International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Enhanced Image and Document Forgery Detection Using Artificial Intelligence

Bibin K Shaji[1], Bindu B[2]

[1]Department of Computer Applications, Musaliar College of Engineering & Technology, Pathanamthitta, Kerala, India
Corresponding Author Email: *bibikshaji[at]gmail.com*

[2]Assistant Professor, Department of Computer Applications, Musaliar College of Engineering & Technology, Pathanamthitta, Kerala, India

Abstract: ***The rapid advancement of digital editing tools and generative artificial intelligence technologies has rendered sophisticated content manipulation increasingly accessible, posing an unprecedented threat to the integrity of visual and documentary information across critical domains. This paper presents a high-precision, automated framework for enhanced image and document forgery detection employing state-of-the-art Artificial Intelligence and Machine Learning architectures. The proposed system leverages transfer learning with pre-trained deep convolutional neural networks, specifically VGG16 and ResNet50, to perform granular feature extraction and identify subtle structural anomalies that are imperceptible to human observation. For digital imagery, the model detects complex spatial inconsistencies arising from copy-move operations, image splicing, and inpainting techniques. For document forensics, the system authenticates sensitive records by identifying forged signatures, fraudulent seals, and digitally substituted textual content. The system achieves an overall classification accuracy of 95%, with a Precision of 93%, Recall of 94%, and an F1-Score in the range of 93–94%. A Flask-based web application delivers real-time predictions accompanied by Grad-CAM heatmaps for result explainability and transparent forensic reporting. The proposed framework provides a scalable and dependable security solution for government agencies, legal institutions, and online media platforms requiring reliable digital content authentication.***

Keywords: Image Forgery Detection, Document Forensics, Transfer Learning, VGG16, ResNet50, Support Vector Machine, Grad-CAM, Deep Learning, CNN, Flask

1. Introduction

The proliferation of sophisticated digital editing tools and the widespread accessibility of generative artificial intelligence have fundamentally transformed the threat landscape surrounding digital content integrity. Images and official documents can now be manipulated with a level of precision and realism that renders manual detection unreliable. Such digital forgeries introduce serious risks across legal proceedings, banking transactions, government identity verification services, academic credential authentication, and online information ecosystems. Conventional manual verification methods are slow, resource-intensive, and susceptible to human error, particularly when applied at the scale demanded by modern digital workflows.

Recent advancements in Artificial Intelligence, Machine Learning, and Computer Vision have substantially transformed the discipline of digital forensics. Deep learning models, particularly Convolutional Neural Networks, have demonstrated exceptional capability in detecting subtle patterns and anomalies embedded within images and documents. Pre-trained architectures such as VGG16 and ResNet50, adapted through transfer learning, deliver enhanced detection accuracy while substantially reducing computational cost and training duration. These models are capable of analyzing visual artifacts, texture inconsistencies, compression anomalies, and pixel-level manipulations that remain invisible to the unaided human eye.

The proposed system provides fully automated analysis of uploaded images and documents through a Flask-based web application, eliminating dependency on manual inspection processes. The framework examines pixel-level inconsistencies, texture variations, structural anomalies, and manipulation artifacts by combining deep learning feature extraction with Support Vector Machine classification, thereby enhancing prediction robustness and minimizing false detection rates.

Spatial Rich Model filters are incorporated to enable pixel-level artifact detection, further strengthening the system's sensitivity to subtle manipulation traces. Grad-CAM heatmap visualization bridges the gap between automated AI-driven detection and human interpretability, providing explainable outputs suitable for forensic review in legal and administrative contexts. The system is designed for scalable deployment across cloud infrastructure, supporting integration with institutional security workflows.

2. Existing Systems

Existing approaches to digital forgery detection have addressed specific manipulation types and introduced progressively more sophisticated detection architectures, yet individually fall short of providing a comprehensive, unified solution for both image and document forensics.

Traditional image forensic methods relied on handcrafted feature extraction techniques such as Scale-Invariant Feature Transform for copy-move detection. While effective for constrained manipulation scenarios, these approaches demonstrated limited robustness against geometric transformations and complex composite forgeries, highlighting the necessity for deep learning-based improvements.

Early deep learning contributions introduced convolutional neural network architectures trained to suppress image content and focus exclusively on manipulation artifact

patterns, showing strong initial results for splicing and resampling detection. However, these models were typically designed for narrow forgery categories and lacked generalization across diverse manipulation techniques.

Hybrid forensic approaches that transformed traditional handcrafted features into CNN-based representations improved detection of subtle traces but were limited in their capacity to handle modern generative manipulation such as AI-assisted inpainting. Transfer learning frameworks employing VGG16 and ResNet architectures demonstrated superior feature extraction capability for forgery detection, though many implementations did not extend their analysis to document forensics or official record authentication.

