A Sensor Pattern Noise (SPN) and Statistical Consistency Framework for Forgery Detection in Scanned Documents

Deepika Dubey[1], Dr. Richa Rohatgi[2*], Dr. Seema R. Pathak[3*]

[1] Research Scholar, Department of Chemistry, Biochemistry and Forensic Science, Amity School of Applied Sciences, Amity University, Haryana, India

[2]Assistant Professor, LNJN NICFS, National Forensic Science University, Delhi Campus, Rohini, New Delhi

[3]Professor & HOD, Department of Chemistry, Biochemistry and Forensic Science, Amity School of [1]Applied Sciences, Amity University, Haryana, India

Abstract: *The increasing use of digital documents in legal, academic, and administrative settings has made it a common target for forgery, especially when scanned documents are involved. Many traditional methods for detecting forgery, such as checking metadata or looking for visual artifacts, often fail when the original file is unavailable or when the forgery is distinctive. This study outlines a statistical approach that tries to address that gap. The focus here is on Sensor Pattern Noise (SPN), a kind of noise that naturally comes from imperfections in scanners or cameras. Since SPN tends to stay consistent for each device, the idea is to use it like a fingerprint to check whether different parts of a document truly belong together or not. We suggest breaking down the image into small patches and then analyzing each patch individually and in comparison, to its neighbors. Features like lighting, edge sharpness, noise consistency and crosspatch analysis are used to flag suspicious regions. Although the framework hasn't been tested on real data yet, it's designed to be practical and adaptable. Later work will involve creating test datasets, running comparisons with existing methods, and checking how well this model holds up in actual forensic use.*

Keywords: **Digital Documents, Forgery, Manipulations, Scanned Documents, Forensic Examination, Sensor Pattern Noise (SPN).**

1. INTRODUCTION

The increasing shift from paper-based documentation to digital systems has, unexpectedly, brought with it a new generation of challenges particularly around document's authenticity and forgery detection. Digital documents, whether they are created digitally or scanned from paper, have increasingly became important for everything from legal status to academic credentials. But this ease of use comes with some risk. Tools for modifying and altering documents have come a long way in just a short period of time, frequently exceeding the defensive capability of the forensic systems. What once required specialist software is now possible via free apps on a smartphone or a laptop. This shift has made it easier for individuals with malicious intent to alter official records in a way that is not immediately perceptible. Fake documents, whether they're forged certificates, altered financial reports, or manipulated IDs threaten to undermine trust in institutions. These forgeries become risky when they're done with skill. For example, a stamp moved a bit, a signature swapped out just right or a tiny shift in alignment, these small alterations often slip past our eyes but can break up the entire forensic trail. With scanned versions of these documents, things get even more challenging. The usual forensic clues like metadata or hidden watermarks often vanish or get overwritten

[1] *Corresponding Authors: Dr. Richa Rohatgi & Dr. Seema R. Pathak

when scanned, which makes the old ways of spotting fakes much less useful. [1] [2] It’s within this context that we're looking at Sensor Pattern Noise (SPN). SPN is a type of noise that comes from tiny flaws in scanner or camera sensors. The important point to note is that this noise stays pretty much the same across scans from the same device, like a fingerprint. Although it is not visible directly, prior studies by Lukás et al., 2006 suggested that it can be used to trace the source of a scanned document and check whether all parts of a document really belong together. [3] It’s not easy to imitate either, which makes it a promising tool in detecting tampering when other clues are missing.

Our proposed framework takes advantage of this capability. It doesn't try to restore missing metadata or chase cryptographic signatures that may never exist in scanned or converted documents. Instead, it breaks down the scanned image into smaller units i.e. 64x64 pixel patches and checks for local inconsistencies in SPN, lighting, and textural alignment. Each patch is evaluated both in isolation and in relation to its neighbours. The idea is straightforward that if parts of a document were tampered with, they would likely betray their foreign origin in terms of scanner fingerprint, edge sharpness, or illumination profile. Our objective with this framework is that we lay out the rationale and architectural logic behind our framework. The methodology that follows is, for now, theoretical we've not yet implemented it fully. But the scaffolding is in place. And after the real world testing, this could be a significant step toward reliable digital forgery detection in scanned documents, a problem that is increasingly urgent but remains surprisingly under-addressed.

2. RELATED WORK

Forged digital documents are not a new problem, yet the sophistication of such forgeries has evolved faster than our ability to detect them. Over the years, several researchers have attempted to address this issue from several perspectives some with success, others with limitations that became more evident over time. To make sense of this growing body of work, it is useful to look at five main categories of approaches: (i) Metadata and File Structure Analysis, (ii) Copy-Move and Content Duplication Detection, (iii) Device-Based Forensics using Sensor Pattern Noise (SPN), (iv) Deep Learning-Based Methods, and (v) Hybrid or Multi-modal Systems.

