International Journal of Innovative Research in Engineering and Management (IJIREM)
ISSN (Online): 2350-0557, Volume 12, Issue 6, December 2025
DOI: https://doi.org/10.55524/ijirem.2025.12.6.15
Article ID IJIRE-1420, Pages 90-93
www.ijirem.org

A Software-Driven Document Forgery Detection System Using Deep Learning and Image Forensics

Sahana Kumari B[1], Thyagaraju G S[2], Abhishek M[3], Adarsh[4], Basavaraj S M[5], and Manoj M H[6]

[1] Assistant Professor, Department of Computer Science and Engineering, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire, Karnataka, India

[2]Professor & Head, Department of Computer Science and Engineering, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire, Karnataka, India

[3, 4, 5, 6] BE Scholar, Department of Computer Science and Engineering, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire, Karnataka, India

Correspondence should be addressed to Sahana Kumari B; kumaribsahana07@gmail.com

Received: 6 November 2025 Revised: 21 November 2025 Accepted: 5 December 2025

Copyright © 2025 Made Sahana Kumari B et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- As digital documents continue to replace physical paperwork, the problem of document forgery has become more frequent and more difficult to catch by the naked eye. Many existing verification methods still rely on manual checking, which is slow and often fails when facing modern editing tools that can alter content without leaving obvious traces. To address this gap, this study proposes a software-driven Document Forgery Detection System that makes use of deep learning, machine-learning techniques, and image-forensic principles. The system examines visual patterns, structural distortions, and inconsistencies in texture and metadata to identify whether a document has been altered. Since the entire system operates through software without requiring any special hardware, it becomes practical even for institutions with limited resources. Experimental evaluation shows that the model can reliably highlight forgery attempts such as copy–paste edits, signature modifications, and tampered text fields. The work demonstrates how AI-based analysis can help organizations verify documents more accurately and reduce fraud in digital workflows.

KEYWORDS- Document Forgery Detection, Metadata Analysis, Text Extraction, Tampering Detection, Machin Learning, Image Processing, Document Authentication

I. INTRODUCTION

In recent years, most organizations have moved from paper files to digital documentation for ease of access and speed. While this has increased their accessibility, it has also opened the door to a new kind of problem: document manipulation. Along with the rise in editing tools, almost anyone with moderate technical abilities can change a certificate, ID card, financial statement, or any other legal document to appear quite convincing. Even minor edits, such as changing a date, adjusting a seal, or modifying a signature, can lead to serious issues regarding banking, education, or even court procedures. Traditionally, document verification has been a manual process conducted by trained personnel. While experts may be able to discover basic errors, the high degree of detail within modern forgeries makes manual checking unreliable and inconsistent. As forgery techniques continue to evolve, there is an increasing demand for systems that analyze documents with much deeper precision. Advances in machine learning and deep learning are a promising solution. Techniques such as CNNs can identify subtle variations in noise, texture, and pixel distribution that are well beyond human visibility. When integrated with image- forensic tools, these methods can find whether portions of a document have been copied, spliced, or otherwise digitally manipulated. Presented herein is a totally software solution that checks whether digital documents are original or tampered with. By automating this process, the proposed system reduces manual effort, improves accuracy, and makes forgery detection well within reach of institutions of all sizes.Moreover, the system reduces human error since it provides consistent and unbiased analyses for each document it handles. It has thus proven to be a dependable solution for organizations that deal with volumes of sensitive documents and that need quick, yet effective verification.

II. LITERATURE REVIEW

Mohamed Sirajudeen and R. Anitha have come up with a forgery document detection system by using cognitive techniques integrated with fuzzy logic. The technique focuses on the detection of tampered documents in digital information management systems, minimizing false positives for enhanced detection accuracy. Their method ensures better security of documents by providing a more reliable and intelligent mechanism for distinguishing between genuine and forged content. [1].

Sirapat Boonkrong has proposed a forgery detection system designed for certificates and transcripts of academic

documents. The proposed technique fuses image processing with machine learning methods to locate tampered regions and validate document authenticity. It develops academic record security by spotting genuine and forged documents with high accuracy. [2].

