**Blockchain-Based Certificate Validation System**

1 Shivam Panjolia, 2 Priyatosh Kumar, 3 Arpit Yadav

1 School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab

shivampanjolia8@gmail.com

2 School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab

priyatoshkumar11@gmail.com

3 School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab

Arpityadav7021@gmail.com

***Abstract*-This paper integrates machine learning and blockchain technology in a certificate validation system. A two-pronged ML model was created: one for doing the validation of certificate elements (ID, type, information on the recipient and issuer, dates and hash), while the other detects tampering in visual elements-signature, logo, watermark, and QR code of hash-while cross-verifying the name, through OCR. The front-end presents two modules: "Issue Certificate," which allows new certificate creation with particular metadata, together with embedding visual security elements (signature, logo, watermark, QR code), and "Verification," which has the certificate being anchored to the Sepolia test network through IPFS and Pinata. One can upload existing certificates to integrate features and put them into the blockchain. The "Validation" module validates certificates by retrieving metadata from the blockchain and applies the first ML model. The second ML model studies the uploaded certificate image for inconsistencies found on visual elements, and OCR checks for name mismatches with regard to metadata. This unique solution adds to the security and authenticity portion in certificate validation.**

***Keywords-:*** ***Certificate Validation, Blockchain, Machine Learning, Metadata Analysis, Anomaly Detection, Signature Verification, Logo Verification, Watermark Verification, QR Code, OCR, IPFS, Sepolia Test Network.***

1. INTRODUCTION

Today, probably every organization depends on issuing and authenticating different types of certificates to prove achievements or competencies. The management of these credentials needs to be efficient and secure. All traditional methods are manual, heavy, laborious, and hard to maintain fraud-free while keeping any possible scalability reduced. To fill these gaps, the paper presents an innovative approach on an organizational credit-or merit-based framework in the form of a certificate management and validation system that employs both blockchain and machine learning (ML).

This will most probably enable a one-stop solution: Work with the entire lifecycle from issuance up to verification of the certificates. All issued certificates will have a unique hash that is recorded and stored using the inherent security and immutability of the blockchain technology ensuring its authenticity and preventing tampering. More so, the system uses two different specialized ML models in the validation process. The first model would be focusing on the structured metadata associated with each certificate-including recipient information, issuer, certificate types, and validity periods-to set out the base validation of the certificate. In contrast, the second model has its focus on the actual visual integrity of the certificate using newly developed techniques for tampering and mismatch/noncompliance detection in also critical constituents like signatures, logos, watermarks, and embedded QR codes carrying the blockchain hash. Optical Character Recognition (OCR), too, aligns the text info on the image of the certificate with its metadata.

Features of the new proposed system include a user-friendly interface for the organizational users. The "Issue Certificate" module is used to create and issue fresh digital certificates while automatically embedding visual security features and associating them with complete metadata-before anchoring them to the blockchain. A very interesting function of the module is that it allows preexisting certificates to be incorporated into the system. Organizations can scan and upload their historical records, and the ML models will analyze them for probable artifacts. Validated preexisting certificates can then be linked with metadata, and hashed to the blockchain to bring them under the roof of the system for security management. Last but not the least, the system will facilitate the management of the credit/merit-based system, which is really an important thing for organizations since it has potential to link points or merit value with the certificates that are issued. The "Verification" module will present a strong mechanism for verification of any certificate for the system, newly issued or older. All such certificates will have their blockchain records checked and will undergo analysis for both metadata and visual integrity based on ML for validation. This will pool together a comprehensive solution that gives organizations a safe, efficient, as well as auditable way for managing and validating their valuable credentials and merit-based achievements.:

1. EXISTING WORK

Many researchers in the past have worked in the areas of certificate validation on the blockchain and document forgery detection through machine learning. These works create the groundwork toward a robust system integrating metadata validation, visual anomaly detection, and decentralized storage. A brief depiction of some of their work goes as follows:

Kumutha Parthiban and Jeyalaksshmi [1] reviewed the application of blockchain technology for academic certificate authenticity. Their work emphasized the use of blockchain’s immutable ledger to store certificate metadata securely, ensuring tamper-proof academic records. The system leverages distributed consensus to enhance trustworthiness and reduce fraudulent activities in certificate issuance and verification processes.

