Detection and Localization of Tampering in Medical Images

A midterm capstone project report submitted in partial fulfillment of the requirement for the award of the degree of

Bachelor of Engineering

in

Computer Engineering / Computer Science and Engineering

Submitted By

Muskan Chalana (102203274)

Ishita Jindal (102203668)

Asmi Gaurav (102203253)

Sudikshya Nyachhyon (102217201)

Bhavuk Gupta (102203981)

Group No 125

Under Supervision of

Dr. Shalini Batra (Professor and Dean of Faculty Affairs)

Dr. Geeta Kasana (Associate Professor)

Department of Computer Science and Engineering

THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY, PATIALA, PUNJAB

Aug 2025

ABSTRACT

The integrity of medical images is critical for accurate diagnosis, treatment planning, and patient safety. With the increasing availability of digital imaging systems, the risk of tampering and unauthorized modifications has become a major concern, potentially leading to misdiagnosis and compromised trust in healthcare technologies. This project addresses the problem by developing a **dual-pipeline tamper detection and localization framework** for CT scans and mammographs. The system leverages **deep learning architectures**, including a ResNet-based classifier and a U-Net segmentation model for CT scans, and an EfficientNetB3 model integrated with CBAM attention mechanisms for mammographs. A dedicated **user interface** was designed to facilitate real-time interaction, allowing clinicians to upload medical images and receive both **binary classification results** (tampered vs. untampered) and **visual localization maps** for suspected manipulations. Performance was evaluated using metrics such as Accuracy, IoU, Dice Score, Precision, Recall, and F1-Score, ensuring robustness and reliability. The outcomes demonstrate the system's ability to combine automated classification with interpretable visual outputs, thereby offering a practical and trustworthy solution for safeguarding the authenticity of medical imaging data.

DECLARATION

We hereby declare that the capstone project group report title "Detection and Localization of Tampering in Medical Images" is an authentic record of our own work carried out at "Thapar Institute of Engineering and Technology, Patiala" as a Capstone Project in seventh semester of B.E. , under the guidance of **Dr. Shalini Batra** and **Dr. Geeta Kasana**, during January to December 2025.

Date: 23th August 2025

Muskan Chalana	102203274	Signature
Ishita Jindal	102203668	Signature
Asmi Gaurav	102203253	Signature
Sudikshya Nyachhyon	102217201	Signature
Bhavuk Gupta	102203981	Signature

Mentor's Signature :

ACKNOWLEDGEMENT

I would like to extend my sincere gratitude to my project mentors, Dr. Shalini Batra and Dr. Geeta Kasana for their invaluable guidance, constructive feedback, and unwavering support throughout this capstone project. Their expertise and insightful suggestions were instrumental in shaping the direction of this research and overcoming numerous challenges.

I am also deeply thankful to TIET Patiala for providing me with the necessary resources, academic environment, and foundational knowledge that made this project possible. The curriculum and faculty have prepared me exceptionally well for this undertaking.

Date: 23th August 2025

Roll No.	Name
102203274	Muskan Chalana
102203668	Ishita Jindal
102203253	Asmi Gaurav
102217201	Sudikshya Nyachhyon
102203981	Bhavuk Gupta

TABLE OF CONTENTS

CHAPTER / SECTION **Page No.**

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AI	Artificial Intelligence
AE	Autoencoder
CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DFD	Data Flow Diagram
ELA	Error Level Analysis
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HIPAA	Health Insurance Portability and Accountability Act
IoU	Intersection over Union
MRI	Magnetic Resonance Imaging
PACS	Picture Archiving and Communication System
PRNU	Photo-Response Non-Uniformity
SRS	Software Requirement Specification
SSIM	Structural Similarity Index Measure
TIET	Thapar Institute of Engineering and Technology
U-Net	CNN for Biomedical Image Segmentation
WHO	World Health Organization

