National College of Ireland

ScreenGuard: AI-Powered Image Authenticity Detection for Financial Transaction Fraud Prevention

MSc Research Project
MSc in Cyber Security (MSCCYB1_JAN25)

Satish Reddy Navuluri
Student ID: X23436913

School of Computing
National College of Ireland

Supervisor: Raza Ul Mustafa

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Satish Reddy Navuluri
Student ID:	X23436913
Programme:	Master of Science In Cyber Security 2025-2026 **Year:**
Module:	MSc (Research) Practicum/Internship Part 2 Information
Supervisor:	Raza Ul Mustafa
Submission Due Date:	28-01-2026
Project Title:	ScreenGuard: AI-Powered Image Authenticity Detection for Financial Transaction Fraud Prevention
Word Count:	7206 **Page Count** 25

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	N.Satish Reddy
Date:	28-01-2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

ScreenGuard: AI-Powered Image Authenticity Detection for Financial Transaction Fraud Prevention

Satish Reddy Navuluri
X23436913

Abstract

The rising use of online payment platforms has increased the risk associated with fraud using altered screenshots of finance transactions. This has made detecting tampering with a high degree of accuracy extremely important. The paper presents a multimodal deep-learning model that combines the characteristics of RGB images, Error Level Analysis, edge maps, metadata cues and optical character recognition derived results to identify images of forged financial transactions. As the methodology, a labelled dataset was built, modality-specific feature extractors were created, and a multimodal convolutional architecture was built and trained on benchmarked against existing transfer-learning baselines such as EfficientNetV2B0, ResNet50V2 and MobileNetV3Small. The suggested model demonstrated the best validation with an AUC of 1.0, and better accuracy, and the confusion-matrix analysis showed significantly better balance between false positives and false negatives than with image-only models. However, there are still weaknesses in terms of the size of the datasets, OCR sensitivity to domain-specific structure, and capability to withstand increasingly complex adversarial manipulations. This work demonstrates the feasibility of multimodal image-forensic fusion for tampering detection using a publicly available document-tampering dataset. While the experimental evaluation is conducted on generic document images rather than proprietary financial transaction screenshots, the findings should be interpreted as a proof-of-concept indicating potential applicability to fintech fraud-screening workflows. Validation on real financial transaction screenshots remains an important direction for future research.

Keywords: Multimodal image forensics, Financial fraud detection, OCR-based tampering analysis, Error Level Analysis, Deep learning authentication.

1 Introduction

1.1 Background and Context

The fast growth of online payment systems and mobile banking programs and wallet systems have radically transformed the ways people and organisations make financial transactions. Through this expansion, there has been a rise of fraud that takes advantage of the flaws in the verification of digital evidence such as the use of altered screenshots and images of transaction confirmation. With image editing software that is available to most people, fraudsters are able to alter the time or recipient names or transaction values to produce realistic yet fake evidence of payment that can be hard to separate authentic artefacts with (Gupta, 2024).

Such image verification in most banks and financial technology organisations still mostly depends on manual inspection. The human eyes need to do a visual comparison of the screenshots with the backend transaction records, which is not only time consuming, but also expensive in addition to being subject to fatigue effects error particularly when the volumes are high and the different applications may have a different layout. The issue with automated detection is that financial screenshots differ in the resolution, font, language, user interface theme and compression options. These variations both influence the visual representation and the pixel statistics below, so it is not sufficient to have simple heuristic rules. Proper image authenticity detection should then act rationally in the combination of low-level pixel patterns, compression-related forensic artefacts, textual, and simple metadata to pinpoint low-level indicators of interference (Shan, 2025).

1.2 Motivation

According to reports by industries, there is always massive loss of money due to digital payment fraud, fraudulent scams where falsified screenshots are employed to acquiring goods or services without actually settling the money. Nonetheless, the detection systems that have been implemented have concentrated on the transactional and behavioural indicators instead of the forensic examination of the submitted images. Automated tools used to detect tampering of financial screen shots are still scarce, and available commercial solutions are designed to identify forgery in documents in general and not focusing on payment confirmation peculiarities (Ajax, Luz, and Bamigboye, 2025).

Studies in image forensics have yielded effective image tampering detection techniques (including error level analysis and noise-based analysis and deep convolutional models trained on benchmark datasets of forgery), which are generic, i.e. applicable regardless of the forgery technique employed. Nevertheless, such methods are not often considered in terms of financial transaction visualizations and often overlook the rich textual and metadata content of such artefacts. An obvious gap then exists to a framework of genuine detecting authenticity based on transaction images, which binds visual, written and metadata characteristics into one model (Sholademi, 2025).

1.3 State of the Art

The current state-of-the-art in image forensics has shifted towards the application of deep learning and multimodal solutions because traditional single-modal solutions are trivial in the face of sophisticated forgeries such as those created by GANs and recompression forgery. The current solutions fall short in their assessment of financial screenshots, where there are semantic texts and platform-dependent layouts involved. It also disregards the semantic validation process conducted by OCR tools. There is a need for a tool known as the ScreenGuard that carries various attributes in a multimodal solution architecture.

1.4 Research Aim

This research aims to develop and evaluate a multimodal deep learning framework that integrates visual, forensic, textual and metadata features to reliably detect tampered financial transaction images, addressing limitations in existing single-modality forensic methods used in fintech fraud prevention.

1.5 Research Question

This project addresses the following central question:

RQ: *Can a multimodal AI model combining RGB information, forensic features derived from error level analysis and edge maps, basic image metadata and text extracted using optical character recognition improve the accuracy of detecting forged financial transaction images?*

1.6 Research Objectives

To address this question, the project pursues five objectives:

1. Investigate on the state of the art in the field of digital image forensics and financial fraud detection.
2. Develop a multimodal deep learning system to detect authenticity of transaction images.
3. Implement feature extraction pipelines for RGB preprocessing, error level analysis, edge detection, metadata extraction and optical character recognition.
4. Training and testing of the multimodal model and its performance as compared to image only transfer learning baselines.
5. Evaluate model behaviour based on validation metrics, receiver operating characteristic curves, precision recall curves and confusion matrices.

1.7 Contributions

This research makes the key contributions:

- To the best of the author's knowledge, this work presents an applied multimodal deep-learning framework that integrates RGB image features, forensic artefacts, OCR-derived text and metadata for document-level tampering detection, framed toward financial transaction imagery.
- Demonstrated that multimodal fusion outperforms single-modality and transfer-learning baselines in discriminatory performance.
- Addressed a clear research gap by focusing specifically on financial transaction image authentication, a domain rarely explored in existing forensics literature.
- Delivered ScreenGuard, a functional prototype demonstrating the feasibility of multimodal document-tampering detection within a fintech-oriented use-case scenario.

1.8 Limitations and Assumptions

The research is limited by a fairly small sample of about 404 pictures that can be a limitation to generalisability of the results. The accuracy of optical character recognition is also dependent on font, image and language quality which may introduce noise to the textual feature. Financial institutions did not provide proprietary datasets of transaction images and as such, experimentation used tampering datasets, as opposed to real customer datasets. The method further assumes that manipulation results in anomalies in representation of error levels in the analysis that can be detected, a fact that could not be true given a very advanced forgery. The validation subset consisted of 81 images, which is insufficient for statistically conclusive

evaluation; therefore, reported metrics should be interpreted as preliminary indicators rather than definitive performance estimates.

1.9 Report Structure

This report is organised into seven chapters.

- **Introduction** presents the research background, problem motivation and aims.
- **Related Work** reviews prior research on image forensics, OCR analysis and fraud detection.
- **Methodology** explains the dataset, feature extraction and experimental design.
- **Design Specification** outlines the multimodal system architecture and requirements.
- **Implementation** details the model construction and software components.
- **Evaluation** presents and analyses the experimental results.
- **Conclusion & Future Work** summarises findings, discusses implications and identifies directions for further research.

2 Related Work

This review of the literature explores technical techniques of tampering an image and the use of deep learning, optical character recognition and metadata analysis in forensic detection. The review highlights the relevance of the existing methodology to the particular issue of financial transactions screenshots, and critically evaluates the strengths, weaknesses and empirical assertions of the studies that have been chosen. Based on the institutional guidelines on literature critique, the focus is put on the methodology, datasets, evaluation measures and generalisability.

2.1 Digital Image Tampering Techniques

Regarding image forensics, researchers have identified splicing, copy move, region inpainting and synthetic generation as the main types of tampering. Splicing is used to join parts of different images, copy move to duplicate parts of the same image to cover or duplicate an object, inpainting to fill-in deleted parts of an image with plausible texture, and GAN based methods to create new realistic images. Abdulqader et al. (2023) report the most frequently used tampering modes in the real-world forensic scenario to be copy move and splicing, which is also in line with the forensic surveys that highlight the presence of simple manipulations that consumer editing tools allow. Abdulqader et al. describe the methods of detection in a descriptive manner, but the article lacks the empirical analysis of the methods and the quantitative results on standard criteria, making the validity of their statements more questionable.

