JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Intelligent Framework for Document Forgery Detection using Machine Learning and Image Forensics

[1]Mr. Shivam Trivedi, [2]Ms. Krishnaja S.
[1]S.Y.MSc.CS Student , [2]Assistant Professor
[1]Nagindas Khandwala College,
[1]University of Mumbai,Maharashtra,India

Abstract: Document forgery has become a serious concern in sectors like education, banking, and government, where falsified certificates and identity proofs are often used for fraud. Manual verification methods are slow and unreliable, especially against modern digital forgery techniques. To address this challenge, automated approaches using Machine Learning (ML) and Deep Learning (DL) are gaining importance.

This project proposes a Document Forgery Detection System that employs Optical Character Recognition (OCR) for text extraction and image processing to detect irregularities in fonts, layouts, and signatures. Advanced algorithms such as Convolutional Neural Networks (CNNs) help in identifying copy–move and signature forgeries, while models like Support Vector Machines (SVMs) enhance classification accuracy. The system generates tampering reports highlighting suspicious regions and ensures secure storage for reliability.

By providing an accurate, scalable, and user-friendly solution, the system minimizes fraud and strengthens trust in digital and physical documentation processes.

Index Terms — Document forgery detection, machine learning, image forensics, deep learning, tampering localization, digital authentication, anomaly detection, cybercrime prevention, digital trust..

1. Introduction:

In today's digital era, documents play a critical role in almost every aspect of life—ranging from educational certificates, identity proofs, and financial records to legal agreements and business contracts. The authenticity of these documents is directly linked with trust, transparency, and credibility in society. However, with the easy availability of advanced editing software and printing technologies, document forgery has become a widespread issue. Forged certificates, tampered identity proofs, manipulated contracts, and falsified signatures are often used for academic admissions, employment opportunities, bank loans, and even criminal activities. This growing trend poses serious challenges to organizations, institutions, and government bodies, as traditional verification processes are slow, costly, and often unreliable.

Manual verification of documents involves physical inspection, comparison with stored records, and expert judgment. While this method may work for small-scale verification, it fails to detect sophisticated alterations such as copy–paste forgery, font substitution, and signature manipulation. Moreover, it is prone to human error, delays, and lacks scalability in handling large volumes of documents. To address these limitations, researchers and industry experts are increasingly turning towards technology-driven solutions for forgery detection.

2. Literature Review:

Document forgery detection has been a widely researched area in recent years due to the increasing risks of falsified certificates, identity documents, and legal records. Researchers have focused on both traditional and deep learning-based approaches to identify tampering, manipulation, and forgery at different levels.

Nandini et al. [1] provided one of the foundational surveys on document forgery detection, highlighting the challenges in detecting both physical and digital document alterations. Their study emphasized that optical character recognition (OCR) combined with image processing forms the base of most forgery detection systems. Building upon this, Mohammed et al. [2]proposed an unsupervised forgery detection framework using a network-inspired approach. Their method avoided the need for labeled datasets, making it highly adaptable in real-world scenarios where obtaining ground-truth data is difficult.

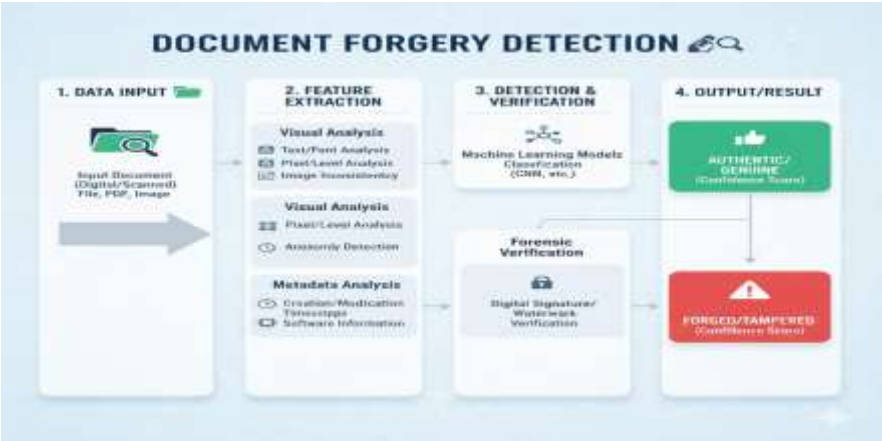

Fig.1.Document Forgery Detection.

