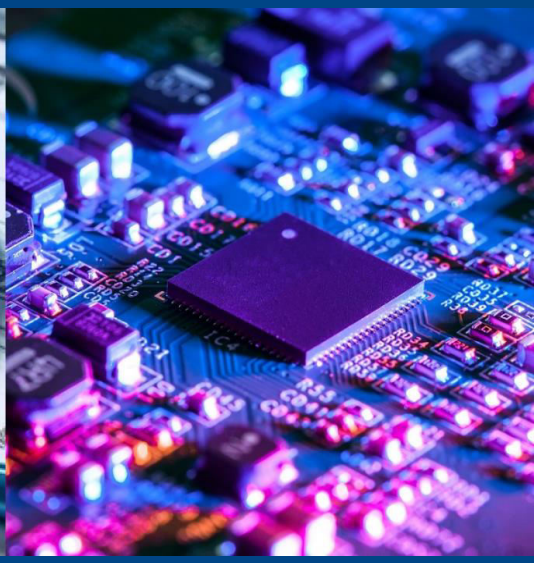
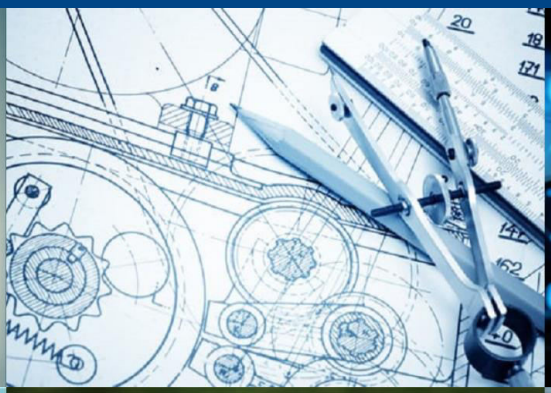
ISSN: 2582-7219

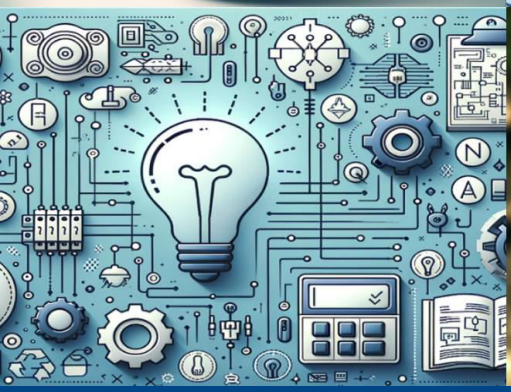




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 2, February 2026**

© 2026 IJMRSET | Volume 9, Issue 2, February 2026| DOI:10.15680/IJMRSET.2026.0902009

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fake Document Detector

Ms. Pratiksha Pradeep Varude [1], Ms. Shravani Santosh Belawale [2], Ms. Sakshi Babaso Gote [3], Ms. Divya Vasanta Hande [4], Ms. Samikasha Sunil Bansode [5], Ms. A. M. Masmardi [6]

Student, Department of Computer Engineering, Sharad Institute of Technology Polytechnic, (Yadrav) Ichalkaranji, Maharashtra, India[1, 2, 3, 4, 5]

Guide, Department of Computer Engineering, Sharad Institute of Technology Polytechnic, (Yadrav) Ichalkaranji, Maharashtra, India[6]

ABSTRACT: The rapid growth of digital documentation has increased the risk of forged and manipulated documents being used for fraud, identity theft, and misinformation. This project presents a Fake Document Detection system designed to automatically identify counterfeit or tampered documents using machine learning and image analysis techniques. The proposed system analyzes structural, textual, and visual features such as fonts, layout consistency, metadata, and image artifacts to distinguish genuine documents from fraudulent ones. Optical Character Recognition (OCR) is employed to extract textual content, while classification models are trained on labeled datasets of real and fake documents to improve detection accuracy. Experimental results demonstrate that the system effectively identifies forged documents with high precision and reliability. This approach reduces the need for manual verification, enhances security, and provides a scalable solution for organizations handling large volumes of sensitive documents.

I. INTRODUCTION

In today's fast-growing digital world, documents play a vital role in every field of human activity. From proving a person's identity to verifying qualifications, ownership, and authorization, documents are considered authentic evidence of truth. Governments, educational institutions, banks, and private organizations depend heavily on official documents such as Aadhaar cards, PAN cards, driving licenses, mark sheets, certificates, property records, and business licenses for their daily operations.