According to Umamaheswari and Karthikeyan (2022), feature extraction like Local Binary Pattern or SURF are still stalwart elements of numerous tamper localisation systems. They suggest improved versions ELBP and ESURF and allege that ESURF produces better robustness and speed. The article presents qualitative improvement but lacks the cross-dataset analysis and statistical significance testing in some cases. The argument that ESURF is more or equally effective than alternatives, in turn, must be viewed with skepticism until it is proven to be valid on larger, more heterogenous data using more standard metrics.

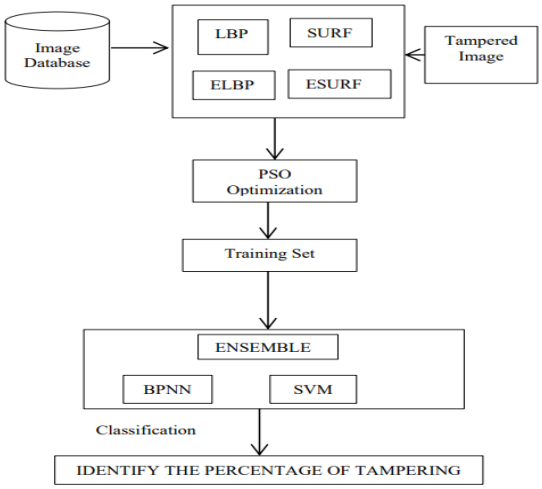

Figure 1: Proposed Workflow Model

[Source: *Umamaheswari and Karthikeyan (2022)*]

Fatima and Banu point out in the framework of advanced forgery the development of deepfakes and GAN generated images that generate content and not simply reassemble existing pixels. They observe that generation of this kind compromises the traditional copy move detectors since synthetic work can be self-consistent in terms of noise and compression. The authors are correct in their identification of the necessity of deep learning models that can recognize less apparent traces of generative behavior, but the review is mostly abstract and lacks experimentative comparisons that can measure the rate of detection with respect to the recent generative models.

According to Chakraborty et al. (2024), a mixture of conventional forensic representations like the error level analysis and learned deep features can provide a good result. An accuracy of 98.55 percent on CASIA dataset is reported using a dual branch network with 16 epochs. Such high precision is indicative of the potential of hybrid systems but also is concerning. Being popular, CASIA is not as diverse as it is in the wild transaction screenshots. Therefore, the 98.55 percent number can be exaggerated expected performance on heterogeneous financial imagery. Moreover, research on the potential overfitting of the paper should be evaluated, and the presence of train test splits and the cross validation of the evaluation should be examined. Concisely, the empirical finding is favorable but not conclusive to the transfer of domains.

Principled detection of manipulation is based on prior work that uses edge maps of the error level analysis and noise residuals. ELA also increases compression artifacts and may render recompression errors as obsolete. Edge analysis and noise residuals represent discontinuities that were brought about by the copy paste operation and synthetic filling. These artifact based techniques however tend to fail when forgeries are reprocessed identically or when generative techniques have a predictable noise behavior. Artifact analysis, therefore, is not sufficient to have effective detection.

2.2 Deep Learning for Image Forensics

According to Castillo Camacho and Wang (2021), convolutional neural networks are the new paradigm of learned forensic features since they produce discriminative representations on raw

pixels by default. Their overall review of the literature sums up architectures and training regimes and points out that CNN based detectors tend to perform better on benchmark datasets than hand crafted features. The review is useful as it contains a taxonomy as well as highlights one significant weakness: CNN models pre-trained on one artifact type can be poorly generalised to unknown tampering modes, or to images of other acquisition pipelines.

Ekhande et al. (2022) also show that CNN methods have proven to be efficient in most forensic activities. According to their review, CNNs are more flexible but they demand large quantities of labeled data and regularisation. Another point of view is the article by Ferreira et al. (2021), who prove that the relatively simple feature extraction and the use of classical classifiers can be very competitive. Ferreira et al. showed that a DFT based feature extraction and an SVM obtained a photo level cross-validation F1 of 99.53 percent. This finding implies that the domain suitable handcrafted features are useful in particular in situations where computational constraints or interpretability are of importance. However, Ferreira et al. also demonstrate that CNNs can be more successful than the classical methods in case sufficient annotated data can be provided.

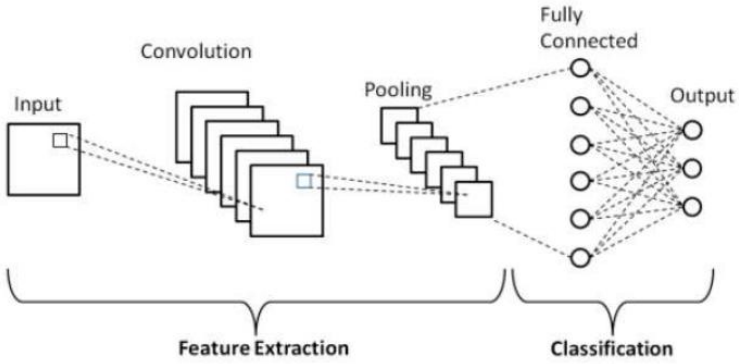

Figure 2: CNN Architecture

[Source: *Ekhande et al. (2022)*]

One of the most important constraints in the entire literature of deep learning forsensics is the propensity to train models using cleaned academic data sets that fail to reflect the diversity of real-world financial screenshots. Different operators' interfaces, fonts renders, language and compression options all vary in financial confirmations, and there are few published datasets that reflect the diversity of financial confirmations. Therefore, high-accuracy claims must be put into context by provenance and diversity of datasets.

2.3 OCR Driven Forensics

Li (2023) says that textual data mining has been found effective at identifying fraud in financial stories, whereby there are semantic hints suggesting risk. Applying this observation to image forensics, OCR-based scene text extraction allows to verify textual consistency against backend records, as well as expected semantic relationships. Li et al. (2025) show that semantic based conflict detection in bilingual text scene can be accurate with 90.03 percent and a F1 of 88.5 percent in their BSTID dataset. This numerical data indicates that semantic consistency is an effective clue to indicate that secondary edits to scene text images. These techniques,

though, rely on the quality of OCR output; in low quality or stylised screenshots OCR errors cause noise which can significantly impair the performance.

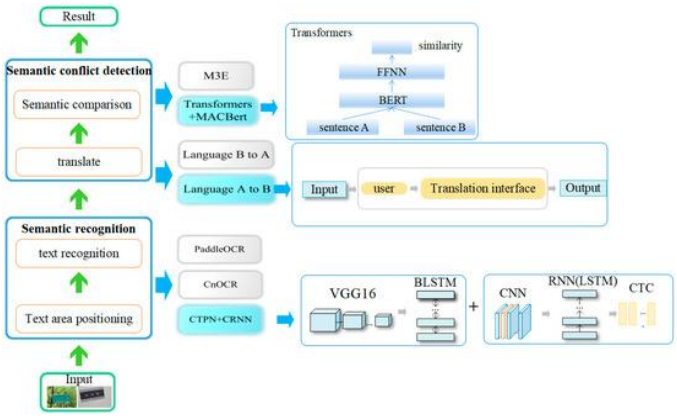

Figure 3: Structure of Semantic-based Text Image Tampering Detection Model

[Source: *Li et al. (2025)*]

2.4 Metadata Based Detection

According to Abraham and Joseph (2023), metadata anomalies like an odd value of DPI, unrealistic file sizes or lack of EXIF fields are informative forensic indicators. Metadata analysis is light and can scan images in a short time to determine whether they are suspicious, and metadata is easily spoofed or removed. Thus, metadata capabilities are useful as a supporting modality and not a decision maker.

2.5 Financial Transaction Fraud Detection Non-Image

Machine learning tools used in transaction anomaly detection are established. According to Maneel et al. (2025), random forest classifiers and rule systems powered with engineered features of transactions perform well in identifying anomalous patterns to identify unusual amounts or frequencies. These are transaction centric systems that are based on structured logs and they do not cover authenticity of images. This separation shows that there is a gap in practical operations whereby the submitted images which is a different channel of evidence, is rarely incorporated into anomaly pipelines.

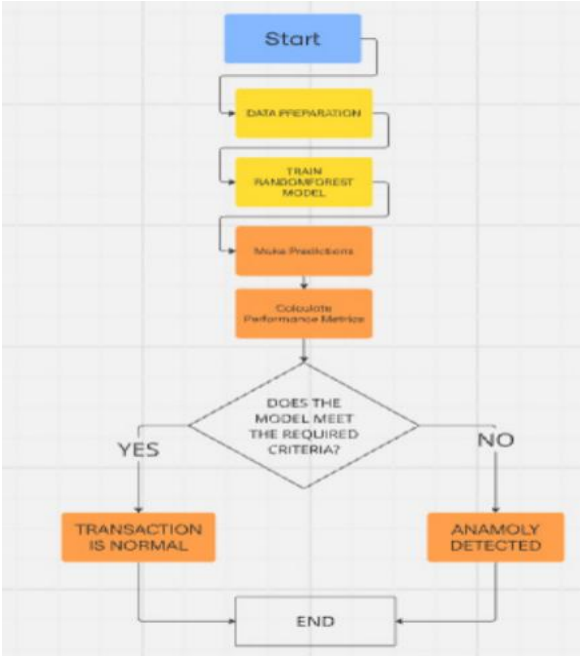

Figure 4: Workflow Diagram

[Source: *Maneel et al. (2025)*]

2.6 Critical Review of State of the Art

The literature shows that single modality solutions are evidently constrained. Artifact based detectors are vulnerable to state-of-the-art generation methods and after filtering pipelines that homogenise compression. CNNs offer powerful learned features and are greatly dependent on the dataset and susceptible to domain shift. The analysis of simple texts or metadatas can be interpreted yet destroyed by adversarial manipulation. The studies with high numeric accuracy can be based on limited benchmarking, e.g. CASIA or curated datasets; thus reported values, e.g. 98.55 percent accuracy or 99.53 percent F1, have to be viewed in the context of data composition and test conditions. There is no consistent provision of cross validation, open test splits and reporting of false positive versus false negative rates, so comparability of results is decreased.

