Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

AI-Powered Certificate Tampering Detection Using Rule-Based and Machine Learning Approaches

Amaan Inamdar
Department of Computer Science
Abeda Inamdar Senior College
Pune, India

Shakila Siddavatam
Department of Computer Science
Abeda Inamdar Senior College
Pune, India

***Abstract*— The use of online learning and e-certification has become a common practice, yet at the same time, it has brought up a number of certificate forgery cases {1}. In light of this, the research titled** "AI-Powered Certificate Tampering Detection System using Rule-Based and Machine Learning Approaches" **proposes an innovative technique to detect such tampering .**

Detecting fraudulent certificates is done using a combination of machine learning and rule-based methods. In the first stage, conventional checks such as PDF metadata inspection, font mismatch detection, layout validation, and QR-code verification are carried out to identify direct inconsistencies . Simultaneously, Random Forest and XGBoost machine learning models analyze layout, text, and metadata features to estimate the probability of tampering . The synergy between these two approaches not only enhances detection accuracy but also provides users with a transparent, reason-based interpretation of the results .

A Flask backend performs document processing, OCR extraction, feature engineering, and model inference, while a React.js frontend enables users to upload certificates and receive detailed verification reports . Each report not only states the tampering probability but also highlights the specific discrepancies. Additionally, the system can instantly validate the URLs embedded in QR codes .

Thus, this study contributes significantly to the field of document forensics by offering a trustworthy and comprehensive verification framework. The hybrid approach outperforms traditional rule-based systems, and future extensions may include blockchain integration for secure and decentralized record verification .

Keywords - Certificate tampering detection, Document forgery analysis, Machine learning verification, Rule-based methods, PDF metadata inspection, OCR feature extraction, QR code validation, Random Forest classifier

The field of digital transformation has created a complete transformation of educational delivery systems and global recognition methods. The growing acceptance of online learning platforms as valid sources for developing skills and granting credentials has allowed Coursera and Udemy to provide learners with access to high-quality courses which they can study from anywhere in the world without facing financial obstacles. The platforms have enabled millions of learners to obtain digital certificates which show their completion of courses while proving their acquired skills and professional development.

Online credentials have improved educational access for students yet they created problems which schools must solve to maintain certificate verification systems that institutions can trust. Educational institutions issue online certificates to students as standard digital files which users can download and share and print. The formats allow people to change names and scores and issuance dates and issuing authorities through basic editing tools which they can easily obtain. The combination of simple certificate manipulation and the absence of standardized verification methods has resulted in an increase of certificate fraud cases together with fake credential distribution. This development has damaged employer faith in online education credentials which online education providers distribute.

Organizations face challenges with traditional verification methods because the process requires institutions to contact their sending institution while using visual inspection methods which require extensive resources and lead to human mistakes. The problem increases because online certificates lack unified verification standards which different platforms and regions use to confirm certificate eligibility..

1.1 Problem Statement

The current digital age has transformed credential acquisition through its widespread acceptance of online education and digital certification because it enables students worldwide to obtain and share certificates with ease. The simple process of obtaining digital certificates has resulted in a major rise of counterfeit and modified digital certificates because hackers can easily change PDF files through basic editing programs and the existing verification process depends on slow and manual methods which produce multiple mistakes. The process of authenticating documents needs institutions to issue documents while verification requires visual inspection which limits its ability to handle high-volume authentication tasks. The absence of common automated verification systems prevents employers and academic institutions and regulatory bodies from validating certificate authenticity which leads to fake credential use and institutional reputation damage and higher verification costs and longer verification times.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

1.2 Significance

The increasing use of online education together with digital certificates has resulted in a corresponding increase of fake and altered credentials which create major problems for educational institutions and employers together with their students. The existing methods used for certificate verification require manual processing which takes a long time and results in high error rates thus making these methods unsuitable for detecting advanced digital file forgery and fraud activities. The online credential protection needs automated systems which provide precise detection capabilities to safeguard institutional trustworthiness and maintain learner competence verification trustworthiness. The proposed system combines rule-based verification with machine learning models to create an efficient verification system which provides better protection against certificate misuse in academic and professional settings.

1.3 Proposed Solution

This study proposes a neighbourhood-based carpooling pick-and-drop system that allows users to share rides with nearby residents traveling along similar routes. By combining verified user identities with route-based ride matching, the system aims to promote safe, affordable, and sustainable shared transportation. The proposed solution focuses on short-distance daily commuting and serves as a practical alternative to conventional ride-hailing services.