However, with the availability of advanced digital editing tools and image manipulation software, creating fake or tampered documents has become alarmingly easy. People can alter or duplicate information such as names, photos, barcodes, or signatures using software like Photoshop, Illustrator, or online PDF editors. As a result, document forgery and fraud have become a major social and security issue. Such forgeries can lead to serious consequences including identity theft, illegal admissions, fake employment, and financial scams.

Therefore, there is a growing need for an intelligent and automated system that can efficiently detect forged documents and verify their authenticity. This need gave rise to the concept of the Fake Document Detector System — a computer-based system that can identify whether a given document is genuine or fake using Machine Learning (ML), Image Processing, Optical Character Recognition (OCR), and Artificial Intelligence (AI) technologies.

II. LITERATURE REVIEW

Document verification has evolved significantly with the growth of digital documentation. Early methods relied on manual inspection of physical features such as paper quality, seals, signatures, and ink consistency. While effective for physical documents, these approaches were slow, subjective, and ineffective against digitally altered documents.

With the rise of scanned and soft-copy documents, researchers began using **digital image processing techniques**. Studies such as Sharma et al. (2016) and Kaur & Gill (2017) applied pixel-level analysis, texture comparison, and feature-matching methods (e.g., SIFT) to detect forged logos, seals, and background inconsistencies. Although these methods improved visual forgery detection, they failed to identify textual manipulations.

To address text-based forgery, **OCR-based systems** were introduced. Patel and Gupta (2018) and Rao et al. (2019) used OCR to extract text and compare it with stored records, enabling detection of altered names, dates, or identification numbers. However, OCR-only systems struggled with image-level tampering.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The introduction of **machine learning and deep learning** significantly improved detection accuracy. Models using SVMs and CNNs (Singh et al., 2019; Chen & Lee, 2020) automatically learned features from document images and detected subtle manipulations. Recent deep learning approaches, including hybrid CNN–OCR models (Das & Mishra, 2021; Ahmed et al., 2022), achieved high accuracy by combining image and text analysis.
Despite these advancements, existing systems often focus on either image-based or text-based forgery and lack real-time performance and ease of deployment. This highlights the need for a **hybrid, efficient, and user-friendly Fake Document Detection system** capable of handling both textual and visual forgery.

III. PROBLEM STATEMENT

With the rapid growth of digital technology, creating or modifying documents using editing software has become easy and widespread. Manual document verification is slow, error-prone, and impractical for handling large volumes of digital documents. As a result, an automated system is required to ensure secure and reliable document verification.
Today, digital and scanned documents are extensively used in education, banking, government, and private organizations for identity and record verification. At the same time, advanced editing tools enable forgers to manipulate text, images, photos, and seals with high precision, making fake documents difficult to detect through manual inspection.

Manual verification depends heavily on human judgment and cannot efficiently process the large number of documents submitted online. Moreover, high-quality forgeries often contain subtle modifications such as font inconsistencies or minor logo alterations that are not easily visible to the human eye.

To overcome these challenges, there is a strong need for an automated Fake Document Detector system. The proposed solution uses Optical Character Recognition (OCR) to verify text content, Image Processing to analyze visual elements, and Machine Learning techniques to classify documents as genuine or fake. This approach reduces human effort, improves accuracy, and enables fast, unbiased, and scalable document verification.

IV. SYSTEM ARCHITECTURE

The Fake Document Detector System is designed to automatically verify document authenticity using Machine Learning, Image Processing, and Optical Character Recognition (OCR). The system analyzes both textual and visual components to accurately detect forged or tampered documents while maintaining efficiency and user-friendliness.
The system architecture consists of a **web-based user interface** that allows users to upload documents in common formats such as PDF, JPG, and PNG. A **pre-processing module** enhances document quality through noise reduction, grayscale conversion, resizing, and contrast enhancement to ensure accurate analysis.

