International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

# Detecting the Source of Digital Images Using Perceptual Hashing and Blockchain Technology

[1]Maeba, Nyegia Beete, [2]Dr. D. Matthias, [3]Dr. O. E. Bennett
[1,2,3]*Department of Computer Science*
*Faculty of Science, Rivers State University*

Date of Submission: 15-07-2024 Date of Acceptance: 25-07-2024

**ABSTRACT**: This research centers on detecting the source of digital images through the integration of perceptual hashing and blockchain technology. The dataset utilized is sourced from the Uncompressed Color Image Database (UCID). Object-Oriented Analysis and Design (OOAD) methodology was used for both the development and evaluation phases of the model. Python was used as the programming language for implementing the perceptual hash algorithm, leveraging its versatility and the availability of machine learning libraries. The Ethereumblockchain serves as the database for storing images and their metadata. Specifically, the image and corresponding metadata are stored within smart contracts, which are written in Solidity. This model incorporates three essential modules to facilitate various aspects of its functionality. The first module is dedicated to the creation and storage of new records within the application's database. Serving as a fundamental component, it enables users or system processes to seamlessly add new data entries. The process commences with user input, as individuals interact with the application, providing the necessary information to be stored as a new record. The second module is tasked with retrieving existing data or records from the application's database or storage system. Its crucial role lies in providing users, system components, or external entities with easy access to information stored within the system, fostering a seamless flow of data retrieval. The third module focuses on the analysis and determination of relationships, differences, or similarities between two or more records within the system. Particularly useful for understanding data interconnections and identifying patterns or variations in datasets, this module is triggered either by user requests or as part of an automated system process requiring the comparison of specific records. The system was tested using the UCID image dataset, which comprises 1,338 authentic images and 9,366 tampered images, providing a comprehensive basis for evaluating the model's performance. Key metrics used in the evaluation include Accuracy, Precision, Recall, and F1 Score, which provide a detailed assessment of the model's effectiveness in detecting the source of images. The model achieved an accuracy of 91.30% with a percent error of 6.69%. The accuracy varied with the degree of tampering, with higher accuracy observed for low tampering rates (1% or 2%) compared to high tampering rates (50%). These results validate the effectiveness of the proposed model in identifying the source of digital images using perceptual hashing and blockchain technology.

**KEYWORDS:** Perceptual hashing, Blockchain, Uncompressed Color Image Database (UCID), Smart contracts, Ethereum

## I. INTRODUCTION

Information is essential to our daily lives and to the success of organizations, businesses, and societies. It allows us to make informed decisions, solve problems, learn new things, and communicate effectively. Information provides the foundation for knowledge, innovation, and progress, and plays a crucial role in shaping our worldviews and perceptions. In today's digital age, the amount of information available is vast and constantly growing, making the ability to effectively gather, process, and use information increasingly important. However, in this era of big data, where large volumes of information are generated at a very high speed and in different formats—most notably digital media. Digital media has become ubiquitous in our daily lives, with the widespread use of digital media, there is an increasing need to detect the source of digital media. The source of digital media can provide important information such as the origin, ownership, and authenticity of the media, it is often

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

difficult to spot the difference between real and fake news. Jang et al., (2018) noted that an important effort that helps to minimize the gross impact of fake news is to track down on the root source of misleading information, other measures can be employed such as literacy education and fact checking.

Fake news refers to intentionally false or misleading information that is spread through various media platforms, such as social media, websites, and news outlets. It is created with the intent of deceiving and misleading the public for various reasons, such as political gain, financial profit, or personal vendettas. Fake news has become a major concern in recent years, as it can lead to harmful consequences, such as social unrest, political instability, and loss of trust in legitimate news sources. It is important for individuals to fact-check information before sharing or acting on it, and for media outlets to prioritize accuracy and impartiality in their reporting.

Fake news can take many forms, from fabricated stories and photoshopped images to misleading headlines and manipulated videos. It is often created and spread for political or financial gain, with the intent of influencing public opinion or generating clicks and ad revenue.

The consequences of fake news can be severe, including the erosion of trust in traditional media sources, the spread of misinformation that can lead to harmful actions or decisions, and the potential for political and social destabilization.

Addressing the issue of fake news demands a comprehensive strategy that encompasses media literacy education, fact-checking, and verification processes. Additionally, the integration of technologies like artificial intelligence and blockchain is crucial for identifying and thwarting the dissemination of misinformation. To effectively combat fake news, indivtoiduals must also embrace a commitment to critical thinking and actively seek out diverse perspectives and information sources.

Overall, fake news is a complex and multifaceted issue that poses significant challenges to our society, and requires a collaborative effort from individuals, media organizations, and technology companies to address. According to Gottfried et al., (2016), asides the use of traditional media, reports have shown that majority of the adult population access their information through digital platforms such as social media.

The topic of fake news gained significant attention during the 2016 U.S. Presidential Election, (Walker, 2017) in his article posted on Birmingham Mail talked about how Donald Trump, the elected President, accused several media outlets of deliberately publishing false information to discredit him. This issue was already in the public discourse, and The New York Times had previously published an article claiming that a prominent supporter of Trump was spreading disinformation. After conducting an investigation, the newspaper found that a photograph used by the Christian Times website to suggest election rigging by President Trump's opponents was, in fact, a picture of ballot boxes used in a U.K. election. The picture did not show fraudulent Clinton votes found in an Ohio Warehouse, as the website claimed.