CHAPTER 1: INTRODUCTION

1.1 PROJECT OVERVIEW

Medical imaging has emerged as one of the most indispensable tools in modern healthcare, playing a crucial role in diagnosis, treatment planning, and long-term monitoring of a wide variety of diseases. Techniques such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), mammography, and X-rays are routinely used by clinicians to identify abnormalities, track disease progression, and guide therapeutic interventions. As the healthcare industry becomes increasingly digitized, the reliance on digital scans has grown substantially, with such images being shared across hospitals, research institutions, and academic databases worldwide. However, this rapid digitization has introduced new risks, as medical images are no longer immune to digital tampering. Malicious alterations to medical scans can have catastrophic consequences, ranging from incorrect diagnoses and unnecessary surgical interventions to erosion of trust in clinical research. With stakes involving patient safety, ethical responsibility, and institutional credibility, the necessity of developing a robust tampering detection framework has become more critical than ever.

In recent years, the manipulation of medical scans has gained attention in both the healthcare and forensic communities. Conventional editing tools as well as more advanced generative techniques can be used to subtly or extensively alter medical imagery without leaving visible artifacts detectable by the human eye. Such manipulations can involve adding artificial lesions, removing existing anomalies, or modifying the intensity of tissues in a manner that may mislead diagnostic procedures. Traditional medical image processing techniques are not sufficient to counter this issue, as they were primarily designed for tasks such as enhancement, segmentation, and classification, not forensic authentication. Hence, dedicated tampering detection mechanisms are essential. These mechanisms must not only differentiate between authentic and tampered scans but also localize and visualize the altered regions to provide interpretable evidence to radiologists and forensic experts. The proposed project addresses this pressing challenge by designing an advanced detection framework that combines traditional forensic methods with modern deep learning architectures.

The approach adopted in this project is grounded in a multi-method investigative strategy. Initially, traditional forensic techniques such as Error Level Analysis (ELA) and noise pattern analysis were explored. ELA is a classical method that highlights compression artifacts introduced during tampering, while noise inconsistency analysis examines variations in sensor-level or compression-level noise that may indicate manipulation. However, while these approaches can sometimes reveal anomalies in natural images, their effectiveness on medical scans was limited due to the high sensitivity of medical imaging to acquisition noise and modality-specific characteristics. To overcome these shortcomings, machine learning-based approaches were investigated. Patch-level classifiers were designed to capture inconsistencies in small regions of images, followed by experimentation with ResNet-18, a deep convolutional neural network pre-trained on large-scale datasets. These models provided stronger baselines but lacked the capacity to accurately localize tampered regions, an essential requirement for medical image authentication.

To address the limitations of the earlier methods, a U-Net–based segmentation framework was developed. The U-Net architecture, originally designed for biomedical image segmentation, is particularly well-suited for pixel-level classification tasks and provides the dual advantage of detecting tampering while simultaneously localizing manipulated regions. In this project, the U-Net model was trained on two benchmark medical datasets: the **Lung CT Diagnosis dataset** (sourced from The Cancer Imaging Archive) and the **CBIS-DDSM mammography dataset** (sourced from Kaggle). A tampered dataset was artificially generated by applying controlled manipulations such as segmentation-based alterations and blob tampering, ensuring a realistic but systematic basis for evaluation. This enabled the model to learn from both authentic and manipulated scans across two very different imaging modalities, thereby improving versatility. During inference, the model was designed to output not only a classification (authentic vs. tampered) but also a segmentation mask highlighting the suspected tampered regions. Such outputs ensure interpretability and facilitate trust in the detection process, as radiologists can visually verify the evidence provided by the model.

The experimental evaluation of the framework was carried out on a test set of 116 CT images and additional mammographic samples. For CT scans, the U-Net–based system demonstrated an accuracy of 85.34%, with a precision of 76.47%, a recall of 98.11%, and an F1-score of 85.95% in image-level classification. More importantly, the system exhibited strong localization capabilities, achieving a mean Intersection-over-Union (IoU) of 0.6165 and a mean Dice coefficient of 0.7536 across tampered slices. These results underscore the robustness of the framework, particularly in terms of recall, which indicates that the system was able to successfully detect the vast majority of tampered images with very few false negatives. In a medical context, this is highly valuable, as missing a tampered case could have severe implications for patient safety and institutional credibility.