2.7 Gap Identification and Justification for ScreenGuard

From literature review, multimodal detection frameworks addressing pixel level forensics tasks using learned visual representations & OCR-derived textual semantics & metadata are clearly called for. Multimodal approach can address failure modes for each other, while textual semantics can indicate semantic discrepancies that ELA and noise residuals fail to identify, ELA & noise residual can unveil traces of recompression that text analysis fails to unveil. Thus, ScreenGuard directly addresses the literature gap identified above and is justified for several theoretical & empirical constraints identified in prior literature.

2.8 Conclusion

The current literature indicates a marked improvement in the area of image forensics in classical approaches, deep learning, and OCR-based methods, but certain important gaps are

noticed in the literature surveyed. The existing techniques mostly work in a controlled environment and hardly leverage the image and metadata together and the semantic and metadata information together, while the banking financial screenshots contain vulnerabilities in this area. The semantic and metadata components are less commonly noticed, and nothing has been performed in the area of banking financial screenshots identification.

Table 1: Literature Review Summary Table

Author(s)	Aim of Study	Methodology	Results	Strengths & Limitations	Notes
Abdulqader et al. (2023)	To review common digital image tampering types and detection approaches.	Descriptive survey of copy–move, splicing, and forensic techniques.	Identifies dominant tampering modes; no quantitative benchmarking.	Strength: Clear taxonomy of tampering. Limitation: No empirical evaluation to verify claims.	Useful conceptual overview; lacks reproducibility metrics.
Umamaheswari & Karthikeyan (2022)	To enhance LBP and SURF features for tamper localisation.	Propose ELBP and ESURF feature variants; qualitative comparisons.	Report qualitative robustness and speed improvements.	Strength: Focus on feature-level optimisation. Limitation: No cross-dataset testing or statistical validation.	Promising but requires stronger empirical support.
Fatima & Banu	To discuss challenges posed by GAN-generated forgery.	Conceptual review of deepfake-era forgeries.	Highlight limitations of classical copy–move detectors.	Strength: Identifies emerging threats. Limitation: No experimental benchmarks.	Highlights need for deep-learning-based approaches.
Chakraborty et al. (2024)	Hybrid learning combining ELA and CNN features.	Dual-branch CNN trained on CASIA for 16 epochs.	Achieved 98.55% accuracy on CASIA.	Strength: Demonstrates hybrid feature effectiveness. Limitation: Dataset lacks diversity; potential overfitting.	Results may not generalise to financial screenshots.
Castillo Camacho & Wang (2021)	To survey CNN-based approaches in forensic detection.	Systematic review of CNN architectures and training strategies.	Conclude CNNs outperform handcrafted features.	Strength: Comprehensive taxonomy. Limitation: Highlights generalisation issues.	Reinforces need for robust multimodal models.
Ekhande et al. (2022)	Evaluate effectiveness of CNNs in image forensics.	Review of deep-learning applications and performance trends.	Report strong CNN performance across tasks.	Strength: Broad evidence base. Limitation: Reliance on curated datasets.	Supports CNN adoption but warns of data dependency.
Ferreira et al. (2021)	Compare handcrafted and deep features in forensics.	DFT-based features + SVM vs CNN architectures.	F1 score up to 99.53% (handcrafted).	Strength: Shows handcrafted features still competitive. Limitation: Evaluated via cross-validation only.	Highlights value of hybrid and multimodal approaches.
Li (2023)	Explore semantic inconsistencies for fraud detection.	Text mining on financial narratives; OCR-driven scene text analysis.	Demonstrates semantic cues predict fraudulent behaviour.	Strength: Strong semantic modelling. Limitation: OCR errors degrade performance.	Motivates OCR integration in image forensics.
Li et al. (2025)	Detect semantic conflicts in bilingual scene text.	OCR + semantic conflict classifier; evaluated on BSTID dataset.	90.03% accuracy, F1 = 88.5%.	Strength: Strong evidence for text-level forensics. Limitation: Depends on OCR accuracy.	Relevant for financial screenshots with text edits.
Abraham & Joseph (2023)	Use metadata irregularities for tampering detection.	Analysis of DPI, filesize and EXIF irregularities.	Metadata anomalies correlate with manipulation likelihood.	Strength: Lightweight and fast. Limitation: Metadata is easily removed or spoofed.	Effective supporting modality, not standalone.
Maneel et al. (2025)	Detect financial anomalies (non-image).	Random forest + rule-based transaction models.	High anomaly-detection accuracy.	Strength: Strong structured-data performance. Limitation: Does not address image fraud.	Highlights operational gap in integrating image forensics.

3 Research Methodology

3.1 Methodological Overview

The proposed study is quantitative based on the scientific method with experimental design as every step is outlined in adequate detail such that it allows other researchers to recreate the research. The sequence of work in this project is built in the form of a pipeline: the collection of data and labelling of original and forged document images used as a proxy for transaction-related artefacts, the preprocessing and cleaning of files, multimodal features, training and hyperparameter optimization of the model, and lastly performance assessment using standard classification metrics. The design of the whole model is to make sure that the model behaviour can be measured, comparable with deep learning baselines and reproducible with fixed random seeds and described preprocessing operations.

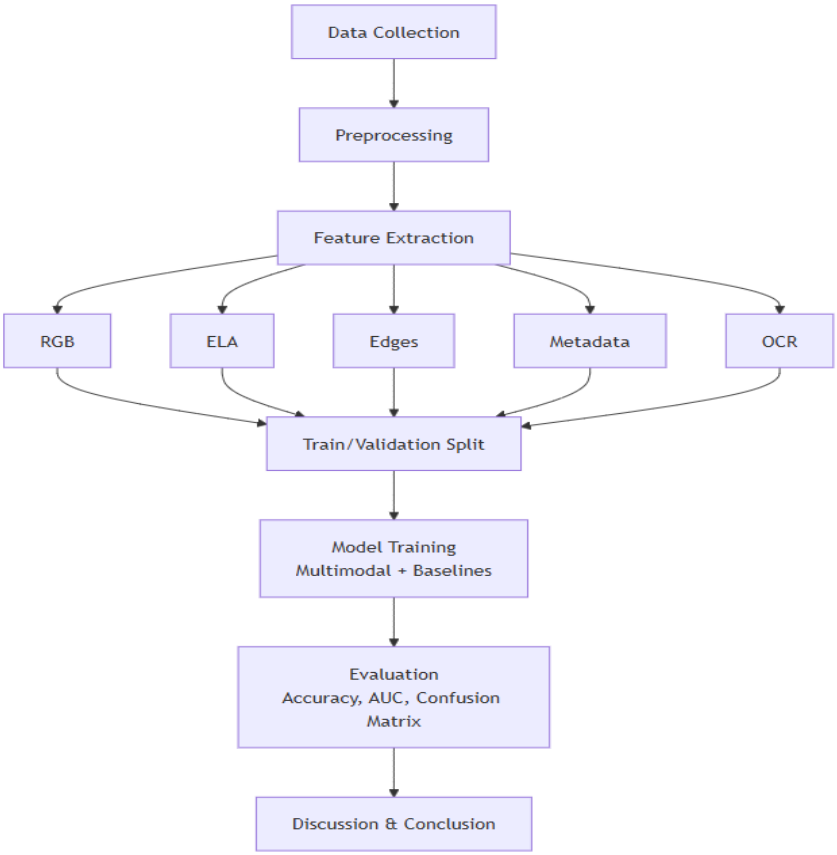

Figure 5: Research Methodology

3.2 Dataset Preparation

The dataset comprises 404 document images sourced from the publicly available AWS Document Tampering Detection repository, consisting of 202 authentic and 202 tampered samples. Although these images are not proprietary financial transaction screenshots, they provide a suitable benchmark for evaluating document-level image tampering detection techniques that are conceptually relevant to transaction-screenshot fraud. This dataset is sourced from the publicly available AWS Document Tampering Detection repository

(https://github.com/aws-samples/document-tampering-detection), which provides 202 authentic and 202 tampered documents for supervised learning. There is also a small predict folder in the repository that provides a number of samples paystub images to be used in inference demonstration, though they are not utilized in training or validation. Such structure enables easy labelling and makes the expansion of the structure in future easier without the need of any architectural modification. It accepts only files with standard image extensions (JPEG, PNG, BMP, TIFF) and this prevents the occurrence of inconsistencies in file formats on loading.

