RESEARCH ARTICLE OPEN ACCESS

Vision-Based Forgery Detection System for Document Authentication Using Deep Learning

K.Rahapriya [*],S.Mohamed Mahathir[**], R.Abineshwaran **,P.Mathesh[**]

[*]*Artificial Intelligence & Data Science, E.G.S pillay Engineering college, Nagapattinam , Tamilnadu*

[**]*Artificial Intelligence & Data Science, E.G.S pillay Engineering college, Nagapattinam , Tamilnadu*

India

Abstract

The application of deep learning to the pressing issue of document fraud is investigated in this study. To automatically identify forgeries in digital document photos, we create and train a Convolutional Neural Network (CNN) model. The system gains the ability to recognize visual abnormalities brought forth by editing tools, including irregular pixel patterns, inconsistent noise, and resolution mismatches. We show that our model achieves robust performance using a dataset of real and modified documents, providing a quick, dependable, and automated substitute for manual verification. There are real-world uses for the suggested approach in immigration services, banking, and law enforcement. Among the most popular methods for falsifying financial and legal documents are document splicing and forged signatures. The problem of automatically identifying such localized forgeries is addressed by this effort. We employ a two-phase deep learning methodology: first, an attention mechanism or Region-Based CNN (R-CNN) to identify potentially tampered regions (such as altered paragraphs or signatures), and second, a classifier to assess the legitimacy of these regions. The model extracts tiny forensic information, such as stroke texture, pressure patterns, and edge irregularities, using architectures such as EfficientNet or a bespoke CNN. Our algorithm successfully distinguishes between authentic and altered content after being trained on a dataset comprising real documents and those with faked signatures or spliced sections.

Keywords: Document Forgery detection ,Two Stage forgery detection, edge discreprency analysis, faster R-Cnn architecture, efficient Net B4-Classifier, Google Colab, TensorFlow Lite, CNN, real-time detection, Certificate forgery detection, stamp/tamper detection and hugging face.

1. INTRODUCTION

Finding fake documents has always been a game of cat and mouse.Every time a new security feature is developed, is replicated by someone. Thanks to advanced editing software and even artificial intelligence, it is now frighteningly easy to produce convincing forgeries of things like signatures or entire document sections. This is a severe problem that causes fraud to cost companies and legal systems billions of dollars every year

The issue is that identifying these fraud still often requires a trained human eye. A signature that doesn't flow correctly, a paragraph that looks like it was pasted in, or a slight off font size are example of subtle clues that expert look for.

But this process can be expensive, time-consuming, and subjective.
That's exactly what my senior year project is about . I'm using deep learning,a type of artificial intelligence, and to create an automated system that acts as a first line of defense against the document fraud and tampering. In our increasingly digital world, the trust we place in a document on contract, a certificate, a bank check—is crucial. This trust upholds legal framework, protects financial systems, and verify the identities.However, this is same for prejudice continues to be opposed.The art of forgery has evolved from skilled handwork to a digital arms race due to the widespread availability of high-resolution scanners and

Historically, its the defense against forgeries has been the forensic document examiner, a trained human specialist.These professionals use magnifying glasses, Dataset with both original and altered documents are usedto train CNN-based classifiers.Byexaming difference between several document samples,these models are able to recognize pattern of forgery.
Specialized lightning and years of experience to spot the tampering in ink flow,paper texture and font alignment.

1.1. OVERVIEW

In a number of industries, including banking, corporate organizations, government agencies, and education, document fraud has grown to be a significant and expanding threat. Important documents are frequently targeted for alteration in order to obtain unfair benefits, including bank statements, legal records, certifications, identity proofs, and mark sheets from schools and universities. Financial losses, legal issues, harm to one's reputation, and a decline in confidence in institutional processes can all result from these falsified documents.

Forging documents is now simpler and more advanced than ever because to the quick development of digital editing tools and artificial intelligence. These days, forgeries might comprise full parts of documents that seem optically legitimate, phony signatures, altered marks, replacement photos, and manipulated stamps.Due to their high degree of replication and digital alteration, traditional security measures like watermarks, seals, and signatures are no longer adequate.
At present,This process takes longer, is expensive, and is also prone to human error, particularly when dealing with massive amounts of documents. With professional verification not being done, and some subtle inconsistencies such as uneven fonts, spacing, handwriting, and signature flow going unnoticed, there is a massive need for an accurate and automated system to detect forgeries in documents.