A significant emphasis was also placed on ensuring interpretability and transparency in the detection pipeline. While machine learning models are often criticized for being “black boxes,” this framework mitigates such concerns by producing explicit segmentation overlays that highlight suspected regions of tampering. These overlays can be directly compared against the original scans, allowing experts to validate or contest the model’s predictions. This feature not only enhances trust but also aligns with the clinical requirement for explainable artificial intelligence (XAI) in medical applications. By visualizing the areas of concern, the system empowers medical professionals to make informed decisions while still relying on their expertise for final interpretation.

The proposed solution contributes not only to the domain of medical imaging but also to the broader field of digital forensics. The methodology demonstrates how a combination of traditional forensic techniques and modern deep learning methods can be synergized to address a complex and high-stakes problem. The curated tampered dataset developed during this project can serve as a valuable resource for further research, while the findings provide empirical evidence of the strengths and limitations of different approaches. Notably, the results achieved with the U-Net model compare favorably with those reported in existing literature, where tamper detection in medical imaging remains a relatively underexplored area. While prior research has demonstrated the use of U-Net and related architectures in tasks such as lesion segmentation and anomaly detection, their application to tampering detection has been limited. The promising results of this project thus provide a foundation for future studies aimed at improving robustness, generalization, and deployment readiness.

From a practical perspective, the successful deployment of such a tampering detection system holds significant potential for integration into healthcare workflows. Hospitals and diagnostic centers could employ this framework as an automated verification layer before radiologists review scans, thereby reducing the risk of medical fraud and misdiagnosis. Research institutions could adopt such tools to safeguard the integrity of datasets used in clinical studies. Furthermore, the interpretability of the framework makes it a suitable candidate for regulatory approval, as healthcare regulators increasingly demand transparency and explainability in AI-driven medical tools.

In conclusion, the project represents a critical step toward enhancing the security and trustworthiness of medical imaging. By systematically evaluating traditional forensic methods, patch classifiers, CNN architectures, and ultimately a U-Net–based framework, a comprehensive approach was developed that balances accuracy, interpretability, and robustness. The results demonstrate that while traditional methods provide limited utility, deep learning–based segmentation offers a highly promising solution for tampering detection in medical scans. Beyond academic contributions, the framework has practical implications for healthcare and medical research, ultimately contributing to safer diagnoses and stronger trust in digital medical systems. Future work could involve extending the framework to additional modalities (e.g., MRI, PET, ultrasound), improving performance through ensemble models, and deploying the system in real-world hospital environments. Through such advancements, the long-term goal of ensuring the authenticity and reliability of medical images can be realized, thus reinforcing both patient safety and the ethical standards of modern healthcare.

1.2 NEED ANALYSIS

Medical imaging stands as the cornerstone of modern healthcare, enabling accurate diagnosis, precise treatment planning, and effective monitoring of disease progression. With the rapid digitization of health records and the integration of AI-driven diagnostic systems, medical scans are now more accessible and widely exchanged than ever before. However, this growing reliance on digital images has introduced a critical vulnerability: the risk of tampering and manipulation. Even minor alterations to diagnostic scans can lead to devastating consequences. For example, the removal of a malignant tumor from a CT scan or the artificial insertion of abnormalities into an MRI may result in false diagnoses, inappropriate therapies, or overlooked life-threatening conditions. In critical scenarios such as early cancer detection or spinal trauma assessment, undetected modifications can delay essential treatment, thereby jeopardizing patient safety and outcomes.

The ethical implications of such manipulations are equally concerning. Altered scans compromise the integrity of medical decision-making and place healthcare professionals in positions of risk, as decisions are made on falsified evidence. Moreover, the credibility of medical research is threatened when manipulated images are incorporated into clinical studies. This may lead to inaccurate findings, ineffective treatment protocols, and wasted resources. Such outcomes erode trust not only in scientific research but also in healthcare institutions that rely heavily on validated imaging evidence for advancement. The absence of robust detection mechanisms therefore poses a direct challenge to the ethical standards, reliability, and accountability of healthcare practices.