During preprocessing, every image is loaded with a standard library of imaging, and then it is converted into three channel RGB image and resized to a standard spatial resolution of 224 by 224 pixels. This standardises the input dimension over devices and applications and it can be used with the transfer learning backbones that are used to compare. Simple cleaning is done through the deletion of unreadable or corrupt files. The arrays of aligned file paths and labels are stored that are subsequently used to assemble the multimodal tensors. The indices of the entire samples are randomized with an initial random number 42 in order to prevent bias in ordering. Using these shuffled indices, the dataset is further divided into a training portion (80 percent of samples) and a validation portion (20 percent). While random shuffling with a fixed seed improves reproducibility, the AWS dataset may contain near-duplicate authentic–tampered document pairs. Consequently, random splitting cannot fully guarantee independence between training and validation sets, and limited content similarity leakage may be present.

3.3 Feature Extraction Methods

The feature extraction algorithm aims at representation of complementary indicators of tampering: visual appearance, forensic artefacts, structural edges, low level metadata and text. In the case of RGB image branch, each resized image is represented as an array of unsigned integer and normalised in the model through rescaling layer, which rescales the intensity of each pixel to the [0, 1] range. Besides, an on the fly augmentation pipeline uses small random rotations, translations, zooms and contrast perturbations. These experiments are able to simulate the variety of screenshots taken on various devices and viewing conditions and decrease the chances of overfitting on the limited sample.

Error Level Analysis (ELA) is a forensic transform that is employed to point out recompression anomalies that may occur locally and tend to be caused in a particular area by an editing and resaving process. On every image, a temporary JPEG representation is stored at a predetermined quality level, reopened and the difference between the pixel values in the original and the recompressed representation is determined. ELA image is scaled with the maximum difference value to enable the subtle artefacts to be displayed and the result is normalised to [0, 1] interval. This generates a three channel ELA image in line with RGB content.

Gradient based edge maps are extracted on a grey scale form of the image through Sobel filters on the horizontal and vertical direction in order to complement ELA. The size of the gradient is calculated and normalised then stretched out into a unit channel map. This edge map is appended to the three channel ELA image along the channel dimension creating a four-channel forensic tensor, which highlights compression artefacts as well as structural discontinuities caused by splicing or copy move actions.

Each image file is derived to consume simple metadata properties to capture global properties that can vary between authentic and fake screenshots. The pipeline measures width and height, aspect ratio, file size in kilobytes, average and standard deviation of the greyscale intensity, and, where metadata contains the information, the dots per inch as to their horizontal and vertical position. These eight values are saved as a fixed length numeric vector of each sample and then the values are standardised in the model.

Textual data is obtained through Tesseract by using Optical Character Recognition. The images are turned to greyscale and upscaled to facilitate recognition then forwarded to the OCR engine. The raw text output is purified by substituting the line breaks with spaces and eliminating space in front and after. Where no text is found, a sentinel token will be used instead, to maintain length of sequence, and to make sure that each image generates an OCR string which is vectorisable.

3.4 OCR Vectorisation

The OCR strings are projected to sparse representation of high dimension with the TextVectorization layer of the TensorFlow in TF IDF mode. The word vocabulary is also learned through training set texts without leakage through the validation set. Unigrams and bigrams are employed in such a way that they may have individual tokens (bank names or currency codes) as well as short phrases (payment successful or UPI reference) that may have a discriminative power. A vocabulary size limit of twenty thousand tokens will be used, which balances expressiveness and the threat of the sparsity of a dataset of 404 images. The resulting TF IDF vectors give a fixed length textual feature space which may be added to visual and metadata branches (Pawar and Gajpal, 2021).

3.5 Dataset Pipeline Construction

The multimodal arrays are enclosed in TensorFlow tf.data pipelines in order to facilitate efficient training. In every partition, we generate a dataset which produces a dictionary of four components; the RGB image tensor, the forensic tensor, the metadata vector, and the raw OCR string, along with the binary label. The training data is randomised on a buffer of the sample size in order to maximise randomisation and batched with a batch size of sixteen andprefetched with AUTOTuner to overlap computing and data loading. The validation dataset is batched and pre-fetched and not shuffled to maintain a deterministic order. This type of pipeline structure minimizes input bottlenecks of the GPU and makes sure that batches have a wide variety of genuine and counterfeit samples.

3.6 Model Training Method

The basic multimodal model is then applied in Keras in the form of a four-branch system. The RGB branch uses a shallow convolutional neural network using batch normalisation, max pooling, global average pooling and dropout regularisation. The four channel ELA plus edge tensor are run through the forensic branch using a different convolutional stack and global pooling. The OCR layer takes input as TF IDF vectors of the text vectorisation layer and processes them by a dense layer with a rectified linear activation and dropout. The metadata branch normalises the eight numerical features and sends them through a dense and dropout layer. The products of the branches are then fused into one, after which two fully connected layers are used, each containing rectified linear activation and dropout, with the final layer

being a sigmoid neuron, which produces the likelihood of an image being forged (Haider and Mustafa, 2025).

The binary cross entropy loss function is used in training and optimised with Adam optimiser with a learning rate of $1e^{-4}$. The accuracy and the area under the ROC curve are two measures of performance. Early stopping is used using validation AUC with patience of five epochs and the optimal weights on the validation set are restored automatically. An epoch limit of twenty is set and a batch size of sixteen to allow reproducibility of the optimisation trajectory is fixed in both NumPy and TensorFlow.

3.7 Baseline Models

Three images only transfer learning baselines are trained using the same RGB inputs: EfficientNetV2B0, ResNet50V2 and MobileNetV3Small, all initialised with ImageNet pretrained weights and run with their regular preprocessing functions. The convolutional backbone is frozen in both instances, then a global average pooling layer, dropout, dense layer, rectified linear activation, additional dropout and a single sigmoid output neuron are used. Such baselines separate the value of high capacity convolutional feature extractors when applied to pixel data only. The comparison of their performance with the suggested multimodal architecture will also enable the study to gauge whether the combination of RGB, forensic transforms, metadata, and OCR features is reasonable in identifying forged financial transaction screenshots.

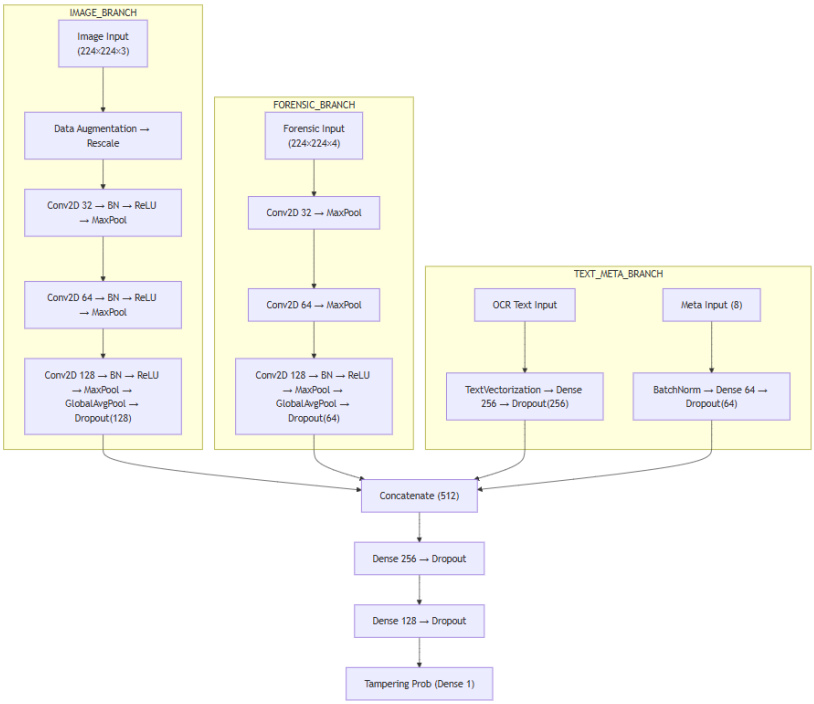

Figure 6: Multimodal Architecture

4 Design Specification

4.1 System Architecture

The ScreenGuard system is envisioned as a multimodal, modular detection pipeline which combines complementary sources of information towards measuring the authenticity of financial transaction images. The high-level architecture consists of four parallel feature streams which are fed into a central fusion and decision module. Image CNN branch is used to process RGB pixel data to represent appearance, layout, and texture information. The Forensic CNN branch takes as input a four-channel forensic tensor consisting of three channel Error Level Analysis image and one channel gradient magnitude (edge) map to indicate recompression artifacts and structural discontinuities. The OCR Text Dense branch takes the output of the optical character recognition branch and transforms it into a dense semantic representation using TF IDF and then dense layers. The Metadata Dense branch takes in an input of a constant length of a numeric array containing width, height, aspect ratio, file size, mean intensity, standard deviation and DPI values. The four-branch outputs are concatenated in a Fusion Layer after which one successively adds fully connected layers with dropout and batch normalisation, and finally a single neuron with a sigmoid activation, which yields the probability of forgery (Pawar and Gajpal, 2021).

