Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/IN05.20250425
Published: November 6, 2025

Document Forgery Verification Using OCR and ConvneXt V2

S.M. Hasanuddin, Shahwar Mahmood and Aryan Asthana
Student, Methodist College of Engineering and Technology
Hyderabad, India
s.hasanuddin20@gmail.com, shahwarmahmood7@gmail.com, aryanasthana05@gmail.com

Umaira Shahneen
Student, Faculty of Engineering
Sharnbasva University, Kalaburagi, India
umairashahneenkhan@gmail.com

Syeda Afeefa Fatima
Student, PDA College of Engineering, Kalaburagi, India
safkhadri6@gmail.com

Qutubuddin Syed Mohammed
Professor, Industrial & Production Engineering
P.D.A. College of Engineering, Kalaburagi, India
syedqutub16@gmail.com

Abstract

Document forgery is a prevalent threat, especially when verifying identities, academic credentials, and during financial transactions. Traditional document authentication methods are heavily reliant on manual reviews or rule-based systems. These systems are time consuming and often fail to notice subtle alterations in the document. Commercial tools like Google Document AI and Microsoft Azure Form Recognizer offer strong OCR features but do not support forgery detection, visual understanding, or personalization. AI solutions for the problem have been researched but they are limited to either extremely specific datasets or fail to utilize better and efficient models for the classification. These solutions are based on datasets that are not available publicly and use basic CNN to classify the images. These papers also fail to include various types of forgery and only focus on certain types specific to their dataset. In consideration of the above issues, we present an AI-based document verification system that brings together Optical Character Recognition (OCR), deep learning forgery detection, and layout-aware document classification. The process starts with image improvement using Local Thresholding and Histogram Equalization (LTHE) and CLAHE. Text is extracted with EasyOCR, and then two main models process the data in parallel: ConvNeXt, which works with the OCR output to detect forgery, and LayoutLMv3, which classifies documents based on their visual structure and content. The output includes a classification label, a forgery confidence score, and a visual explanation through Score-CAM.

Keywords
Document, Forgery, OCR, Convnext

1. Introduction

Digitization of governance processes has required quicker and more secure document verification mechanisms. From academic transcript verification of universities to know your customer documentation checks by banks, the detection of forged or manipulated documents has become a foremost issue. Yet contemporary forgery methods—running from straightforward editing to AI-assisted synthetic alterations—have gone ahead of conventional verification processes. Manual check is quite time-consuming, prone to errors, and not well designed for noticing slight manipulations, particularly when documents look legitimate on the surface.

To fill this void, the PaperTrail project proposes an AI-based system for computerized document forgery verification. It harnesses the strengths of two different yet complementary models: a ConvNeXt-based forgery detection model boosted with EasyOCR text support and LayoutLMv3, a transformer model that can perform layoutaware classification of documents. Combined, these models scan both the visual and structural properties of a document and provide an overall authenticity judgment.

Another key characteristic of PaperTrail is that it is explainable. With Score-CAM, the system marks tampered areas on the document image, enabling users to understand and believe the model's choice. Unlike most black-box models, this visual interpretability provides accountability and auditability.

In lieu of lack of publically available forgery datasets, a custom dataset was synthesized by enhancing and augmenting existing datasets like SIDTD, RVL-CDIP, to create a comprehensive dataset of forged and real documents across multiple document types.

1.1 Objectives

Automates classification and verification of various official documents using content and layout understanding.
Integrates OCR to extract and analyze text from scanned documents for deeper validation.
Employs deep learning to detect forged regions and manipulated text in documents.
Reduces reliance on manual verification, ensuring faster and more consistent processing.
Provides a secure, user-friendly interface with detailed reports, including forgery heatmaps and confidence scores.
Supports audit trails and secure storage for compliance and legal tracking.
Designed to be modular and adaptive, improving over time with exposure to more document types and forgeries.