2. Literature Review

The work performed on falsifying documents and certificates has been done in a variety of fields such as image forensics, QR verification, PDF authentication, and hybrid machine learning-based techniques. Each of them offers a set of important tricks but also the obvious drawbacks. For instance, the image-forensics models are good at detecting visual tampering but they cannot do anything with PDFs, metadata, or QR codes [1]. QR verification systems can provide quick checks against databases but they will be useless if the QR codes are removed, replaced, or changed [2]. They will also not detect any modifications in layout or text. Blockchain technology has been used to protect certificates but only at the moment of their issuance. It offers no protection against tampered PDFs uploaded afterwards [3].

Research on PDF authentication studied metadata plus structure and caught subtle changes but no measures for images and QR checking were available [4]. Image-forensics ML models, such as the one with CNN and GAN detectors, are capable of identifying pixel-level modifications but cannot check the document's metadata or QR codes [5]. Moreover, QR integrity work reveals how QR manipulation can be traced; however, it does not consider features of the certificate as a whole [6]. Collaborated methods like merge rule-based and ML techniques for detection of structural and layout anomalies but still do not create a complete and unified solution.

In conclusion, currently available technologies provide solutions only for a small portion of the certificate verification issue. Visual editing is handled by image forensics, data in QR is validated via QR methods, the structure is checked by PDF authentication, and the issuance is secured by blockchain. But, no single method can uncover all forms of tampering. The situation calls for an integrated, multi-layered system that includes PDF analyzing, image forensics, and QR validation.

Research Gap

The existing literature largely emphasizes scalability, centralized architectures, and algorithmic efficiency, with limited focus on trust-oriented, neighborhood-based carpooling systems. There is a lack of practical implementations that address social familiarity, safety, and routine short-distance commuting. This research addresses this gap by proposing a localized, community-driven carpooling system designed for sustainable and trusted urban mobility.

3. Methodology (Development Process)

3.1. Methodology

The system detects certificate tampering through the combination of two methods which are rule-based verification and machine learning. The system processes uploaded certificates to extract both text content and metadata information while analyzing key features which include layout design, font styles, logos, and QR code elements. The system uses basic rules to identify clear irregularities which include font discrepancies and metadata changes. The system uses structured features to train Random Forest and XGBoost models which have been developed to differentiate between authentic and fake certificates. The system generates an explainable report which combines rule-based system alerts with machine learning outputs to present an authenticity conclusion and identify details that require further investigation.

3.2 Frontend Methodology

The certificate tampering detection system operates its user interface through React.js and HTML5 and CSS3 and JavaScript to enable users to upload certificates and preview documents and verify their results. The React frontend includes components for user authentication, certificate upload, verification status display, and detailed discrepancy reporting. The frontend collects user inputs through certificate file uploads and manual confirmation fields which it transmits to the backend using REST APIs. React.js provides developers with tools to handle application state and update user interface elements while React.js maintains an uninterrupted user experience that delivers verification results and explanations to users during the entire process. The system provides component modularity which enables developers to expand the interface by adding new features that include user profiles and historical verification logs.

3.3 Backend Methodology

The system uses Python with the Flask framework to create a backend that manages all essential application functions which include certificate handling and feature extraction and model evaluation. The Flask backend provides

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

RESTful APIs that allow the frontend to send certificate documents which the system processes to produce organized verification results.

• Certificate Preprocessing: The uploaded certificate (PDF, JPG, or PNG) is first processed. The system uses PyMuPDF library to extract text and metadata from PDF documents which it processes and uses OCR modules to convert images into text.
• OCR & Text Extraction: Optical Character Recognition (OCR) engines extract printed text (e.g., names, course titles) for layout and content analysis.
• Feature Engineering: The process involves extracting metadata and layout features together with font consistency checks and QR code presence/decoding and other certificate attributes to create structured data for analysis.
• Rule-Based Verification: Heuristic checks are applied to detect direct inconsistencies which include mismatched fonts and missing metadata fields and invalid logos and unverifiable QR codes and the system assigns rule-based flags that indicate potential tampering.
• Machine Learning Inference: The system uses structured features to evaluate certificate tampering probability through two classification models which include Random Forest and XGBoost. These models use known authentic and forged samples to analyze patterns which they learned to produce probabilistic scores for detecting tampering.
• Explainable Report Generation: The backend system combines results from rule-based checks together with machine learning predictions to create a detailed report which shows specific discrepancies and tampering probabilities to the frontend display.