Given these risks, the urgent need for a reliable medical image tampering detection framework is evident. Such a solution must not only differentiate between authentic and manipulated scans but also provide clear localization of the altered regions, thereby ensuring transparency and interpretability in diagnostic processes. By integrating image forensics with deep learning architectures, it becomes possible to capture subtle inconsistencies that human experts may overlook. This empowers radiologists and researchers with a trustworthy second line of verification, reducing the chances of clinical misjudgment or research malpractice. Furthermore, ensuring adaptability across multiple imaging modalities—including CT, MRI, and mammography—makes the proposed framework versatile and applicable in diverse medical contexts. Consequently, the implementation of such systems is essential for safeguarding the authenticity, reliability, and ethical standards of digital medical imaging in the future.

1.3 RESEARCH GAPS

Medical image tampering detection has been studied using a variety of approaches, ranging from watermarking and resampling analysis to deep learning–based methods. However, when focusing specifically on tampering in CT and mammography images, several important gaps remain. The following research works illustrate existing methods and their limitations, along with how the present project addresses those gaps.

Gap 1 – Dependence on Watermarking Techniques
Chô et al. (2007) proposed a near-lossless watermarking approach for mammograms, where hidden watermarks enable image integrity verification [Springer]. While effective, such methods are invasive, as they alter the original image by embedding additional information.
This raises concerns in medical imaging, where diagnostic accuracy depends on preserving original pixel data.
Our Approach: We adopt a non-invasive, vision-based forensic model that detects tampering directly from image content without requiring embedded watermarks. This ensures clinical images remain unaltered while still enabling tamper detection.

Gap 2 – Focus on Pathology, Not Tampering
Xi et al. (2018) developed CNN-based weakly supervised methods for abnormality localization in mammograms [arXiv]. Similarly, Zhang et al. (2023) used multi-scale attention networks for lesion classification and localization [Frontiers].
These methods are excellent for pathology detection, but they do not address malicious image tampering.
Our Approach: Instead of detecting lesions or cancer-related abnormalities, our system is explicitly designed to localize tampered regions such as added/removed microcalcifications or nodules. This ensures radiologists can differentiate between true pathology and artificially introduced features.

Gap 3 – Limited Use of High-Resolution Forensic Features
Marra et al. (2019) introduced full-resolution CNNs for generic image forgery detection, demonstrating the importance of preserving high-frequency cues [arXiv]. Similarly, Bunk et al. (2017) leveraged resampling artifacts and Radon transforms for localization [arXiv].
However, these approaches have not been applied to medical imaging, where fine-grained structures (e.g., microcalcifications in mammograms) are crucial.
Our Approach: We extend high-resolution forensic detection concepts to mammography and CT imaging, enabling the detection of micro-level alterations that may go unnoticed in conventional pathology models.

Gap 4 – Generalisation Across Tampering Types
Cozzolino et al. (2018) presented ForensicTransfer, a domain adaptation method that generalizes to unseen forgery types [arXiv]. While effective in natural images, medical image tampering often involves subtle manipulations such as lesion removal/addition rather than large-scale edits. Existing methods lack robustness to these subtle manipulations.
Our Approach: We focus on adaptation to subtle tampering in medical images, ensuring robustness even when manipulations are minimal or adversarial in nature.

Gap 5 – Lack of Explainability in Tamper Detection

Recent medical AI work, such as GMIC (Geras et al., 2020) [NYU/ArXiv], emphasizes explainability by providing global and local interpretations of mammograms.

However, most forensic tamper detection models stop at binary classification (tampered/original) without providing interpretable localization outputs.

Our Approach: Our system integrates explainable localization heatmaps, allowing radiologists to visually verify suspicious areas. This improves trust and usability in clinical workflows.

Gap 6 – Vulnerability to Adversarial Noise

Ma et al. (2023) highlighted the issue of adversarial robustness in medical imaging [Springer]. Most existing tamper detection models are vulnerable to minimal pixel-level perturbations that can mislead the system without being noticeable to the human eye.