An **OCR module** extracts and analyzes text to identify inconsistencies such as font variations, alignment issues, and mismatches with stored records. In parallel, the **image analysis module** examines visual elements like logos, seals, edges, and background textures to detect manipulation. A **machine learning model**, typically a CNN, processes these extracted features to classify documents as genuine or fake and provides a confidence score.The system stores document data and verification results securely in a database and generates a final verification report highlighting authenticity status and suspected forgery regions.Overall, the proposed design delivers an automated, accurate, scalable, and secure solution for document verification, with a modular architecture that supports future enhancements and large-scale deployment.

V.SYSTEM OVERVIEW

The proposed Fake Document Detector allows users to upload scanned images or PDF documents, which are then automatically analyzed using OCR, Image Processing, and Machine Learning techniques. The system determines whether a document is genuine or forged and provides a confidence score indicating the reliability of the result.
The process begins with the **input stage**, where users upload documents such as ID cards, certificates, licenses, or official records. The uploaded document undergoes **pre-processing**, including noise removal, contrast enhancement, resizing, and skew correction to improve image quality for analysis.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

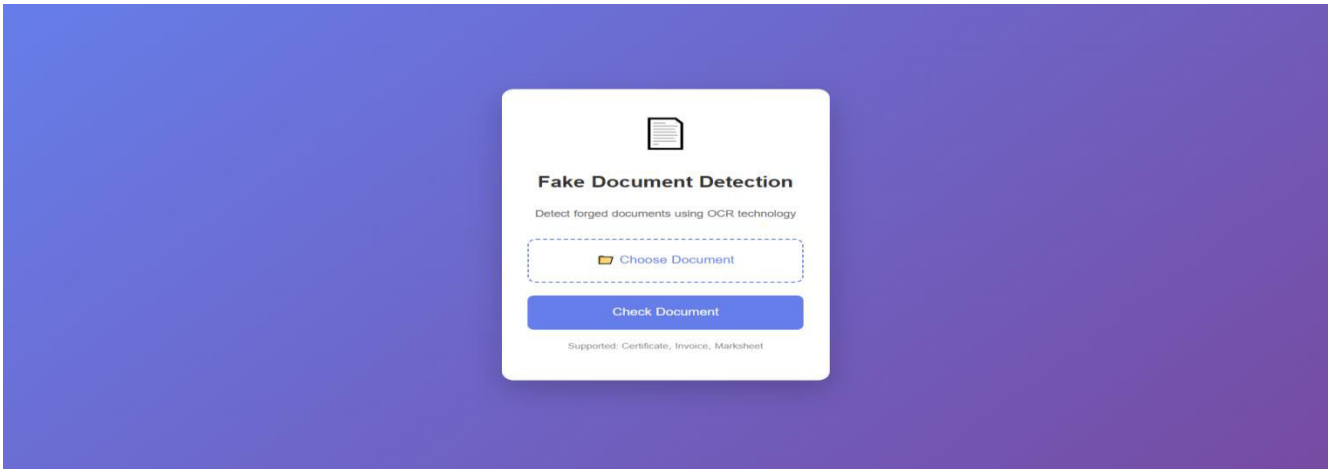
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

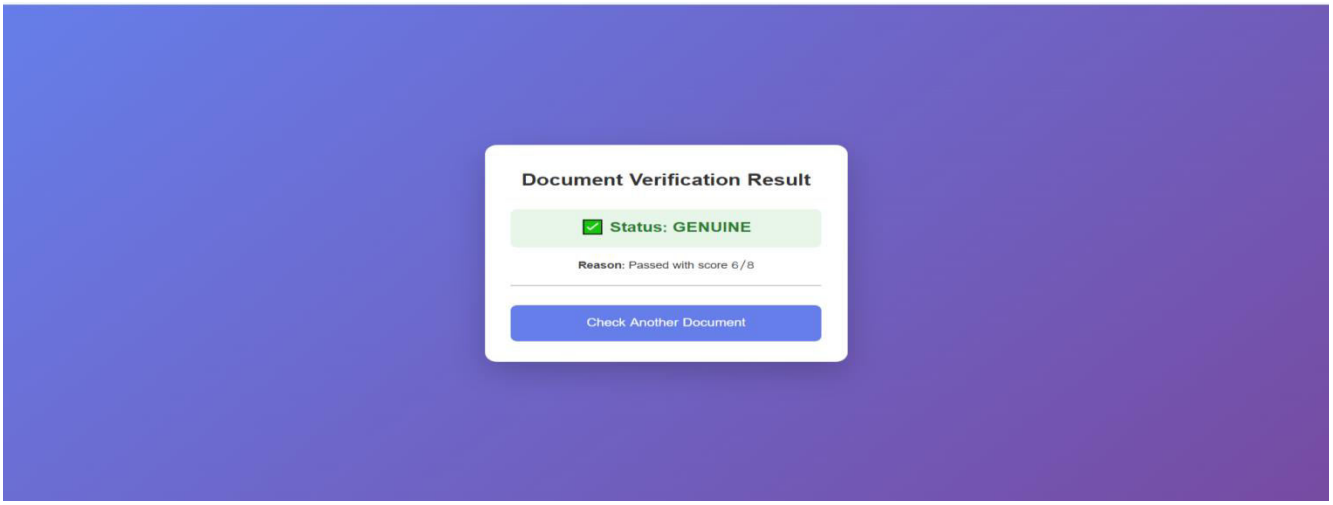
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Next, **OCR** is used to extract textual content from the document. The system checks for text anomalies such as font inconsistencies, misalignment, missing data, or formatting errors and enables cross-verification with stored records when available. In parallel, the **image analysis module** examines visual features to detect manipulation signs like copy-move forgery, splicing, color inconsistency, and blurred edges.