Mrs. R. Suganthalakshmi et al.[2] proposed a blockchain-based certificate validation system utilizing the Ethereum platform. Their approach involves issuing certificates with cryptographic hashes stored on the blockchain, enabling verifiers to authenticate certificates by comparing the hash of the presented document with the blockchain record. The system incorporates smart contracts to automate validation and improve transparency in certificate management.

Mohammed Abdulbasit Ali Al-Ameri et al.[3] developed an unsupervised forgery detection method for documents based on a network-inspired clustering approach. This technique identifies anomalies in visual elements such as signatures and logos without requiring labeled forgery data. Their experiments on scanned document datasets achieved 92% accuracy in detecting forgery regions, demonstrating robustness and adaptability for certificate validation applications.

Garima Jaiswal et al.[4] introduced a deep learning framework employing Convolutional Autoencoders (CAEs) for document forgery detection. The method preprocesses scanned images to enhance ink patterns and text alignment before feature extraction and classification. Their model effectively detects tampering and mismatched ink patterns, contributing to improved forgery identification accuracy.

A. Gayathiri et al.[5] presented a blockchain-based certificate validation system integrating QR codes and cryptographic hashing (SHA-512) for secure certificate issuance and verification. The system stores certificate metadata and hashes on the blockchain via smart contracts, enabling tamper-proof verification by matching the hash of the submitted certificate with the blockchain record. Their work highlights enhanced security and usability in academic credential verification.

M. Y. Kubiy et al.[6] proposed CertLedger, a novel Public Key Infrastructure (PKI) model leveraging blockchain for certificate transparency. By recording TLS certificates on an immutable ledger, CertLedger addresses vulnerabilities in traditional PKI systems, such as split-world attacks, and improves transparency and trust in certificate issuance and revocation.

R.A. Huber Jr. et al.[7] developed tamper detection techniques for identification documents using image enhancement and deep learning models, including Convolutional Neural Networks (CNNs). Their approach detects document alterations such as cutting, copying, and erasing, achieving 94.5% accuracy in an IEEE competition dataset. The model’s preprocessing and feature extraction steps enable effective forgery detection applicable to certificates.

S.B. Samanthula et al.[8] designed CertiSafe, a blockchain-based certificate safety system aimed at preventing counterfeiting. The system generates unique certificate identifiers stored on the blockchain, ensuring secure validation through blockchain’s immutability. Tested on academic certificate datasets, CertiSafe demonstrated its potential to reduce fake certificate issuance and support educational institutions in adopting secure verification frameworks.

1. METHODS AND TECHNIQUES

This project dealt with the construction of a machine-learning and blockchain-assisted certificate validation system. The system accommodates three machine-learning models- XGBoost Classifier for metadata validation, MobileNetV2 CNN for visualization of anomalies, and a computer-vision template-matching technique, with blockchain technologies such as Ethereum Sepolia testnet and IPFS (through Pinata) for secure storage and verification. The front end has two different modules: Certificate issuance and Certificate verification, thus forming a robust and user-friendly solution for ensuring authenticity.

* Machine Learning Models
* Metadata Validation Model

The first ML model assesses the validity of certificate

Methodology: Good old XGBoost Classifier. manage structured, multi-featured data well and are prone to less overfitting. The training has been done on a labeled data set of legitimate and counterfeit certificates' metadata, and the following features have been extracted: string similarity scores date consistency checks, and hashing integrity).

Training Process: The data will be split into 80% for training and 20% for testing. Hyper-parameters have been optimized through grid search using Fitting 5 folds for each of 144 candidates, totaling 720 fits to maximize accuracy. See Fig.1.