2.1 Metadata and File-Based Analysis

Earlier detection approaches depended heavily on the examination of metadata and embedded structural information to identify the forged documents or images. Metadata often contains timestamps, author names, editing software info, or file signature hashes. These can, in theory, reveal unauthorized tampering. For example, Boonkrong (2024) designed a verification system using cryptographic hash functions for academic records, which reportedly achieved very high accuracy. [4] Similarly, Dubey et al. (2024) shared a case study where use of metadata and file signatures were done for detecting inconsistencies in scanned authorization certificates. [5]

However, while appealing for their simplicity and low computational cost, these methods fall apart in practical cases. Documents that are printed and scanned, which is a very common situation often lose metadata altogether, or worse, if they are regenerated by the scanning device, it will effectively wipe out the original forensic trail. In such scenarios, metadata-based systems are essentially blind. [1] Our model considers metadata only as a secondary cue; the real decision-making is based on content-derived features, which persist regardless of file format changes.

2.2 Copy-Move and Duplication Detection

Another early detection technique involves detecting repeated or moved regions within an image or a document i.e. the classic copy-move forgery. This method assumes that a region copied from one part of the document to another will have identical or nearly identical pixel structures. Fridrich et al. (2003) initiated this work using block-matching, and later improvements incorporated colour spaces by Luo et al., (2006) and dimensionality reduction using wavelets and Singular Value Decomposition (SVD) by Li et al., (2007). [6] [7] [8] While these models work well for image-based manipulation, they show high false positive rates when applied to textual documents. Scanned documents, especially those containing forms, templates, or repeated characters (for e.g., multiple zeroes, headers, logos), naturally include many similar-looking regions. Abramova and Böhme (2016) highlighted this problem and showed that even high-performing copy-move detectors often misclassify regular layout elements as tampered regions. [9] Our framework addresses this by shifting focus away from 'what looks similar' to 'what doesn’t belong', using scanner-level noise patterns and texture inconsistencies instead of pure pixel matching.

2.3 SPN and Device-Level Fingerprinting

Sensor Pattern Noise (SPN) detection is a more efficient and promising class of techniques. SPN is essentially a unique fingerprint that arises from small, fixed-pattern defects in a scanner’s imaging sensor. Lukáš et al. (2006) demonstrated that SPN can reliably identify which device captured a particular image, and later studies adapted this concept to document forensics. [3] Abdalla et al. (2018) combined SPN with Patch-Match techniques to detect forged regions in a variety of document formats. [10] Rabah (2022) applied SPN-based methods specifically to flatbed-scanned documents and emphasized the importance of scanner-specific noise like dust trails and light striping. [2] Despite their potential, SPN methods face challenges. SPN can degrade with lossy compression, uneven illumination, or repeated scanning. Dust removal, paper folds, or even updates in scanner firmware may slightly alter the fingerprint. That said, these issues are not insurmountable. Our model mitigates them by combining SPN with statistical consistency checks and spatial coherence analysis, using patch-wise comparisons rather than whole-image signatures. This patch-based approach allows us to localize tampering even when the global fingerprint is partially degraded.

2.4 Deep Learning-Based Approaches

The surge in deep learning has transformed many areas of computer vision, including document forgery detection. Early systems like those proposed by Sarode et al. (2020) used convolutional neural networks (CNNs) paired with Error Level Analysis (ELA) to detect the altered content. [11] But before that Shakir and Zwyer (2018) developed a three-stage pixel-based detection pipeline using neural nets trained on synthetic datasets. However, a major limitation of many early deep learning models is generalization. [12] As Zhao et al. (2021) illustrated, models trained on clean or simulated data often fail when tested against low-resolution, real-world scans. [13] Moreover, their work also showed that deep learning can be used not just for detection but also for generating convincing forgeries. This dual-use nature makes AI a double-edged sword.

Some more recent studies have focused on improving robustness and interpretability. James et al. (2020) proposed graph-based representations of characters to detect minute glyph-level manipulations. [14] Meanwhile, Umadevi and Rao (2021) targeted homogenous document regions using fixed-size windows to flag tampering. [15] Still, these models often require large labelled datasets and substantial computational power. In real forensic scenarios, those conditions are rarely met. Our approach avoids this by not depending on neural networks as a first line of detection. Instead, we use traditional image processing and statistical verification to pre-select suspicious patches, and reserve machine learning (i.e. the Random Forests) for confirmatory classification.