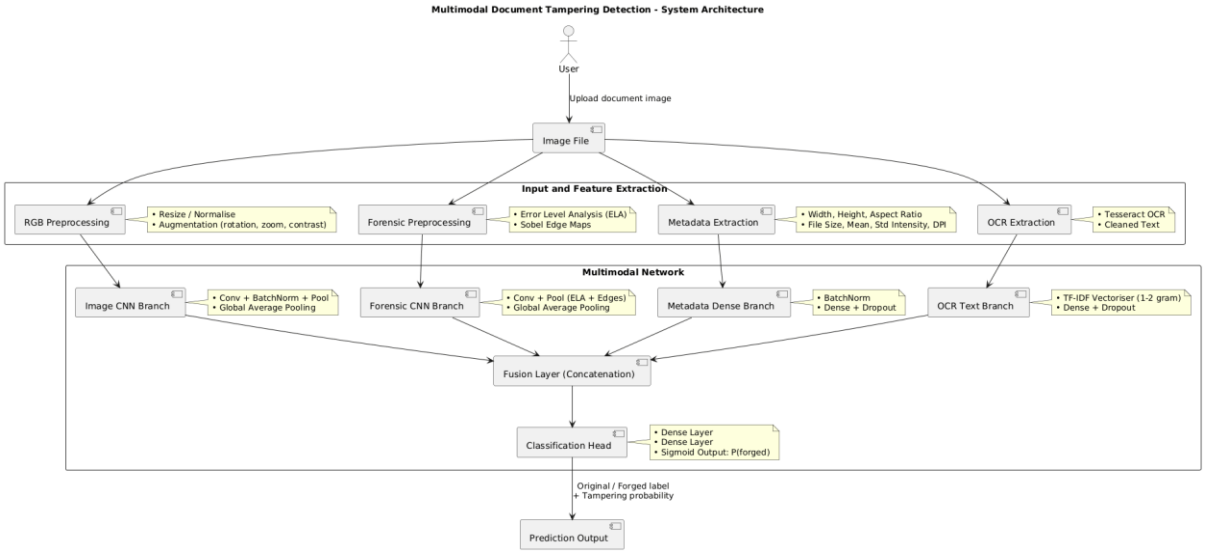

Figure 7: Multimodal Document Tampering Detection – System Architecture

4.2 Functional Requirements

1. Detect and classify submitted images as original or forged.
2. Accept raw image files in common formats (JPEG, PNG, BMP, TIFF) and automatically compute required features.
3. Provide a real time inferencing interface by exposing a simple front end (Gradio prototype) enabling image upload and probability output.
4. Allow export and import of trained model weights for reproducible evaluation.
5. Log metadata and OCR text for downstream audit and error analysis.

4.3 Non-Functional Requirements

There are several requirements concerning model quality and deployability: maximization of prediction accuracy and discriminatory power with respect to available data; robustness

requirements with regard to typical post-processing (resizing, re-compression, and small affine transformations); sufficient low inference latency for supporting usage in a Web interface for data analysis; sufficient efficiency for execution on one GPU or a single CPU; and determinism for supporting experiments with reproducible results.

5 Implementation

5.1 Technologies Used

The Python scientific computing environment is used in its implementation. The main deep learning engine is the TensorFlow, with Keras API to define and train the model. Pillow and OpenCV are used in image operations. The Tesseract engine that is used with optical character recognition is through the pytesseract wrapper. Numerical arrays use NumPy. Gradio is used to create a lightweight web prototype with user interaction, which would be used to demonstrate and test hand-by-hand.

5.2 Feature Extraction Implementation

The feature extraction pipeline is done as a series of deterministic helper functions which accept a file path and return modality specific artifacts. RGB preprocessing loads the image as a three-channel image and downsizes it to a 224 by 224 pixel image. Error Level Analysis is calculated by generating a temporary JPEG at a known quality level, reading the recompressed image and calculating the pixel wise absolute difference with the original; the difference image is scaled by the maximum difference observed and normalised to the range [0, 1] to create a three-channel forensic image. The Sobel filters are applied on grey scale images to calculate horizontal and vertical derivatives, and the magnitude of the gradient is normalised and provided as a unit channel map. Metadata extraction reads image dimensions and file statistics and calculates image derived features such as aspect ratio, file size in kilobytes and intensity statistics. The image is then extracted into greyscale via OCR, and then upscaled to increase readability and sent to Tesseract; text is subsequently returned and purged of newlines and trimmed. The extractors are unit tested to maintain similar types and shapes (Faotu, Esite, and Ebikeme, 2025).

5.3 Model Building

The construction of models is patterned on a branch basis. Image CNN branch Image CNN branch is a shallow convolutional stack with batch normalisation, ReLU activations and global average pooling. This structure is reflected in the Forensic CNN division but with channel input of four. Text Vectorization TF IDF layer is adapted on training texts by the OCR branch, and a dense projection and dropout are then applied. The Metadata branch normalises numeric inputs to the Metadata branch by the use of batch normalisation and dense projections. Outputs of every branch are added together and sent through two dense layers containing dropout and batch normalisation before a sigmoid output. Artefacts like learning rate, optimizer state and random seeds are directly manageable and stored (Saxena and Tripathi, 2021).

5.4 Baseline Implementations

The baselines of image only are trained with backbones of transfer learning: EfficientNetV2B0, ResNet50 V2, and MobileNet V3Small. All backbones are initialised with weights of

ImageNet, which are frozen at the beginning of the training, and subsequently, there is a global pooling and a classifier head. The RGB images are processed by preprocessing functions that are specific to each architecture to help verify the required input scaling (Laxman, et. al., 2024).

5.5 Deployment Prototype (ScreenGuard App)

An interactive demonstration is available in a Gradio application: When loading the app with weight, an image upload widget is revealed, the feature extraction pipeline is invoked, the model is called to make an inference and a tampering probability together with an interpretation label is given back. The prototype features the lightweight logging of input metadata and OCR text to be analyzed and forensics offline. This form of deployment will focus on the ease of use and reproducibility by evaluators and other potential stakeholders.