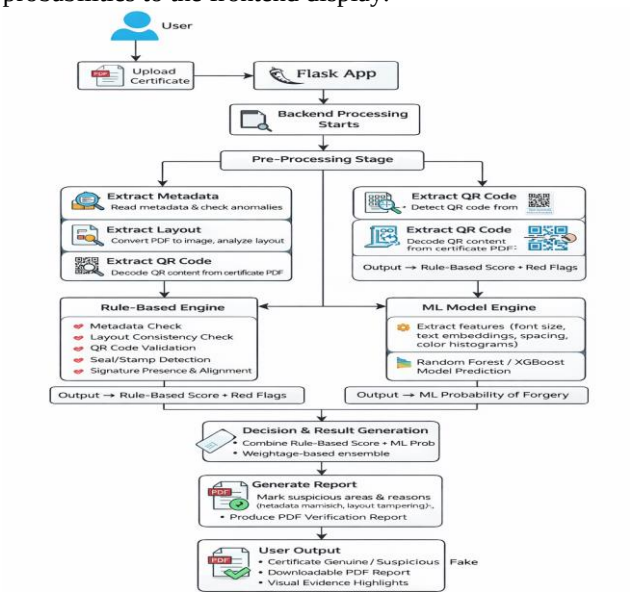

Figure 1 System Architecture

4.1 Technologies Used

The proposed certificate tampering detection system is developed using a modern web and AI technology stack to ensure scalability, reliability, and accurate tampering analysis. The selected technologies support secure user interaction, robust document processing, machine learning inference, and seamless frontend-backend communication. Table 1 presents the major technologies used at different levels of the system.

Table 1: Technology Stack for Certificate Tampering Detection System

Component	Technology Used
Frontend	React.js, HTML5, CSS3, JavaScript
Backend	Python (Flask)
PDF & Metadata Processing	PyMuPDF
OCR Text Extraction	Tesseract OCR
Computer Vision & Layout Analysis	OpenCV
QR Code Decoding & Validation	Pyzbar
Machine Learning Models	scikit-learn (Random Forest)
Rule-Based Engine	Custom Python heuristics
API Communication	REST API
Database	MySQL / any SQL database

4.2 User Interface (UI) & Screenshots

The certificate tampering detection system offers a user interface which enables users to upload digital certificates and verify their authenticity within a simple and straightforward process. Users can select certificate files (PDF, JPG, PNG) for upload and view clear verification results once the system processes the document. The interface presents outcomes like valid, suspicious, or tampered with concise explanations, making it easy for users to understand the result without technical difficulty. Users can easily operate the system on both desktop and mobile devices because all essential functions which include upload buttons and progress indicators and result summaries were designed to provide clear functions. The UI displays metadata inconsistencies and QR-code problems to users, which enhances system usability while enabling users to comprehend verification results.

4.2.1 User Interface Overview

The certificate tampering detection system establishes a complete set of role-based and functional user interfaces which enable both regular users and administrators to proceed through their verification process. The research provides screenshots which show how each interface element functions to help users interact with the system.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

• Welcome (Home) Page:

This page serves as the primary entry point for users to explore the platform while providing links to both login and registration functions, which researchers show in their figures.

• Login Interface:

The system provides users with a secure login page that allows them to enter the system after providing their authenticated credentials. Users must select their role between User and Admin, then they need to provide their login information for system access.

• User Registration & Forgot Password Interfaces:

The registration form allows new users to create their accounts. The "Forgot Password" interface enables users to recover their credentials through a secure process when they forget their passwords.

• Upload Certificate Interface:

The upload screen becomes accessible to users who want to verify their certificates, which they can upload in PDF or JPG or PNG formats, after they log in. Users can use simple controls to select files which they want to submit for analysis of potential tampering.

• Reports Page (User):

Users can access complete reports about their uploaded certificates after verification ends, which show the chance of tampering and all visible differences in layout and metadata and QR validation.

• Admin — View All Reports Interface:

The dedicated screen for admins shows all verification reports which exist throughout the system. The system enables administrators to monitor system activities while they examine questionable certificates and manage all system records.

• Manage Users Interface (Admin):

Admins have the power to view all user accounts which exist on the platform because they can both edit and delete these accounts to control user access and activities on the platform.