It is important to identify the entities involved in disseminating information on social media platforms. To illustrate this, we consider Channels Television as the source of news, a social media platform, Facebook, and three users — Amadi, Belema, and Soibi. Channels TV publishes a news article on its website. Amadi shares the article on his Facebook wall, Belema obtains the article from Amadi's wall, modifies the content, and shares it on his own wall. If Soibi shares the content obtained from Belema's wall, she is spreading fake news. The dissemination of fake news continues if other users obtain the same news content from Soibi's Facebook wall.

What if there was a way for Soibi to verify the original content from the modified one? Blockchain technology can be used to trace the source of digital media by creating a secure and immutable record of ownership and transactions. When a piece of digital media is created, its metadata, such as the date, time, and location of creation, can be recorded on a blockchain. This creates a permanent record that can be accessed and verified by anyone, at any time. (Huckle et al., 2017) introduced an early prototype of a blockchain-based distributed application called Provenator. It uses the trust mechanisms of blockchain technology to validate the originator of digital media sources used in news content. As discussed earlier by Gottfried et al., (2016), social media is the primary source of information for majority of the adult population. Huckle et al., (2017) in their work used Provenator to store source metadata in a blockchain, this enables content creators to prove without doubt the origins of their digital media resources

## II. LITERATURE REVIEW

Saad et al., (2019) in their paper proposed an innovative blockchain system to tackle present challenges and curb the spread of misinformation within the network. The researchers detailed their analysis of information flow in social networks and introduced a streamlined detection system, underscoring its capacity for efficient deployment

with minimal resource overhead. The authors explained that their system involved the news source and platform sharing a blockchain with read and write access. They clarified that the information source could be a group of publishers, such as CNN, BBC News, etc. Similarly, the platform could be a group of social networks, such as Facebook, Twitter, LinkedIn, etc. They added that there was also a group of social network users who could share information on their timelines and invoke transactions to the blockchain. However, these users were unable to participate in the consensus mechanism or write to the blockchain. The researchers noted that by restricting access privileges to both the source and the platform, there was a decrease in the number of participants involved in the consensus process. Consequently, this led to quicker transaction finality, heightened throughput, and a diminished ledger maintenance overhead for end users.

Jang et al., (2018) in a recent inquiry highlighted that researchers examined the issue of fake news in the context of the 2016 US presidential election, utilizing a recent advancement in computational network science, particularly evolution tree analysis. The study involved retrieving 307,738 tweets about 30 fake and 30 real news stories to examine the root content, producers of the original source, and evolution patterns. According to the findings, root tweets about fake news were primarily generated by ordinary users who often included a link to non-credible news websites. The study also observed significant differences between real and fake news stories in terms of evolution patterns. The tweets about real news had a wider breadth and shorter depth than those about fake news in the evolution tree analysis. Additionally, tweets about real news spread widely and quickly, while tweets about fake news underwent a greater number of modifications in content over the spreading process.

Gilda, S. (2017) investigated the use of natural language processing techniques for the detection of 'fake news', which referred to misleading news stories that came from non-reputable sources. The researchers applied term frequency-inverse document frequency (TF-IDF) of bi-grams and probabilistic context-free grammar (PCFG) detection to a corpus of about 11,000 articles using a dataset obtained from Signal Media and a list of sources from OpenSources.co. They tested their dataset on multiple classification algorithms including Support Vector Machines, Stochastic Gradient Descent, Gradient Boosting, Bounded Decision Trees, and Random Forests. The researchers found that TF-IDF of bi-grams fed into a Stochastic Gradient Descent model identified non-credible sources with an accuracy of 77.2%, with PCFGs having slight effects on recall.

Ahmed, et al., (2017) focuses on the problem of deceptive content such as fake news and fake reviews, which has become increasingly dangerous for online users. They clarify that the issue of fraudulent reviews has had repercussions for both consumers and businesses. Additionally, the matter of fake news came into focus in 2016, particularly following the most recent U.S. presidential elections. Since both fake reviews and fake news involve the dissemination of inaccurate information, the issue of opinion spam rapidly emerged as a burgeoning research area, driven by the prolific nature of user-generated content. The authors note that it is challenging to distinguish between real and fake reviews, and even humans often struggle with this task. They present a novel n-gram model designed for the automatic detection of fake content, specifically targeting fake reviews and fake news. Their approach involves a comparison of two distinct feature extraction techniques and six machine learning classification methods. Through experimental evaluation using established public datasets and a newly introduced fake news dataset, the results demonstrate promise and an enhancement compared to state-of-the-art methods.

Wei, et al., (2014) introduces a novel approach where images are divided into 2×2 blocks, termed mini-squares, and searched for one of nine types of patches likely to recover a mini-square altered by seam carving. Our method evaluates the patch transition probability among three-connected mini-squares, achieving detection accuracies of 92.2% for images with 20% seam carving and 95.8% for those with 50% seam carving. We also explore additional applications of our patch analysis method, such as identifying hot regions frequently traversed by carved seams.