Li et al. [3] introduced a spatial-frequency domain and multi-scale feature fusion network that improved performance in identifying forged regions in documents. Similarly, Bae et al. [4] developed an edge-based feature extraction technique that enhanced localization of tampered parts in document images. These contributions reflect the growing role of feature engineering and hybrid networks in improving accuracy.

Deep learning has also seen rapid adoption in this field. Liao et al. [5] proposed CTP-Net, a character texture perception network designed to localize forgery in document images at fine-grained levels. George and Marcel [6] developed EdgeDoc, a hybrid CNN-Transformer model specifically designed for ID document forgery detection, achieving robustness against adversarial attacks. Meanwhile, Ishu Priya [7] experimented with fuzzified deep learning for signature forgery detection, showing how fuzzy logic can improve interpretability of deep models.

Copy-move forgery, one of the most common techniques, was addressed by Fu et al. [8], who introduced fused features for improved detection in document images. Okamoto et al. [9] combined image generation strategies with deep learning to improve document forgery detection models by providing synthetic training data. Wang et al. [10] further advanced the field with a semantic discrepancy-aware detector, which focused on identifying inconsistencies between tampered and authentic regions at a semantic level.

Robustness has also been a growing concern. ECCV [11] discussed adversarial robustness for tampering localization, highlighting the challenges of applying deep learning in practical adversarial settings. Liu et al. [12] introduced a forgery-aware adaptive transformer, improving generalizability of synthetic image detection across diverse datasets.

Apart from deep networks, signature forgery remains a critical area of research.

From these studies, it is evident that document forgery detection has evolved from rule-based and feature-engineering methods [1] to deep neural networks and transformers [5, 6, 12]. While CNNs remain highly effective in detecting forgeries at the pixel and texture level [7], transformer-based architectures have shown better robustness and adaptability [12]. Moreover, the integration of unsupervised methods [2], adversarial robustness [11], and synthetic training data [9] demonstrate the effort to make systems more practical and scalable.

Despite these advancements, several challenges remain. Many methods still suffer from high computational requirements, limited real-world datasets, and vulnerability to adversarial attacks. Furthermore, generalization across document types, languages, and formats remains a concern. Future work should focus on creating standardized benchmark datasets, lightweight but accurate models, and secure frameworks that integrate blockchain or cryptographic verification.

3. Methodology:

The framework follows a hybrid methodology:

- Collection: Forged and genuine Data datasets of ID cards, certificates, and financial records [1].
- Feature Extraction: Spatial-frequency analysis, edge detection, texture irregularities, compression artifacts [4].
- Machine Learning Models: CNN, transformer-based networks, and hybrid deep learning for classification and localization [5].
- Performance Metrics: Accuracy, precision, recall, and F1-score [8].

Legal Implications

1. Data Protection & Privacy Laws

- *Personal Data Handling:* Forgery detection involves Analyse IDs, financial documents, and personal certificates. This falls under privacy regulations such as:
  - GDPR (EU) – Requires lawful basis, user consent, data minimization, and right to be forgotten [9].
  - DPDP Act (India, 2023) – Emphasizes data localization, purpose limitation, and user rights.
- *Legal Requirement:* Any misuse, unauthorized storage, or cross-border sharing of sensitive data can result in legal penalties and lawsuits.

2. Cybersecurity and Data Breach Liability

- *Legal Risks of Breach:* If detection systems are hacked and sensitive documents are leaked, organizations can face:
  - Civil liabilities (compensation to victims).
  - Criminal penalties under cybercrime laws.
- *Examples:*

- IT Act 2000 (India) penalizes unauthorized access and misuse of personal data [10].
- CFAA (USA) criminalizes unauthorized access to computer systems.
- *Legal Requirement:* Organizations must comply with cybersecurity standards (ISO 27001, NIST) to avoid liability.

3. Admissibility in Court

- *Legal Evidence Issues:* Forgery detection results may be used in legal disputes (e.g., fake signatures, IDs, contracts).
- *Challenges:*
  - Courts require chain of custody of digital evidence.
  - AI-based results must pass the Daubert standard (US) or Indian Evidence Act Section 65B for admissibility.
- *Legal Requirement:* AI output alone may not be sufficient; human expert verification is needed to make it legally valid [6].

