Document Manipulation Detection and Authenticity Verification Using Machine Learning and Blockchain

Shantanu Sarode[1], Utkarsha Khandare[2], Shubham Jadhav[3], Avinash Jannu[4], Vishnu Kamble[5], Digvijay Patil[6]

[1,2,3,4]*Student, Dept. of Information Technology Engineering, P.E.S's Modern College of Engineering, Pune, Maharashtra, India*
[5,6]*Asst. Professor, Dept. of Information Technology Engineering, P.E.S's Modern College of Engineering, Pune, Maharashtra, India*

---***---

Abstract - *A system to detect the manipulated images and documents using neural networks and verify the authenticity of the individual's identity documents (Aadhar card, License, Passport, etc.) using Blockchain and Image Processing. In the present condition people forge the documents that is misuse the individual's identity documents and images in illegal ways and also various online portals require the repeated submission of identity documents for various purposes which is time consuming. So to avoid such illegal use the system helps in detecting such manipulated images and documents and helps in reducing the repeated submission of documents. The problem with the existing system is that they detect manipulated documents and images on the basis of metadata of the particular document which are edited through specific manipulating methods like splicing and coloring using simple editing softwares such as Adobe Photoshop, etc. So the proposed system will help in detecting such manipulated documents by using the concept of Neural Network under Image Processing through Error Level Analysis method and also will help in reducing the repeated submission of documents by providing a method of submission of documents at least once that is by verifying a single document and providing information related to other documents that is simply checking the authenticity of the document as well as user by using the concept of Blockchain. The system is basically used in Organizational, Institutional and Government Systems as well as Student related fields such as Government Scholarship portals, e-KYC procedures, University Certificate Verification and Notary systems.*

Keywords: **Document, Blockchain, security, Image Processing, Decentralized, ELA, Neural Network.**

1. INTRODUCTION

Developing a system to detect the manipulated images using neural networks and verify the authenticity of the individual's identity documents (Aadhar card, License, Passport, etc.) using Blockchain and Image Processing. Nowadays, many people forged that is misuse the individual's identity documents and images in illegal ways. So to avoid such illegal use the system helps in detecting such manipulated images and documents. Also, in the present condition various online portals require the repeated submission of identity documents for various purposes which are inconsistent and time consuming. So the system would help in reducing the repeated submission of documents.

Nowadays, people edit the documents or images using simple editing softwares such as Adobe Photoshop, etc. The documents or images are edited by coloring or splicing some parts of the documents as well as many people forged the individual's identity documents such as Aadhar card or pan card by just adding a text field on the unique id of the document and giving it a fake id which is not possible to be catched by human eye. Also these documents have to be uploaded on various web portals repeatedly, and if the document is not correct then it has to be submitted again which is time consuming. So to detect such forged documents and reduce repeated submission of documents the proposed system is very relevant and useful.

The problem with the existing system is that they detect manipulated documents and images which are edited through specific manipulating methods like splicing and coloring. There are some software's which help in making alterations to the documents which are not possible to be catched by human eye. So the proposed system will help in detecting such manipulated documents which are used in improper ways by using the concept of Neural Network under Image Processing. Also the second problem which will be solved through the system is that repeated submission of documents, so the proposed system will help in reducing this factor by providing a method of submission of documents at least once that is by verifying a single document and providing information related to other documents that is simply checking the authenticity of the document as well as user by using the concept of Blockchain. The system is needed at the time of submission of individual's identity documents on various web portals like Scholarship and Educational systems where it checks whether the document is real or not and then if found real then on the basis of the submitted document submits the information related to other required documents on the web portal. So this system is needed in such cases where the user submits the forged that is manipulated documents on the web portal.

1.1 Literature Review

In the existing system the algorithms and the techniques used were complex and difficult to understand and results were not much accurate and consistent and also there were security issues generated while storing the documents. The proposed system consists of methods and techniques which have accurate and consistent results and also are easy to understand. Complexity in the solving the problems is reduced due to consideration of the proposed system.

The problem with the existing system is that they detect manipulated documents and images on the basis of metadata of the particular document which can be edited through specific manipulating methods using simple editing softwares. Also in existing system, various web portals require repetitive submission of documents due to some problems in verification of documents.

So the proposed system will help in detecting such manipulated documents by using the Convolutional Neural Network under Image Processing with Error Level Analysis method and also will help in reducing the repeated submission of documents using concept of Blockchain. In the existing systems the algorithms and the techniques used were complex and difficult to understand and results were not much accurate and consistent and also there were security issues generated while storing the documents. The proposed system consists of algorithms and techniques which have accurate and consistent results and also are easy to understand. Complexity in the solving the problems is reduced due to consideration of the proposed system.