Kanoksak, et al., (2015) in their paper proposed a method inspired by the Blocking Artifact Characteristics Matrix (BACM) to detect tampering caused by seam modifications in JPEG retargeted images without requiring the original image. The BACM block matrix reveals that the original JPEG image has regular symmetrical data, whereas the symmetry is disrupted in blocks modified by seam carving.

Huckle, et al., (2017) In their research, they presented an application based on blockchain technology with the capability to verify the authenticity of digital media. This application is designed to reveal the origin of any digital media, including instances where images are utilized out of context to deceive users, utilizing the trust

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

mechanisms inherent in blockchain technology. In essence, the study asserts that the tool can unequivocally establish the authenticity of digital media. The framework proposed in this paper builds upon the approach introduced by (Huckle et al., 2017) by incorporating a method known as perceptual hashing, addressing the limitations mentioned in the previously outlined problem statement.

## III. SYSTEM DESIGN

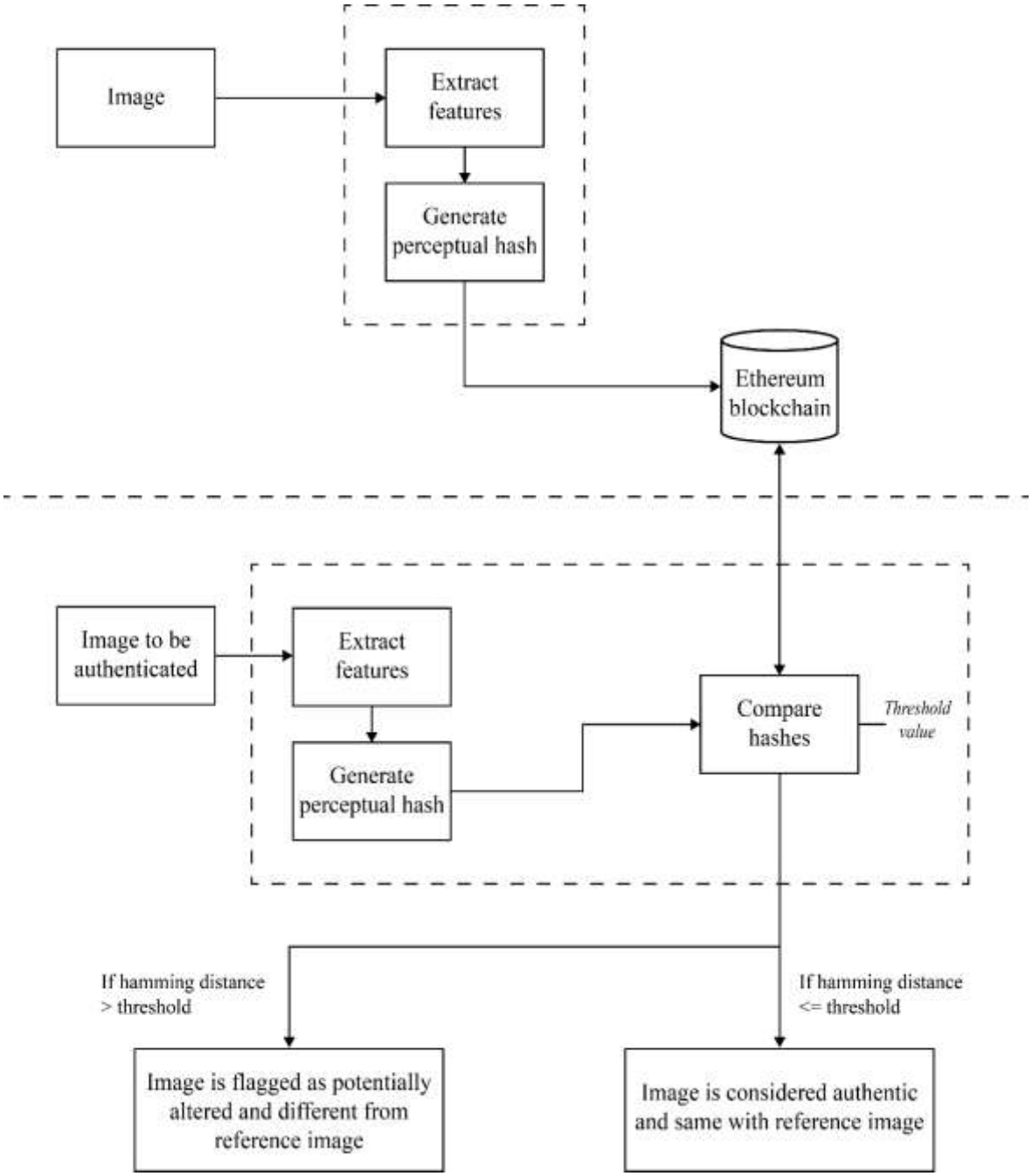



**Figure 1:System Architecture**

System design is the process of defining the architecture, interfaces, and data for a system that satisfies specific requirements. It defines the components, their interactions, and the relationships