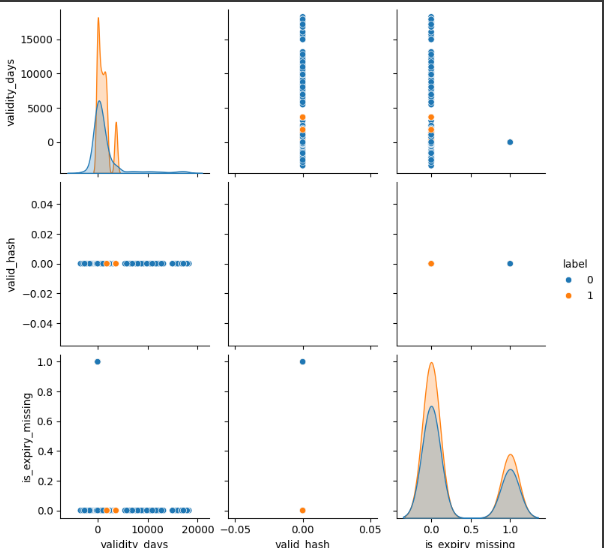
|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | **Hyperparameter** |  |  | | --- | |  | | **Value** |
| classifier\_\_colsample\_bytree | 0.8 |
| classifier\_\_learning rate | 0.05 |
| |  | | --- | |  |   classifier\_\_subsample | 5 |
| classifier\_\_n\_estimators | 100 |
| classifier\_\_subsample | 0.8 |

Fig .1

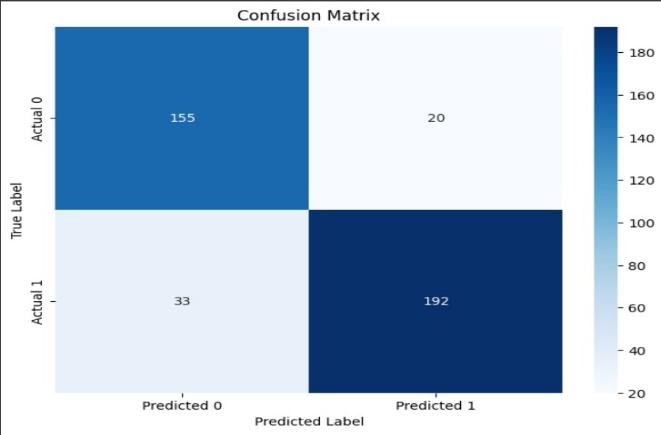
Hyperparameter optimized using GridSearchCV on the XGBoost model inside an imbalanced pipeline

Performance: 88% of the test data were classified correctly. The model showed a balanced performance across the valid and invalid certificate classes. For class 0 (valid), the precision was 0.82, recall of 0.89, and F1-score of 0.85; for class 1 (invalid), the precision was 0.91, recall of 0.85, and F1-score of 0.88. The overall macro and weighted averages of the performance were the same: precision, recall, and F1-score all at 0.88.

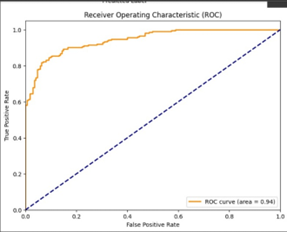
Application: During validation, metadata extracted from a certificate (using OCR or by manual input) are passed to the model and classified in binary as either valid or invalid, therefore ensuring the metadata conforms with defined expected patterns.

 Fig.2

pair plot of the Metadata’s Extracted features

Fig .3

confusion matrix having Low false positive (20)

Fig.4

ROC curve showing high true positive rate (TPR) and low false positive rate (FPR)

* Visual Anomaly Detection Model (MobileNetV2 CNN)

The second machine learning model detects the signs of tampering in visual elements-signatures, logos or hallmarks, watermarks, and QR codes encoding the hash-in images from certificates.

- Methodology: MobileNetV2, a lightweight CNN pre-trained on ImageNet, is now fine-tuned for this purpose, as it delivers an excellent performance and highly accurate processing for image classification. The model "learns" what "normal" patterns of visual elements look like from the authentic certificates and detects anomalies (for example, distorted signatures and fake logos) as deviations.

- Training Process: Application of synthetic forgeries (for example, altered signatures and blurred watermarks on the custom created certificates ) with a combination of rotation, scaling, and Gaussian noise enhancement in training process of the model. The model is trained with categorical cross-entropy loss, Adam optimizer (learning rate=0.001), and a batch size of 32, achieving an accuracy of 96.88% on the validation set.

