GLOBAL JOURNAL
OF COMPUTER SCIENCE AND TECHNOLOGY - D: NEURAL & AI
Volume 26 | 2026 Article ID: GJCST-276449 Language: English

AI Tools and Techniques for Detecting Document Forgery

Online First

Dr. Yasser Elmalik Ahmed Seleman [§ ϑ *]
ϑ Associate Professor
ORCID 0009-0007-2052-0408
*Corresponding Author

Dr. Mohamed Osman Mohamed Gasm Elseid [‡ ϑ]
ϑ Associate Professor
ORCID 0009-0006-4395-5001

Dr. Mohaned Moatasem Apass Karmalah [§ ξ]
ξ assistant professor

§ Faculty of Computer Science and Information Technology, University of Technology (Sudan)
‡ Department of Computer and Information Systems, University of Technology (Sudan)

RECEIVED	ACCEPTED	ONLINE PUBLISHED	PEER REVIEW
2026-08-16	2026-08-26	2026-09-22	Double Blind

Abstract

His rapid growth of digital documentation has increased the risk of document forgery in several industries, including public services, banking, and education. A significant part of traditional document verification methods is manual inspection, which is often inefficient and prone to errors. To detect forged documents, an automated technique that utilizes machine learning and image processing is presented in this paper. The system examines structural and content-based features to identify unauthorized changes, such as altered text areas and visual elements. The proposed model uses preprocessing and feature extraction techniques to differentiate between genuine and fraudulent document types. The results demonstrate that the system improves verification accuracy while decreasing the need for human intervention. This strategy offers a practical means of enhancing the security of documents.

Artificial intelligence | Document Forgery | Neural Network (CNN)

AI USE STATEMENT
No generative AI was used for analysis or results.

FUNDING
No external funding was declared for this work.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY
Not applicable for this article.

ETHICS
No ethics committee approval was required for this article type.

CONSENT
Not applicable for this article.

TRIAL REG.
Not applicable.

Crossref DOI: 10.34257/GJCSTD276449

How to Cite: Seleman et al. (2026). AI Tools and Techniques for Detecting Document Forgery. Global Journal of Computer Science and Technology. Ahead of Print. DOI: 10.34257/GJCSTD276449

AR Experience Web Page DOI

Print ISSN 0975-4350
Online ISSN 0975-4172

Under the strict compliance and defined process of

ARCHIVAL RECORD

GJCST · Ahead of Print · 2026

Article ID GJCST-276449 · DOI 10.34257/GJCSTD276449

Print ISSN 0975-4350 · Online ISSN 0975-4172

Global Journal of Computer Science and Technology - Section D
GJCST

AI Tools and Techniques for Detecting Document Forgery

Dr. Yasser Elmalik Ahmed Seleman[§ ϑ * iD], Dr. Mohamed Osman Mohamed Gasm Elseid[‡ ϑ iD], and Dr. Mohaned Moatasem Apass Karmalah[§ ξ]

Affiliations
§ Faculty of Computer Science and Information Technology, University of Technology (Sudan)
‡ Department of Computer and Information Systems, University of Technology (Sudan)

Qualifications / Designations
ϑ Associate Professor
ξ assistant professor

Abstract

His rapid growth of digital documentation has increased the risk of document forgery in several industries, including public services, banking, and education. A significant part of traditional document verification methods is manual inspection, which is often inefficient and prone to errors. To detect forged documents, an automated technique that utilizes machine learning and image processing is presented in this paper. The system examines structural and content-based features to identify unauthorized changes, such as altered text areas and visual elements. The proposed model uses preprocessing and feature extraction techniques to differentiate between genuine and fraudulent document types. The results demonstrate that the system improves verification accuracy while decreasing the need for human intervention. This strategy offers a practical means of enhancing the security of documents.

Keywords: *Artificial intelligence, Document Forgery, Neural Network (CNN)*

*** Corresponding Author**
Dr. Yasser Elmalik Ahmed Seleman

DOI
10.34257/GJCSTD276449

1. Introduction

Artificial intelligence is used to analyze documents with high precision, it aids in detecting document forgery and tampering, provides advanced tools, techniques for analyzing images and text converts scanned images into analyzable text.

It reveals invisible text alterations Example: Using OCR to analyze identity documents and verify their authenticity OCR in forgery detection. Document forgery remains a growing challenge in digital transactions, education, and identity verification. Traditional Optical Character Recognition (OCR) systems identify text accurately but fail to detect tampered or falsified content. This paper presents a lightweight hybrid approach that combines OCR-based preprocessing with a Convolution Neural Network (CNN) to detect document forgery in scanned images. The pipeline first extracts textual and structural regions using Tesseract OCR, followed by CNN-based classification to differentiate between genuine and forged document patches. A synthetic dataset of 1,200 documents containing both real and manipulated versions was created using OpenCV for forgery simulation. The proposed model achieved 93.8 percent accuracy and 89 percent recall, outperforming classical threshold-based image comparison techniques. The study demonstrates how combining text recognition and deep visual analysis enhances the reliability of automated document verification systems.