TECHNIQUES	TRADITIONAL	PROPOSED
Document Storage Type	Central Server	Distributed Server
Document Storage	Document PDF / Image	Document Hash Value
Document Manipulation Detection	Through Metadata	Through Error Level Analysis & Neural Network
Document Security	Less Security	More Security
Accuracy of Prediction	Medium	High

Table -1: Comparison of Techniques

2. DESIGN AND ARCHITECTURE
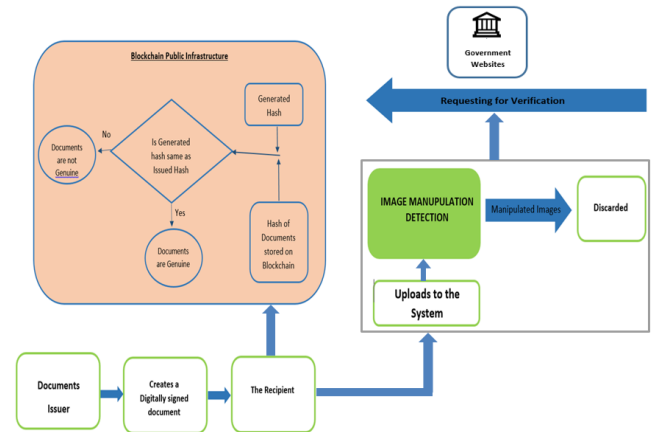
2.1 Architecture of the System

Figure 2.1: Architecture of the system

The above diagram specifies the architecture of the proposed system. It specifies the integration of the two modules: Image Processing and Blockchain. The proposed system works in a manner such that when a user visits any online web portal like scholarship portal when he submits a document for uploading the system checks whether the document is forged that is manipulated or not by passing it to the Image Processing module for scanning through Neural Network by Error Level Analysis method. So the document image is passed to two level analysis method because the metadata of a document image can be changed through some softwares so for more accuracy it is passed through Neural Network analysis. Once the document is identified as real, it is passed to the second module implemented through Blockchain.

First a user space is created by the user through Blockchain as per Government aspects where the user stores the hash values of the authenticated individual's identity documents by scanning the document to avoid repetitive references to the documents which also provide the security to the documents information through Blockchain. So while submitting the document the second module scans the document and a hash value of that document is obtained and compared with the hash value of document stored in the Blockchain space and if it is same then provides the information related to the other documents and hence completes the document manipulation detection and authenticity verification process of the particular web portal.

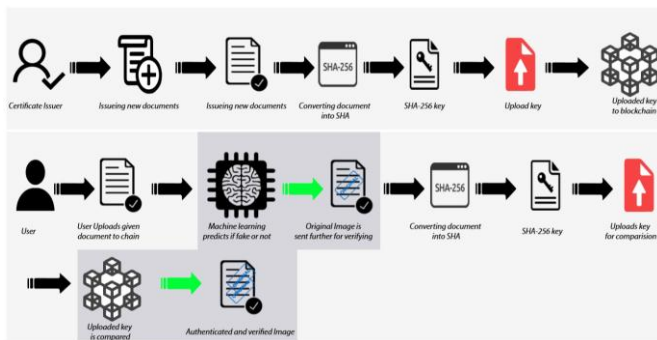## 2.2 Workflow of the System

Figure 2.2: Workflow of System

The diagram above specifies the workflow of the proposed system that is how the controls flows from one sub module to another sub module and completes the two step verification process for Document Manipulation Detection and Authenticity Verification.

3. Technologies used in proposed system

The domains included in the proposed system include:
[1] Image Processing
[2] Blockchain

Following are the latest techniques used for getting accurate and efficient results:
[1] Error Level Analysis
[2] Convolutional Neural Network

The sections under technologies include:
[1] Image Processing for Document Manipulation Detection
[2] Blockchain for Document Authenticity Verification
[3] Algorithms of proposed system

The proposed system specifies an idea in the form of a two-step verification process for detecting whether the document uploaded by the individual user is authentic for the desired purpose.

3.1 Image Processing for Document Manipulation Detection

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/ features associated with that image.

Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too. In medical field physicians and researchers make diagnosis or research based on imaging. Multimedia technology consists of both images and video. With the increase of free cost and availability of advanced image manipulation tools everyone can gain privileges and show their ability to tamper the image in order to falsify it. Image tampering is a digital art which needs understanding of image properties and image manipulation abilities. One tampers images for various reasons either to enjoy fun of digital works or to produce false evidence.