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

between them in order to achieve the desired functionality of the system. Below is a detailed list of the components:

1. **User Interface:** The User Interface component allows users to interact with the system, providing inputs and receiving outputs. It includes functionalities for image submission, viewing authentication results, and accessing source metadata.
2. **Image Submission:** The Image Submission component handles the process of accepting image inputs from users. It may involve options such as uploading images from local storage, providing URLs, or capturing images using a camera.
3. **Feature Extraction and Perceptual Hashing:** The Feature Extraction and Perceptual Hashing component analyzes the submitted images to extract relevant visual features. Perceptual hashing techniques are applied to generate unique hash values representing the visual characteristics of the images.
4. **Blockchain Integration:** The Blockchain Integration component facilitates the integration of blockchain technology into the system. It interacts with the blockchain network to store and retrieve perceptual hashes, image metadata, and authentication records securely. Smart contracts may be used to manage transactions and enforce the rules and logic of the system on the blockchain.
5. **Source Detection and Authentication:** The Source Detection and Authentication component compares the perceptual hash of the submitted image with stored hashes on the blockchain. It utilizes similarity measures, such as Hamming distance, to determine the authenticity of the image based on the similarity between hashes.
6. **Source Metadata Retrieval:** The Source Metadata Retrieval component retrieves and provides access to the source metadata associated with authenticated images. Metadata may include information such as author, date, location, or other relevant details that indicate the image's origin and credibility.

### 1.1 Class Diagram

This class diagram in figure 3.2 shows the following classes and the relationship between the classes involved in the proposed system.

### 1.2 Activity Diagram

This activity diagram in figure 3.3 shows the steps involved in the proposed System. The user uploads an image, which is then processed by the system to generate a perceptual hash. The perceptual hash is then stored on the blockchain. The Image Owner can then verify the authenticity of the image by comparing the perceptual hash of the image they have to the perceptual hash stored on the blockchain. The system will return a verification result, indicating whether the image is authentic or not.

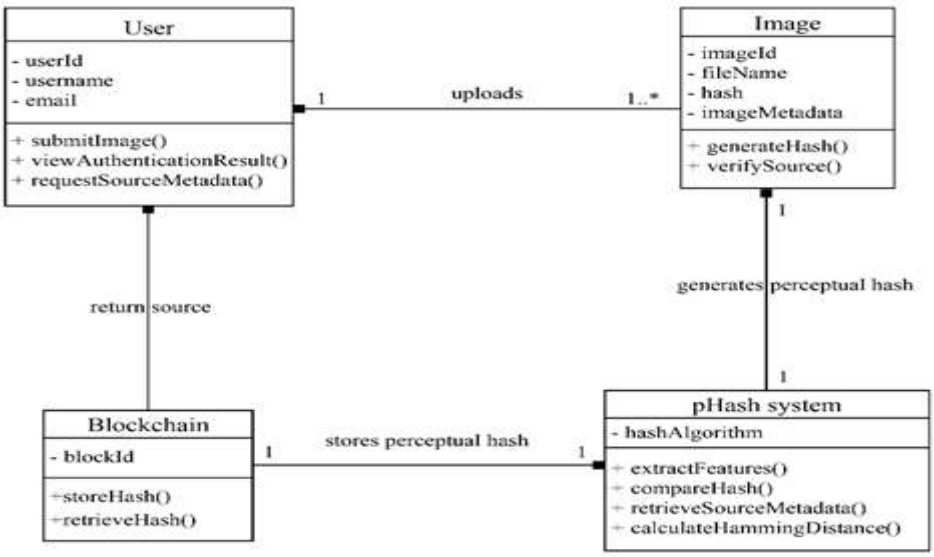
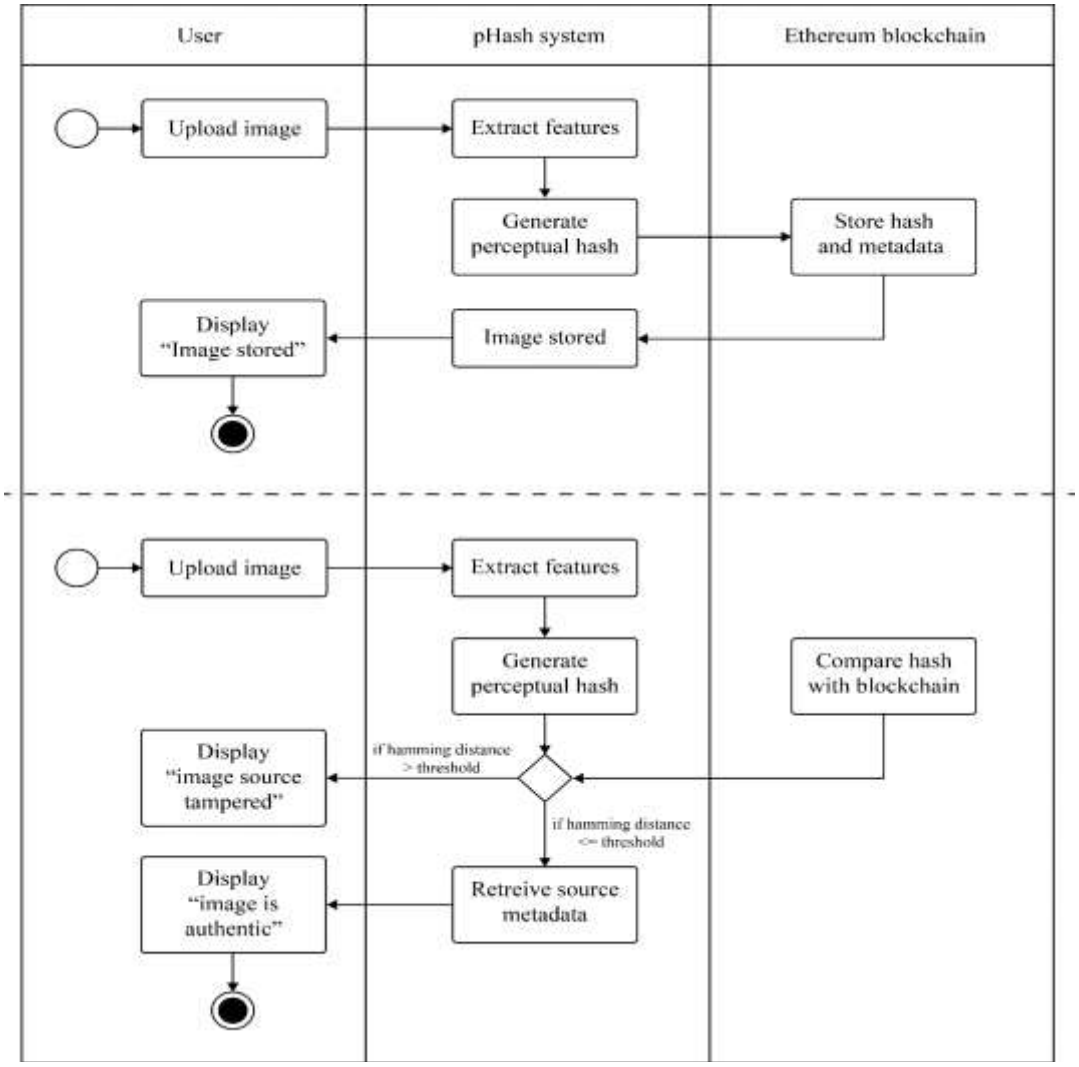