2.5 Hybrid and Multi-Modal Strategies

Recent studies consistently show that no single analytical tool, be it metadata probing, structural inspection, or signal-analysis can universally uncover forgeries spanning the wide array of document-tampering tactics now in play. Recognizing the limitations of individual methods, several researchers have proposed hybrid systems that combine two or more techniques. For example, Abdalla et al. (2018) crafted a multimodal detector that marries sensor-pattern-noise extraction, block-matching algorithms, and template-matching searches to pinpoint suspect edits in digital-scanned documents. By correlating noise-signature decay with spatial congruencies, their system isolates copied or reconfigured areas even when re-compression or dual-scan blur obscures the pixel fingerprints of the original mischief. [10] Such a hybrid design, although effective, still leans heavily on exhaustive archival fingerprints and calibrated threshold settings, which can compromise field deployment when operating with unknown or dynamic document populations. Similarly Sarode et al. (2020) proposed a web-based tool that unites neural network classifiers and classical error-level analysis (ELA), enabling detection of anomalies in pixel distribution and compression artifacts while also uncovering deeper spatial patterns in document layout. The system can flag inconsistencies that might suggest forgery. Yet its dependence on high-resolution scans and extensive training datasets limits its utility in low-resource settings or when working with older, degraded records that often exist in archival environments. [11]

More recent studies combine convolutional neural networks (CNNs) with carefully designed handcrafted features. Nath and Naskar (2023), for example, built a detection pipeline that couples CNN-based splicing detection method with checks for metadata consistency, allowing their system to pinpoint forged regions in digital images. By fusing content signs with broader contextual information, they enhanced the model's efficiency. Still, how well the pipeline handles low-fidelity scanned documents remains to be fully validated. [16] Prior to them Bayar and Stamm (2016) had advanced the field with a novel constrained convolutional neural network explicitly designed to tone down main image content and boost forensic details during the learning phase. Their approach underscores the value of steering deep architectures toward features that actually matter for detection, rather than letting them memorize irrelevant pixel noise that could mislead the analysis. [17]

When viewed collectively, these hybrid and multi-modal platforms depicts a promising path forward. By stacking different detection layers such as spectral noise pattern checks, frequency domain anomalies, layout analysis, and deep feature learners they yield models that are harder to fool. Yet the current designs either presume pristine data or demand high compute loads that limit portability. Our new framework addresses this limitation while preserving the modular spirit of hybrid methods. It starts with spectral noise pattern estimation as a core signal, layers in patch-coherence consistency checks, and calls on classifiers only when the signal-to-noise ratio justifies the computational cost. By doing so, the design harmonizes detection accuracy with on-the-ground deployability, a compromise too many prior methods overlooked.

After carefully reviewing the current body of research, it becomes evident that while the field has advanced significantly, each direction comes with its own set of blind spots. The range of techniques is wide, often innovative, but very few manage to hold up under the specific constraints of scanned document forensics where data is often degraded, context is minimal, and manipulation can be subtle. What struck us during this review was not just the creativity in these methods and techniques, but also how often they were developed in silos. There’s a tendency to optimize for ideal conditions, clean datasets, clear artifacts, or predictable forgeries while the real-world scenarios remain challenging and more unpredictable. Some of the techniques are technically elegant but fail when documents are poorly scanned or partially obscured. Others generalize well but offer little interpretability, making them hard to defend in forensic or legal contexts. Our work does not claim to be a universal solution. Rather, it emerged from the gaps we kept noticing like, the absence of mid-level features that link visual and statistical cues, the lack of modularity in many deep learning systems, and the need for forensic methods that can adapt on the fly without re-training from scratch. What we propose is not just a combination of past ideas, but a rethinking of how those components interact and can be re- designed to perform more efficiently, less for academic benchmarking and more for the real world forgeries that often forensic experts actually deal with.

The below table outlines how research in digital document forgery detection has evolved over the past two decades. It not only just lists methods but it also reflects the shifting focus of the field, from early experiments with block-matching and metadata checks to the more recent adoption of SPN-based models, deep learning and multi-modal strategies. For each study, we’ve noted the core technique, classification strategy, and its reported impact. Taken together, these studies show how researchers have responded sometimes reactively, sometimes innovatively to the evolving nature of document manipulation in real-world scenarios.