Our Approach: We incorporate robust training and evaluation strategies to ensure that the detection model can withstand subtle adversarial-style tampering, making it more clinically reliable.

1.4 PROBLEM DEFINITION AND SCOPE

Problem Definition

The reliability of medical imaging is increasingly at risk in the digital era. Modalities such as **CT scans and mammograms** play a central role in disease diagnosis, treatment planning, and long-term monitoring. However, because these scans exist in digital form, they are susceptible to tampering. Malicious or unintentional alterations can fabricate or conceal abnormalities, which in turn may cause misdiagnosis, inappropriate treatment, or missed early disease detection. In sensitive cases such as cancer screening through mammography or trauma detection through CT imaging, even subtle modifications can result in severe clinical consequences. Furthermore, the use of falsified medical images in clinical research compromises scientific credibility, leading to unreliable findings and potentially harmful medical practices. Despite these critical risks, no standardized tampering detection mechanism is currently integrated into medical imaging workflows. Visual inspection by radiologists often fails to detect subtle manipulations, underscoring the urgent need for automated and interpretable tampering detection systems.

Scope of the Project

This project seeks to design and implement a **tampering detection framework** tailored for medical imaging, focusing on **both CT scans and mammograms**. The framework integrates classical image forensic techniques with advanced deep learning approaches, ensuring robust detection across modalities. The proposed system is designed to perform two tasks:

1. **Classification** — determining whether a medical scan is authentic or tampered.
2. **Localization** — identifying and segmenting regions of manipulation for interpretability.

To simulate real-world tampering, a large dataset of CT scans and mammograms was collected, and a tampered subset was generated using segmentation-based and blob-based manipulation methods. Multiple detection strategies were evaluated, including Error Level Analysis (ELA), noise pattern analysis, patch-based classification, ResNet18-based image classification, and U-Net–based segmentation. Among these, the **U-Net architecture achieved the strongest results**, providing high classification accuracy (85.34%) and reliable localization performance (Mean Dice Score: 0.7536, Mean IoU: 0.6165).

The project's scope, while currently demonstrated on CT scans and mammograms, is extendable to other modalities such as MRI and PET scans. This ensures adaptability in diverse diagnostic contexts. Although deployment in clinical practice would require integration with hospital PACS systems and large-scale validation, this project establishes a strong proof of concept. It demonstrates how tampering detection methods can enhance **trust, reliability, and security** in medical imaging workflows by assisting radiologists and researchers in verifying the authenticity of scans.

1.5 ASSUMPTIONS AND CONSTRAINTS

Assumptions

1. **Data Authenticity**
   It is assumed that the **Lung CT Diagnosis dataset** (sourced from The Cancer Imaging Archive) and the **CBIS-DDSM dataset** (sourced from Kaggle) are authentic, clinically validated, and free from prior manipulation. These datasets serve as the ground truth for generating tampered data.

2. **Tampering Simulation**
   The tampered datasets created through segmentation-based editing and blob manipulation are assumed to be representative of real-world tampering scenarios. These artificially generated manipulations are treated as valid proxies for malicious alterations.

3. **Single-Modality Input**
   Each input to the system is assumed to belong to a single imaging modality (either CT or mammogram) and not mixed images. The model is expected to function independently across modalities without requiring modality-specific preprocessing during inference.

4. **Standardized Preprocessing**
   It is assumed that all input scans (from both datasets) are preprocessed into a consistent format (grayscale, standardized resolution, and normalized pixel intensity). This ensures that the model generalizes well across both CT scans and mammograms.

5. **Ground Truth for Evaluation**
   During performance evaluation, manually annotated tampered regions (on both CT and mammogram data) are assumed to be accurate representations of ground-truth manipulation for segmentation metrics (IoU and Dice).

6. **Deployment Context**
   It is assumed that in real-world usage, the system would act as an assistive tool for radiologists and not as a standalone decision-making authority. Human verification will remain a necessary component of diagnosis.