With the growing digitization of administrative and financial operations, document authenticity has become increasingly critical. Forgeries involving altered text, modified seals, or forged signatures can lead to serious legal and financial repercussions. While OCR systems are effective at extracting text, they lack the capability to detect manipulations in the underlying image structure. This motivates the use of computer vision and machine learning to address document forgery detection through visual intelligence. In this study, we propose a hybrid pipeline that integrates OCR preprocessing with a CNN-based classifier to distinguish genuine from tampered documents. The approach focuses on detecting texture inconsistencies, font mismatches, and signature manipulation patterns that OCR alone cannot capture.

2. Research Problem

Document forgery detection is crucial in today's digital age, as it has evolved from simple photocopies to sophisticated digital manipulations.

The problem is compounded by the increasing sophistication of fraudulent activities, which have evolved from simple photocopies to complex digital manipulations. Increased opportunities for fraud, leaving businesses more vulnerable to document manipulation. Fake bank account details, forged invoices, falsified payslips or counterfeit identity documents can result in substantial financial losses, damage to a company's reputation.

3. Importance of Research:

1. Definition of document forgery and its implications.
2. The role of Artificial Intelligence (AI) in combating forgery.
3. Overview of common forgery types: forged signatures, altered text, fake stamps, and manipulated images.

4. Research Objectives:

Research helps find new facts and expands human understanding of different topics through the following:

1. Understand the importance of detecting forged documents.
2. Overview of AI's role in enhancing detection accuracy.
3. Common types of document forgery encountered.

5. Research Methodology: The Analysis Employed three main AI Techniques:

Image Forensics and Pixel Analysis: Using AI-driven image processing software to detect alterations in the document's scanned image, such as inconsistent pixel patterns, layering, or cloning artifacts.

Metadata Examination: Extracting and analyzing metadata from the digital file to check for discrepancies in creation dates, editing history, or software used.

Linguistic and Stylometric Analysis: Applying natural language processing (NLP) tools to evaluate writing style, grammar, and vocabulary against known authentic samples.

6. Research Questions:

1. How can artificial intelligence enhance the accuracy of document forgery detection?
2. How can these tools be developed to counter future forgery?
3. What are the ethical challenges associated with using artificial intelligence in this field?

7. Research Introduction

Document forgery refers to documents that have been altered, falsified, or created with the intent to mislead, deceive, or commit fraud. Forging documents can take many forms, including altered text, fake images, manipulated signatures, or fully digital creations designed to appear authentic. Criminals may also manipulate signatures to imitate approvals or authorizations, which create serious legal and financial risks.

Beyond signatures, image manipulation allows people to alter photos or scanned documents to misrepresent information, while digital alteration modifies electronic files in ways that leave no obvious traces. Together, these techniques create a broad spectrum of threats, making it increasingly difficult for institutions to trust documents without robust verification systems.

8. Matching Detection Methods to Document Type and Risk Level

Detection method selection should be driven by two factors: the document type being verified and the risk associated with a false negative.

1. **Low-risk, high-volume** scenarios: Automated digital verification and OCR-based checks are appropriate for efficiency.
2. **High-risk, low-volume scenarios**: Combined human and technology review provides the most reliable outcome.
3. **Physical documents with embedded security features**: Physical inspection should be the first line of verification.
4. **Digital or scanned documents**: Metadata analysis and forensic imaging are the most applicable starting points.