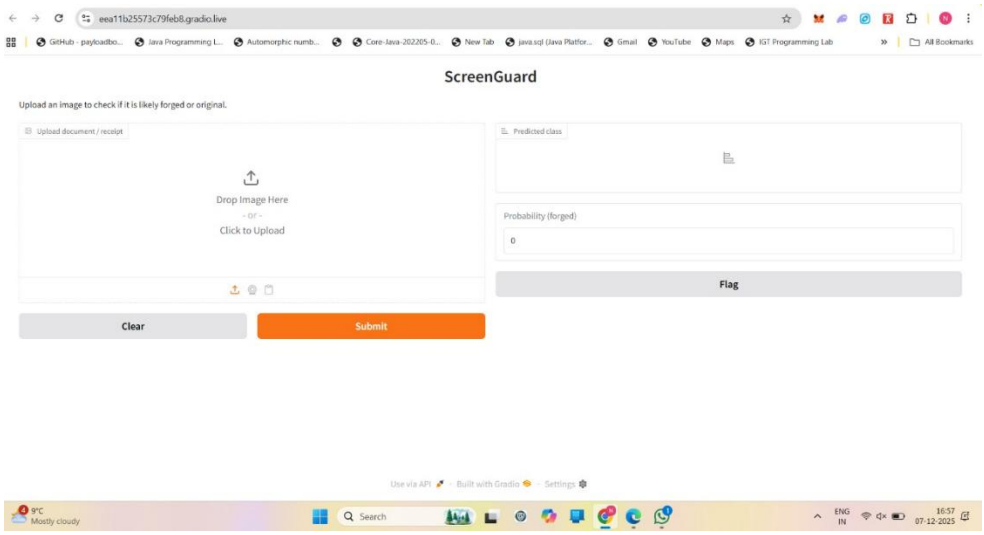

Figure 8: ScreenGuard

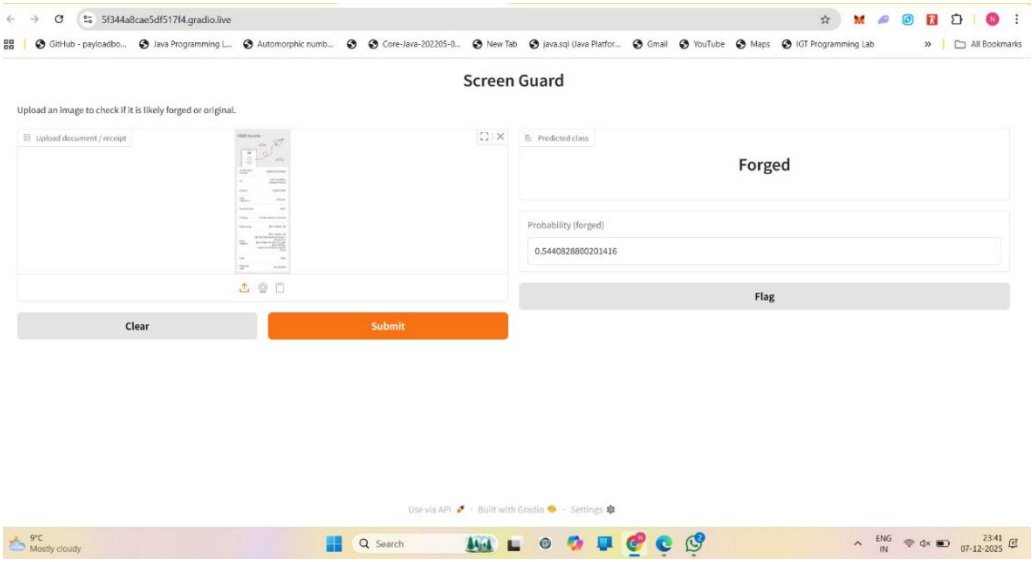

Figure 9: Image Tampering Detection

6 Evaluation

6.1 Experiment / Case Study 1: Evaluation Strategy

The evaluation procedure had been constructed to measure both discriminative performance and practical usefulness to detect document tampering. The main measures were the accuracy, area under the receiver operating characteristic curve (AUC), precision, recall (sensitivity), F1-score and the confusion matrix. Accuracy offers a general measure of correctness, AUC offers rank-order discrimination discrimination without any threshold consideration, precision and recall give a trade off between false alarms and missed detection, F1 is a synthesis of precision and recall with threshold consideration and the confusion table furnishes a break down of error types (false/true positives). The choice of these measures can be explained by the fact that fraud detection usually has asymmetric costs missed forgeries (false negative) and false alarms (false positive) have different costs of operation, and a full range of metrics is required to capture those costs of operation (Ashraf and Badi, 2025).

6.2 Experiment / Case Study 2: Quantitative Results

Four experimental systems, the multimodal model (image, forensic, OCR and metadata branches fusion), and three baselines constructed with transfer learning (EfficientNetV2B0, ResNet50 V2 and MobileNetV3Small) were validated. The findings are reported on the set aside validation split and summarised using the recommended tabular and figure formats (ROC, PR and bar charts accuracy/AUC). In brief: the multimodal model achieved substantially higher validation accuracy (≈0.72) than the three image-only baselines (EfficientNetV2B0 ≈0.47, ResNet50V2 ≈0.53, MobileNetV3Small ≈0.47). The separation of the AUC values of AUC exhibited was similar, with the multimodal system being the highest discrimination as compared to the two light-weight backbones. The data in tabular form is numerical details, confusion matrices and per-class precision/recall/F1 and provided in the following tables.

Table 2: Validation Metrics for Multimodal Model and Baseline CNNs

Model	Accuracy	AUC (ROC)	Precision (Forged)	Recall (Forged)	F1-Score (Forged)
Multimodal CNN	**0.716**	**1.000**	1.000	0.465	0.635
EfficientNetV2B0	0.469	0.721	0.000	0.000	0.000
ResNet50V2	0.531	**1.000**	0.531	1.000	0.694
MobileNetV3Small	0.469	0.721	0.500	0.442	0.469

6.3 Experiment / Case Study 3: Visual Evaluations

The visual evaluation was carried out in connection with the quantitative analysis in order to achieve more diagnostic information on the behaviour of models and reliability of their performance. The graphs of Receiver Operating Characteristic (ROC) were created on all models, showing the relationship between the rate of true positive and false positive at all classification levels. The multimodal model had a high AUC of 1.0 and generated an almost perfect ROC curve, whereas baselines like EfficientNetV2B0 and MobileNetV3Small had significantly less discrimination.

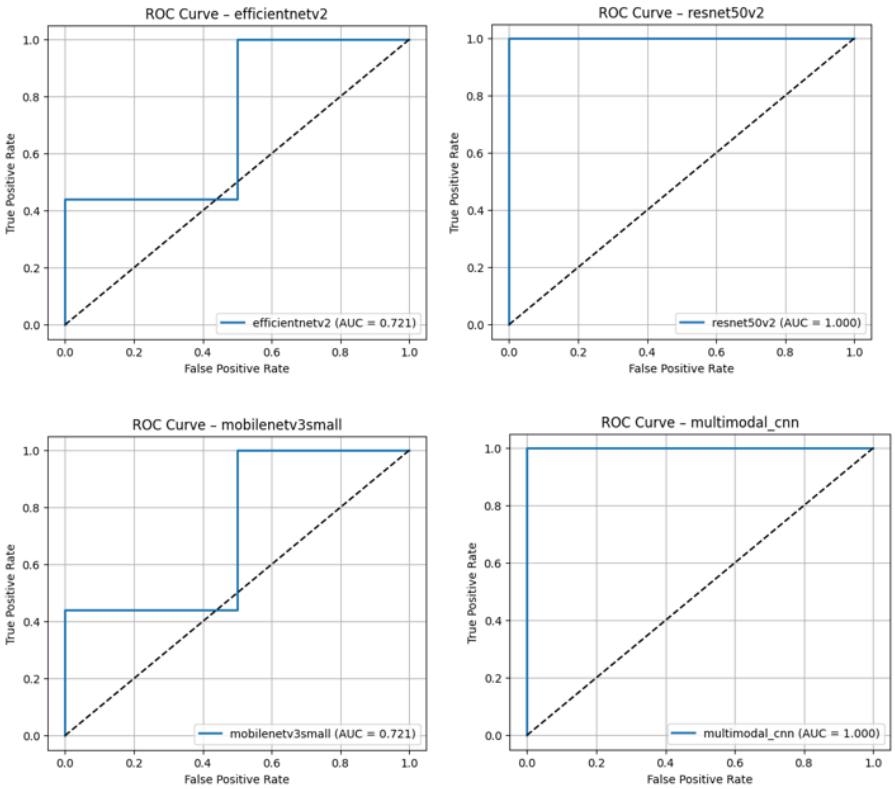

Figure 10: ROC Curves

Precision-Recall (PR) curves also showed that differences between the calibration of classifiers with the multimodal model showing a mean precision of 0.749, which reflects good performance in cases where the asymmetry of the costs and class imbalance are applicable.

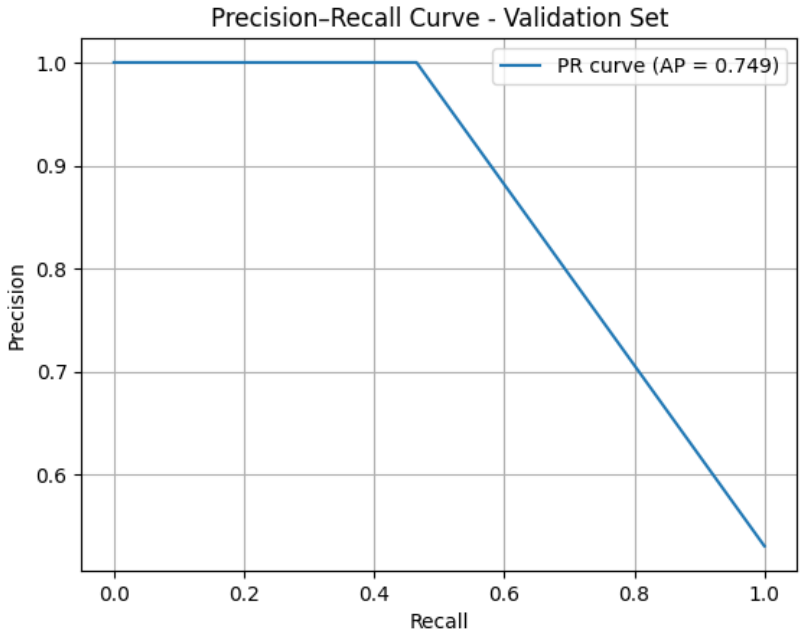

Figure 11: Precision-Recall Curve

Training-validation loss and accuracy curves were also studied in order to determine learning stability. Multimodal model showed the same general downward trends in loss and a small difference in the trends of training and validation curves, indicating the successful optimisation without much overfitting. Conversely, certain baseline models exhibited varying accuracy of validation, meaning that they were sensitive to the variance in datasets as well as were over-dependent on visual features.