Table 1. Summary of key research contributions in digital document forgery detection (2001–2025)

S.No.	Author	Method	Classifier	Result
1	Deringas (2001) [18]	Presented a case study where he studied detection of forgery in digitally altered documents through analysis of background and seal impressions.	Reconstructed Background & Seal Impressions	Successfully detected alterations in digitally altered documents.
2	Fridrich Jessica et al. (2003) [19]	Developed a block-based algorithm that uses accurate and strong matching methods to find copy-paste forgeries.	Exact Match & Robust Match	The robust matching method proved effective for retouched regions.
3	Luo et al. (2006) [7]	Proposed an overlapped block-based algorithm using colour features from RGB and YCbCr	Overlapped Block-Based Algorithm	Effective for small regions, but human intervention needed for

		channels for forgery detection.		large smooth regions.
4	Li Guohui et al. (2007) [8]	Used wavelet transform and SVD for dimensionality reduction and sorted neighbourhood approach for finding duplicated regions.	DWT (Discrete Wavelet Transform) and SVD (Singular Value Decomposition)	Reduced computational complexity but failed on geometrical transformations.
5	Mahdian Babak & Saic Stanislav (2007) [20]	Developed overlapping block-based detection with blur invariants and k-d tree for similarity analysis.	Overlapping Block-Based Detection	Detected copy-move forgery but had long computational time (approx.40 min).
6	Erkilinc et al. (2011) [21]	Proposed a page layout segmentation technique with pre-processing, text/photo detection, and edge detection.	Page Layout Segmentation	Achieved 85% accuracy in detecting text and photo segmentation.
7	Saini & Kaur (2015) [22]	Researched forensic analysis of system-generated documents altered by image-processing technologies.	Image Manipulation Characteristics	Identified manipulation using image processing applications like Adobe Photoshop.
8	Sameria et al. (2015) [23]	Investigated alterations in offline scanned documents and digital images using MATLAB software.	Offline Scanned Document Alteration	Highlighted the need for more research on alterations in offline scanned documents.
9	Abramova & Bohme (2016) [9]	Analysed block-based near-duplicate detection in scanned text documents, considering similar-looking glyphs.	Near-Duplicate Detection	Moderate success in detecting copy-move fraud in scanned text.
10	Abdalla et al. (2018) [10]	Developed a fusion strategy based on Patch-Match and sensor pattern noise to detect the copy-move forgery.	Patch-Match Enhanced Fusion	Showed high efficacy in both passive and active forgery detection.
11	Shakir & Zywer (2018) [12]	Proposed a pixel-based forgery detection technique for scanned documents which they tested with a MATLAB-based system.	Pixel-Based Detection	Successful forgery detection through a three-stage process using a dataset.
12	Sarode et al. (2020) [11]	Presented a web application design for document manipulation detection using Neural Networks and Error Level Analysis.	Error Level Analysis	Web-based solution detected manipulated documents effectively.
13	James et al. (2020) [14]	Used graph comparison for forgery detection by extracting character features from business documents.	Graph Comparison Strategy	Successfully detected forgery in real business document datasets.

14	Darem et al. (2020) [24]	Suggested methods for detecting forgeries in JPEG-compressed domains using DCT coefficients and template matching.	DCT Coefficients & Template Matching	Detected and located forgery regions in JPEG images accurately.
15	Umadevi & Rao (2021) [15]	Developed a method for detecting altered documents using uniform document regions and fixed window size.	Homogenous Document Region Analysis	Effective only for documents with uniform regions; limited in scope.
16	Lin Zhao et al. (2021) [13]	Created a low-cost document forgery algorithm using deep learning to alter text in identity documents.	Low-Cost Forgery Algorithm	Successfully manipulated text in identity documents using a deep learning-based approach.
17	Rabah C. B. (2022) [2]	Worked on scanned document integrity and authenticity verification using flatbed scanners and image forensics.	Scanned Document Forensics	Developed a secured environment for trusted document transactions and falsification prevention.
18	Dubey et al. (2024) [5]	Explored challenges in detecting forgery in digitally manipulated documents using file signature, metadata, and hash value comparison.	File Signature & Metadata Comparison	Successfully identified forged authorization certificates using advanced forensic techniques.
19	Boonkrong (2024) [4]	Proposed a cryptographic hash function to verify authenticity of digital academic documents with 100% accuracy.	Cryptographic Hash Function	Achieved 100% accuracy with fast verification speed, outperforming other methods like CNN and blockchain.
20	Li Li, Yu Bai, Shanqing Zhang, Mahmoud Emam (2025) [26]	Proposed a multi-category tamper detection algorithm using spatial-frequency features and multi-scale feature fusion with attention-based multi-supervision and a multi-category detection head.	Multi-category Detection Head, Attention-based Multi-Supervision, HRNet-based Multi-scale Feature Extraction	Extensive experiments show improved F1 score by 5.73%, accurate localization of tampered regions, and correct identification of multiple forgery types.