In recent year, devising and deploying effective approaches to detect the authenticity of digital images is the new field of forensic science, which cases are increased yearly. Usually the image forensic focuses on the following issues: Confirm origin of the image: judge that the image comes from the specific imaging device or is generated by computer. If the image is obtained by the device, determine the devices a reference parameters including imaging equipment types, time, location, etc.

Image Tampering Techniques:
[1] Copy-Move (Cloning)
[2] Image Splicing
[3] Noising or Blurring
[4] Blending

Image Manipulation Detection Techniques:
[1] Metadata Analysis
[2] Error Level Analysis
[3] Watermarking
[4] Digital Signature
[5] Principal Component Analysis

Advantages of Image Processing:
[1] Images can be automatically sorted.
[2] It is used to analyze medical images.
[3] Unrecognizable Features can be made prominent.
[4] It is used to analyze cells and their composition.
[5] Minor errors can be rectified
[6] It allows robots/Self-Driven Cars to have a Vision
[7] It allows industries to remove defective products

The Image Processing Module basically includes of two parts: Error Level Analysis and Neural Network. These parts in combination help to detect whether the document image is manipulated by any means or not.

3.2 Blockchain for Document Authenticity Verification

As the Blockchain module is going to be used for various authenticity checking of the documents and it includes various security hash functions. Integration of these modules creates an web application which uses both the modules for a single way of use for increasing the efficiency of the certain problems.

The Architecture of the system explains the whole system flow that is how the system is going to process the input and

output of the system. So the Project Architecture explains how both the Image Processing module and Blockchain module work together. It also explains where the actual implementation of the Blockchain and Image Processing technologies is done.
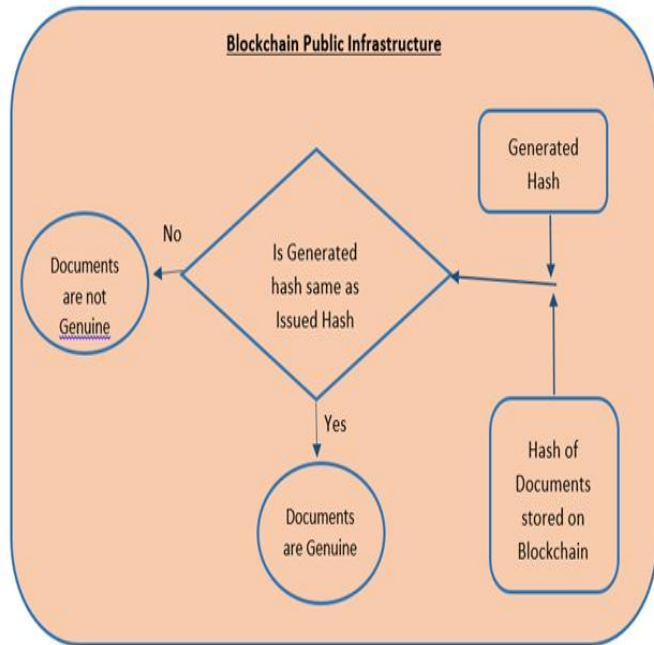

Figure 3.1: Blockchain Storage Architecture

Drawbacks of Existing System:

In the existing systems the documents were usually stored in a centralized database at a single location which creates problems such as:

[1] System Crash
[2] Data loss
[3] Less data security
[4] Load on single server
[5] Less efficiency in retrieval and storage of data
[6] Results are not accurate that is not 100 percent perfect
[7] Less accuracy for image manipulation detection

So the above points explain how the existing system has drawbacks in its implementation in real life.
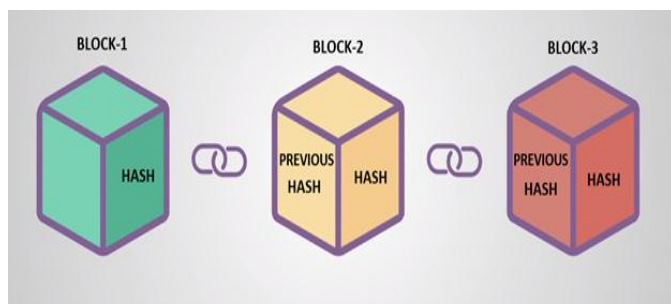

Figure 3.2: Simple Blockchain Structure

Deployment phase of the system is the main part that is how the system is to be used in the real life. A web application of the system is created which runs at the back end of the whole system.