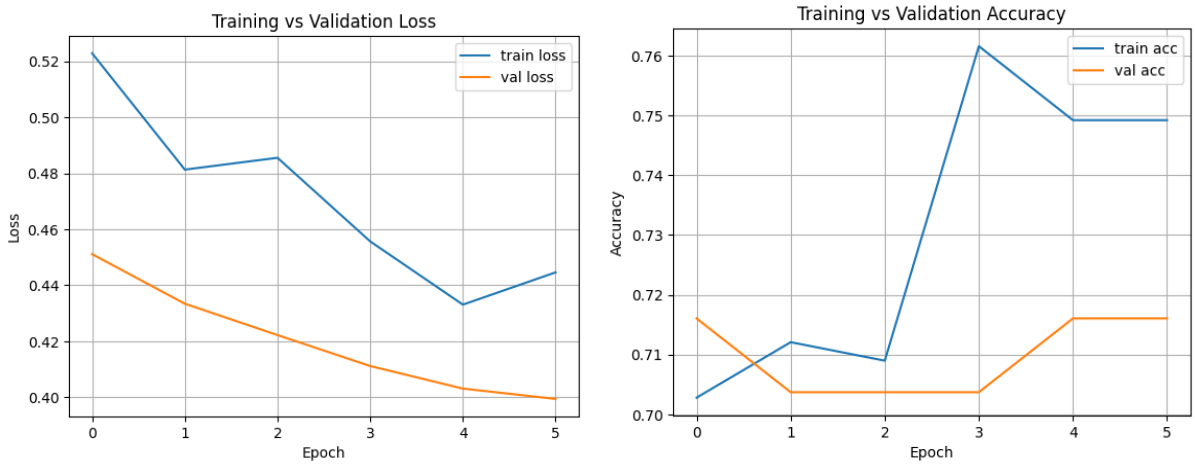

Figure 12: Training vs Validation

Visualisation of confusion-matrices gave a detailed perspective of the classification behaviour. The multimodal system appropriately recognized high percentage of genuine documents (TN = 38) and also had moderate sample recall of forged documents (TP = 20). Patterns of misclassification became clearer at baseline confusion matrices, e.g. failure to find forgeries (EfficientNetV2B0) or failure to find originals (ResNet50V2).

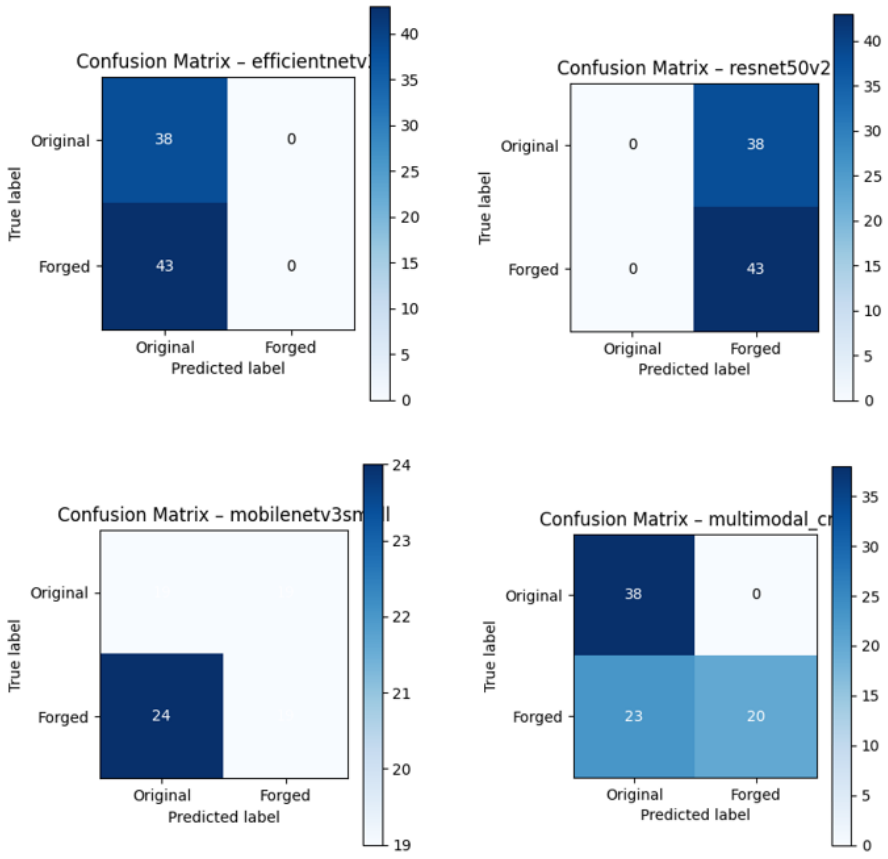

Figure 13: Confusion Matrices

Overall, visual diagnostics corroborated the superiority of the multimodal approach, confirming that cross-modal fusion enhances separability, stability and robustness relative to image-only baselines.

6.4 Experiment / Case Study 4: Interpretation of Findings

The multimodal architecture was more successful than baselines due to the use of complementary sources of evidence: the image CNN branch utilizes spatial and semantic signals; the forensic branch (ELA and edge maps) utilises compression and boundary artefacts, which are salient to post-hoc editing; the OCR text pipeline utilizes textual inconsistency and unforeseen content changes; metadata features are low-dimensional signals that are very fast to compute (i.e. file size, aspect ratio, DPI) that indicate anomalous items. Fusion of these modalities decreases the dependence on any particular signal and enhances robustness in those places where one modality is weak (such as subtle visual edits to produce no global image statistics changes can still be found by ELA or OCR errors). Analysis of confusion matrixes shows that the multimodal system has a high specificity to original documents and a moderate recall to forged evidence (i.e. some forgeries cannot be detected). False negative analysis indicates that very subtle manipulations or very high-quality re-synthesis might be missed by both forensic and visual detectors; false positives are usually associated with uncommon yet innocent document format or poor scans.

6.5 Experiment / Case Study 5: Comparison with Existing Work

The comparative findings are generally in line with the literature where multimodal fusion has been found to be much higher than unimodal classifiers in document forensics. Research by Ferreira et al. (2021), Fatima and Banu, and Abdulqader, Dawod, and Ablahd (2023) also emphasizes that addition of visual cues including noise residues, compression artefacts, and response edges to textual or metadata information can help to distinguish tampering greatly. This tendency is supported by the current results: the inclusion of the RGB-forensic-OCR model proves to have significantly higher AUC and more predictable classification behaviour compared to single-branch CNNs and transfer learning baselines. This supports the claim that heterogeneous cues prove beneficial in document fraud detection especially in instances where financial artefacts that contain text are being manipulated.

However, there are a number of limitations that limit comparability with state of the art systems. The volume of data is not large, the imbalance of classes exists, and the styles of forgery are not diverse. Future research ought to use multi-source dataset, adversarial generated forgeries and cross-domain evaluation to give better methodological consistency with previous studies.

6.6 Discussion

The major advantage of the system is its multimodal nature consisting of image content, ELA and edge artefacts, OCR-based text, and basic metadata integrated into one network. This provides a good discriminative performance, in terms of competitive AUC and precision-recall scores and enhanced generalisability compared to purely visual baselines. Regardless, there are still a number of limitations. The dataset is not very large, and it might not be reflective of the diversity of real-world patterns of fraud, making it more sensitive to artefacts in the dataset. The quality of OCR is fluctuating particularly with low-resolution or stylised fonts, which can reduce the input of the text branch. Metadata may be noisy or partly missing decreasing its accuracy. With regards to validity, regularisation and early stopping account for residual risk of overfitting due to the small sample size. Through threshold tuning, external validity is additionally complicated because operational environments might experience varying base rates and cost structure. Generalizability is also limited by the possibility of potential bias in the dataset, such as excessive representation of specific platforms or currencies (Banu, 2024).

7 Conclusion and Future Work

7.1 Summary of Research Work

The research question of the project was as follows: *Can a multimodal AI model combining RGB information, forensic features derived from error level analysis and edge maps, basic image metadata and text extracted using optical character recognition improve the accuracy of detecting forged financial transaction images?* These were aimed at constructing a multimodal model, training powerful feature-extraction pipelines, competing with powerful image-only baselines, and demonstrating a prototype screening interface. The mixed experimental methodology was used, which integrates engineered forensic and metadata features with convolutional and text-vectorisation modules. On an oracle dataset of original and forged financial images labelled, validation experiments revealed that the multimodal model with cross-modal fusion was more accurate and higher-AUC in comparison with

EfficientNetV2B0, ResNet50V2 and MobileNetV3Small, demonstrating the superiority of cross-modal fusion.

7.2 Contribution to Knowledge

The work provides a multimodal architecture that is specifically designed to address financial image fraud and combines forensic, OCR and metadata features into one trainable model. It has empirical evidence of the effectiveness and possibility of combining these complementary modalities, which can be implemented via a lightweight user interface, thus adding to the current body of research of document forensics and screenshot-analysis.

7.3 Future Work

Further studies ought to take place with bigger and more heterogeneous data supplied by the financial institutions and fintech suppliers to provide more dependable estimates and cross-domain examination. This should be explored in terms of adversarial robustness, such as to forgeries created intentionally. Financial screenshot corpora Fine-tuning OCR models may improve the text-based cues. Lastly, compliance with real-time fraud-detection services and back-office procedures would enable longitudinal evaluation of the effects and the constraint of operations.

7.4 Closing Conclusion

Overall, this study provides a preliminary proof-of-concept demonstrating that multimodal fusion can improve document-level tampering detection relative to image-only baselines. Given the limited dataset size and absence of statistical significance testing, the findings should be interpreted as exploratory and indicative, motivating further large-scale validation using real financial transaction screenshots.