3. METHODOLOGY

This section lays out the thinking behind the framework we've developed for detecting forgeries in digital documents specially focusing on documents that were scanned using mobile phone cameras. Rather than treating the problem as a black-box classification task, the goal here has been to ground each component in observable, forensic evidence i.e. the things that have come up repeatedly across past studies and casework. From what we've seen, forgery in scanned documents doesn't generally show up as one significant anomaly. Instead, it is a mixture of minor irregularities like metadata that doesn't entirely make sense, noise patterns that don't quite match with the device's usual output, or subtle misalignments in layout that are easy to miss at first glance but feel off when looked at closely. These inconsistencies minute but cumulative are what shaped the design of our method. It's important to clarify that this is still a theoretical model. We are in progress of its deployment or field-testing, but the framework has been shaped with a very practical problem in mind i.e. how to detect tampering in documents that were never captured under ideal conditions. Scans might be skewed, compressed, converted multiple times, or even taken via mobile phones with inadequate light. In such conditions, sophisticated algorithms alone are not enough and you will need a methodology that is modular, transparent, and grounded in how these artifacts behave in the real world.

3.1 Metadata Screening

The first thing we look at is metadata, while the metadata alone cannot confirm the authenticity, it can still offer a useful entry point. This is patterns or anomalies (like software signatures inconsistent with the file type) often serve as early red flag. They do not confirm the tampering, but they can help you to dig deeper. For this step, we extract whatever is available like the timestamps, author names, editing software, file version histories, and embedded tags. As others have also pointed out, like Nguyen et al. (2020), sometimes the metadata is inconsistent and sometimes, it's completely missing. [1] And it can also be said that even the absence can be informative because clean or untouched documents usually leave some kind of trace. A document stripped bare of any metadata may be trying to hide where and how it was created.

1.2 Pre-processing of Scanned Documents

To avoid skewed results due to poor scan quality or lighting variations, we apply a series of pre-processing steps:

- Convert the document images to grayscale, using ITU-R BT.601 weighting. This is to avoid unnecessary distraction from colour information.
- Apply a bilateral filter, this step reduces unwanted visual noise while preserving the edges, which is required for identifying the tampered signatures or stamps (Zhang & Ma, 2011). [25]
- Resize the document to a width of 1024 pixels while maintaining the aspect ratio. This makes the next steps consistent across the document.

1.3 Patch Based Document Segmentation

As a next step we will slice the pre-processed document image into smaller patches of 64x64 pixels. This size was chosen because it strikes a balance and also captures the small tampered zones like a few letters or a signature, without losing any layout context. Each patch is treated independently in the following analysis steps.

1.4 Feature Extraction in Each Patch

For each patch we will extract and analyse on four set of features:

- **SPN Residuals**: Here we compute Sensor Pattern Noise by removing predictable signal via denoising. What's left behind reflects the device's unique fingerprint (Lukás et al., 2006). [3]

- **Edge Sharpness**: Using the Sobel filters, we will check for odd edges. If a patch was copied and pasted, the edges often flag the abnormal edge transitions (Fridrich et al., 2003). [3]
- **Lighting Irregularities**: Uneven or variations in light distribution are another important clue to detect the forgery. Here we analyse the patch's illumination distribution which is computed using the local histogram equalization.
- **Frequency Defects**: We use Fast Fourier Transform (FFT) filters to detect scan-line artifacts or specks that don't belong.

1.5 Comparison with Scanner Fingerprint Dataset

To check whether a particular patch genuinely belongs to the same device that captured the rest of the document/image, we compare its extracted features mainly the SPN trace and some texture cues with a reference database of known scanner fingerprints. For the comparison itself, we use standard similarity measures like cosine distance and Euclidean distance. If a patch shows that it is a mismatch it means its characteristics are too far from what the scanner's fingerprint would typically produce and this raises a red flag and is tagged as potentially manipulated. This step isn't about absolute proof but about spotting inconsistencies that merit closer inspection.

1.6 Cross-Patch Consistency Check & Coherence Analysis

A document may not exhibit overt signs of forgery in certain sections, yet the overall impression might still seem alarming. To address this we will apply:

- **Block Transition Analysis:** To check for abrupt discontinuities in SPN or illumination features across adjacent patches.
- **Neighbour Similarity Tests:** Patches that are statistically out-of-sync with surrounding patches are given higher tampering scores.
- **Coherence Deviation Estimation**: To validate the alignment and structural flow across patch borders, disruptions in geometric or spectral continuity suggests non-authentic alterations.

This generates a heatmap showing areas with high tampering probability. These local scores are added up to calculate an overall forgery likelihood.

1.7 Classification Thresholds and Decision Rule

We can confirm that it is a forged document if more than 10% of the patches are flagged as inconsistent (value adjustable), the document is classified as suspicious. This threshold was chosen based on pilot analysis, but will be re-tuned during real-world validation.