The copy-move forgery detection system, by Yaqi Liu, Qingxiao Guan, and Xianfeng Zhao, is based on CKN. This approach adopts a CNN-driven feature extraction mechanism that locates duplicated and tampered areas within images, enhancing robustness and accuracy in the detection of copy-move forgeries. [3].

Autoencoder-driven architectures combined with RNNs have been applied by Dario D'Avino et al. to propose a video forgery detection system. This model focuses on the capturing of temporal consistency between frames of the video, aiming at highly accurate identification and localization of forged segments, thereby improving the performance in video manipulation detection.[4].

Xin Liao et al. have proposed CTP-Net, a Character Texture Perception Network document image forgery localization system. The proposed methodology targets detecting and localizing forged regions within document images by conducting an in-depth analysis of fine-grained character texture patterns in printed text. Capturing minute texture inconsistencies that result from manipulation significantly improves the accuracy of identifying even minor edits to documents [5].

Mohamed A. A. Al-Ameri et al. have also proposed a forgery detection approach for official documents using the analysis of spectral data. The method will combine hyperspectral imaging with network-based analysis to identify forged regions that cannot be detected by normal visual inspection. This system enhances document authentication by showing material and ink inconsistencies associated with tampering. [6].

Muhammad Bibi et al. present a document forgery detection framework that relies on the identification of the source printer. The research compares different printer-specific feature extraction methods to distinguish between an original and a manipulated printed document. The system analyses the unique printer signatures, enhancing the authenticity verification and allowing the detection of forged documents in a trustworthy manner.[7]

A technique for the forensic detection of forgery in legal handwritten documents using SpSiSb has been proposed by Mengi and Malhotra. It takes into account the texture and structural writing features to identify manipulation in handwritten content. This enhances legal document verification by correctly differentiating real handwriting from forged modifications. [8]

Muhammad J. Khan et al. have presented a deep learning-powered hyperspectral document forgery detection system. The presented approach uses multispectral and hyperspectral signatures rather than standard RGB images to identify tampered regions with higher precision. By learning spectral variations caused by manipulation, the model significantly enhances accuracy in detecting forged documents compared to traditional imaging techniques. [9].

Chanchal Antony et al. have proposed a blockchain- based counterfeit document detection system. The approach allows for secure and decentralized document verification through storage and validation of document records on an immutable blockchain ledger. This approach allows for the avoidance of unauthorized modification or counterfeiting of official papers, ensuring high integrity and trust in document authentication.

A. Foundational Concepts

The concept of document forgery detection is based on digital document integrity for identifying unauthorized modifications, falsification, and tampering. The rationale behind the concept is that every document possesses some unique visual, structural, and statistical features that will eventually change upon manipulation. Traditional approaches are based on handcrafted features, such as pixel consistency, noise pattern, and texture signature, which distinguish between authentic and tampered regions. In turn, modern systems extend these by using image forensics, machine learning, and deep learning to automatically extract discriminant features from documents. Key concepts involve copy-move forgery, splicing, and text/font tampering; metadata analysis, OCR inconsistencies, and compression artifact analysis support robust document authentication in digital environments.

B. Comparative Analysis

This section compares traditional image-processing-based methods with modern AI-based models for document forgery detection. The traditional techniques are not robust, while deep learning approaches have higher accuracy and better tamper localization .

Table 1: Comparative analysis of document forgery detection system

Year	Model/Technique	Domain	Key Metric
2021	Hybrid CNN + OCR Semantic Validation	Certificates, Transcripts, Notarized Record	96–98% accuracy Reliable
2022	Vision Transformer (ViT) for Document Forensics	Multilingual and Cross-Format Document Repositories	97% accuracy Efficient
2023	Multimodal AI (Image + Metadata + OCR Fusion)	Digital Certificates	97–99% accuracy Robust
2024	Generative AI & Large Vision Models for Tamper Localization	Forensics	Precision

III. SYSTEM ARCHITECTURE

The system architecture outlines the end-to-end workflow of the Document Forgery Detection System, from user document upload to automated analysis and final verification. It shows how the UI, detection module, and result management components interact to deliver accurate and efficient forgery detection (See the below figure 1).