Two parts in deployment module:

[1] Image Processing Module: In this module the software which detects the manipulated documents runs at the back end of the particular web portal.
[2] Blockchain Module: In this module a central authority stores the hash values of the particular individual as identity documents.

Real Life Applications:

The topic explains the real life applications related to Document Manipulation Detection and Authenticity Verification, that is how the system can be used in various ways in various fields. The applications are:

[1] Government Scholarship Websites
[2] e-KYC Procedures
[3] University Certificate Verification
[4] Notary Systems

3.3 Algorithms of Proposed System

Neural Network:

Neural Networks is generally a part of deep learning approach. Neural Networks is a class of deep neural networks, most commonly applied to analysis and recognition of visual images. Neural Networks use relatively little preprocessing. Neural Networks can automatically learn the characteristics of images.

Structure of Convolutional Neural Network:

A Convolutional Neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a Neural Network typically consist of: Convolutional Layer, RELU layer, Pooling Layer, Fully Connected Layers.
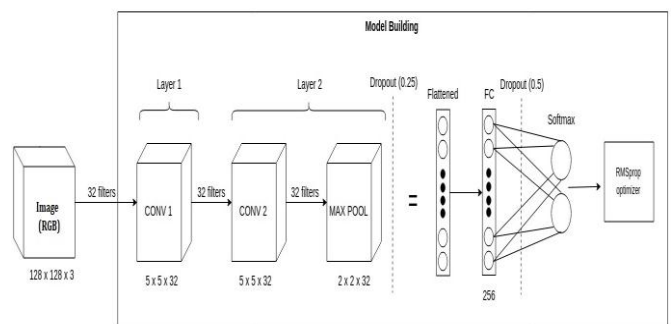

Figure 3.3: Model Structure

Error Level Analysis Algorithm:

Error level analysis (ELA) works by intentionally resaving the image at a known error rate, such as 95 percent, and then computing the difference between the images. Even with a complex neural network, it is not possible to determine whether an image is fake or not without identifying a common factor across almost all fake images. So, instead of giving direct raw pixels to the neural network, we gave error level analyzed image.
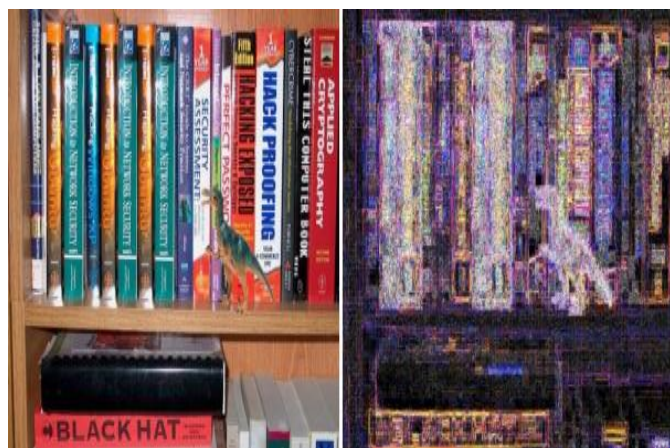

Figure 3.4: Original and ELA Image
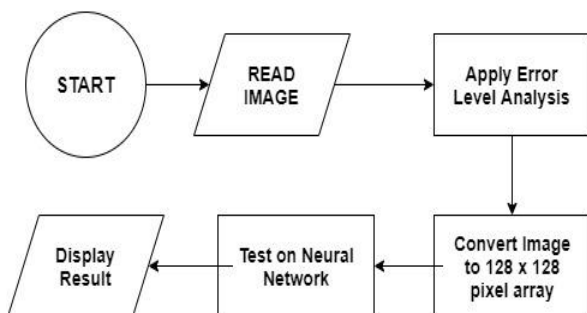
Integration of Neural network and ELA:

Figure 3.5: Flow of Image Processing Module

SHA-256 Algorithm for Blockchain:

For storage of hash values of documents in the Blockchain the Secure hashing algorithms are used. There are two SHA Algorithms: SHA256 and SHA 512.

SHA-256 Characteristics:

One-way function: That is from the initial message; it is easy to create a hash value. However, from the hash value, there is no way to restore the original message.

This algorithm accepts a message M of any length as the input and delivers 256 bit hash value. Blockchain network can withstand the attack by linking blocks together using a cryptographic hash function.

4. RESULTS

The proposed system is able to detect whether the document image is authentic or not in an accurate and efficient manner. Combination of Image Processing and Blockchain works together very efficiently and the results obtained are accurate. The Image Processing includes: Error Level Analysis and Neural Network which gives the high accuracy and Blockchain ensures high data security for the proposed system through SHA algorithm.