A **machine learning model**, typically a CNN trained on genuine and forged documents, analyzes extracted features and classifies the document as real or fake. Finally, the system presents the **output**, including document status, confidence score, and highlighted suspicious regions, along with a downloadable verification report.
The system is implemented using Python with tools such as OpenCV, TensorFlow/Keras, and Tesseract OCR, supported by a web-based interface and secure database storage. Overall, the proposed system offers an automated, accurate, scalable, and secure solution for digital document verification, helping organizations prevent fraud and ensure trust in online documentation.

VI. RESULTS AND OUTCOMES



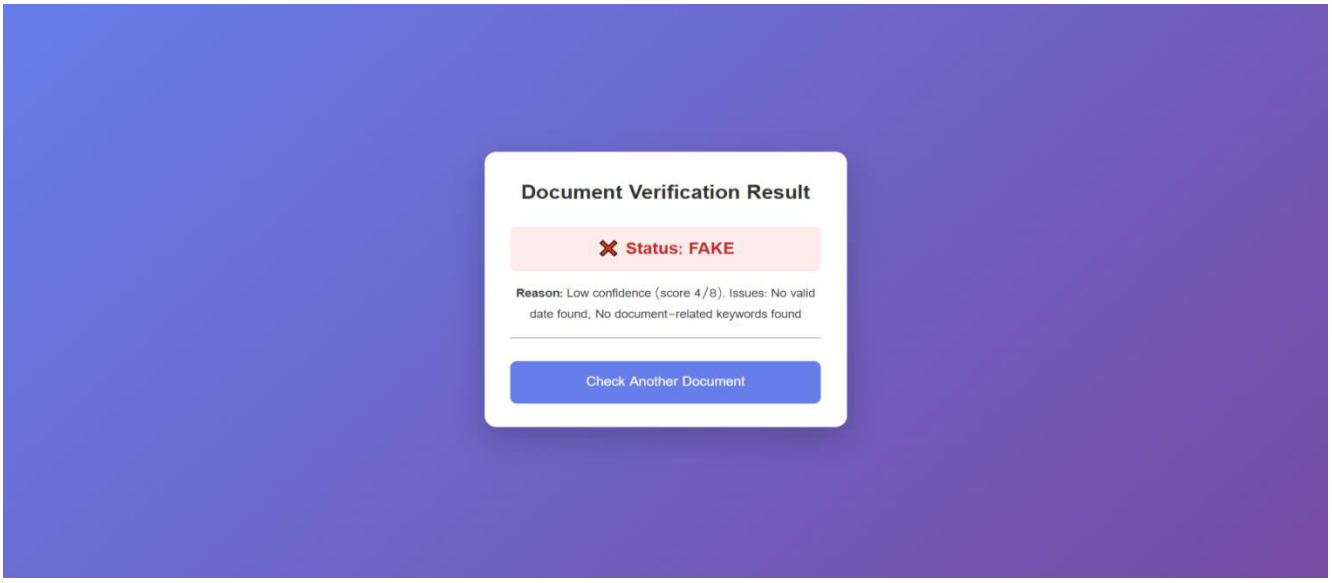


© 2026 IJMRSET | Volume 9, Issue 2, February 2026| DOI:10.15680/IJMRSET.2026.0902009