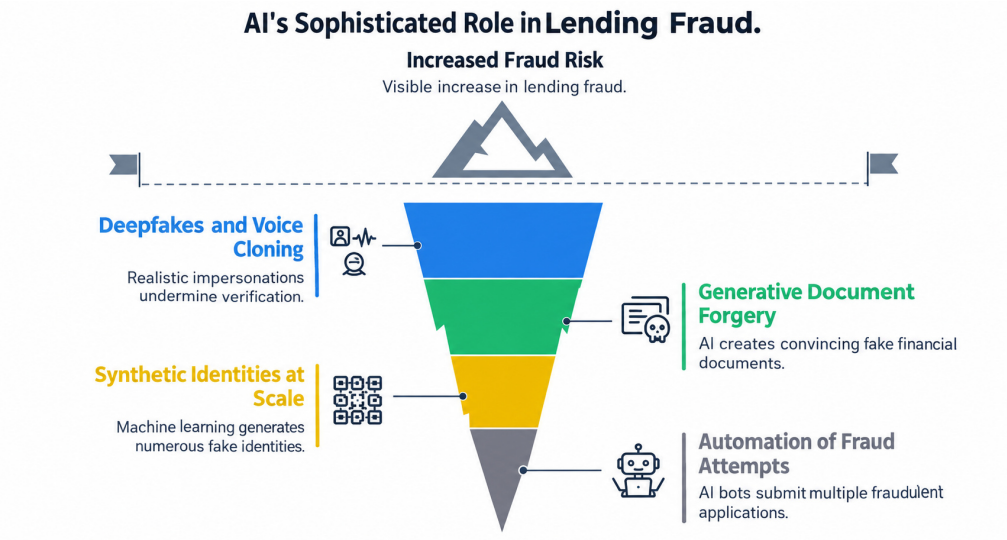

Figure 1. The forgery steps

9. AI techniques for Forgery Detection:

9.1. Machine learning approaches:

1. **Feature Extraction**: Identifying key characteristics such as font inconsistencies, ink color variations, and signature irregularities.
2. **Classification Algorithms**: Using supervised learning models (e.g., SVM, Random Forest) to classify documents as genuine or forged.
3. **Anomaly Detection:** Unsupervised learning to detect unusual patterns that indicate forgery.

10. Deep learning approaches:

1. **Convolution Neural Networks (CNNs)**: For image-based forgery detection (e.g., detecting manipulated images or scanned documents).
2. **Natural Language Processing (NLP):** To analyze textual inconsistencies or suspicious language patterns.

3. **Tools and Platforms** Open-source libraries: Tensor Flow, PyTorch, scikit-learn.
4. **Specialized software:** Adobe's AI tools, forensic document examination software.

An explanation of the use of Convolution Neural Networks (CNNs) For image processing and Natural Language Processing (NLP) techniques.

AI-based networks detect digital document forgery by analyzing visual, structural, and semantic inconsistencies that traditional rule-based tools miss. Core AI Architectures Used Convolution Neural Networks (CNNs): Act as the cornerstone for image patch classification, analyzing spatial features, pixel discrepancies, and compression anomalies to spot tampered regions. Dual-Stream / Hybrid Networks (e.g., CAT-Net): Combine RGB pixel analysis with frequency-domain inspection (like Discrete Cosine Transform coefficients) to uncover hidden manipulation traces. Optical Character Recognition (OCR) & Intelligent Document Processing (IDP): Extract and digitize text to cross-check structural data, fonts, layouts, and semantic logic against known templates. Generative Adversarial Networks (GANs): Used both defensively to model synthetic manipulation patterns and to train discriminators in identifying overly "perfect" AI-generated layouts. Key Detection Mechanisms Metadata & Forensic Analysis: Scrutinizes digital fingerprints, compression levels, and lighting consistency. Semantic Cross-Checking: Reads numbers, dates, and names in unstructured files (like invoices or bank statements) to verify internal logic and catch mismatched data fields.

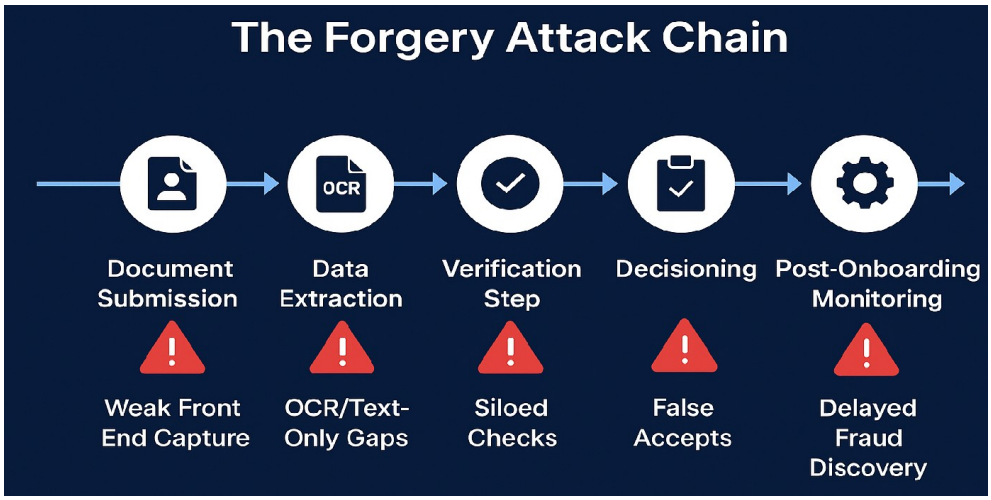

Figure 2. The forgery attack chain

Use of convolution neural networks (CNNs) to detect image tampering Techniques for identifying altered signatures and stamps Role of AI in analyzing document textures and backgrounds NLP techniques to detect inconsistencies in language and style Automated comparison with known authentic documents Detection of anomalies in fonts and formatting.