2. Literature Survey

The earliest work we reviewed targets the academic domain, where Sirapat Boonkrong (2024), proposes. a hash-function-based pipeline to authenticate digital transcripts and certificates, benchmarking it against CNN and blockchain baselines and reporting 100 % verification accuracy on test documents. Although effective, the scheme presumes the verifier has access to pristine originals and is confined to academic layouts, limiting its utility in heterogeneous, real-world settings. PaperTrail removes that dependency by operating directly on single, possibly tampered images and extending coverage to many document classes.

To amplify subtle artefacts before deep learning, Yong-Yeol Bae, Dae-Jea Cho, and Ki-Hyun Jung (2025), introduce Log-Transform Histogram Equalization (LTHE) [2], a symmetry-aware enhancement that increases local contrast and preserves edge detail. Their Symmetry-journal study shows consistent accuracy gains across three CNN backbones, confirming that better low-level features translate into stronger forgery detection. PaperTrail adopts LTHE—followed by CLAHE—so ConvNeXt receives sharper inputs and can recognise faint copy-move seams or blurred overwrite zones.

For a broader criminological perspective, B. Karimov (2024) classifies manipulation tactics—including erasure, overwriting, and copy-move—and stresses that reliable detection must fuse visual analysis with semantic checks such as field-to-field consistency. This aligns with PaperTrail's hybrid stack, where EasyOCR extracts text, LayoutLMv3 verifies structural semantics, and ConvNeXt inspects pixel-level integrity, jointly tackling the spectrum of tampering modes Karimov catalogues.

A CAPSULE-Net/ELA framework by Nandini N. et al (2024), tackles signature and copy-move forgery through compression-level inconsistencies. While novel, its heavy reliance on JPEG error maps makes it fragile under aggressive down-sampling or re-save operations. PaperTrail mitigates such brittleness by employing feature maps

learned by ConvNeXt—which are less sensitive to compression noise—and by providing Score-CAM heat-maps for transparent decision support.

Deep-forgery generation is explored by Yamato Okamoto et.al.2023, who synthesize diverse attacks and use self-supervised pre-training to bridge domain gaps. They demonstrate that large-scale synthetic corpora can improve robustness but also highlight the residual mismatch between artificial and genuine artefacts. PaperTrail similarly augments data but grounds it in SIDTD and MIDV2020 templates before adding realistic noise, achieving a balance between synthetic diversity and authentic texture statistics.

Datasets underpinning modern research remain scarce. MIDV-2020 provides 72 k annotated frames of 1 000 mock IDs under varied capture conditions, supplying a benchmark for detection, localisation, and classification tasks; however, it contains no explicit forgeries. SIDTD, introduced by Carlos Boned, Maxime Talarmain, Nabil Ghanmi et al (2024), extends MIDV-2020 with systematically altered documents and forged text/photo regions, addressing the imbalance between genuine and tampered samples and offering predefined training-validation partitions. PaperTrail draws on both: LayoutLMv3 is first tuned on MIDV-2020 for layout comprehension, while ConvNeXt is refined on SIDTD-derived and custom-augmented forgeries to learn pixel-level anomalies.

Commercial OCR suites—Google Document AI, Azure Form Recognizer, and Amazon Textract—deliver high-quality text extraction and form understanding but lack native forgery detection, provide little model transparency, and incur usage costs. PaperTrail closes this functional gap by integrating open-source OCR, CNN, and transformer modules within an explainable workflow that runs locally or on-prem, satisfying domains where auditability and cost control are paramount.

Recent work on language-specific challenges demonstrates that a one-size-fits-all approach to document-forgery detection can falter when confronted with non-Latin scripts. Bae, Cho, and Jung (2025) analyse visual complexity in Korean forms and show that CNN models trained solely on English datasets under-perform by more than 15 percentage points when applied to Hangul documents, chiefly because line density, character spacing, and glyph complexity differ markedly from Latin-alphabet layouts. By introducing a Korean-language benchmark and highlighting these domain gaps, they argue for localized datasets and adaptive pre-training. PaperTrail meets this call by allowing document-type-specific fine-tuning and by keeping its OCR layer modular so that language-specific models can be plugged in without revisiting the entire pipeline.