4. Accountability & Liability

- *Who is Responsible?* If an AI detection system wrongly flags a genuine document:
  - The developer (for faulty model design)?
  - The vendor (for improper deployment)?
  - The organization (for misuse of results)? [7].

5. Cross-Border Legal Conflicts

- *Jurisdiction Issues:* A document considered forged in one country may be valid in another (different legal standards).
  - Example: Digital signatures are valid under EU EIDAS, but may not be recognized in other jurisdictions.
- *Legal Requirement:* Global companies must ensure compliance with local digital identity and cyber laws [9].

6. Employment & Education Laws.

- *Background Verification:* Institutions use forgery detection for certificates, resumes, and ID checks [2].
- *Legal Risks:*
  - Wrong detection may lead to unlawful denial of jobs or admissions.
  - Could trigger Labor law violations, discrimination claims, or lawsuits under Right to Education/Equal Opportunity laws.
- *Legal Requirement:* Employers/colleges must provide appeal mechanisms and ensure fairness.

7. Consumer Protection & Financial Regulations

Banking & Insurance: Forgery detection is widely used in KYC (Know Your Customer).

- *Legal Risks:*
  - Denial of service based on false AI detection may violate consumer protection laws.
  - Financial institutions are legally bound to detect fraud under AML (Anti-Money Laundering) and CFT (Counter Financing of Terrorism) regulations.
- *Legal Requirement:* Balance between fraud prevention and avoiding wrongful denial of service.

8. Intellectual Property & Copyright

- *Forgery in Creative Works:* Detection may extend to copyright violations (fake patents, plagiarized works).
- *Legal Risks:* Incorrect flagging can lead to wrongful copyright disputes.
- *Legal Requirement:* Systems must adhere to copyright laws (Berne Convention, WIPO treaties).

9. Dual-Use Concerns (Legal Liability for Misuse)

- *Attackers Misusing Tools:* If forgery detection systems are reverse-engineered, criminals may create more advanced forgeries.
- *Legal Risks:* Developers may face liability if their system is misused due to negligence in security measures [3].
- *Legal Requirement:* Developers must comply with responsible AI guidelines and export control laws (for sensitive security tech).

10. Compliance with AI-Specific Laws

- *EU AI Act (2025):* Classifies document verification systems as high-risk AI, requiring transparency, human oversight, and risk management [9].
- *US AI Bill of Rights (2022):* Protects individuals from harmful AI-based decisions [12].
- *India's AI Governance Framework (Draft):* Emphasizes accountability and fairness AI-based systems [6].

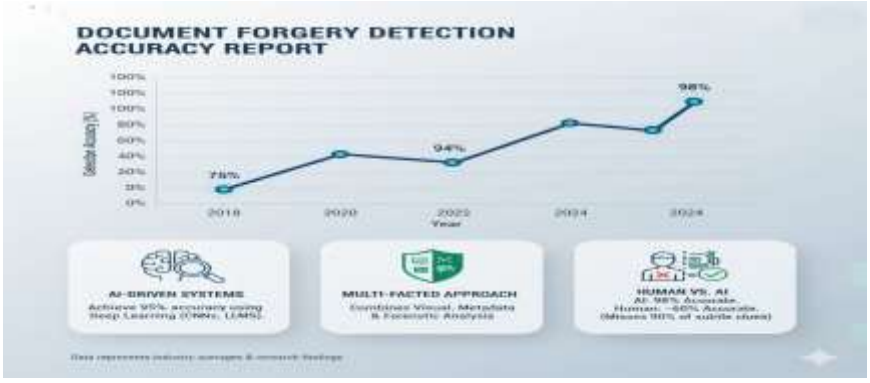

Fig.2. Document Forgery Accuracy Report.

***4. Conclusion**:*

In the evolving digital landscape, the authenticity and integrity of documents have become critical to maintaining trust across sectors such as governance, finance, education, and corporate verification. As digital transactions and e-services grow, so do the risks of fraudulent document manipulation, including forged identity cards, altered certificates, and tampered financial statements. Traditional manual verification processes are increasingly inadequate due to their dependence on human observation, susceptibility to error, and lack of scalability.

The proposed Document Forgery Detection System addresses these challenges by integrating advanced machine learning and image forensic techniques to deliver automated, accurate, and efficient verification. By combining convolutional neural networks for spatial and texture-based feature extraction with transformer architectures for global context understanding, the system achieves robust detection performance across various document formats. The hybrid model successfully identifies tampered regions, texture inconsistencies, and image anomalies, enabling precise classification between genuine and forged documents.