Figure 3.2 Class diagram showing the structure and relationships of classes, interfaces and interactions of the current system

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252




## IV. RESULTS AND DISCUSSION

The proposed system uses the Uncompressed Colour Image Database (UCID) for its image source detection feature, the dataset is available at https://nas.minelab.tw:5001/sharing/CXQLPrbCr

The aim of the UCID is to provide a benchmark dataset for image retrieval. The database has over 1300 images together with a ground truth (predefined query images with corresponding model images that should be retrieved). It is envisaged that the dataset is used for the evaluation of image retrieval techniques.

### 4.1 Seam Carving

Seam carving is an image processing technique used for content-aware resizing. It involves removing or adding seams (connected paths of pixels) in an image to resize it while preserving important content and features.

### 4.2 Quality Factor

Quality factor is often associated with image compression, especially in the context of JPEG compression. In JPEG compression, the QF determines the trade-off between image quality and file size. A higher QF generally results in better image quality but larger file sizes.

### 4.3 Percentage Value

The percentage value refers to a percentage of the original image size or scaling factor. It may suggest that the seam carving operation is performed to resize the image, and the resulting image size is reduced to a set percentage of the original size.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

| Test Case No. | Test Case Name | Precondition | Test Case Input | Test Case Output | Test Case Description |
|---|---|---|---|---|---|
| 1.0 | Seam carving Image tamper | Seam carving with quality factor of 100 (QF100) and a percentage value of 1% is applied on the UCID image dataset. | Store original UCID images (before seam carving) in the database | Original UCID images (before seam carving) are successfully stored in the database and ready for source detection with tampered images | **Input:** Load UCID images (after seam carving) for source detection **Output:** Loaded UCID images (with seam carving) is ready for source detection by the proposed system |

Table 4.1: Test Case 1.0 (Seam Carving Tampering at 1%)

| Test Case No. | Test Case Name | Precondition | Test Case Input | Test Case Output | Test Case Description |
|---|---|---|---|---|---|
| 2.0 | Seam carving Image tamper | Seam carving with quality factor of 100 (QF100) and a percentage value of 2% is applied on the UCID image dataset. | Store original UCID images (before seam carving) in the database | Original UCID images (before seam carving) are successfully stored in the database and ready for source detection with tampered images | **Input:** Load UCID images (after seam carving) for source detection **Output:** Loaded UCID images (with seam carving) is ready for source detection by the proposed system |

Table 4.2: Test Case 2.0 (Seam Carving Tampering at 2%)

| Test Case No. | Test Case Name | Precondition | Test Case Input | Test Case Output | Test Case Description |
|---|---|---|---|---|---|
| 3.0 | Seam carving Image tamper | Seam carving with quality factor of 100 (QF100) and a percentage value of 5% is applied on the UCID image dataset. | Store original UCID images (before seam carving) in the database | Original UCID images (before seam carving) are successfully stored in the database and ready for source detection with tampered images | **Input:** Load UCID images (after seam carving) for source detection **Output:** Loaded UCID images (with seam carving) is ready for source detection by the proposed system |