Colour information can also be leveraged as a cue for tampering. Gornale, Patil, and Benne (2022) propose an RGB-channel statistical framework that detects forged regions via abnormal distribution shifts between red, green, and blue layers. Tested on scans of certificates and licences, their method excels at spotting copy-move edits that disturb channel consistency, yet it struggles when forgers re-sample the whole image to hide chromatic artefacts. PaperTrail absorbs this insight by incorporating colour-space features alongside texture and frequency cues in ConvNeXt's training data, thereby strengthening resilience against colour-based manipulations.

Early foundational work by Harley, Ufkes, and Derpanis (2015) explored the application of deep convolutional neural networks for document image classification and retrieval. Their study, conducted before the rise of layout-aware transformers, demonstrated that CNNs could outperform traditional handcrafted feature-based approaches on document type classification tasks. The research established that visual cues alone—such as spacing, font density, and layout structure—contain rich semantic information for classifying scanned documents. This insight underpins PaperTrail's use of ConvNeXt and also informed the decision to fine-tune transformer models like LayoutLMv3 with visually diverse datasets such as MIDV-2020 and RVL-CDIP.

2.1 Datasets Used

The success of any document forgery detection system hinges on the quality and variety of its training data. Since no single public dataset met the dual needs of document classification and forgery detection, the PaperTrail project adopts a hybrid approach, combining public datasets with custom augmentations to simulate real-world document fraud scenarios.

2.1.1 SIDTD (Synthetic ID and Travel Document Dataset)

The SIDTD dataset, introduced by Boned et al. (2025), was used as the secondary source for training the forgery detection model. SIDTD builds upon MIDV-2020 [10] and introduces systematic manipulations such as fake field insertions, image swaps, and layout tampering. These synthetic forgeries mirror real-world techniques used in document fraud. For PaperTrail, SIDTD was further augmented with compression artifacts, Gaussian noise, low-resolution blurring, and printer-scan distortions to increase the model's robustness.

2.1.2 RVL-CDIP (Ryerson Vision Lab – Center for Document Image Processing Dataset)

The RVL-CDIP dataset (Harley, Ufkes, and Derpanis 2015) was incorporated to broaden the system's classification capabilities beyond IDs. It includes over 400,000 real-world scanned documents categorized into 16 types such as letters, reports, forms, and invoices. While not forged, this dataset enhances the model's generalization by introducing diverse print structures and office-style layouts. A subset of this dataset was used as real images for which forgeries were generated for the purpose of training the model.

2.1.3 Custom Synthetic Augmentation

Recognizing the limitations of available datasets, additional synthetic forgeries were generated using real and synthetic templates. This involved:

- Programmatic field replacement using mismatched fonts.
- Copy-paste tampering between document samples.
- Artificial scan distortions (blur, resolution loss).
- Compression and chromatic noise artifacts.

These augmentations were essential to teaching the forgery detection model how to distinguish genuine texture patterns from digitally manipulated regions. The final dataset consisted of approximately 28,000 labeled real and forged samples, which were split across training, validation, and testing sets for ConvNeXt V2.

3. Methodology

The implementation phase of PaperTrail is arguably the most essential phase. It begins with the gathering of data and ends with the entire web-app being integrated with the trained models.

3.1 Modules

The implementation of the PaperTrail system is broken down into several distinct modules, each responsible for a specific part of the functionality.