11. Summary of key AI tools and their effectiveness:

1. Emerging trends like block chain integration and explainable AI.
2. Encourage ongoing learning and critical evaluation of AI methods.
3. Machine learning algorithms for pattern recognition.
4. Deep learning for image and text analysis.
5. Natural Language Processing (NLP) for content verification.

12. Research Results

Document Forgery: The act of creating, altering, or falsifying documents with the intent to deceive. AI Role: AI systems analyze patterns, inconsistencies, and anomalies in documents to detect forgery. Common Techniques: Image analysis, text analysis, pattern recognition, and anomaly detection. AI-powered tools provide a comprehensive approach to detecting document forgery by integrating image analysis, metadata scrutiny, and language evaluation. In this case, the evidence indicates a high likelihood of forgery,

13. Some Key Points to Consider:

Types of Forgery: Document forgery can involve various forms, including forged signatures, altered contracts, and fake identification documents. Detection Methods: Advanced detection methods include metadata analysis, image forensics, and data cross-checking to identify forged or manipulated documents. AI and Machine Learning: The use of AI and machine learning techniques is increasingly being employed to detect forgery, as they can identify patterns and anomalies that human reviewers may miss. Regulatory Expectations: Financial institutions are expected to demonstrate how they ensure document authenticity as part of Customer Due Diligence (CDD). Online Tools: There are various online tools available for detecting forgery, such as TrueDoc, which analyzes documents for tampering and AI-generated signals. By integrating these methods and technologies, organizations can enhance their ability to detect and prevent document forgery, thereby protecting against financial and reputational risks.

14. Conclusion and Recommendations

14.1. Conclusion

AI technologies have significantly advanced the field of document forgery detection by enabling automated, accurate, and scalable

analysis. Techniques such as machine learning, deep learning, and natural language processing effectively identify anomalies in text, signatures, and document structures. AI models trained on large datasets can detect subtle forgeries that traditional methods might miss, improving reliability and reducing human error. However, challenges remain in handling diverse document formats, evolving forgery tactics, and ensuring data privacy.

15. Recommendations

1. Adopt Hybrid AI Approaches: Combine multiple AI techniques (e.g., image analysis with NLP) to enhance detection accuracy across various forgery types.
2. Continuous Model Training: Regularly update AI models with new forgery examples to adapt to emerging manipulation methods.
3. Integrate Human Expertise: Use AI as an assistive tool alongside expert review to validate findings and handle complex cases.
4. Focus on Explain ability: Develop AI systems that provide transparent reasoning for their decisions to build trust among users.
5. Ensure Data Security: Implement robust data protection measures when handling sensitive documents during AI processing.
6. Promote Standardization: Encourage the creation of standardized datasets and benchmarks to facilitate consistent evaluation of AI forgery detection tools.
7. Invest in User Training: Educate end-users on AI tool capabilities and limitations to maximize effectiveness and reduce misuse.

REFERENCES

[1] Qin, X., Tian, J., Sheng, J., Xia, T., Wang, Y., Li, C., & Zeng, G. (2025). Towards fine-grained document tampering detection: new dataset and benchmark. In *Pattern Recognition and Computer Vision – PRCV 2025, Lecture Notes in Computer Science*, Vol. 16278. Note: 16,479 tampered images, 12 source datasets, 3 languages, 8 tampering methods. https://link.springer.com/chapter/10.1007/978-981-95-5676-2_1 Cited by: https://arxiv.org/html/2602.20569v1#S2.SS1.SSS0.Px2.p1.1.

[2] Gupta, A., & Gupta, R. (2021). Semantic analysis for detecting forged textual documents using BERT. *Journal of Information Security and Applications*, 58, 102724.

[3] Sharma, P., & Singh, S. (2020). Forgery detection in documents using OCR and deep learning. *International Journal of Computer Applications*, 176(10), 1–7.

[4] Tancik, M., Mildenhall, B., & Ng, R. (2020). StegaStamp: Invisible hyperlinks in physical photographs. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2117–2126.

[5] Li, Y., & Li, J. (2019). Detecting forged documents through semantic inconsistency analysis. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 1234–1243.

[6] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.

[7] Mohammed, R. A., Nabi, R. M., Mahmood, S. M., & Nabi, R. M. (2019). State-of-the-art in handwritten signature verification system. *Proceedings of the International Conference on Advanced Science and Engineering (ICOASE)*, 174–179.

[8] Wu, Y., AbdAlmageed, W., & Natarajan, P. (2019). ManTra-Net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9543–9552.