- Compatibility with OCR: Text (recipient name and further information) is extracted from images of certificates through Tesseract OCR, compared with other metadata, thus enriching itself for further progress in the detection of anomalies.

- Performance: The model thus gives a score of 96.88% accuracy, judged against a confusion matrix and metrics on anomaly detection rate, based on binary classification for each visual element, tampered/not-tampered.

- Application: While examining certificate images, the model would point out those specific areas that appear tampered with and cross-references it with the metadata to validate the text extracted from OCR.

|  |  |
| --- | --- |
| Component | **Details** |
| Base Model | MobileNetV2 (pre-trained on ImageNet) |
| Input Image Size | 224 × 224 × 3 |
| Base Model Trainable | No (frozen during training) |
| Data Augmentation | Rotation (±10°), Zoom (±10%) |
| Rescaling | Yes (1./255) on all images |
| Additional Layers | GlobalAveragePooling2D → Dense(64, ReLU) → Dropout(0.3) → Dense(1, Sigmoid) |
| Loss Function | Binary Crossentropy |
| Optimizer | Adam |
| Metrics | Accuracy |
| Training Epochs | 10 |
| Batch Size | 32 |
| Class Mode | Binary |
| Training Directory | dataset/train |
| Validation Directory | dataset/val |

Fig .5

CNN Architecture and Training Configuration

* Template-Matching Model (Computer Vision)

A third model is to apply template matching to verify signatures and logos against their corresponding references.

- Methodology: A threshold of 0.8 is set to determine a match, which could be presence or similarity in comparing certificate signatures and logos with genuine reference templates using OpenCV's `cv2.matchTemplate` with normalized cross-correlation.

- Training Process: Templates for this model come from a small handpicked pool of authenticated certificates, for example, issuer logos; standard signatures. There is not much training needed, as it's deterministic.

- Performance: However, the model may not generalize well as it works with exact matching, although this would be high precision. Thus, it can serve as a baseline for CNN's probabilistic approach.

A certificate with qr code

AI-generated content may be incorrect.- Application: This model confirms that expected visual elements are actually present during verification and serve as a complement (deterministic check) to the MobileNetV2 CNN.

Fig .6

Template Matching Output using CV

* Blockchain Integration

The certificates can be stored and verified in an immutable manner within the system using blockchain technology.

- Store: The images of the certificates are uploaded to IPFS via Pinata. That way it is available decentralized and preserved forever. Metadata and SHA-256 hash from the image and metadata were stored using a smart contract in the Ethereum Sepolia test net.

Smart Contract: Written in Solidity, it manages the issuance of certificates and validation. It stores the hash, metadata, and CID relating to IPFS content. It has methods for adding certificates (issueCertificate) and retrieving data (verifyCertificate).

Method: When a certificate is issued, it is uploaded onto IPFS, a hash is generated, and the smart contract records the data at the Sepolia testnet. Otherwise, during verification, the hash of the uploaded certificate is recomputed and compared with the chain record to check for alterations.

Safety: Sepolia testnet, a cheap Ethereum test ground, provides immutability, and IPFS diversifies and makes it accessible.

* Front End Modules of Two web-based front end modules built with React.js help the user interact with the system.
* Issue Certificate Module

This module enables issuers to create new certificates or complement already existing ones with security features and blockchain integration. The procedure involves:

New Certificates: Users fill out the metadata fields using forms (name, e.g., "recipient\_name," "issuer\_id"). A custom template will embed these fields onto the certificate image as text overlays (using, for example, Pillow to position, "certificate\_id" at the top, "recipient\_name" below). The following visual security elements will then be added: signature (digital or pre-stored), logo (issuer hallmark), watermark (semi-transparent pattern), and QR code (encoding the hash).

Pre-existing certificates such as Upload Existing Certificate: A user uploads an already existing certificate image. Tesseract OCR extracts parallel text patterns from that (e.g. recipient name, dates) with a consistency check with user-provided metadata (e.g., using string if similar). If it is consistent, a watermark, logo, signature, and QR code will be overlaid on the image; if it is inconsistent, an exception will be flagged.

Create a unique "certificate\_id" and "uuid" for each certificate.