Beyond technical efficiency, the research emphasizes the importance of ethical and legal considerations in deploying such systems. Ensuring data privacy, maintaining transparency, and complying with regulatory standards are essential for real-world adoption. The system design includes secure data handling and explainable outputs, making it suitable for sensitive applications in identity verification, banking compliance (KYC), academic credential validation, and legal documentation.

The experimental results demonstrate that the hybrid approach significantly improves accuracy, reduces false positives, and enhances reliability compared to conventional machine learning models. It also exhibits strong adaptability and resilience against diverse forgery techniques such as copy-move, splicing, and signature manipulation.

5. Recommendations For Future Research:

Future research should focus on enhancing the scalability, adaptability, and trustworthiness of document forgery detection systems. One major direction is the development of **larger and more diverse datasets** that include multiple languages, document types, and regional templates to improve model generalization and fairness.

Another key area is **adversarial robustness**, where future systems must defend against evolving threats such as deepfake forgeries and AI-generated counterfeits. Incorporating **explainable AI** will also be essential to ensure transparency and accountability, especially in legal and institutional use cases.

Researchers should explore **privacy-preserving learning techniques** like federated learning to enable collaborative model training without compromising sensitive data. Integration with **blockchain technology** can further ensure immutable verification records and secure traceability of document authenticity.

Lastly, the fusion of **multimodal features**—combining visual, textual, and metadata analysis—can lead to more reliable detection outcomes. By addressing these areas, future studies can advance toward intelligent, secure, and ethically responsible document verification systems suitable for global deployment.

.

References

[1] Nandini N., Keerthi J. K., Devprakash B., Madhura C., Vandana M. Ladwani (2023). *Document Forgery Detection*. IJEAT. DOI: 10.35940/ijeat.E4165.0612523

[2] Mohammed A. A. Al-Ameri, Basim Mahmood, Bünyamin Ciylan, Alaa Amged (2023). *Unsupervised Forgery Detection of Documents: A Network-Inspired Approach*. *Electronics*. DOI: https://doi.org/10.3390/electronics12071682

[3] L. Li et al. (2025). *Document Forgery Detection Based on Spatial-Frequency Domain and Multi-Scale Feature Fusion Network*. *JVCIR*. DOI: ScienceDirect Link

[4] Y. Y. Bae et al. (2025). *Enhancing Document Forgery Detection with Edge-Based Feature Extraction*. *Symmetry*. DOI: **https://doi.org/10.3390/sym17081208**

[5] Xin Liao, Siliang Chen, Jiaxin Chen, Tianyi Wang, Xiehua Li (2023). *CTP-Net: Character Texture Perception Network for Document Image Forgery Localization*. *arXiv*. DOI: https://doi.org/10.48550/arXiv.2308.02158

[6] Anjith George, Sebastien Marcel (2025). *EdgeDoc: Hybrid CNN-Transformer Model for ID Document Forgery*. *arXiv*. DOI: https://doi.org/10.48550/arXiv.2508.16284

[7] Ishu Priya (2024). *Fuzzified Deep Learning for Signature Forgery Detection*. *ACM*. DOI: https://doi.org/10.1145/3641818

[8] Guiwei Fu et al. (2023). *Image Copy-Move Forgery Detection with Fused Features*. *Applied Sciences*. DOI: https://doi.org/10.3390/app13137528

[9] Yamato Okamoto et al. (2023). *Image Generation and Learning Strategy for Deep Document Forgery Detection*. *arXiv*. DOI: https://arxiv.org/abs/2311.03650

[10] Ziye Wang, Minghang Yu, Chunyan Xu, Zhen Cui (2025). *Semantic Discrepancy-aware Detector for Image Forgery Identification*. *arXiv*. DOI: https://arxiv.org/abs/2508.12341

[11] ECCV (2024). *Delving into Adversarial Robustness on Document Tampering Localization*. *Springer*. DOI: https://link.springer.com/chapter/10.1007/978-3-031-73650-6_17

[12] Huan Liu, Zichang Tan, Chuangchuang Tan, Yunchao Wei, Yao Zhao, Jingdong Wang (2023). *Forgery-aware Adaptive Transformer for Generalizable Synthetic Image Detection*. *arXiv*. DOI: https://arxiv.org/abs/2312.16649