1.8 Future Implementation Plans

As of now, we are in the simulation phase and this model hasn't been implemented yet. The plan is to:

- Build a test dataset of known forged and original scanned documents.
- Try this system on real hardware: flatbed scanners, MFPs, and mobile phones.
- Compare our performance against existing models like Error Level Analysis (ELA) and CNN-based detection.
- Use standard metrics like precision, recall, and F1-score to report how we're doing.

Because of its modular structure, we can later swap out individual steps. If a better SPN estimator comes along, we can plug it in. Or if a different patch size works better on mobile images, we can adjust.

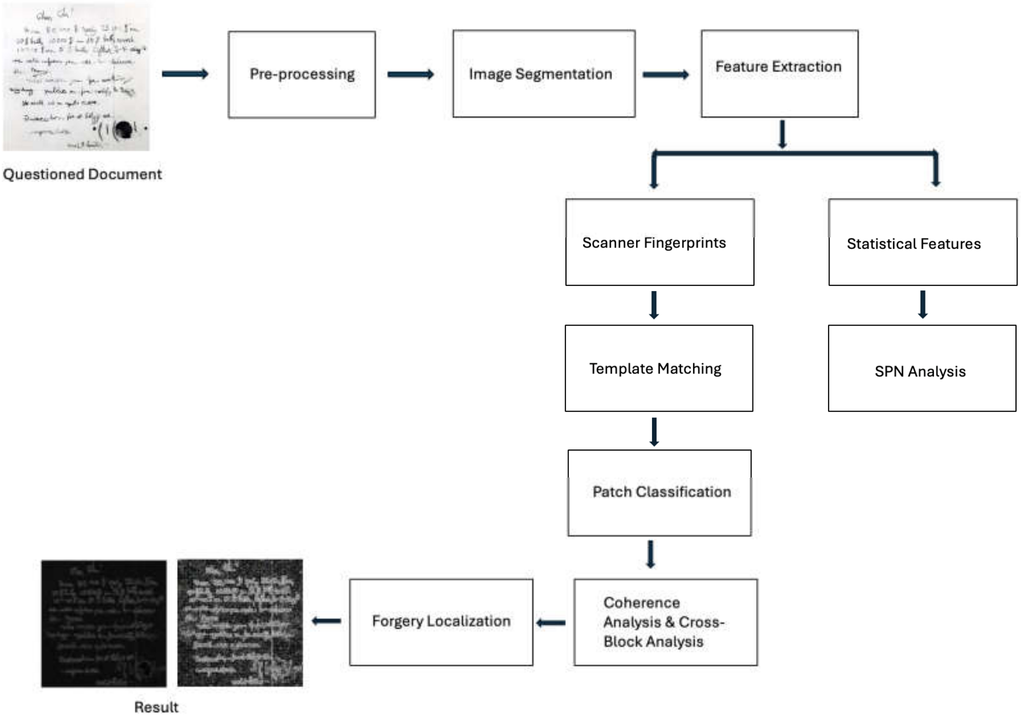

Figure 1. Proposed workflow for digital document forgery detection using SPN and statistical features

The model begins by cleaning and segmenting the document image, then extracts two types of features device-specific and statistical. Using SPN analysis and template matching, each patch is evaluated. Finally, coherence checks across regions help flag suspicious areas, allowing precise localization of possible forgeries.

4. CHALLENGES IN DETECTING FORGED DIGITAL DOCUMENTS

Our approach gives a structured method to identify forgeries, however there are still several challenges that make both detection and verification more complicated.

4.1 The Forge-and-Recapture Documents

One of the hardest problems is identifying forgeries in documents that have been printed, tampered with physically, and then re-scanned. These “forged-and-recaptured” documents often show no metadata inconsistencies, and even visual signs of tampering can be very minute or lost. The absence of original digital markers makes these cases particularly resistant to conventional forensic approaches.

4.2 Tracing the Document’s Source

To identify which device (scanner, mobile phone) created or produced the digital document is vital in establishing chain of custody. Unfortunately, the rapid turnover in hardware and software makes this task increasingly difficult. For example, a forged certificate scanned at a cybercafe may never be traceable, especially if the device is not registered or its fingerprint is not in the database.

4.3 The Ever-Changing Toolkit of Forgers

As we've seen, the forgery detection technologies have come a long way. Today's forgers are not just masters in Photoshop; they have also learned new and better ways to make forgeries. With AI-based tools now available for free, one can fabricate entire documents in minutes. The sheer diversity of manipulation techniques from font tampering to structural layout reconstruction makes it impractical to rely on a single detection technique.