Constraints

1. **Dataset Availability**
   The project is constrained to two publicly available datasets: **Lung CT Diagnosis (TCIA)** for CT scans and **CBIS-DDSM** for mammograms. While these datasets are widely used in research, their size and population diversity may limit the generalizability of the model to broader clinical settings.

2. **Synthetic Tampering vs. Real Tampering**
   Since tampered datasets were artificially generated (via segmentation and blob-based editing), a constraint exists in comparing system performance against actual maliciously altered scans, which may employ more sophisticated manipulation techniques.

3. **Computational Resources**
   Training deep learning models such as ResNet18 and U-Net required high-performance GPUs. Due to resource limitations, experiments were conducted on college-provided or cloud-based GPUs with constraints on training time, batch size, and model complexity.

4. **Modality Limitation**
   Although the framework was tested on **CT scans and mammograms**, constraints of time and data availability prevented extending the system to other modalities such as MRI and PET. The system's performance outside the tested modalities remains unverified.

5. **Evaluation Metrics**
   While standard classification metrics (Accuracy, Precision, Recall, F1-score) and segmentation metrics (IoU, Dice) were employed, clinical validation with practicing radiologists was not performed due to time and ethical constraints.

6. **Scalability in Clinical Workflows**
   Real-world deployment requires integration with PACS (Picture Archiving and Communication Systems) in hospitals. Such integration, along with compliance with medical data privacy standards (HIPAA, GDPR), was beyond the project's current scope.

1.6 STANDARDS

The development of a tampering detection framework for medical imaging must align with established technical, medical, and ethical standards to ensure reliability, interoperability, and compliance. The following standards are considered relevant to this project:

1. **Data Representation Standards**

- **DICOM (Digital Imaging and Communications in Medicine):**
  All medical images (CT scans from TCIA's *Lung CT Diagnosis* dataset and mammograms from *CBIS-DDSM*) follow the DICOM format, which is the international standard for storing, transmitting, and managing medical imaging information. This ensures compatibility with existing hospital systems such as PACS (Picture Archiving and Communication Systems).
- **Grayscale Normalization Standards:**
  Image preprocessing adheres to standard practices of intensity normalization (e.g., scaling to 0–255 or 0–1 range) and resizing to maintain consistency across datasets and imaging modalities.

2. **Performance Evaluation Standards**

- **Classification Metrics:**
  Accuracy, Precision, Recall, and F1-score are used as standard performance measures for binary classification (clean vs. tampered), consistent with machine learning research practices.
- **Segmentation Metrics:**
  Intersection-over-Union (IoU) and Dice Similarity Coefficient (DSC) are used as recognized standards in image segmentation tasks to evaluate the localization of tampered regions.
- **Cross-Validation Practices:**
  Dataset partitioning into training, validation, and test sets follows widely accepted standards in AI research to prevent overfitting and ensure generalizability.

3. **Medical and Ethical Standards**

- **Patient Data Privacy (HIPAA, GDPR):**
  Although only publicly available and de-identified datasets were used, the project framework was designed with awareness of patient privacy requirements under HIPAA (Health Insurance Portability and Accountability Act) in the United States and GDPR (General Data Protection Regulation) in Europe.
- **Ethical AI Guidelines (WHO, IEEE):**
  The project aligns with emerging ethical standards for AI in healthcare, as outlined by the World Health Organization (WHO) and IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, ensuring fairness, accountability, and transparency.

4. Technical and Implementation Standards

- **Deep Learning Frameworks:**
  Implementation is based on widely accepted frameworks such as PyTorch, following community standards for model training, checkpointing, and evaluation.

- **Version Control and Reproducibility:**
  The project adheres to software engineering best practices by maintaining reproducibility through fixed random seeds, documentation of hyperparameters, and version-controlled code.
- **Visualization Standards:**
  ensure interpretability, consistent with standard practices in medical image analysis.