• About Us Page:

This interface provides information about the system objectives, core features, and benefits to both users and administrators.

4.2.2 UI Screenshots

The following figures illustrate the key user interface screens of the Neighborhood-Based Carpooling Pick-and-Drop System for Sustainable Urban Mobility, highlighting the main functional components of the application.

Table 2: Description of System Interface Screens and Database Components

Figure No.	Description
Figure 2	Welcome (Home) page displaying an overview of the platform and navigation links for login and registration.
Figure 3 a	User registration interface for creating a new user account on the platform..
Figure 3 b	Login interface where users and admins securely enter credentials and select their role for system access.
Figure 4	"Forgot Password" interface allowing users to recover access securely if they forget their credentials.
Figure 5	Certificate upload interface where users can upload PDF, JPG, or PNG certificates for tampering analysis.
Figure 6	User reports page showing detailed verification results, tampering probability, and highlighted discrepancies.
Figure 7	Admin "View All Reports" interface where administrators can view all certificate verification results across the system.
Figure 8	Admin "Manage Users" page enabling administrators to edit or delete registered users and control platform access.
Figure 9	About Us page providing information on the system objectives, features, and benefits for users and admins.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

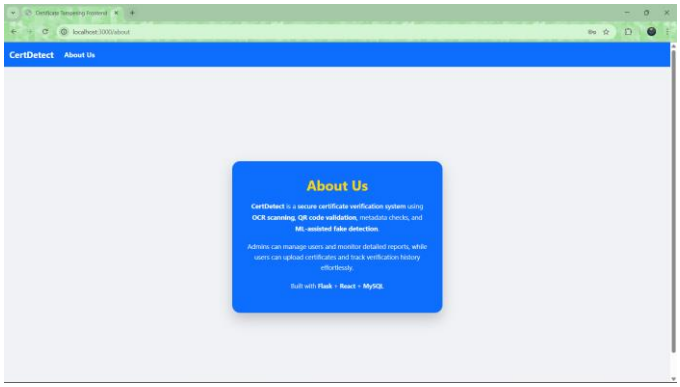

Figure 2

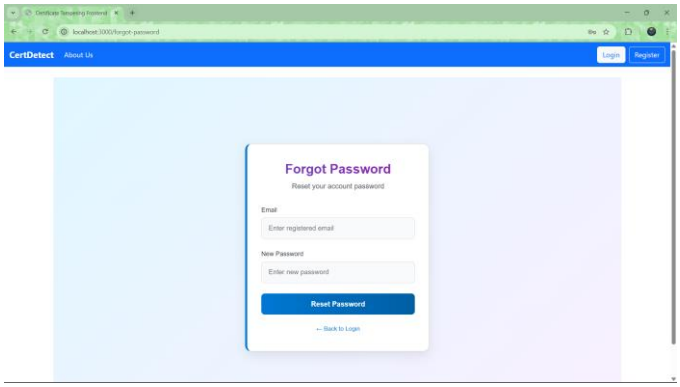

Figure 4

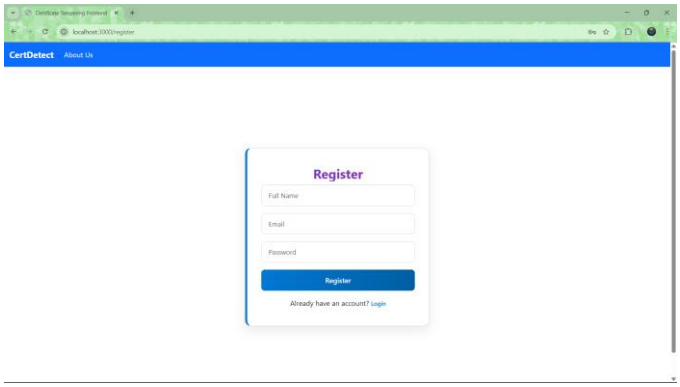

Figure 3 a

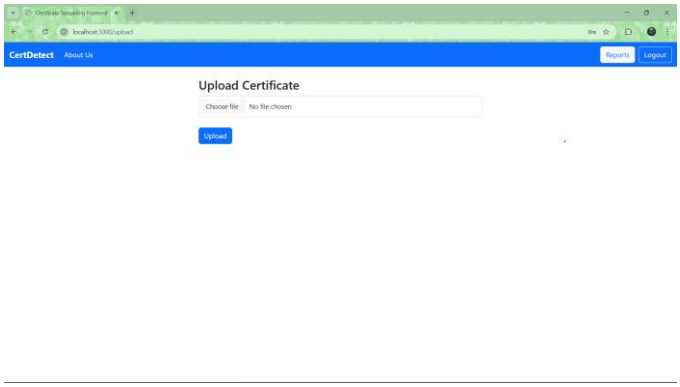

Figure 5

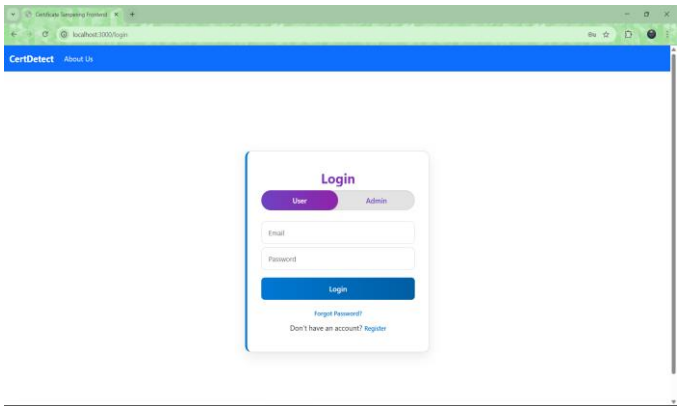

Figure 3 b

Figure 6

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

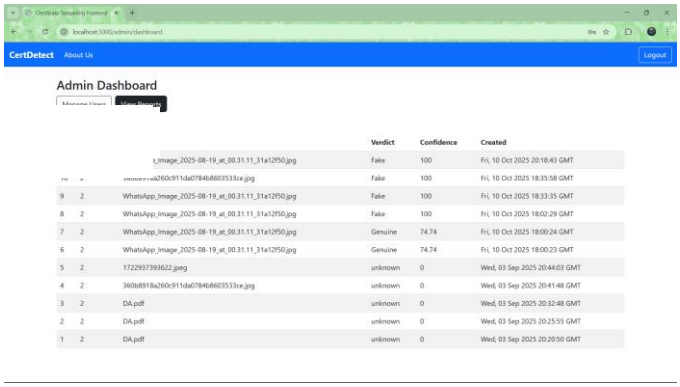

Figure 7

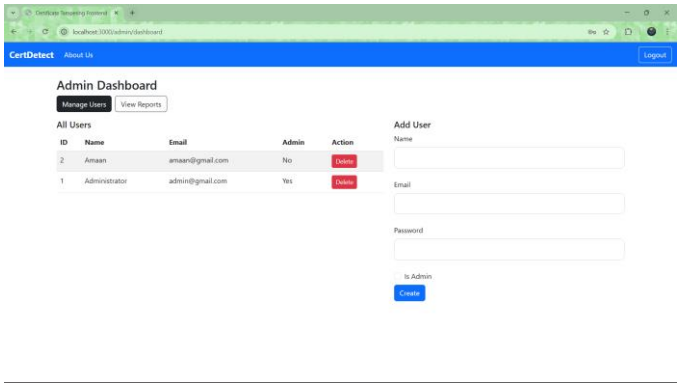

Figure 8

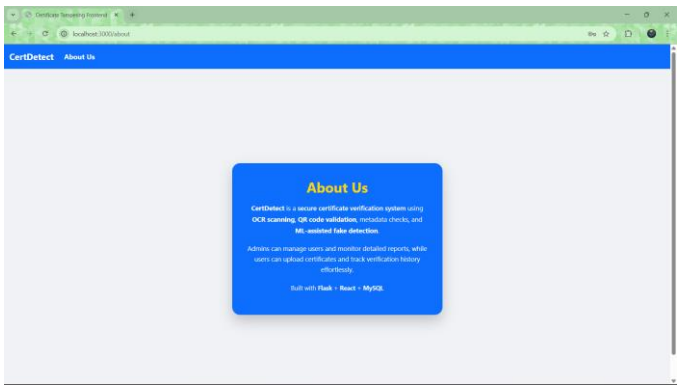

Figure 9

5. DISCUSSION

5.1 Strengths of the System

• Enhanced Security and Fraud Prevention:
The system automatically identifies tampered or forged certificates by analysing visual features, metadata, and embedded codes, which helps to decrease the chances of both fraudulent activities and unauthorized document modifications. The AI-based checks improve their detection precision through their ability to learn new forgery techniques which emerge in the future.