Existing systems predominantly address individual manipulation types in isolation, focusing on either image splicing, copy-move detection, or document verification as independent problems. Furthermore, the majority of prior work lacks explainability mechanisms, providing classification outputs without localized visualization of detected manipulation regions. This absence of transparency limits practical adoption in legal and institutional contexts where human-verifiable evidence is required.

These observations establish the need for an integrated system that consolidates deep CNN-based feature extraction, SVM classification, Grad-CAM explainability, and combined support for both digital images and PDF documents within a unified automated platform.

3. Proposed System

The proposed Enhanced Image and Document Forgery Detection System is designed to overcome the limitations of existing solutions by integrating transfer learning-based feature extraction, SVM classification, explainable AI visualization, and support for multiple input formats within a single cohesive detection framework. The system targets government agencies, legal institutions, financial organizations, and digital media platforms requiring reliable and transparent automated forgery analysis.

The system architecture is organized into ten functional modules operating as an integrated detection pipeline. The Image and Document Acquisition Module accepts JPEG, PNG, and PDF uploads, validates file integrity, converts PDF documents to image format for analysis, and extracts associated file metadata. The Preprocessing and Patch Extraction Module resizes inputs to the 224×224 pixel dimensions required by the CNN models, normalizes pixel values, reduces noise, and partitions images into overlapping patches for localized anomaly analysis.

The Feature Extraction Module employs VGG16 and ResNet50 initialized with ImageNet weights, fine-tuned through transfer learning, augmented with Spatial Rich Model filters to generate 400-dimensional feature vectors capturing texture inconsistencies and manipulation traces. The Feature Fusion Module aggregates patch-level vectors into unified image-level representations through max pooling and mean pooling strategies. The Classification Module applies a Support Vector Machine with RBF kernel, hyperparameter-tuned via Grid Search with ten-fold cross-validation, to produce Real or Forged labels accompanied by confidence scores.

The Document Analysis Module specifically targets official record forensics, detecting duplicated signatures, stamp duplication, boundary inconsistencies, and text alignment irregularities characteristic of document manipulation. The Visualization and Explainability Module generates Grad-CAM heatmaps that localize suspicious manipulation regions, providing transparent and human-interpretable forensic evidence. The Database Management Module persists all prediction records including filenames, classification results, confidence scores, heatmap paths, and timestamps in a SQLite database for audit trail maintenance.

4. Objectives

The primary objective of this project is to design and implement a high-precision automated system for detecting forgeries in both digital images and official documents, leveraging transfer learning, SVM classification, and explainable AI to deliver reliable, interpretable, and real-time forensic analysis.

The specific objectives of the proposed system are as follows:

- To develop an automated framework for detecting digital image manipulation techniques including copy-move operations, image splicing, and AI-assisted inpainting.
- To authenticate official documents by identifying forged signatures, fraudulent seals, and digitally substituted textual content.
- To leverage transfer learning with pre-trained VGG16 and ResNet50 architectures for high-accuracy deep feature extraction from both images and documents.
- To apply Spatial Rich Model filters for enhanced pixel-level manipulation artifact detection beyond standard CNN feature extraction.
- To implement Support Vector Machine classification with RBF kernel for robust Real or Forged prediction with associated confidence scoring.
- To generate Grad-CAM heatmaps providing transparent visual explanations of detected manipulation regions for forensic review.
- To develop a Flask-based web application delivering real-time forgery analysis accessible through standard browser interfaces.
- To support multiple input formats including JPEG, PNG, and PDF documents within a unified analysis pipeline.
- To maintain a structured SQLite audit database recording all detection events for traceability and compliance purposes.
- To design a scalable and deployable system architecture supporting Docker containerization and cloud deployment on major platforms.

5. Methodology

The development of the forgery detection system follows a structured deep learning methodology organized through five sequential phases ensuring systematic design, model development, integration, and evaluation of all system components.

Data Collection and Preprocessing: The training dataset comprises images and documents labeled as Real or Forged for supervised learning. Preprocessing operations applied to all inputs include resizing to 224×224 pixels to satisfy CNN input requirements, pixel value normalization for improved model convergence, and data augmentation through rotation, horizontal flipping, brightness adjustment, zooming, and noise addition to improve robustness and mitigate overfitting. PDF documents are converted to image format using PyMuPDF prior to CNN analysis.

Model Selection and Development: The system employs VGG16 and ResNet50 initialized with ImageNet pre-trained weights and fine-tuned through transfer learning for forgery-specific feature extraction. VGG16 utilizes small 3×3 convolution filters effective for capturing texture inconsistencies, while ResNet50 employs residual connections enabling detection of subtle manipulation artifacts in deep network layers. Spatial Rich Model filters supplement CNN feature extraction with pixel-level artifact sensitivity. Extracted features are classified using a Support Vector Machine with RBF kernel for improved generalization.