3.1.1 Dataset Preparation

The basis of PaperTrail's ability to detect forgeries lies in its robust dataset architecture:

- SIDTD (Synthetic ID and Travel Documents) offers systematically created, forged documents, such as edited fields and exchanged images.
- RVL-CDIP offers more than 400,000 actual scanned documents in 16 classes to serve a variety of document categories.
- Custom Synthetic Augmentation includes:
    - Copy-paste manipulations
    - Font and style inconsistencies
    - Compression, noise, and scanner artifacts

These datasets were used to create a comprehensive dataset which was preprocessed and separated into training, validation, and test sets of approximately 28,000 samples (real and counterfeit). This method guarantees layout variability and authentic forgery strategies.

3.1.2 Image Preprocessing

We preprocess document images before we feed them into OCR and CNNs using:

- LTHE and CLAHE to increase contrast and detail
- OpenCV for resizing, denoising, and normalization

This preprocessing process ensures a uniform quality of input from various sources, i.e., camera scans and PDFs.

3.1.3 OCR Extraction

We apply EasyOCR to extract multilingual text from the improved images. The OCR layer yields:

- Raw extracted text
- Bounding boxes and layout information
- Structured field-wise data (e.g., Name, Date, ID Number)

This output is subsequently passed on to downstream models for reasoning and validation.

3.1.4 Forgery Detection Model (ConvNeXt V2)

Having been trained on the augmented datasets, the ConvNeXt V2 model:

- Takes preprocessed image tensors as input
- Identifies forged regions based on texture, color, and layout irregularities

Output consists of a Forgery Label, Confidence Score, and Activation Map. The model is optimized using Focal Loss and includes Score-CAM for visual explanations.

3.1.5 Document Classification (LayoutLMv3)

The LayoutLMv3 model is trained on RVL-CDIP and MIDV-2020 layouts: ● It takes both visual layout and OCR content into account

● It predicts document types (e.g., Aadhaar, Invoice, Bank Letter)

The output contains a Document Class Label with confidence. The model assists in steering documents through custom validation pipelines.

3.1.6 Visual Explainability with Score-CAM

Score-CAM generates a heatmap for every document being scored:

- Identifies forged regions
- Displays model attention for improved understanding
- Aids manual verification by auditors

These heatmaps are returned with classification labels in the frontend.

3.2 Architecture

PaperTrail's three-tier system architecture ensures scalability, modularity, and a clear division of responsibilities (Figure 1)

Figure 1: High-Level System Architecture

3.3 Flowchart

The workflow of the system follows a clear, structured path from input to output, as detailed by the flowchart below (Figure 2).

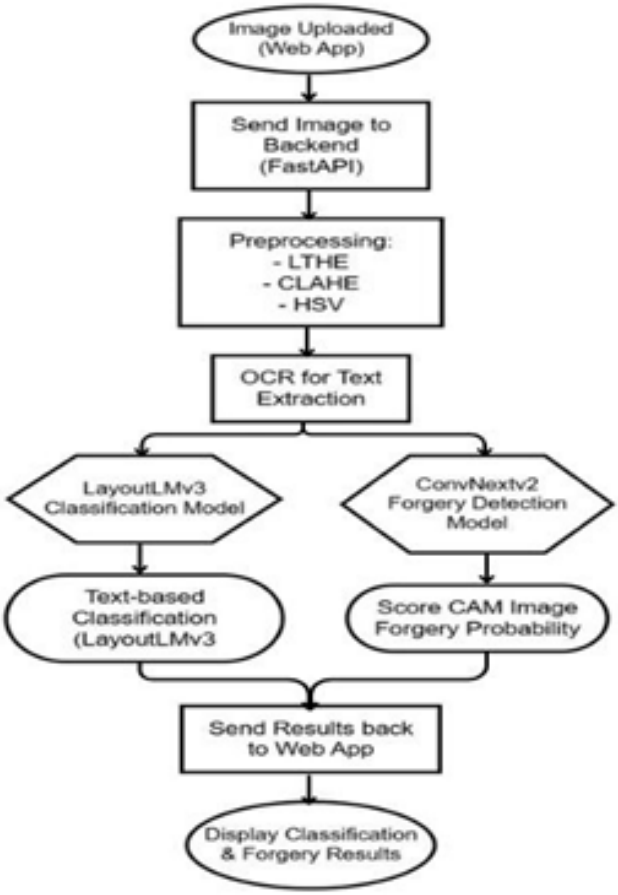

Figure 2: Flowchart

4. Testing

Testing is a vital phase in the development of PaperTrail. This phase aims to verify the performance, stability, and reliability of the main features of the system - forgery detection, document classification, OCR quality and the integration of the entire system.