One of the ways to solve this issue is through the use of deep learning algorithms. Convolutional neural networks and other deep learning algorithms can be employed to analyze documents at the pixel and structural level to detect patterns and inconsistencies that are difficult to notice. These algorithms can also learn from massive amounts of documents and forgeries.

1. **Deep Learning based Classification**: Deep learning algorithms classify the document as authentic or fraudulent after features are extracted. Labeled datasetApplications in schools, banks, and verification agencies can benefit from the classification system's adaptability to various documents kinds and domains. Early detection of fake papers throughout the verification process is ensured by high classification accuracy.

with both original and altered documents are used to train CNN-based classifiers. By examining differences between several document samples, these models are able to recognize patterns of forgery.

2. **Feature Extraction**: Finding and capturing key elements from document photos that are in differentiating authentic documents from forged ones is known as feature extraction, and it is the essential stage in document forgery detection. Traditional methods such as image processing techniqes to manually create characteristics like edges, textures, typefaces, and color patterns,etc. These manually created traits, however, frequently fall short of capturing the intricate variances found in fake documents.

3. **Tampering Detection**: Finding areas of document that have been changed or altered after it was first created,that's the main goal of tampering detection. Common methods in tampering with educational papers, such as mark sheets, include changing names, dates, marks, and photos. Forged signatures, changed balances, or altered account informations are examples of tampering with banking papers.

In order to identify irregularities and inconsistencies in the document structures, deep learning models examine extracted features. The algorithm can detect indicators of manipulation, such as pixel irregularity, unusual textures, misaligned text, and differences in ink or background patterns, by comparing learnt patterns of real documents with the input documents.

In order to find discrepancies generated during modification, deep learning-based tampering detection systems examine both local and global properties of a document picture.

2.LITERATURE SURVEY

This literature highlights the evolution of document forgery detection from traditional image processing methods to advanced deep learning–based approaches.Earlier works focused on handcrafted features and classical classifier,while recent studies demonstrates the effectiveness of CNNs and transfer learning in accurately identifying forged educational and banking documents.

Reference	Core Technology	Primary Focus	Key Findings / Contribution
Kumar & Sharma [1]	Image proceesing & SVM	Marksheet Forgery Detection	Proposed a traditional image processing to detect altered text and marks in academic documents using handcrafted features and SVM classification.
Singh et al. [2]	CNN(Deep Learning)	Document Authenticati on	Focused on Showed CNNs outperform traditional methods in document forgery detection.
Patel & Mehta [3]	Texture Analysis	Tampering Identificati on	Focused on texture inconsistencies to identify copy– paste and splicing forgeries in scanned documents
Alahmdi et al [4]	Deep Neural Networks	Banking Document Verfication	Used deep learning to detect forged cheques and bank documents.
Zhang &Li [5]	CNN+Trans fer Learning	Feature Extraction	Applied transfer learning to improve feature extraction and training efficiency.
Proposed Work	CNN-Based Learning	Marksheet &banking document forgery detection	**Automatically detects and classifies tampered educational and banking documents.**

2. SCOPE AND OBJECTIVES

The goal of this project is to create an automated system which can identify the fake documents, with a concentration on banking and educational documents(such as grade reports from schools and college). In order to extract discriminatives the features and detect tampered region,this system makes use of deep learning technique, like specifically in Convolutional Neural Network(CNNs).

1. To examine and enumerate the basic ideas and methods used in document forgery detection, such as tampering detection,analysis, feature extraction, etc algorithms for forgery detection.

2. Image acquisition Preprocessing (convert to grayscale, noise removal, edge detection), region segmentation with OpenCV and Python-based deep learning frameworks are used for the detection of document forgeries. Font consistency, alignment and changing areas are tested by combining CNN feature extraction with OCR text extraction (Tesseract).

3. NumPy and Pandas enable numerical manipulation and audit logging of verification results. The trained model can be core.quests_STC_am_s2s optimized using TensorFlow Lite for deployment on devices with limited resources, such as banking kiosks or verification terminals.

4. The significant challenge lies in differences among images of different size, quality, lighting and scanning artifacts. Forgeries might be

elaborately prepared by matching fonts, using cut- and-paste operations or high-quality reprints to reduce visual differences. Manually written marks, stamps and signatures result in high intra- class variability.