Table 4.3: Test Case 3.0 (Seam Carving Tampering at 5%)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

| Test Case No. | Test Case Name | Precondition | Test Case Input | Test Case Output | Test Case Description |
|---|---|---|---|---|---|
| 4.0 | Seam carving Image tamper | Seam carving with quality factor of 100 (QF100) and a percentage value of 10% is applied on the UCID image dataset. | Store original UCID images (before seam carving) in the database | Original UCID images (before seam carving) are successfully stored in the database and ready for source detection with tampered images | **Input:** Load UCID images (after seam carving) for source detection **Output:** Loaded UCID images (with seam carving) is ready for source detection by the proposed system |

Table 4.4: Test Case 4.0 (Seam Carving Tampering at 10%)

| Test Case No. | Test Case Name | Precondition | Test Case Input | Test Case Output | Test Case Description |
|---|---|---|---|---|---|
| 5.0 | Seam carving Image tamper | Seam carving with quality factor of 100 (QF100) and a percentage value of 20% is applied on the UCID image dataset. | Store original UCID images (before seam carving) in the database | Original UCID images (before seam carving) are successfully stored in the database and ready for source detection with tampered images | **Input:** Load UCID images (after seam carving) for source detection **Output:** Loaded UCID images (with seam carving) is ready for source detection by the proposed system |

Table 4.5: Test Case 5.0 (Seam Carving Tampering at 20%)

| Test Case No. | Test Case Name | Precondition | Test Case Input | Test Case Output | Test Case Description |
|---|---|---|---|---|---|
| 6.0 | Seam carving Image tamper | Seam carving with quality factor of 100 (QF100) and a percentage value of 30% is applied on the UCID image dataset. | Store original UCID images (before seam carving) in the database | Original UCID images (before seam carving) are successfully stored in the database and ready for source detection with tampered images | **Input:** Load UCID images (after seam carving) for source detection **Output:** Loaded UCID images (with seam carving) is ready for source detection by the proposed system |

Table 4.6: Test Case 6.0 (Seam Carving Tampering at 30%)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

| Test Case No. | Test Case Name | Precondition | Test Case Input | Test Case Output | Test Case Description |
|---|---|---|---|---|---|
| 7.0 | Seam carving Image tamper | Seam carving with quality factor of 100 (QF100) and a percentage value of 50% is applied on the UCID image dataset. | Store original UCID images (before seam carving) in the database | Original UCID images (before seam carving) are successfully stored in the database and ready for source detection with tampered images | **Input:** Load UCID images (after seam carving) for source detection **Output:** Loaded UCID images (with seam carving) is ready for source detection by the proposed system |

Table 4.7: Test Case 7.0 (Seam Carving Tampering at 50%)

| SN | Perceptual Hash | Filename | Timestamp | Hamming Distance | Remark |
|---|---|---|---|---|---|
| 1 | afafd450888d9d58 | | | 8 | Image source tampered |
| 2 | a6b3d1cc399d1ca4 | | | 4 | Image source tampered |
| 3 | c725d2970db698e4 | | | 6 | Image source tampered |
| 4 | ccef680f71862c2d | | | 10 | Image source tampered |
| 5 | 858598cb77536257 | | | 6 | Image source tampered |
| 6 | 858598cb77536257 | | | 20 | Image source tampered |
| 7 | 8a9ccd95e2464e6b | | | 10 | Image source tampered |
| 8 | 8c6393896cd76dc4 | | | 8 | Image source tampered |
| 9 | a5a5269224babeda | | | 20 | Image source tampered |
| 10 | a5a5269224babeda | | | 10 | Image source tampered |
| 11 | ccb3f3580ca2f392 | | | 14 | Image source tampered |
| 12 | e8f6d4f0914655d1 | | | 10 | Image source tampered |
| 13 | cceda7121a53c8dc | | | 16 | Image source tampered |
| 14 | d9dfcea49212073a | | | 10 | Image source tampered |
| 15 | cdcbca8ce543c6a4 | | | 4 | Image source tampered |

Table 4.8: Results from Test Case 5.0 (Table 4.5)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

| SN | Perceptual Hash | Filename | Timestamp | Hamming Distance | Remark |
|---|---|---|---|---|---|
| 1 | afafd452888dad50 | | | 8 | Image source tampered |
| 2 | a6b6d1d9399c1ca4 | | | 4 | Image source tampered |
| 3 | c524d2970fb698e9 | | | 6 | Image source tampered |
| 4 | ccef790e71862c29 | | | 10 | Image source tampered |
| 5 | 8585b8c37717625e | | | 6 | Image source tampered |
| 6 | 8a9ccd95f2464e69 | | | 20 | Image source tampered |
| 7 | c561938864f76ce6 | | | 10 | Image source tampered |
| 8 | a5a4369034babbda | | | 8 | Image source tampered |
| 9 | ccb3f34c0ca2f392 | | | 20 | Image source tampered |
| 10 | e8f2d5f0914555d1 | | | 10 | Image source tampered |
| 11 | dddfcea49010073b | | | 14 | Image source tampered |
| 12 | cfcb5a8cedc2c484 | | | 10 | Image source tampered |
| 13 | ec85e7f8002ffc60 | | | 16 | Image source tampered |
| 14 | e9ade0e9e92c1d06 | | | 10 | Image source tampered |
| 15 | a0e1de160169f8fe | | | 4 | Image source tampered |

