WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Document Forgery Detection using Machine Learning.

Shravani Jadhav[1], Vedika Jadhav[2], Arati Naigade[3], Utkarsha Nangare[4].

[1,2,3,4,]Department of Computer Science and Engineering DACOE, Karad.

ABSTRACT:

Robust detection of forged documents is essential in an increasingly digital world, especially as personal identification and government documents like Aadhaar cards are frequently targeted for fraud. This review synthesizes contemporary machine learning techniques applied to document forgery detection—ranging from traditional classifiers like SVM and Random Forests to advanced deep learning approaches such as Convolutional Neural Networks (CNNs)—with a focus on Indian identity verification systems. Unique challenges addressed include detecting copy-move, splicing, and signature manipulations in Aadhaar cards, leveraging features like optical character recognition (OCR), image forensics, and error level analysis (ELA). The discussion covers comparative metrics, real-world implementation barriers, dataset issues, and adaptive algorithms needed for evolving fraud patterns. The survey also highlights the gaps in publicly available datasets for ID forgery detection and suggests directions for future research in scalable, real-time document authentication systems.

Keywords: Document Forgery Detection , Aadhaar Card Verification , Machine Learning Deep Learning , Convolutional Neural Networks (CNN) , Support Vector Machine (SVM) , Error Level Analysis (ELA) , Character Recognition (OCR) , Digital Document Authentication , Capsule Neural Networks , Fraud Detection , Document Image Analysis , ID Document Security.

Introduction :

The rapid digitization of governance and financial systems has increased the reliance on electronic identity documents, among which the Aadhaar card serves as India's most critical proof of identity. However, this widespread dependence has also made Aadhaar documents frequent targets of fraud and counterfeiting, such as manipulated photographs, altered text fields, forged signatures, and tampered QR codes. These forgeries pose severe threats to data security, financial integrity, and personal privacy, as they can be used to create fake identities or gain unauthorized access to services. Traditional manual verification or OCR-based checks are inadequate against sophisticated tampering techniques like AI-generated morphing and deepfake attacks. To address this, machine learning and deep learning approaches have emerged as effective solutions, capable of automatically identifying minute visual inconsistencies and detecting manipulations in scanned Aadhaar images with high accuracy. By leveraging CNNs, feature extraction, and error-level analysis, these systems can differentiate genuine documents from forged ones, ensuring secure, efficient, and scalable authentication across digital ecosystems .

Literature Survey :

The table below provides a comparative analysis of research papers focused on automated subjective answer evaluation using web development. It highlights the authors, titles, publication years, along with the advantages (pros) and limitations (cons) of the proposed methods and technologies in each study.

Table 1 – Comparative Analysis of Research

Sr. No.	Author(s)	Title	Year	Pros	Cons
1	Musab Al-Ghadi, Zuheng Ming, Petra Gómez-Krämer, Jean-Christophe Burie	Identity Documents Authentication based on Forgery Detection of Guilloche Pattern	2022	Focuses on the guilloche security-pattern in identity documents; uses CNN feature extraction and similarity measure.	Not specific to Aadhaar; practical dataset may be limited; real-world deployment details missing.
2	Nandini N, Madhura C, Keerthi Joshi K, Devprakash B, Vandana M Ladwani	Document Forgery Detection	2023	Proposes a novel capsule-layer network + ELA to detect forgery (signature / copy-move) in documents.	Not Aadhaar-specific; journal not IEEE; focuses more on general documents than ID cards