5. CONCLUSION

This paper proposes the design, idea, architecture and solution of an advanced as well as secured web application for detecting whether the document image is authentic or not using the two-step verification process named: Document Manipulation Detection and Authenticity Verification. A simple, affordable, easy to use and most secured system is proposed to solve the document forgery issues, document security and storage issue, repetitive submission of document issue, and illegal use of fake document issue.

REFERENCES

[1] Osmal Ghazal and Omar S. Saleh, "A Graduation Certificate Verification Model via Utilization of the Blockchain Technology", School of Computing, Universiti Utara Malaysia, (2016)

[2] Khuat Thanh Son, Nguyen Truong Thang, Le Phe Do, and Tran Manh Dong, "Application of Blockchain Technology to Guarantee the Integrity and Transparency of Documents", Institution of Information Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam, (2018)

[3] Teddy Surya Gunawan, Siti Amalina Mohammad Hanafiah, Mira Kartiwi, Nanang Ismail, Nor Faradilah Za'bah, Anis Nurashikin Nordin, "Development of Photo Forensics Algorithm by Detecting Photoshop Manipulation using Error Level Analysis", Indonesian Journal of Electrical Engineering and Computer Science (2017)

[4] M. Wang, M. Duan, and J. Zhu, "Research on the Security Criteria of Hash Fucntions in the Blockchain", In proceedings of BCC'18, Incheon, Republic of Korea, (2018)

[5] Nor Bakiah Abd Warif, Mohd. Yamani Idna Idris, Ainuddun Wahid Abdul Wahab, Rosli Salleh, "An Evaluation of Error Level Analysis in Image Forensics",

Image Forensics, JPEG compression, Error Level Analysis (2015)

[6] Hak-Yeol Choi, Han-UI Jang, Dongkyu Kim, Jeongho Son, Seung-Min Mun, Sunghee Choi, Heung-Kyu Lee, "Detecting Composite Image Manipulation based on Deep Neural Networks", School of Computing, KAIST, Daejeon, South Korea (2017)

[7] Ms. Jayashri Charpe, Ms. Antara Bhattacharya, "Revealing Image Forgery through Image Manipulation Detection", Proceedings of 2015 Global Conference on Communication Technologies(GCCT 2015)

[8] Muhammed Afsal Villan, Kuncheria Kuruvilla, Johns Paul, Prof. Eldo P Elias, "Fake Image Detection using Machine Learning", Image Forensics, Metadata Analysis, Error Level Analysis, Deep Neural Networks (2017)

[9] https://www.blockchainexpert.uk/blog/document-verification-using-blockchain.

[10]   https://www.martinstellnberger.co/document-certification-through-the-blockchain

[11]   https://ieeexplore.ieee.org/document/6625374

[12]   https://towardsdatascience.com/introduction-to-artificial-neural-networks-ann-1aea15775ef9

[13]   https://www.analyticsvidhya.com/blog/2014/10/ann-work-simplified/

[14]   https://design.tutsplus.com/tutorials/10-techniques-that-are-essential-for-successful-photo-manipulation-artwork--psd-242

[15]   https://iopscience.iop.org/article/10.1088/1742-6596/1193/1/012013/pdf

[16]   https://blog.paperspace.com/intro-to-optimization-in-deep-learning-gradient-descent/

[17]   https://blog.paperspace.com/intro-to-optimization-in-deep-learning-gradient-descent/

[18]   https://www.digitalocean.com/community/tutorials/how-to-deploy-python-wsgi-apps-using-gunicorn-http-server-behind-nginx

[19]   https://towardsdatascience.com/simple-way-to-deploy-machine-learning-models-to-cloud-fd58b771fdcf

[20]   https://docker-curriculum.com/

BIOGRAPHIES

Shantanu Sarode
Student at Dept. of Information Technology, P.E.S's Modern College of Engineering, Pune, Maharashtra, India

Utkarsha Khandare
Student at Dept. of Information Technology, P.E.S's Modern College of Engineering, Pune, Maharashtra, India

Shubham Jadhav
Student at Dept. of Information Technology, P.E.S's Modern College of Engineering, Pune, Maharashtra, India

Avinash Jannu
Student at Dept. of Information Technology, P.E.S's Modern College of Engineering, Pune, Maharashtra, India

Vishnu. S. Kamble
Asst. Professor at Dept. of Information Technology, P.E.S's Modern College of Engineering, Pune, Maharashtra, India

Digvijay. A. Patil
Asst. Professor at Dept. of Information Technology, P.E.S's Modern College of Engineering, Pune, Maharashtra, India