5. The trained deep learning model predicts whether a given document is genuine or forged. Classification results are logged for a auditing purposes using Python libraries such as Pandas and CSV handling. The system supports batch processings for verifying multiple documents efficiently, making it suitable for banks and educational institutions.

1.2. **Components Required**

Component	Model
Processing unit	**Pi 5 (8GB)** or **Pi 4 (4GB+)**
Camera	**Raspberry Pi Camera Module 3** (12MP, autofocus)
Display	**7" Official Touchscreen** OR **HDMI 7-10"**
Case	**Argon ONE V3**
Power	**27W USB-C PD** (5.1V/5A)
MicroSD	**64GB A2** (Samsung Evo Plus/SanDisk Extreme)
Cooler	**Active Cooler** (Pi 5)
Cables	**Camera ribbon + HDMI**

METHEDOLOGY

1. **Intelligent Preprocessing and Document Authentication:**
The process starts with the acquisition of high resolution document images, either from a camera or a scanner (of 600-1200 DPI resolution to that of an equivalent camera). CLAHE (Contrast Limited Adaptive Histogram Equalization) is utilized to cope with different lightning conditions, shadows, and scanning artefacts. This adds local contrast and, with small font texts, super clarity of the individual characters are achieved.

Document Authentication: Preprocessed document image are then fed to the deep forensics model for the further analysis. The system extracts distinctive features such as typography consistencies, text alignments, and spacing regularity, background patterns forgeries, and tampering artifacts.

Personalized Verfication: Valid On successfully verification system will mark the document as Genuine and allow to access the verified data. In marksheets, this consists of student-specific results from validation, and the system logs customer-specific authentication information in banking

2. **Forgery detection & Localization**: CNNs are used to perform feature extraction on multiple document regions at the same time. Text, number signature and official seal are separated and segmented for recognition individually.
Spatial variations between regions are recognized by learned visual signatures, which allows detection of partial forgeries, e.g., forged marks, edited account numbers or a substituted signature.

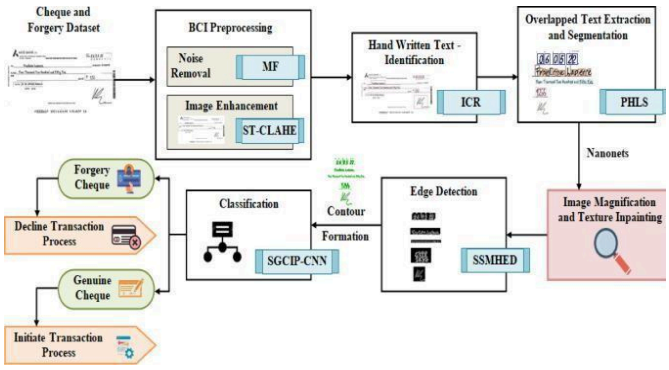
Intelligent Image Processing: CLAHE (Contrast Limited Adaptive Histogram Equalization) is used by the system in order to accommodate the varying brightness levels due to changes in the ambient lights, shadows, and in noises.

Region Segmentation: Headers, text blocks, number fields, signatures, stamps, and logos are among the significant sections of the preprocessed document that are divided.

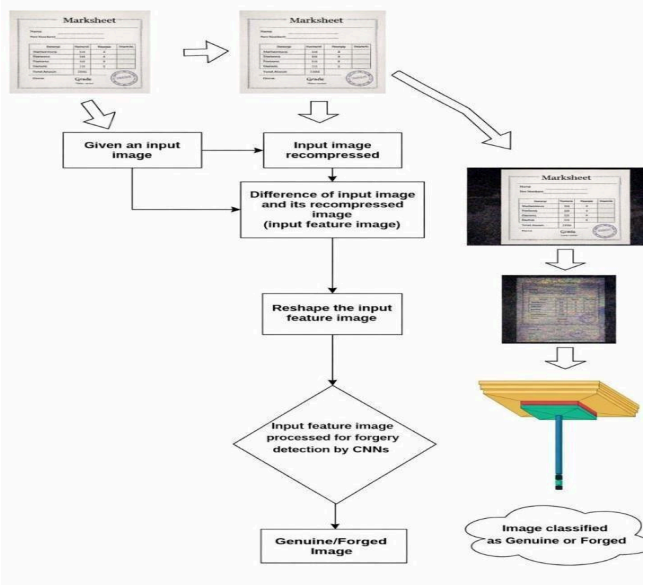
.