Sr. No.	Author(s)	Title	Year	Pros	Cons
3	Lin Zhao, Changsheng Chen, Jiwu Huang	Deep Learning-based Forgery Attack on Document Images	2021	Studies how forgery attacks on document images are executed (text editing, etc.) — helps you understand the adversarial side.	Attack-side focus rather than detection system; not Aadhaar-specific; may not cover full system implementation..
4	N. Veena & S. Thejaswini	Aadhaar Block: An Authenticated System for Counterfeit Aadhaar Enrolment in Citizen Services Using Blockchain	2023	Specifically targeted at Aadhaar-card counterfeit enrolment, uses blockchain for authentication — very relevant to your Aadhaar forgery domain.	More about counterfeit enrolment than image-forgery detection; may not have deep CV/ML methods described.
5	Chandana D. & Kumar S.	Enhancing Aadhaar Card Image Security with Machine-Learning-Based Face Morphing Detection	2024	Focuses on face-morphing attacks in Aadhaar-card images; proposes ML algorithm that improves accuracy under real-world distortions.	dataset and code may not be public-accessible, so reproducibility could be limited..

Problem Statement :

Developing an accurate fraud detection system for Aadhaar cards is complex due to evolving forgery techniques, high document diversity, and stringent accuracy requirements. The challenge lies in ensuring real-time, scalable detection across varied forgery types, such as image manipulation, signature replacement, and content alteration, while minimizing false positives and adapting to new fraud strategies.

Objectives :

1. To review and compare machine learning methods for document forgery detection, emphasizing Aadhaar card verification.
2. To identify the strengths and shortcomings of current algorithms, datasets, and system architectures.
3. To highlight practical deployment challenges and future research avenues in scalable and real-time forgery detection systems

Discussion :

The State-of-the-art machine learning techniques, particularly deep convolutional neural networks (CNNs) and hybrid architectures, have shown superior performance in document forgery detection. In Aadhaar card verification, VGG16 and similar CNN models excel in identifying subtle manipulation patterns and feature extraction. Some practical deployments now process large-scale authentication tasks daily. However, challenges remain: model robustness against novel forgery, dataset diversity, adaptability to new fraud types, and minimizing resource overhead are critical bottlenecks.

Conclusion :

Machine learning has revolutionized the field of document forgery detection, providing automated and intelligent systems capable of identifying even the most subtle forms of tampering in identity documents such as the Aadhaar card. Recent advancements in deep learning architectures, including Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and Generative Adversarial Networks (GANs), have significantly enhanced both the accuracy and robustness of forgery detection models. These systems can efficiently analyze complex features such as texture inconsistencies, pixel-level noise, and geometric distortions that are often invisible to the human eye. Moreover, large-scale datasets and transfer learning approaches have improved scalability, enabling models to adapt across varied document types and imaging conditions. However, as forgery tools evolve—with AI-driven image synthesis, morphing, and smart editing—the need for real-time detection mechanisms, continuous model retraining, and diverse dataset augmentation becomes critical. Integrating such adaptive learning strategies ensures that machine learning systems remain resilient against new attack patterns, thereby maintaining trust, authenticity, and security in Aadhaar-based and other identity verification processes.

7.REFERENCES:

1. Solanki, A., & Pittalia, N. (2024) : AI-driven virtual try-on using deep learning and 3D body reconstruction. Advances in Computer Vision Systems, 18(2), 123–139.
2. Pandey, A. et al., “A Review On Deep Learning Techniques For Image Forgery Detection,” IEEE, 2022[6].
3. “A Review on Image Forgery Detection Techniques Using Machine Learning,” SSRN, 2023[1].
4. Kandel, S., “Machine Learning-Based Signature Forgery Detection in Document Authentication Systems,” IJFMR, 2023[7].
5. Jaiswal, G., “Deep feature extraction for document forgery detection with ink mismatch,” ScienceDirect, 2022[5].
6. “Document Forgery Detection,” JETIR, 2024[3].
7. “Forgery Detection in Aadhaar-based KYC using Deep Learning,” ACM Digital Library, 2024[2].

8. "Forged Document Detection," IJPREMS, 2024[9].
9. Mehrjardi, FZ., "A survey on deep learning-based image forgery detection," ScienceDirect, 2023[8].
10. "Ai Powered Image Processing System For Automated Document Verification," IJCRT, 2025[10].
11. "Image forgery detection using deep learning," Zenodo, 2024[4].