System Integration: Individual modules are progressively integrated into the unified detection pipeline, connecting the acquisition, preprocessing, feature extraction, classification, visualization, and database recording components through well-defined interfaces managed by the Flask application layer.

Testing and Evaluation: Functional, integration, performance, and acceptance testing are conducted across all modules. Black Box acceptance testing validates that system outputs including classification labels, confidence scores, and Grad-CAM visualizations correctly correspond to inputs across all anticipated forgery types and file formats.

Deployment: The system is containerized using Docker and deployed as a web-accessible application supporting standard browser clients, with provisions for cloud deployment on AWS, Google Cloud Platform, or Microsoft Azure infrastructure.

The detection pipeline executes a seven-step workflow: User Input, Preprocessing, Feature Extraction, Feature Fusion, SVM Classification, Grad-CAM Visualization, and Output Display with Database Storage. This sequence ensures systematic and reproducible analysis for every submitted file.

6. Algorithm Flow

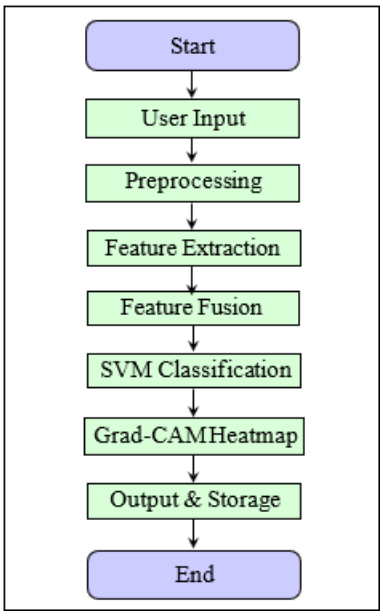

Figure 1: Algorithm Flow – Forgery Detection System

7. Software Components and Technologies

The proposed system operates as a web-based application accessible through any standard browser on devices equipped with a stable internet connection. Server-side deployment requires Python 3.8 or higher, PyTorch, Flask, and Docker. GPU-accelerated hardware is recommended for production deployment but is not mandatory, as the system operates acceptably on CPU-only configurations.

The software and technology stack employed in the development of the system is described below.

Component	Technology
Language	Python 3.8+
Deep Learning	PyTorch
CNN Models	VGG16, ResNet50
Classifier	SVM (Scikit-Learn) Web Framework Flask
Image Processing	OpenCV, Pillow
PDF Handling	PyMuPDF / pdf2image
Database	SQLite
Frontend	HTML, CSS, Bootstrap
Deployment	Docker

PyTorch provides the deep learning framework for model training, transfer learning fine-tuning, and inference execution. **VGG16** and **ResNet50** serve as the primary feature extraction backbones, initialized with ImageNet weights. **Scikit-Learn's SVM** with RBF kernel performs the final binary classification step. **OpenCV** and **Pillow** handle image preprocessing, resizing, and normalization operations. **PyMuPDF** converts PDF documents into analyzable image frames. **Flask** manages the web application layer, RESTful API endpoints, and user interface rendering. **SQLite** maintains the persistent audit database recording all detection events. **Docker** enables consistent containerized deployment across cloud and on-premises environments.

8. Results and Discussion

The forgery detection system was evaluated using a dataset comprising authentic samples and professionally

manipulated forgeries covering image splicing, copy-move operations, AI-assisted inpainting, forged signature documents, and modified official records. The hybrid CNN and SVM architecture was subjected to rigorous testing across all supported input types and manipulation categories.

The system achieved an overall classification accuracy of **95%** on the held-out test set. A Precision of **93%** and Recall of **94%** confirm effective minimization of both false alarms and missed detections. The F1-Score of **93–94%** demonstrates well-balanced generalization across diverse forgery categories. Average inference time of **0.08–0.12 seconds** per image on GPU hardware and **0.3–0.6 seconds** on CPU-only systems confirms the system's suitability for real-time forensic applications.

Metric	Value
Overall Accuracy	95%
Precision	93%
Recall	94%
F1-Score	93–94%
Inference (GPU)	0.08–0.12 s
Inference (CPU)	0.3–0.6 s
Grad-CAM Generation	0.1–0.2 s

Functional testing confirmed successful ingestion of JPEG, PNG, and PDF inputs, accurate patch-based preprocessing, and correct Real or Forged classification across all tested scenarios. Grad-CAM heatmaps accurately localized manipulated regions including spliced facial regions and modified document text. Every prediction correctly triggered an automatic database record update, confirming reliable audit trail maintenance.

