inspired image grouping approach. Multimedia Tools and Applications, 78 (11), 14355-14377.

Gedara, K., Nguyen, M., Yan, W. (2022) Visual blockchain for intelligent surveillance in a smart city. Blockchain Technologies for Sustainable Development in Smart Cities, Book Chapter, IGI Global.

Gedara, K., Nguyen, M., Yan, W. (2023) Video blockchain: A decentralised approach for secure and sustainable networks with distributed video footages from vehicle-mounted cameras in smart cities. Electronics (journal).

Gedara, K., Nguyen, M., Yan, W. (2023) Enhancing privacy protection in intelligent surveillance: Video blockchain solutions. BLOCKCHAIN'23.

Gedara, K., Nguyen, M., Yan, W., Li, X. (2024) Advancing video data privacy preservation in IoT networks through video blockchain. Information 2024, 15(3), 171.

Gulzar, N., Abbasi, B., Wu, E., Ozbal, A., Yan, W. (2013) Surveillance privacy protection. Intelligent Multimedia Surveillance, 83-105.

Gupta, M., Yan, W. (2022) Video watermarking with digital signature and fingerprinting. Applications of Encryption and Watermarking for Information Security, IGI Global.

Han, T., Xie, W., & Zisserman, A. (2019). Video representation learning by dense predictive coding. In *ICCV Workshop*.

Heidari, A., Navimipour, N. J., Dag, H., Talebi, S., & Unal, M. (2024). A novel blockchain-based deepfake detection method using federated and deep learning models. *Cognitive Computation*.

Hu, R. (2019) Visual blockchain using Merkle Tree. Master's Thesis. Auckland University of Technology, New Zealand.

Hu, R., & Yan, W. (2020). Design and implementation of visual blockchain with Merkle tree. In *Handbook of Research on Multimedia Cyber Security* (pp. 282–295).

Hussain, Z. F., & Ibraheem, H. R. (2023). Novel convolutional neural networks-based Jaya algorithm approach for accurate deepfake video detection. *Mesopotamian Journal of Cybersecurity*, **2023**, 35–39.

Kataoka, H., Wakamiya, T., Hara, K., & Satoh, Y. (2020). Would mega-scale datasets further enhance spatiotemporal 3D CNNs? https://doi.org/10.48550/arXiv.2004.04968%20Focus%20to%20learn%20more

Korshunov, P., & Marcel, S. (2022). Improving generalization of deepfake detection with data farming and few-shot learning. *IEEE Transactions on Biometrics, Behaviour, and Identity Science*, **4**(3), 386–397.

Korshunov, P., Jain, A., & Marcel, S. (2022). Custom attribution loss for improving generalization and interpretability of deepfake detection. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 8972–8976).

Li, Y., & Lyu, S. (2018). Exposing deepfake videos by detecting face warping artifacts. https://doi.org/10.48550/arXiv.1811.00656

Ling, H., Wang, L., Zou, F., Yan, W. (2011) Fine-search for image copy detection based on local affine-invariant descriptor and spatial dependent matching. Multimedia Tools and Applications 52 (2), 551-568.

Ling, H., Cheng, H., Ma, Q., Zou, F., Yan, W. (2011) Efficient image copy detection using multi-scale fingerprints. IEEE Multimedia, 19, 60–69

Ling, H., Feng, H., Zou, F., Yan, W., Lu, Z. (2010) A novel collusion attack strategy for digital fingerprinting. International Workshop on Digital Watermarking, 224-238.

Liu, J., Ling, H., Zou, F., Yan, W., Lu, Z. (2012) Digital image forensics using multi-resolution histograms. Crime Prevention Technologies and Applications for Advancing Criminal.

Liu, P., Zhou, S., Yan, W. (2022) A 3D cuboid image encryption algorithm based on controlled alternat quantum walk of message coding. Mathematics, 10 (23), 4441.

Liu, Z., Yang, M., Yan, W. (2017) Image encryption based on double random phase encoding. International Conference on Image and Vision Computing New Zealand.

Liu, Z., Yang, B., Yan, W. (2021) A framework for image encryption on frequency domain. Research Anthology on Artificial Intelligence Applications in Security (pp.328-338)

Ma, B., Wu, J., Lai, E., Yan, W. (2023) A privacy-preserving word embedding text classification model based on privacy boundary constructed by deep belief network. Multimedia Tools and Applications.

Ma, B., Yan, W., Lai, E., Wu, J. (2021) A new noise generating method based on Gaussian sampling for privacy preservation. International Symposium on Geometry and Vision.

Moodley, E., Huo, G., Hsieh, M., Cai, S., Yan, W. (2013) Password security and protection. Managing Trust in Cyberspace, 449.

Mi, B., Wu, B., Huang, D., Liu, Y., Chen, L., & Wan, S. (2022). Privacy-oriented transaction for public blockchain via secret sharing. *Security and Communication Networks*.

Moolikagedara, K., Nguyen, M., Yan, W. Q., & Li, X. J. (2023). Video blockchain: A decentralized approach for secure and sustainable networks with distributed video footage from vehicle-mounted cameras in smart cities. *Electronics (Switzerland)*, **12**(17).

Ramachandran, S., Nadimpalli, A. V., & Rattani, A. (2021). An experimental evaluation on deepfake detection using deep face recognition. In *International Carnahan Conference on Security Technology*.

Raman, R. K., & Varshney, L. R. (2018). Distributed storage meets secret sharing on the blockchain. In *Information Theory and Applications Workshop (ITA)*.

Shu, Y., Yu, J., Yan, W. (2019) Blockchain for security of a cloud-based online auction system. Exploring Security in Software Architecture and Design.

Shu, Y., Yu, J. Yan, W. (2020) Blockchain for security of cloud-based online auction. Research Anthology on Blockchain Technology in Business, Healthcare.

Singh, P., Chaudhary, K., Chaudhary, G., Khari, M., Rawal, B. (2022) A machine learning approach to detecting deepfake videos: An investigation of feature extraction techniques. Journal of Cybersecurity and Information Management. Volume 9 , Issue 2 , PP: 42-50.

Wang, H., & Liao, J. (2021). Blockchain privacy protection algorithm based on Pedersen commitment and zero-knowledge proof. In *International Conference on Blockchain Technology and Applications* (pp. 1–5).