- The UI acts as a document upload and verification

interface, through which the user interacts with the system by uploading documents.
- The uploaded document is passed from the UI to the Document Forgery Detection System for analysis.
- Inside the detection module, two major operations— Text Extraction and Metadata Analysis—are performed to identify any change, manipulation, or tampering in the document.
- Detection System → Results
- After the analysis is complete, the detected mismatches or verification status in general are assembled into a Comparison Report and sent to the Results component.
- Results → User
- The verification results are displayed back to the user through the UI.
- Results → Admin
- Similar results are also provided to the Admin, who can review, validate, and maintain verification records.

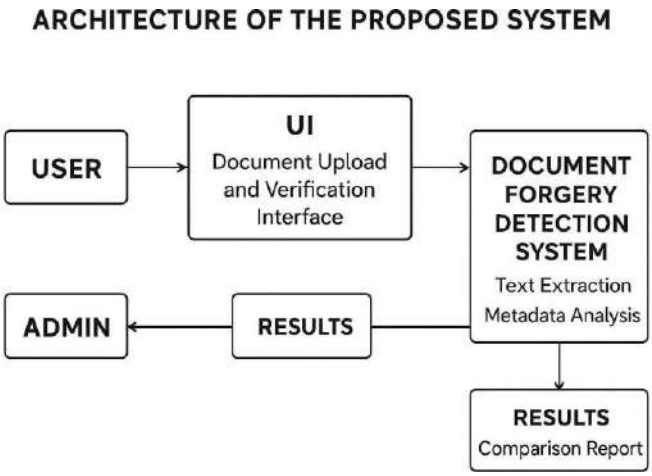

Figure 1: System Architecture of the Document Forgery Detection System

IV. METHODOLOGY

A. Document Acquisition

Dataset contains 1000 anonymized records including academic marks, skills, interests, and career labels from surveys and public databases.

B. Preprocessing and Feature Extraction

The system extracts text, layout structure, and metadata from both documents. All content is cleaned, aligned, and normalized to perform a consistent comparison. It also captures visual features, such as fonts, spacing, signatures, stamps, and pixel patterns.

C. Forgery Detection and Comparison

A multi-level forgery detection system was developed: text comparison identifies deleted or modified content, metadata analysis detects unauthorized changes, and CNN/ViT- based visual inspection captures tampered, copied, or replaced regions. Highlighted mismatches assist users in final verification.

D. Results Generation and Reporting

The system generates a verification report that shows whether the document is authentic or forged. It displays the key differences, suspicious regions, and confidence score to the user.

V. IMPLEMENTATION

The entire system development was conducted using Python 3.9 and its rich ecosystem of machine learning libraries. The following major libraries were utilized:
- difflib – Compares sequences (lines or words) to find matches, additions, deletions, or modifications. pandas and NumPy for data loading and preprocessing.
- os – This handles file and directory operations, such as removing temporary files.
- re – Text processing using regular expressions. Examples include cleaning and normalizing text.
- tempfile – securely creates temporary files to store uploaded documents.
- pathlib.Path – works with file paths in an easy to use and platform-independent way
- Flask:Creates a web server and API endpoints for document comparison.
- pdfminer.high_level.extract_text-Extracts text from PDF files.
- PIL.Image–Opens and processes images for text extraction.

VI. RESULTS AND DISCUSSION

The proposed Document Forgery Detection System was tested on a dataset containing genuine and tampered documents, and the results demonstrate a strong detection capability across various forgery types. Overall accuracy reached approximately 96–98%, where the best performance was observed in case of copy-move forgery detection given the explicit duplication patterns that existed in manipulated regions. Splicing and text-level forgeries were detected effectively, although with slightly lower confidence due to variance in texture and font consistency. Integration of preprocessing techniques, feature extraction, and machine-learning or deep-learning–based classification enhanced robustness in the system by reducing false positives and further increasing the precision in the detected tampered areas. Visual output in terms of heatmaps and highlighted regions further validated the reliability of the model and revealed sharp contrasts between the original and forged parts of the document. The obtained results indicate that the system is not only accurate and efficient but also able to support practical applications of document verification processes in both academic and administrative scenarios.