4.4 Lack of Benchmarks and Standardization

Based on the literature reviewed and while searching for the appropriate samples/ datasets we observed that the field currently suffers from a lack of standardized datasets, making it hard to compare the efficacy of different detection systems. There is also inconsistency in reporting accuracy. Some studies report only on curated lab datasets with perfect scans, which is far from representative of the real-world conditions.

5. RESULTS AND DISCUSSION

The proposed framework has been designed with a specific objective including, to identify document forgeries that are too minute to be detected by conventional techniques relying solely on metadata, superficial visual inspection, or static layout analysis. While a comprehensive quantitative evaluation is forthcoming, the following scenario-based illustration allows us to demonstrate the operational strengths and internal logic of this model.

Let us take the example of a scanned certificate being examined for signs of tampering. As the document is broken down into uniformly sized patches, the framework begins its assessment by scanning for irregularities at a microstructural level. In the early stages of this process, certain patches begin to stand out not because of any glaring defect, but due to subtle deviations in how edges appear or how light seems to distribute across the area. These are often signs that something is slightly off. To probe further, the framework compares the features extracted from these questionable patches such as SPN traces and texture statistics against a reference dataset comprising known scanner fingerprints. In this comparison, the deviation of these patches from expected scanner-specific norms is statistically significant, suggesting they may not have originated from the same device as the rest of the document. Further to this when the framework evaluates the spatial consistency across neighbouring patches, it notices a disruption specifically in the lower part of the page where adjacent blocks lose their typical alignment in terms of SPN patterns and frequency-domain signals. This kind of break is not something easily visible to the human eye. But when the document is rendered into a visual anomaly map, this region appears clearly demarcated, revealing a deviation that challenges the structural coherence of the original scan. Our model assigns a tampering probability score of 0.81 on a normalized 0-1 scale, exceeding the defined threshold for flagging suspicious activity. The result is not presented in isolation, the system outputs a report comprising a forensic summary of the detected anomalies along with annotated visual overlays for manual review by a trained examiner. This synthesis of statistical evidence and spatial analysis enhances the interpretability of the model's verdict. What sets this framework different from previous approaches is not a single breakthrough, but the way it combines multiple forensic elements such as sensor-based noise analysis, inter-block consistency checks, and the statistical anomaly detection into one unified system. Many existing techniques rely heavily on metadata examination and or assume a fixed document structure, which often doesn’t hold true in real-world conditions. This model, by contrast, is built to operate even when such cues are missing or corrupted, making it more adaptable to low-quality or inconsistently scanned inputs.

While the empirical validation and real-world testing are part of the next phase of this research, the present framework represents a considered step toward practical utility. Each stage of our workflow has been designed to be interpretable and modifiable and not hidden behind black-box logic, but structured so that its inner workings can be examined and improved. This emphasis on transparency and modularity is intentional, allowing future researchers or forensic practitioners to improve or build on it as needed. Its modular architecture has been built with flexibility in mind, so that improvements say, in SPN extraction techniques or classification strategies can be integrated down the line without needing to overhaul the entire system. To summarize, this effort represents a conscious attempt to translate core principles from forensic theory into a practical approach that can withstand the unpredictable conditions of real-world scanned or altered documents.

6. CONCLUSION

The framework proposed in this study addresses a real and growing problem in digital document forensics identifying tampered scanned documents in the absence of obvious visual or metadata signs. By focusing on Sensor Pattern Noise (SPN), patch-level consistency, and cross-region coherence, the model offers a layered, modular approach to forgery detection.

While the current version remains a blueprint, future implementation and validation could yield a system that is both interpretable and scalable. As the arms race between forgery and detection accelerates, tools that combine device fingerprinting with structural and statistical analysis may be our best defence.

Acknowledgments

The authors would like to express their sincere gratitude to Amity University for providing the resources and support necessary for the completion of this research.

REFERENCES

[1] V. T. Nguyen, H. T. Do, and H. X. Nguyen, “Detecting forged metadata in scanned documents using machine learning,” Proc. IEEE Int. Conf. Digit. Forensics Cyber Crime, 2020, pp. 55–62.

[2] C. B. Rabah, “Analysis of scanned documents for integrity and authenticity checking,” HAL Open Science, 2022; HAL Id: tel-03516239.

[3] J. Lukáš, J. Fridrich, and M. Goljan, “Digital camera identification from sensor pattern noise,” IEEE Trans. Inf. Forensics Secur., vol. 1, no. 2, 2006, pp. 205–214.