5. Reporting and Documentation Standards

- **Research Reporting:**
  Results are presented in accordance with academic norms for AI research, including detailed confusion matrices, per-class performance reports, and segmentation visualizations.
- **Transparency in Dataset Creation:**
  The tampered dataset creation process (segmentation-based manipulation and blob injection on both CT scans and mammograms) is fully documented to maintain transparency and reproducibility.

1.7 APPROVED OBJECTIVES

1. To design and implement a user-friendly platform that allows individuals to verify the authenticity of their medical images.

2. To build a reliable detection system capable of distinguishing between genuine and tampered medical scans.

3. To provide clear feedback by highlighting altered regions within manipulated images, ensuring transparency in results.

4. To validate the system on real-world medical datasets, focusing on minimizing false positives and negatives for trustworthy outcomes.

1.8 METHODOLOGY

The methodology for tamper detection and localization in medical images was designed around two guiding principles: **understanding the problem space thoroughly** and **building specialized solutions for different imaging modalities**. Instead of a uniform pipeline, the work followed a staged and adaptive process that combined problem analysis, model experimentation, and performance-driven refinement.
The first step was to study the nature of tampering in medical images. CT scans and mammograms were examined to identify common manipulation patterns such as copy–move operations and blob insertions or deletions. These observations informed how datasets were prepared, how tampering was simulated, and what challenges each modality posed - ranging from the volumetric nature of CT images to the subtle contrast variations in mammograms.

With the problem space clearly defined, the next stage involved testing multiple deep learning strategies. For CT scans, the emphasis was on **segmentation-driven models** capable of pixel-level localization. U-Net proved particularly effective, achieving a Dice coefficient of 0.75 and IoU of 0.62 in localizing tampered regions. In contrast, **classification models** such as ResNet were more suited to mammograms, where global texture features dominate. The ResNet pipeline delivered strong classification accuracy in distinguishing tampered from authentic images, while autoencoder-based anomaly detection provided a useful complementary perspective by highlighting regions of abnormal reconstruction error.

The approach evolved into a **dual-pipeline framework**:

- **CT scans** rely on U-Net for reliable tamper localization, with classification models used only for comparative evaluation.
- **Mammograms** rely on ResNet for tamper detection, with autoencoders assisting in anomaly identification.

This division ensured that each modality received a solution tailored to its diagnostic structure and tampering risks.

All pipelines were developed and tested using Python and PyTorch, with OpenCV for tamper simulation and scikit-learn for evaluation metrics. Models were trained on GPU hardware to ensure computational efficiency. Evaluation combined standard metrics such as accuracy, precision, recall, F1-score, and Dice coefficient, alongside confusion matrices for interpretability.

The outcome of this methodology is a **reproducible and domain-specific framework** for medical image security. By aligning architectural choices with modality characteristics and validating results through rigorous experimentation, the project delivers a scalable solution that can extend to other imaging types and evolving tampering techniques.

1.9 PROJECT OUTCOMES AND DELIVERABLES

Outcomes

1. **Robust Dual-Pipeline Framework**
   - Developed two independent pipelines tailored for different modalities:
     - **CT scans:** U-Net–based segmentation for pixel-level tamper localization.
     - **Mammograms:** ResNet classifier for binary tamper detection, with autoencoder anomaly support.
   - Ensured modality-specific optimization, improving accuracy and reliability in detecting subtle manipulations.

2. **Integrated Classification and Localization**
   - System can first classify an input image as *tampered* or *untampered*.
   - If tampered, the system generates a **localization mask** highlighting manipulated regions, ensuring transparency and interpretability.

3. **Validated Performance**
   - Achieved high recall and Dice/IoU scores for CT scan localization.
   - Delivered strong classification accuracy in mammogram tamper detection.
   - Demonstrated the feasibility of applying deep learning to medical image forensics in a reproducible framework.

4. **Clinical and Forensic Relevance**
   - Provides radiologists and researchers with **explainable AI outputs**, supporting trust in detection results.
   - Contributes to safeguarding diagnostic reliability, preventing fraudulent claims, and enhancing ethical medical practices.