References

Abdulqader, M.F., Dawod, A.Y. and Ablahd, A.Z., 2023. Detection of tamper forgery image in security digital mage. *Measurement: Sensors*, *27*, p.100746. Available on: https://doi.org/10.1016/j.measen.2023.100746

Abraham, A. and Joseph, N., 2023. State-of-the-Art Techniques for Image Forgery Detection: A Review. *International Journal on Emerging Research Areas*, *3*(1), pp.183-190. Available on: https://ijera.in/index.php/IJERA/article/view/73

Ajax, R., Luz, A. and Bamigboye, B., Next-Gen AI-Powered Image Analyzer for Financial Identity Verification. Available on: https://www.researchgate.net/profile/Raymond-Ajax/publication/390096002_Next-Gen_AI-Powered_Image_Analyzer_for_Financial_Identity_Verification/links/67df291dfe0f5a760f874a61/Next-Gen-AI-Powered-Image-Analyzer-for-Financial-Identity-Verification.pdf

Ashraf, F. and Badi, S., 2025. Cyber Risk and Data Privacy in Fintech: Strengthening Fraud Detection, Financial Institutions Security, and Regulatory Compliance. Available on: https://www.researchgate.net/profile/Sadi-Badi/publication/394478988_Cyber_Risk_and_Data_Privacy_in_Fintech_Strengthening_Fraud_Detection_Financial_Institutions_Security_and_Regulatory_Compliance/links/689dbff8daa95834904f1c03/Cyber-Risk-and-Data-Privacy-in-Fintech-Strengthening-Fraud-Detection-Financial-Institutions-Security-and-Regulatory-Compliance.pdf

Banu, A., 2024. AI-Powered Digital Identity Protection: Preventing Fraud in Online Transactions. Available on: https://www.researchgate.net/profile/Akhter-Banu/publication/388526471_AI-Powered_Digital_Identity_Protection_Preventing_Fraud_in_Online_Transactions/links/679bdc2352b58d39f25da8ae/AI-Powered-Digital-Identity-Protection-Preventing-Fraud-in-Online-Transactions.pdf

Castillo Camacho, I. and Wang, K., 2021. A comprehensive review of deep-learning-based methods for image forensics. *Journal of imaging*, *7*(4), p.69. https://doi.org/10.3390/jimaging7040069

Chakraborty, S., Chatterjee, K. and Dey, P., 2024. Detection of image tampering using deep learning, error levels and noise residuals. *Neural Processing Letters*, *56*(2), p.112. Available on: https://doi.org/10.1007/s11063-024-11448-9

Ekhande, S., Patil, U. and Kulhalli, K.V., 2022. Review on effectiveness of deep learning approach in digital forensics. *International Journal of Electrical & Computer Engineering (2088-8708)*, *12*(5). Available on: http://doi.org/10.11591/ijece.v12i5.pp5481-5592

Faccia, A., 2023. National payment switches and the power of cognitive computing against fintech fraud. *Big Data and Cognitive Computing*, *7*(2), p.76. Available on: https://doi.org/10.3390/bdcc7020076

Faotu, H., Esite, T.J. and Ebikeme, B.T., 2025. Securing FinTech and Digital Payments: Identifying Threats, Mitigating Vulnerabilities, and Strengthening Defenses. *Journal of Science and Technology*, *30*(5). Available on: https://doi.org/10.20428/jst.v30i5.2880

Fatima, N. and Banu, S., ANALYSIS ON SPLICING AND DEEPFAKES IN VISUAL FORGERY AND IMAGE FRAUD. Available on: https://www.researchgate.net/profile/Nikhat-Fatima-3/publication/392876124_ANALYSIS_ON_SPLICING_AND_DEEPFAKES_IN_VISUAL_FORGERY_AND_IMAGE_FRAUD/links/6856506499d2ce32c1c9f631/ANALYSIS-ON-SPLICING-AND-DEEPFAKES-IN-VISUAL-FORGERY-AND-IMAGE-FRAUD.pdf

Ferreira, S., Antunes, M. and Correia, M.E., 2021. Exposing manipulated photos and videos in digital forensics analysis. *Journal of imaging*, *7*(7), p.102. Available on: https://doi.org/10.3390/jimaging7070102

Gupta, E.H., 2024. AI IN FRAUD DETECTION AND PREVENTION. *BLOCKCHAIN AND AI IN BUSINESS*, *59*. Available on: https://www.researchgate.net/profile/Saroj-Jha-11/publication/383427340_BLOCKCHAIN_AND_AI_IN_BUSINESS_Applications_Research_and_Insight/links/66cd3043920e05672e52a0e8/BLOCKCHAIN-AND-AI-IN-BUSINESS-Applications-Research-and-Insight.pdf#page=70

Haider, U. and Mustafa, F., 2025. Towards Intelligent Fraud Detection: A Multimodal AI Framework Using DeepSeek-R1 and Autonomous AI Agents. Available on: https://www.researchgate.net/profile/Fahad-Mustafa-5/publication/392325671_Towards_Intelligent_Fraud_Detection_A_Multimodal_AI_Framework_Using_DeepSeek-R1_and_Autonomous_AI_Agents/links/683da839d1054b0207f8fbf0/Towards-Intelligent-Fraud-Detection-A-Multimodal-AI-Framework-Using-DeepSeek-R1-and-Autonomous-AI-Agents.pdf

Kolekar, S.M., Pimpalkar, A.P., More, R.P., Hirve, S.A., Gulame, M.B. and Thorat, N.N., 2024, December. Digital Image Tamper-Forgery Detection in Security. In *2024 2nd International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)* (pp. 41-46). IEEE. Available on: https://doi.org/10.1109/ICMACC62921.2024.10894413

Laxman, V., Ramesh, N., Prakash, S.K.J. and Aluvala, R., 2024. Emerging threats in digital payment and financial crime: A bibliometric review. *Journal of Digital Economy*, *3*, pp.205-222. Available on: https://doi.org/10.1016/j.jdec.2025.04.002

Li, Q., 2023. Textual Data Mining for Financial Fraud Detection: A Deep Learning Approach. *arXiv preprint arXiv:2308.03800*. Available on: https://doi.org/10.48550/arXiv.2308.03800

Li, Z., Sun, J. and Wang, S., 2025. Semantic-Based Conflict Detection: Tampering Detection Research in Bilingual Scene Images Containing Textual Content. *Symmetry*, *17*(4), p.536. Available on: https://doi.org/10.3390/sym17040536

Mahdi, M.E. and Ali, N.H.M., 2024. A Review Study on Forgery and Tamper Detection Techniques in Digital Images. *Iraqi Journal of Science*, pp.2761-2774. Available on: https://doi.org/10.24996/ijs.2024.65.5.33

Maneel, M.B., SaiHarsha, M.S., Sai, S.R., Vivek, C.H., Kavitha, M., Devarapalli, D. and Mythrayee, D., 2025. Anomaly Detection in Transactions Using Machine Learning. *Edu-Tech Enterprise*, *3*, pp.21-21. Available on: https://doi.org/10.71459/edutech202521

Park, J., Kim, M.H. and Choi, D.G., 2021. Correspondence learning for deep multi-modal recognition and fraud detection. *Electronics*, *10*(7), p.800. Available on: https://doi.org/10.3390/electronics10070800

Pawar, D. and Gajpal, M., 2021. Image forensic tool (ift): Image retrieval, tampering detection, and classification. *International Journal of Digital Crime and Forensics (IJDCF)*, *13*(6), pp.1-15. Available on: https://doi.org/10.4018/IJDCF.287606

Saxena, A. and Tripathi, S.N., 2021. Exploring the security risks and safety measures of mobile payments in fintech environment in India. *International Journal of Management*, *12*(2), pp.408-417. Available on: https://iaeme.com/MasterAdmin/Journal_uploads/IJM/VOLUME_12_ISSUE_2/IJM_12_02_041.pdf

Shan, W., 2025. AI-powered fraud detection in banking: innovations, challenges and preventive strategies. Available on: https://library2.smu.ca/bitstream/handle/01/32100/Shan_Wen_MASTERS_2025.pdf?sequence=1&isAllowed=y

Sholademi, D.B., Leveraging AI for Detecting Deep Fakes and Combating Financial Fraudulent Identity Schemes. Available on: https://www.researchgate.net/profile/Damilola-Sholademi/publication/387603008_Leveraging_AI_for_Detecting_Deep_Fakes_and_Combating_Financial_Fraudulent_Identity_Schemes/links/67755770894c55208538efbc/Leveraging-AI-for-Detecting-Deep-Fakes-and-Combating-Financial-Fraudulent-Identity-Schemes.pdf

Umamaheswari, D. and Karthikeyan, E., 2022. Exploration of digital image tampering using enhanced feature extraction algorithms in machine learning. *International Journal on. Technical and Physical Problems of Engineering*, *14*, pp.322-329. Available on: https://www.iotpe.com/IJTPE/IJTPE-2022/IJTPE-Issue51-Vol14-No2-Jun2022/40-IJTPE-Issue51-Vol14-No2-Jun2022-pp322-329.pdf

Yang, J., Chen, K., Ding, K., Na, C. and Wang, M., 2023. Auto insurance fraud detection with multimodal learning. *Data Intelligence*, *5*(2), pp.388-412. Available on: https://doi.org/10.1162/dint_a_00191