4.1 Testing Strategy

Every element within the system was individually tested to ensure it functioned properly independent of others. We coded unit tests with Python's inbuilt unittest and pytest modules. The following modules were tested:

- Image preprocessing modules (CLAHE, LTHE)
- OCR extraction and formatting rules
- ConvNeXt and LayoutLMv3 model loading and inference functions
- Score-CAM generation rules
- API utility functions and error checks

These tests verify input-output consistency, data type integrity, and exception handling at the module level.

4.3 Test Results

Confusion Matrix: The confusion matrix shows that the model correctly identified 1740 real documents and 1765 forged documents, demonstrating a balanced performance across both classes (Table 1, Figure 3)).

Table1. Classification Report

Class	Precision	Recall	F1-Score	Support
Real	0.81	0.86	0.83	2017
Forged	0.86	0.81	0.83	2186
Accuracy			0.83	4203
Macro Avg	0.83	0.84	0.84	4203
Weighted Avg	0.84	0.83	0.83	4203

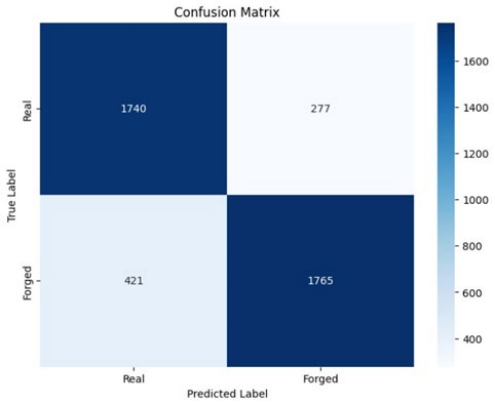

Figure 3. Confusion Matrix

Validation vs Training Scores: The scores show a steady rise in accuracy accompanied by a steady fall in the loss which starts too slow after a certain point (Figure 4 and 5).

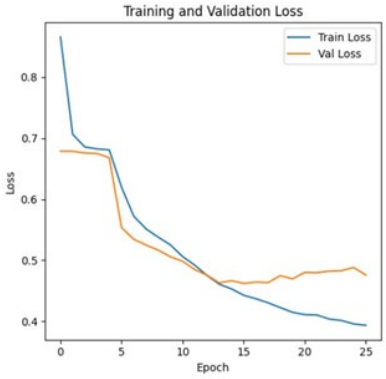

Figure 4: Training vs Validation Loss

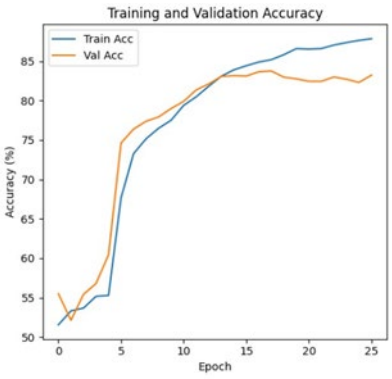

Figure 5: Training vs Validation Accuracy

5. Conclusion

The PaperTrail system delivers a robust and agile solution to ensure document forgery verification. It applies AI methods in image recognition, document categorization, and text extraction. With the age of digital and scanned documents booming across industries such as banking, education, governance, and HR, the urgency of rapid and trusted document authentication grows by the day. PaperTrail addresses this need by streamlining the verification process through a multi-model deep learning methodology.