• Reliable and Explainable Verification Results:
The system uses rule-based rules which include font and layout checks together with machine learning models such as Random Forest and XGBoost to deliver precise results which come with full disclosure of the reasoning behind certificate flagging.

• Time and Effort Savings:
The automated analysis system takes over for verification tasks which used to require manual work thus it eliminates the requirement for human experts to check each certificate while it speeds up the process of validating authenticity.

• Improved Trust and Credibility:
The process of verifying digital certificates delivers results which help employers and educational institutions plus other stakeholders to trust that academic credentials maintain their authentic value.

• Scalability and High Volume Support:
The system uses AI-based processing together with rule-based systems to process multiple certificate verifications at once which makes it ideal for organizations that need to issue many credentials.

• Clear Discrepancy Highlighting:
The system not only classifies certificates but also shows users the exact differences between layout and metadata and QR validation, which enables them to determine the specific areas which have been tampered with.

5.2 Challenges and Limitations

• Potential for False Positives/Negatives:

AI-based document verification systems identify incorrectly all certificates which results in system errors that detect fake documents because they fail to recognize actual cases of document forgery when dealing with uncommon

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

document formats. The public perception of automated verification systems will suffer from this issue.

• Data Privacy and Security Concerns:

The system must handle sensitive personal information (such as names, IDs, and course data) during processing. Organizations face significant challenges when they need to secure their data while also adhering to privacy laws such as GDPR.

• Dependence on High-Quality Training Data:

Effective machine learning models require large, diverse datasets of both genuine and tampered certificates. The process of building these datasets faces challenges because of their restricted availability and the need to protect private information which results in reduced model performance.

• Adaptability to Evolving Forgery Techniques:

The model needs regular training updates because forgery techniques keep developing through advanced editing tools and adversarial methods which create new document tampering methods.

• Scalability and Performance:

The system needs optimization work because it struggles to manage high certificate verification demands yet requires optimization work to maintain operational efficiency.

5.3 Future Scope

- **Mobile Application Development:**

  The development of a dedicated mobile application for both Android and iOS platforms will enable users to upload certificates while accessing verification results through their smartphones. The mobile application will enable users to use their device's camera for live scanning and authentication of physical certificates.

- **Blockchain Integration for Decentralized Verification:**

  The implementation of blockchain technology creates a system that maintains certificate hashes and verification data in an unchangeable format which decentralized storage can secure. The blockchain records a certificate's cryptographic hash which enables immediate identification of any unauthorized modifications that occur after that point.

- **Advanced Analytics & AI Enhancements:**
  Machine learning and deep learning techniques will develop new capabilities that enable system operators to detect anomalies and forecast security breaches while the system learns from fresh verification data.

6. Conclusion

The growing popularity of online education has created new opportunities for people to learn and earn credentials, but this development has made it difficult to protect the trustworthiness of digital certifications. The process of verifying documents through conventional methods which include direct institution verification results in slow operations that lack effectiveness while containing security risks that decrease trust from employers and academic institutions and students. Standard editing tools enable users to produce fake or altered certificates however the absence of effective validation systems leads to undetected fraud which results in expensive verification delays.

The project develops a verification platform that combines machine learning tampering detection with rule-based assessment methods to create a more dependable and expandable verification system. The rule based component detects obvious anomalies by identifying inconsistent metadata and altered layout elements and invalid QR codes while the machine learning models Random Forest and XGBoost study certification patterns to discover hidden tampering indicators. The system achieves better detection results through its two methods which also produce understandable outputs that show particular differences between each certificate.

The system uses automated processes together with improved verification methods to validate actual achievements while minimizing fraudulent behavior which results in faster decision processes for academic and professional functions

.7. References

1. TienChi. (2021). MFAN for Image Forgery Detection.
2. IJERT. (2021). QR-based Certificate Validation. International Journal of Engineering Research & Technology (IJERT).
3. IJRASET. (2020). Blockchain for Tamper-proof Certificates. International Journal for Research in Applied Science and Engineering Technology (IJRASET).
4. Sharma, R., & Patel, K. (2021). Automated PDF Authentication Systems. IEEE Access.
5. Kim, J., & Lee, S. (2022). Machine Learning for Image Forensics. Elsevier.
6. Singh, A. (2023). QR Code Integrity Verification in Academic Certificates. Springer.