[4] S. Boonkrong, “Design of an academic document forgery detection system,” Int. J. Inf. Technol., 2024. https://doi.org/10.1007/s41870-024-02006-6

[5] D. Dubey, R. Rohatgi, and S. R. Pathak, “Unveiling digital document manipulation: A case study in forensic examination,” Indian J. Forensic Med. Toxicol., vol. 18, no. 2, 2024, pp. 46–52. https://doi.org/10.37506/w909cd74

[6] J. Fridrich, D. Soukal, and J. Lukáš, “Detection of copy-move forgery in digital images,” Proc. Digital Forensic Research Workshop, 2003.

[7] W. Luo, J. Huang, and G. Qiu, “Robust detection of region-duplication forgery in digital image,” Proc. 18th Int. Conf. Pattern Recognit. (ICPR), vol. 4, 2006, pp. 746–749.

[8] G. Li, Q. Wu, D. Tu, and S. Sun, “A sorted neighbourhood approach for detecting duplicated regions in image forgeries based on DWT and SVD,” Proc. IEEE Int. Conf. Multimedia Expo (ICME), 2007, pp. 1750–1753. https://doi.org/10.1109/ICME.2007.4285025

[9] S. Abramova and R. Böhme, “Detecting copy–move forgeries in scanned text documents,” Electron. Imaging, vol. 2016, no. 8, 2016, pp. 1–9.

[10] Y. Abdalla, M. T. Iqbal, and M. Shehata, “Fusion approaches system of copy-move forgery detection,” Am. J. Comput. Sci. Eng. Surv., vol. 6, no. 1, 2018, pp. 1–12.

[11] S. Sarode, U. Khandare, S. Jadhav, A. Jannu, V. Kamble, and D. Patil, “Document manipulation detection and authenticity verification using machine learning and blockchain,” Int. Res. J. Eng. Technol., vol. 7, no. 5, 2020.

[12] S. H. Shakir and N. Zwyer, “Forgery detection based image processing techniques,” Int. J. Sci. Eng. Res., vol. 9, no. 11, 2018.

[13] L. Zhao, C. Chen, and J. Huang, “Deep learning-based forgery attack on document images,” IEEE Trans. Image Process., vol. 30, 2021, pp. 7964–7979. https://doi.org/10.1109/TIP.2021.3112048

[14] H. James, O. Gupta, and D. Raviv, “OCR graph features for manipulation detection in documents,” arXiv:2009.05158v2 [cs.CV], 2020.

[15] M. Umadevi and R. C. Rao, “Identification of tampered document based on homogeneous regions of image: a forensic perspective,” Int. Res. J. Eng. Technol., vol. 8, no. 11, 2021.

[16] S. Nath and R. Naskar, “Automated image splicing detection using deep CNN and metadata consistency,” Multimed. Tools Appl., vol. 82, 2023, pp. 33691–33720. https://doi.org/10.1007/s11042-023-16229-w

[17] B. Bayar and M. C. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” Proc. 4th ACM Workshop Inf. Hiding Multimedia Secur. (IH&MMSec), 2016, pp. 5–10.

[18] A. Deringas, “Traces of forgery in digitally manipulated documents,” Probl. Forensic Sci., vol. 46, 2001, pp. 375–382.

[19] J. Fridrich, D. Soukal, and J. Lukas, “Detection of copy-move forgery in digital images,” Proc. Digital Forensic Research Workshop, 2003.

[20] B. Mahdian and S. Saic, “Detection of copy–move forgery using a method based on blur moment invariants,” Forensic Sci. Int., vol. 171, no. 2–3, 2007, pp. 180–189.

[21] S. Erkilinc, M. Jaber, E. Saber, P. Bauer, and D. Depalov, “Analysis and classification of complex scanned documents,” Int. Soc. Opt. Photonics, 2011. doi:10.1117/2.1201107.003819

[22] K. Saini and S. Kaur, “Forensic examination of computer-manipulated documents using image processing techniques,” Egypt J. Forensic Sci., vol. 6, 2015, pp. 317–322.

[23] S. Sameria, V. Saran, and A. K. Gupta, “Analysis of offline scanned document and digital images for alteration through digital image processing,” Int. J. Soc. Relev. Concern, vol. 3, no. 8, 2015, pp. 25–27.

[24] A. Darem, A. A. Alhashmi, M. Javed, and A. B. AbuBaker, “Digital forgery detection of official document images in compressed domain,” Int. J. Comput. Sci. Netw. Secur., vol. 20, no. 12, 2020, pp. 115–121.

[25] Y. Zhang and J. Ma, “Bilateral filtering: theory and applications,” IEEE Signal Process. Mag., vol. 28, no. 1, 2011, pp. 61–72.

[26] L. Li, Y. Bai, S. Zhang, and M. Emam, “Document forgery detection based on spatial-frequency and multi-scale feature network,” J. Vis. Commun. Image Represent., 2025.