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



VII. ADVANTAGES & DISADVANTAGES

- **Advantages:**

1. **Automated Verification:**
- The system reduces the need for manual checking, saving time and effort.
2. **High Accuracy:**
- Combines OCR, AI, and image processing to detect both textual and visual tampering accurately.
3. **Time-Efficient:**
- Large volumes of documents can be verified quickly, which is ideal for banks, universities, and government offices.
4. **Consistency and Reliability:**
- Reduces human error and provides consistent results for every document.
5. **Confidence Score and Highlighted Tampering:**
- Shows suspicious areas and assigns a confidence score, allowing better decision-making.
6. **Security and Privacy:**
- Encryption and secure APIs protect sensitive data during verification.

- **Limitations:**

1. **Dependence on Quality of Input:**
- Poorly scanned documents or blurred images may reduce the accuracy of OCR and image analysis.
2. **Complex Handwritten Documents:**
- Handwritten text is difficult for OCR and AI to process accurately, limiting detection in such cases.
3. **Database Dependency:**
- Verification of document authenticity is limited by the availability of reference databases.
4. **Processing Large Files:**
- Very high-resolution documents or PDFs with multiple pages may increase processing time.
5. **New Forgery Techniques:**
- Advanced forgery techniques not included in the training dataset may evade detection until the system is retrained.

VIII. CONCLUSION AND FUTURE SCOPE

The **Fake Document Detector** project successfully demonstrates the design and implementation of an **automated system** for verifying the authenticity of documents. In today's digital world, where document forgery and tampering are increasingly common, such a system is essential for ensuring **security, reliability, and trust** in various sectors like banking, education, government, and corporate organizations.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

By integrating **Optical Character Recognition (OCR)**, **Artificial Intelligence (AI)**, and **image processing techniques**, the system can efficiently detect **textual inconsistencies** and **visual tampering** in uploaded documents. The combination of **AI-based classification** and **database verification** allows the system to provide results with a **confidence score**, reducing human effort and minimizing errors in manual verification.

The use of modern **software technologies** like Python, FastAPI/Django, OpenCV, TensorFlow, and secure databases, along with **frontend frameworks** such as React or Flutter, ensures a **user-friendly, scalable, and secure platform**. Additionally, encryption and secure APIs guarantee the protection of sensitive data during upload and verification.
While the system shows high efficiency and accuracy, there is scope for future enhancements, such as **integration with government databases**, **support for handwritten documents**, **real-time verification**, and **mobile deployment**, which can further expand its usability.

In conclusion, the Fake Document Detector is a **robust, reliable, and intelligent solution** for tackling document forgery. It not only simplifies the verification process but also enhances security and trust, making it a valuable tool for organizations and institutions handling sensitive documents.

REFERENCES

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
2. Gonzalez, R. C., & Woods, R. E. (2018). Digital Image Processing (4th Edition). Pearson.
3. Smith, R. (2007). An Overview of the Tesseract OCR Engine. Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR), 629–633.
4. Chollet, F. (2017). Deep Learning with Python. Manning Publications.
5. OpenCV Documentation. (2023). Retrieved from https://opencv.org/
6. TensorFlow Documentation. (2023). Retrieved from https://www.tensorflow.org/
7. FastAPI Documentation. (2023). Retrieved from https://fastapi.tiangolo.com/
8. Django Software Foundation. (2023). Django Documentation. Retrieved from https://www.djangoproject.com/
9. Tesseract OCR GitHub Repository. (2023). Retrieved from https://github.com/tesseract-ocr/tesseract
10. MongoDB Documentation. (2023). Retrieved from https://www.mongodb.com/
11. PostgreSQL Documentation. (2023). Retrieved from https://www.postgresql.org/
12. Jain, A. K., & Bhattacharjee, S. (2020). Document Forgery Detection Techniques: A Review. Journal of Information Security and Applications, 55, 102612.
13. Oliveira, L., & Farid, H. (2019). Detecting Forged Documents Using Image Processing and Machine Learning Techniques. International Journal of Computer Vision.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com