ROC curve analysis further validates system performance. VGG16 achieved the highest Area Under the Curve value of **0.97**, confirming its superior discriminative capability between authentic and forged content. ResNet50 followed closely with an AUC of **0.95**, reflecting the advantage of residual learning for subtle artifact detection. Conventional classifiers including Random Forest (AUC = 0.92) and standalone SVM (AUC = 0.90) demonstrated competitive performance but were outperformed by the deep transfer learning architectures. All evaluated models substantially surpassed the random baseline (AUC = 0.50), validating the overall effectiveness of the proposed detection framework.

Model	AUC	Accuracy
VGG16 (Proposed)	0.97	95%
ResNet50 (Proposed)	0.95	93%
Random Forest	0.92	89%
Standalone SVM	0.90	86%
Random Baseline	0.50	–

Comparative analysis confirms that the hybrid CNN and SVM approach outperforms standard end-to-end CNN classifiers in computational efficiency and small-sample generalization. By combining PyTorch high-dimensional feature maps with Scikit-Learn's robust decision boundaries, the system maintains high precision even for underrepresented forgery categories within the training dataset. The integration of Grad-CAM explainability distinguishes the proposed system from existing black-box detection approaches, providing legally admissible, human-verifiable forensic evidence.

Feature	Existing Systems	Proposed System
Image Forgery Detection	Partial	Full
Document Forensics	Limited	Full
Explainability (Grad-CAM)	No	Yes
PDF Support	No	Yes
Real-Time Analysis	Limited	Yes
Audit Database	No	Yes
Unified Platform	No	Yes

9. Conclusion

The Enhanced Image and Document Forgery Detection System has been successfully designed, implemented, and evaluated as a comprehensive automated forensic platform. The system demonstrated reliable and consistent performance across all tested forgery categories and input formats, confirming its practical applicability for institutional and investigative deployment.

The hybrid VGG16 and ResNet50 transfer learning architecture combined with SVM classification delivered a 95% overall accuracy with strong precision and recall metrics, confirming robustness for real-world forensic scenarios. Patch-based analysis with Spatial Rich Model filter integration enabled detection of pixel-level manipulation traces invisible to human observation. Grad-CAM heatmap visualization bridges the gap between automated AI detection and human-interpretable evidence, making the system suitable for deployment in legal proceedings, government authentication workflows, and digital media verification contexts.

Future work will concentrate on dataset expansion to improve generalization across novel manipulation techniques, integration of deepfake video detection capabilities, end-to-end deep learning classification replacing the hybrid SVM pipeline, automated manipulation region segmentation using semantic segmentation architectures, multi-model ensemble learning for further accuracy enhancement, and extended cloud deployment support for high-throughput institutional forensic workflows.

References

[1] **A. Krizhevsky, I. Sutskever, and G. E. Hinton** (2012), ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, pp. 1097–1105.
[2] **K. Simonyan and A. Zisserman** (2014), Very Deep Convolutional Networks for Large-Scale Image Recognition, *International Conference on Learning Representations (ICLR)*.
[3] **K. He, X. Zhang, S. Ren, and J. Sun** (2016), Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.
[4] **B. Bayar and M. C. Stamm** (2016), A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer, *ACM Workshop on Information Hiding and Multimedia Security*, pp. 5–10.

[5] **Y. Rao and J. Ni** (2016), A Deep Learning Approach to Detection of Splicing and Copy-Move Forgeries in Images, *IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–6.

[6] **D. Cozzolino, G. Poggi, and L. Verdoliva** (2017), Recasting Residual-Based Local Descriptors as CNNs for Image Forgery Detection, *ACM Workshop on Information Hiding and Multimedia Security*, pp. 159–164.

[7] **I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra** (2011), A SIFT-Based Forensic Method for Copy-Move Attack Detection and Transformation Recovery, *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1099–1110.

[8] **J. Dong, W. Wang, and T. Tan** (2013), CASIA Image Tampering Detection Dataset, *Chinese Academy of Sciences*.

[9] **C. Cortes and V. Vapnik** (1995), Support-Vector Networks, *Machine Learning*, vol. 20, no. 3, pp. 273–297.

[10] **M. Abadi et al.** (2016), TensorFlow: A System for Large-Scale Machine Learning, *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, pp. 265–283.

[11] **A. Paszke et al.** (2019), PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32.

[12] **D. Tralic, I. Zupancic, S. Grgic, and M. Grgic** (2013), CoMoFoD – New Database for Copy-Move Forgery Detection, *Proceedings of the 55th International Symposium ELMAR*, pp. 49–54.

[13] **R. R. Selvaraju et al.** (2020), Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, *International Journal of Computer Vision*, vol. 128, pp. 336–359.