The model incorporates ConvNeXt V2 for visual anomaly detection and LayoutLMv3 for document layout and structure comprehension. This guarantees that content and context are both analyzed. EasyOCR enables the system to

read text from documents of all kinds, even if noisy or degraded. Preprocessing techniques such as CLAHE and LTHE enhance document quality and enhance the performance of OCR and classification models.

One of the system's strongest features is that it utilizes Score-CAM as an explainability mechanism. In contrast to most black-box systems, PaperTrail provides transparency in the form of visual heatmaps. These heatmaps inform users which areas of a document impacted the model's forgery determination. This establishes user trust, facilitates easier audits, and provides an additional layer of verification support.

The project also prioritizes user usability and practicality in the real world. It has a FastAPI-based backend and a minimalist web frontend, so users can use the system in real time. Documents are processed and results are received in less than 10 seconds on normal hardware, making the system simple to integrate into normal workflows.

In short, PaperTrail bridges the old manual checks of documents to new intelligent systems of verification. It enhances accuracy, reduces human error, and gives results in a format that is easily understandable. Its modularity and the ability to accommodate growing datasets and document types make it a firm foundation for future expansion and use within industries.

References

B. Karimov, "Methods of Document Forgery and Their Detection Problems," The American Journal of Political Science, Law and Criminology, vol. 6, no. 02, pp. 50-54, 2024. https://www.theamericanjournals.com/index.php/tajpslc/article/view/4853

C. Boned, M. Talarmain, N. Ghanmi, G. Chiron, S. Biswas, A. M. Awal, O. R. Terrades, J. Llados, and T. Paquet, "SIDTD: Synthetic Dataset of ID and Travel Documents," arXiv preprint arXiv:2401.01858, 2024. https://arxiv.org/abs/2401.01858

K. Bulatov, E. Emelianova, D. Tropin, N. Skoryukina, Y. Chernyshova, A. Sheshkus, S. Usilin, Z. Ming, J.-C. Burie, M. M. Luqman, and V. V. Arlazarov, "MIDV-2020: A Comprehensive Benchmark Dataset for Identity Document Analysis," arXiv preprint arXiv:2107.00396, 2021. https://www.researchgate.net/publication/359393635_MIDV-2020_a_comprehensive_benchmark_dataset_for_identity_document_analysis

N. Nandini, K. J. Joshi, B. Devprakash, C. Madhura, and V. M. Ladwani, "Document Forgery Detection," International Journal of Engineering and Advanced Technology, vol. 12, no. 5, pp. 39-44, 2023. https://journals.blueeyesintelligence.org/index.php/ijeat/article/view/270?articlesBySa meAuthorPage=3

S. Boonkrong, "Design of an Academic Document Forgery Detection System," International Journal of Information Technology, 2024. https://link.springer.com/article/10.1007/s41870-0.\ 24-02006-6

S. S. Gornale, G. Patil, and R. Benne, "Document Image Forgery Detection Using RGB Color Channel," Transactions on Machine Learning and Artificial Intelligence, vol. 10, no. 5, pp. 1-14, 2022. https://journals.scholarpublishing.org/index.php/TMLAI/article/view/13126

W. Harley, A. Ufkes, and K. G. Derpanis, "Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval," in Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR), 2015, pp. 991–995. https://arxiv.org/abs/1502.07058

Y. Okamoto, G. Osada, I. Yahiro, R. Hasegawa, P. Zhu, and H. Kataoka, "Image Generation and Learning Strategy for Deep Document Forgery Detection," arXiv:2311.03650, 2023. https://scholar.google.com.vn/citations?view_op=view_citation&hl=id&user=9gAuuL0AAAAJ&citation_for_view=9gAuuL0AAAAJ:2osOgNQ5qMEC

Y.-Y. Bae, D.-J. Cho, and K.-H. Jung, "A New Log-Transform Histogram Equalization Technique for Deep Learning-Based Document Forgery Detection," Symmetry, vol. 17, no. 3, p. 395, 2025. https://www.mdpi.com/2073-8994/17/3/395