Table 4.9: Results from Test Case 6.0 (Table 4.6)

| Images | 1% | 2% | 5% | 10% | 20% | 30% | 50% |
|---|---|---|---|---|---|---|---|
| Image 1 | 96.88 | 96.88 | 96.88 | 90.63 | 87.50 | 87.50 | 84.38 |
| Image 2 | 100.00 | 96.88 | 93.75 | 93.75 | 93.75 | 84.38 | 78.13 |
| Image 3 | 96.88 | 96.88 | 100.00 | 100.00 | 90.63 | 81.25 | 75.00 |
| Image 4 | 96.88 | 90.63 | 87.50 | 90.63 | 84.38 | 81.25 | 78.13 |
| Image 5 | 93.75 | 90.63 | 90.63 | 93.75 | 90.63 | 93.75 | 81.25 |
| Image 6 | 100.00 | 96.88 | 68.75 | 84.38 | 84.38 | 87.50 | 87.50 |
| Image 7 | 100.00 | 96.88 | 100.00 | 96.88 | 87.50 | 71.88 | 65.63 |

Table 5.0: Source detection result for Seam Carving at different tamper rates at QF100 using UCID images

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 573-584 www.ijaem.net ISSN: 2395-5252

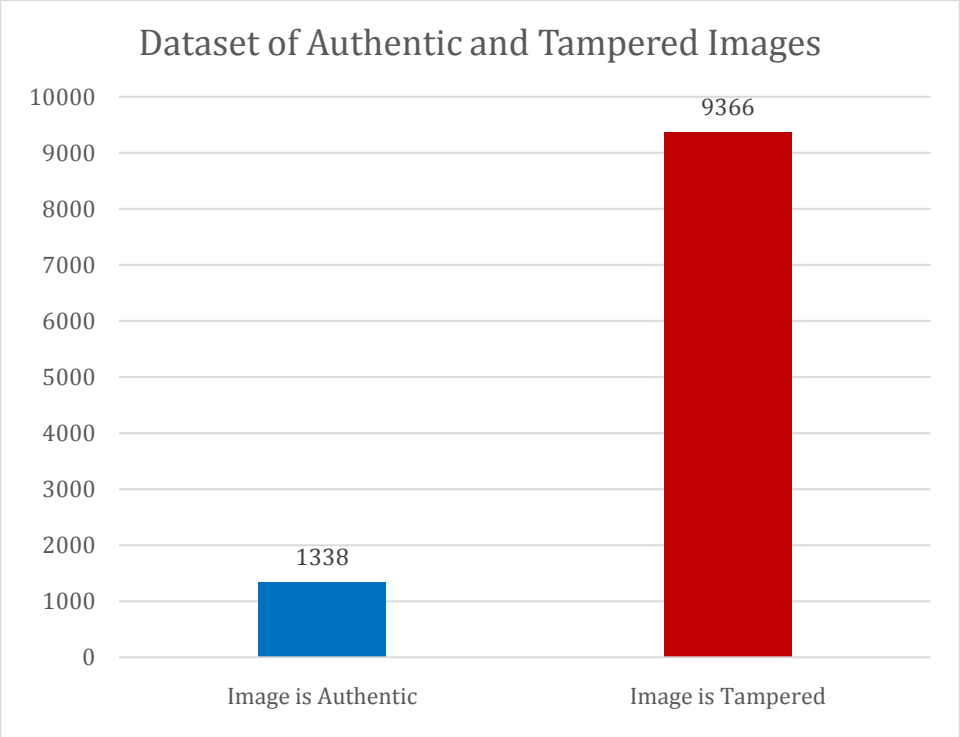



Fig 4.1: Authentic and Tampered Datasets

## V. DISCUSSION

The system was tested using the UCID image dataset, which comprises 1,338 authentic images and 9,366 tampered images, providing a comprehensive basis for evaluating the model's performance. Key metrics used in the evaluation include Accuracy, Precision, Recall, and F1 Score, which provide a detailed assessment of the model's effectiveness in detecting the source of images. The model achieved an accuracy of 91.30% with a percent error of 6.69%. The accuracy varied with the degree of tampering, with higher accuracy observed for low tampering rates (1% or 2%) compared to high tampering rates (50%). These results validate the effectiveness of the proposed model in identifying the source of digital images using perceptual hashing and blockchain technology.