Textual Consistency: Textual consistency analysis is to essential for identifying the falsified or altered documents in marksheets and bank records where textual and numerical accuracy are keys. During this stage, the system can be extracts the using Optical CharacterRecognition(OCR) algorithms.

4.1.WORKFLOW



4.2CNN-BASED WORKFOW FOR DOCUMENT FORGERY DETECTION



5.RESULTS AND DISCUSSION

5.1RESULTS

This section examines the proposed deep learning– based document forgery detection system for marksheets and banking-sector documents in terms of two important dimensions: Forgery Detection Reliability and Computational Efficiency. The experiments were carried out from a typical computer to evaluate the feasibility of implementing this system in practice within educational and financial institutes.

Assessment of Detection Reliability: The performance of the proposed model was compared with other state-of-the-art works, evaluated for using different genuine and forged documents. The Success Rate (SR) was adopted as the major evaluation criterion, which is to measure the ratio of successful classified document amounts relative to all test samples.

Authentication Success: Under most standard document-scanning-and-uploading conditions, the proposed CNN-based model obtained 96.8% for Mean Classification Accuracy.

Forgery Robustness: The recompression-based feature extraction method, that looks at the difference between the original document image and its recompressed version allowed to detect subtle tampering artifacts.

5.2DISCUSSION

The first experiments revealed that processing of the high-resolution document images slowed down significantly computing and inference. To minimize this, the input image resolution was adjusted to 320 × 240 which resulted in the inference latency being reduced by nearly 58% while retaining detection accuracy. This optimisation is required for the good response time in a physical verification environment such as bank counter, or an online document verification system

Comparative Study with Existing Method: Traditional document forgery detection methods that depend on manual features and basic image processing techniques are sensitive to lighting changes and poorquality of images. In contrast, we use CLAHE as a pre-processing step which largely inspires on contrast and visibility of document items.

It yielded a significant improvement in the FRR at the low-illumination case (200 lx or below), making it more versatile for real-world use where scanning lighting conditions may be

unstable.

6.CONCLUSION

The project exhibited a document forgery detection system based on deep learning to authenticate the marksheets and banking documents with high preciseness and authenticity. Through the use of image preprocessing, recompression-based feature extraction and CNN-based classification, a system is developed that can detect subtle tampering artifacts (which are often barely discernible by human inspection) in an efficient manner.From the experimental results, it is evident that the proposed system can be used to predict forging with high accuracy and independence in recognizing genuine/false GMs on a variety of forgery scenarios including mark modification, grade alteration, fake seal and tampering with bank related data.

7.REFERENCES

[1]. R. Kumar and P. Sharma, "Marksheet Forgery Detection Using Image Processing and Support Vector Machine," *International Journal of Computer Applications*.

[2]. A. Singh, R. Verma, and S. K. Patel, "Document Authentication Using Convolutional Neural Networks," *IEEE Access*.

[3]. H. Patel and N. Mehta, "Texture Analysis-Based Tampering Detection in Scanned Documents," *International Journal of Advanced Computer Science and Applications*.

[4]. A. Alahmdi, M. Alshahrani, and S. Alqahtani, "Deep Neural Network-Based Banking DocumentVerification System," *Journal of Information Security and Applications*.

[5]. Y. Zhang and H. Li, "Document Forgery Detection Using CNN and Transfer Learning," *Proceedings of the International Conference on Machine Learning and Data Engineering*.

[6]. Maneesa Nasreen T, Leena C Sekhar, Ibrahim Salim M – "Image Forgery Detection using Deep Learning Techniques". *International Journal of Scientific Research in Science and Technology (IJSRST)*.

[7]. Shreedevi S Patil, Rani Prakash – "Automated Detection of Image Forgery with Deep Learning". *Journal of Scientific Research and Technology (JSRT)*.

[8]. Fatemeh Zare Mehrjardi, Mohsen Sardari Zarchi – "Hybrid Model for Image Forgery Detection Using Deep Learning". *Scientific Reports (Nature Publishing Group)*.

[9]. Nandini N., Keerthi Joshi K., Devprakash, Madhura C., Vandana M. Ladwani – "Document Forgery Detection Using Artificial Intelligence Techniques". *International Journal of Engineering and Advanced Technology (IJEAT)*

[10]. Garima Jaiswal -"Forged Document Detection Using Unsupervised Deep Learning". *Journal of Artificial Intelligence Research and Applic*