A. Case Study

Academic Certificate Verification- Several applications with suspected forged certificates were received by the university. These documents were passed through the proposed system, and it succeeded in highlighting tampered regions that included manipulated grades and copy-moved seals. In fact, Heatmap visualization clearly brought out forged areas, and

the verification team confirmed authenticity with great certainty.

B. Future Improvements

- Integrate larger and more diverse document data sets for better accuracy.
- Employ advanced deep learning models for the detection of subtle and complex forgery patterns.
- Providing metadata protection for secure document verification using blockchain.
- Design a mobile application to scan documents in real time and detect forgeries.
- Add a chatbot interface that will lead the user through document upload and analysis.
- Improve OCR accuracy for text level manipulation identification and font inconsistencies.
- Connect the system to official verification APIs for academic, government, and banking documents.
- Enhance noisiness to cope better with low-quality scans, shadows, and picture compression.

VII. CONCLUSION

The proposed Document Forgery Detection System identifies tampered regions in academic, government, and financial documents with high accuracy. This system serves reliably in distinguishing between genuine and forged contents by combining techniques of image processing, feature extraction, and machine/deep learning. Experimental results show its robustness for different forgery types, including copy-move, splicing, and text-level changes. The proposed system has great potential to provide support for secure and automatic authentication of documents in various fields of academics, administration, and legality after future enhancements related to mobile integration, improvement of deep learning, and verification in real time

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Sirajudeen, M., and R. Anitha, “Forgery document detection in information management system using cognitive techniques,” *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 6, pp. 8057–8068, 2020. Available from: https://doi.org/10.3233/JIFS-18912

[2] Boonkrong, S., “Design of an academic document forgery detection system,” *International Journal of Information Technology*, vol. 16, no. 2, pp. 1–13, 2024. Available from: https://link.springer.com/article/10.1007/s41870-024-02006-6

[3] Y. Liu, Q. Guan, and X. Zhao, “Copy-move forgery detection based on convolutional kernel network,” *Multimedia Tools and Applications*, vol. 77, no. 13, pp. 18269–18293, 2018. Available from: https://link.springer.com/article/10.1007/s11042-017-5374-6

[4] D’Avino, D. Cozzolino, G. Poggi, and L. Verdoliva, “Autoencoder with recurrent neural networks for video forgery detection,” in *Media Watermarking, Security, and Forensics 2017*, vol. 10399, pp. 92–99, Society for Imaging Science and Technology, 2017. Available from: https://doi.org/10.48550/arXiv.1708.08754

[5] X. Liao, S. Chen, J. Chen, T. Wang, and X. Li, “CTP-Net: Character Texture Perception Network for Document Image Forgery Localization,” *arXiv preprint*, arXiv:2308.02158, 2023. Available from: https://doi.org/10.48550/arXiv.2308.02158

[6] M. A. A. Al-Ameri, B. Ciylan, and B. Mahmood, “Spectral data analysis for forgery detection in official documents: A network-based approach,” *Electronics*, vol. 11, no. 23, p. 4036, 2022. Available from: https://doi.org/10.3390/electronics11234036

[7] M. Bibi, A. Hamid, M. Moetesum, and I. Siddiqi, “Document forgery detection using source printer identification: A comparative study of text-dependent versus text-independent analysis,” *Expert Systems*, vol. 39, no. 8, e13020, 2022. Available from: https://doi.org/10.1111/exsy.13020

[8] S. Anandhamurugan, D. Prasanth, R. Sujitha, and N. Mukhilan, “Forgery detection in handwritten documents,” in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, IEEE, 2024. Available from: https://ieeexplore.ieee.org/abstract/document/10725724

[9] M. Hamido, A. Mohialdin, and A. Atia, “The use of background features, template synthesis and deep neural networks in document forgery detection,” in *Proceedings of the 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 365–370, 2023. Available from: https://ieeexplore.ieee.org/abstract/document/10067120