## VI. CONCLUSION

The primary objective of this research was to establish a robust blockchain database capable of securely storing file hashes and associated metadata. Achieved through the creation and deployment of smart contracts using Solidity, our blockchain solution provides a decentralized and tamper-proof environment for file storage. This decentralized architecture ensures data integrity, immutability, and transparency, aligning with the core principles of blockchain technology. To enhance user interaction with the blockchain database, a well-designed and intuitive interface was crucial. Leveraging the capabilities of ReactJS, we successfully crafted a user-friendly program interface. Users can seamlessly upload, retrieve, and interact with files stored on the blockchain, creating a seamless bridge between the complexities of blockchain technology and a user-centric environment. In addition to blockchain integration, a key aspect of this project involved designing a flask server endpoint to implement perceptual hashing using Python. This endpoint allows for the efficient calculation of perceptual hashes for uploaded images by making a POST request, contributing to the overall functionality and versatility of the system. By choosing an appropriate perceptual hashing algorithm and designing a robust Python API, our research ensures the accurate representation of image content within the blockchain database.

A comprehensive testing and evaluation framework was implemented to validate the effectiveness and reliability of the developed system. By employing hamming distance as a metric for assessing perceptual hash similarity, we rigorously tested the functionality of the blockchain database, React interface, and Python flask server endpoint. The results of these tests not only verify the correctness

of the implemented components but also provide valuable insights into the system's performance under different scenarios.

## REFERENCES

[1]. Ahmed H, Traore I, Saad S. Detecting opinion spams and fake news using text classification,Security and Privacy, 2017;e9. https://doi.org/10.1001/spy2.9

[2]. Buntain, C., and Golbeck, J. (2017). Automatically identifying fake news in popular twitter threads. 2017 IEEE international conference on smart cloud (smartcloud). IEEE 208–215.

[3]. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52 (1): 1 – 4.

[4]. Dwivedi, A. D., Singh, R., Dhall, S., Srivastava, G., & Pal, S. K. (2021). Tracing the Source of Fake News using a Scalable Blockchain Distributed Network. In Proceedings of 2020 IEEE International Conference on Mobile AdHoc and Sensor Systems, 38 - 43. IEEE. https://doi.org/10.1109/MASS50613.2020.00015

[5]. Gilda, S. (2017). Evaluating machine learning algorithms for fake news detection. 2017 IEEE 15th student conference on research and development (SCORED). IEEE110–115.

[6]. Gottfried, J. and Shearer, E. "News Use Across Social Media Platforms 2016," 26 May 2016. [Online]. Available: https://www.pewsearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/.

[7]. Guacho, G. B., Abdali, S., &Papalexakis, E. E. (2018). Semi-supervised content-based fake news detection using tensor embeddings and label propagation. Proc. socal NLP symposium.

[8]. Haya, R. and Hasan, K. S. (2019). Combating deepfake videos using blockchain and smart contracts. 10.1109/ACCESS.2019.2905689.

[9]. Huckle S, White, M. (2017). Fake news: a technological approach to proving the origins of content, using blockchains. Big Data 5:4, 356–371, DOI: 10.1089/big.2017.0071

[10]. J.D. Wei, Y.J. Lin, Y.J. Wu, "A patch analysis method to detect seam carved images," in PatternRecognition Letters 36, 2014, 100 –106.

[11]. Kanoksak W., Timothy K.S., Senior Member, IEEE, Wen-Lung Chang and Hon-Hang Chang"Tamper Detection of JPEG Image Due to Seam Modifications." DOI 10.1109/TIFS.2015.2464776, IEEE Transactions on Information Forensics and Security

[12]. Morin, R. Trump: New York Times is 'fake news'. POLITICO. 17AD. Available online at http://politi.co/2jAdgwn (last accessed 13 March, 2023).

[13]. Nwachukwu, O.J. (2017). "Why I said President Buhari is dead-Ex British lawmaker, Eric Joyce" 26 May 2017. [Online]. Available: https://dailypost.ng/2017/05/26/said-president-buhari-dead-ex-british-lawmaker-eric-joyce/.

[14]. Okoro, E.M., Abara, B. A., Umagba, A. O., Ajonye, A. A. and Isa, Z. S. "A hybrid approach to fake news detection on social media" Nigeria Journal of Technology (NIJOTECH) 2 (37): 454 - 462, 2018

[15]. Ozbay, F.A. and Alatas, B. "Fake news detection within online socialmedia using supervised artificial intelligence algorithms", Physica A (2019), doi:https://doi.org/10.1016/j.physa.2019.123174.

[16]. PREMIS Editorial Committee. The PREMIS Data Dictionary Version 3.0. 2015. Available online at www.loc.gov/standards/premis/v3/premis-30-final.pdf (Last accessed 6 March, 2023).

[17]. Qian Chen, et al., Information Processing and Management, https://doi.org/10.1016/j.ipm.2020.102370. Recognition Letters 36, 2014, 100 –106.

[18]. Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. Proceedings of the 2017 ACM on conference on information and knowledge management. ACM797–806.

[19]. S. Mo Jang, TiemingGeng, Jo-Yun Queenie Li, Ruofan Xia, Chin-Tser Huang, Hwalbin Kim, Jijun Tang. (2018). A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis, Computers in Human Behavior. (doi: 10.1016/j.chb.2018.02.032

[20]. Saad, M., Ahmad, A., and Mohaisen, A. (2019). "Fighting fake news propagation with blockchains," in 2019 IEEE Conference on Communications and Network Security (CNS), 1–4.