Uploads the finalized certificate to IPFS using Pinata and records the metadata and hash on Sepolia Testnet using the smart contract.

Security measures include cross-platform OCR, performing consistency checks while issuing cards to ensure that the metadata conforms to visible text and does not allow false entries. The custom template and overlays are understood in introducing verifiable features, which build resistance to tampering.

Output: A template-based certificate downloadable with embedded security features and a blockchain record. For already existing certificates, the output is the enhanced image with overlaid elements and updated metadata.

* Verification Module

Functionality: The user verifies the certificate by uploading an image, which enables the performance of multi-layered authenticity checking:

Hash Comparison: The hash of the uploaded certificate is first calculated using SHA-256 and compared to the hash stored on the blockchain in Sepolia testnet.

Metadata Validation: A Random Forest Classifier (with 88% accuracy) analyzes the metadata extracted through OCR or manual methods for valid/invalid classification.

Visual Anomaly Detection: Using a MobileNetV2 CNN and template-matching model, signatures, logos, watermarks, and QR codes will be assessed for the presence of tampering; the former gives an accuracy of 96.88%.

OCR Cross Check: Tesseract OCR extracts text (recipient names, etc.) from the image which is matched with the existing metadata with the idea of consistency checking.

Output: A detailed report on authenticity, with tags for hash mismatches, inconsistencies in metadata, and evidence of visual tampering. In the case of pre-existing certificates, the results of verification would be based on comparison of the hash in the blockchain and alignment of OCR/metadata.

* System Workflow

That is New User: Metadata inputs from one or uploads one's certificate.

For new certificates, a custom template embeds metadata; for pre-existing certificates, OCR obtains text, earns consistency checking per metadata, overlays watermark, logo, signature, and QR code if validated.

Finalized elements added visually and metadata.

Certificate is stored on IPFS, and metadata/hash recorded on Sepolia testnet.

Verification:

The user uploads a certificate.

Comparison of hashes against a blockchain record.

Metadata is validated using Random Forest (88% accuracy).

Visuals are verified with the help of MobileNetV2 CNN (96.88% accuracy) and template matching.

Subsequently, OCR ensures that text is consistent with the metadata. The results are output on the frontend. Below is Provided a tentative workflow for the proposed technique.

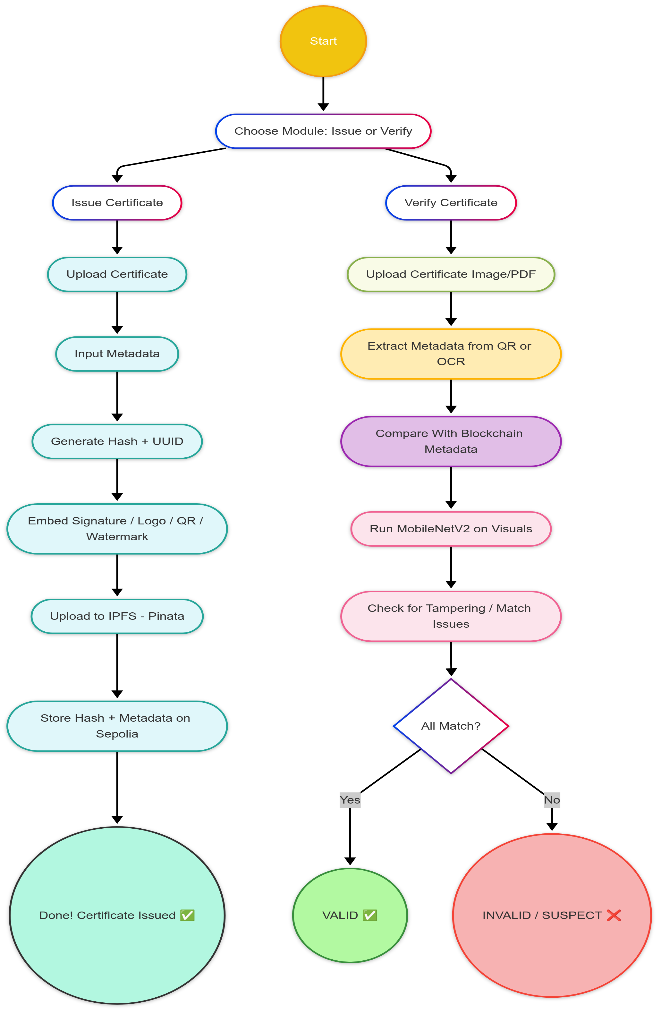


Fig.7

flow diagram of the working sample

IV CONCLUSION

The Blockchain-Based Certificate Validation System introduces machine learning and blockchain technology to face certificate issuance and verification challenges. The prime conclusions of the report are as given below:

Better Security and Authenticity:

* The certificate on the blockchain cannot be changed or tampered with.
* The machine learning models authenticate thoroughly metadata against visual integrity.
* A Composite Solution:
* The system provides complete lifecycle management of the certificate from issuance to verification.
* It honors both the issued certificates and pre-existing certificates, making it adjustable to different organizational needs.
* Greater Efficiency:
* The automation of validations resulted in reduced manual work and less error-prone.
* An ergonomic interface allows easy adoption by organizations.
* Scalability and Flexibility:
* The integration with IPFS promotes decentralized storage and takes care of scalability.
* The machine learning models can be fine-tuned for other types of certificates or purposes.
* Possible Applications:
* All the authorities in charge of awarding credentials, such as educational institutions, professional organizations, and government bodies, may utilize this product to secure credential management.
* This system would detect forgery and hence will put an end to counterfeiting of certificates.

To summarize, the new approach combines state-of-the-art technology to provide a secure, effective, and scalable solution for certificate management and validation, bridging a gap in conventional systems and setting the platform for the next technological advancements.

REFERENCES

[1]Kumutha Parthiban, Jeyalaksshmi, "Blockchain Technology and Academic Certificate Authenticity-Review," MIT Media Lab, accessed Apr. 08, 2025. [Online].

Available: https://www.researchgate.net/publication/352441968\_Blockchain\_Technology\_and\_Academic\_Certificate\_Authenticity-Review

[2] [Mrs. R. Suganthalakshmi, Mrs. G. Chandra Praba, Mrs. K. Abhirami, Mrs. S. Puvaneswari, "BLOCKCHAIN BASED CERTIFICATE VALIDATION SYSTEM" GitHub, accessed jul. 07, 2022.](https://www.irjmets.com/uploadedfiles/paper/issue_7_july_2022/28889/final/fin_irjmets1659003745.pdf)

[3] Mohammed Abdulbasit Ali Al-Ameri ,Basim Mahmood,Bünyamin Ciylan, Alaa Amged "Unsupervised Forgery Detection of Documents: A Network-Inspired Approach," Electronics, vol. 12, no. 7, 2023. [Online].

Available:<https://www.mdpi.com/2079-9292/12/7/1682#academic_editors>

[4] Garima Jaiswal, Arun Sharma, Sumit Yadav,"Deep feature extraction for document forgery detection with convolutional autoencoders," Computers & Electrical Engineering, vol. 101, 2022. [Online].

Available: https://www.researchgate.net/publication/358320112\_Deep\_feature\_extraction\_for\_document\_forgery\_detection\_with\_convolutional\_autoencoders

[5] A. Gayathiri,J. Jayachitra,S. Matilda, "Certificate validation using blockchain"

Available: https://ieeexplore.ieee.org/document/9201988

[6]M. Y. Kubiy, Mehmet Sabir Kiraz , Hacı Ali Mantar ,”CertLedger: A new PKI model with Certificate Transparency based on blockchain”,Computers&Security,vol. 85,2019,https://www.sciencedirect.com/science/article/abs/pii/S0167404818313014

[7]R.A. Huber Jr.,S.P.Mallick,M.W. Flagg,K. Singhal,”Tamper detection for identification documents"

Available: https://patents.google.com/patent/WO2018071768A1/en

[8] S.B. Samanthula,R. Kumari,Ch. Pavangeethanjali, "CertiSafe: A Blockchain Based Certificate Validation and Safety System," IJRASET,

Available: https://www.ijraset.com/research-paper/certisafe-a-blockchain-based-certificate-validation-and-safety-system