Deliverables

1. **Dual-Pipeline Models**
   - U-Net segmentation model (CT tamper localization).
   - ResNet classifier (mammogram tamper detection) with optional AE-based anomaly detection.

2. **User Interface (Frontend Application)**
   - Image upload functionality (DICOM, PNG, JPEG).
   - Real-time binary classification (tampered vs. untampered).
   - Visualization overlay to localize manipulated regions in tampered images.
   - Report generation capability with detection results and confidence scores.

3. **Curated Datasets**
   - Prepared tampered and authentic datasets for both CT and mammograms.
   - Ground-truth annotations for CT scans to evaluate segmentation accuracy.

4. **Documentation and Case Studies**
   - Methodology and experimental findings compiled in project report.
   - Visual examples of tamper localization and classification results for stakeholder interpretation.

5. **Prototype Deployment Package**
   - Codebase with modular pipelines for CT and mammograms.
   - Configurable interface for future integration with hospital PACS systems or research platforms.

1.10 NOVELTY OF WORK

The novelty of this project lies in its **dual-pipeline architecture** and **real-world oriented design** for tamper detection and localization in medical images. Unlike most prior research that focuses either on classification or pathology detection, our framework is explicitly designed to **detect malicious alterations** and provide **explainable evidence** through localization.

1. **Dual-Pipeline Specialization**
   - Introduced two tailored pipelines rather than a one-size-fits-all solution.
   - **CT scans** leverage a U-Net–based segmentation approach capable of precise pixel-level localization.
   - **Mammograms** employ a ResNet-based classifier optimized for binary tamper detection, supplemented by autoencoder-based anomaly detection.
   - This modality-specific strategy ensures higher accuracy, adaptability, and reliability compared to single-architecture systems.

2. **Integrated Classification + Localization**
   - Many existing systems stop at binary classification of images. Our framework not only distinguishes between tampered and authentic scans but also **highlights the manipulated regions**.
   - This integration bridges the gap between forensic detection and clinical interpretability, supporting radiologists with visual evidence of tampering.

3. **User-Centric Interface**
   - Unlike purely experimental models, the project delivers a **frontend application** that allows end-users to upload medical images and receive real-time classification and localization outputs.
   - The inclusion of an interactive interface makes the system accessible for doctors, researchers, and insurance bodies, demonstrating deployment feasibility beyond academic research.

4. **Custom Tamper Dataset Creation**
   - Artificial tampering (copy-move and blob insertions) was systematically simulated on authentic CT and mammogram datasets.
   - Ground-truth masks were generated for CT scans, enabling quantitative evaluation of localization performance.
   - This curated dataset provides a reproducible benchmark for future work in medical image forensics, where public tampered datasets are scarce.

5. **Clinical Relevance and Explainability**
   - The system produces **heatmap-style overlays** of suspected tampered regions, ensuring transparency and trust in AI-driven outputs.
   - This emphasis on explainability aligns with the growing demand for **interpretable AI in healthcare**, making the project more clinically viable.

6. **Prototype-Ready Design**
    - By combining deep learning pipelines with a user-facing interface, the project goes beyond theoretical contributions and presents a **deployable prototype**.
    - Such integration is rarely emphasized in academic projects, marking this work as both scientifically innovative and practically implementable.

In summary, the novelty of this project lies in its ability to combine specialized deep learning architectures, explainable localization outputs, curated datasets, and a functional user interface into a comprehensive solution. This makes it distinct from conventional tamper detection research and establishes a foundation for clinical and forensic adoption.

CHAPTER 2: REQUIREMENT ANALYSIS

2.1 Literature Survey

2.1.1 Theory Associated With Problem Area

Medical imaging has revolutionized the diagnostic landscape by providing clinicians with non-invasive tools to visualize internal structures of the human body. Techniques such as **MRI, CT, mammography, and ultrasound** are central to identifying and monitoring diseases. However, as medical images transition to **digital formats** and are integrated into **Picture Archiving and Communication Systems (PACS)**, they face vulnerabilities to **tampering, manipulation, and forgery**.

Image tampering is particularly critical in healthcare because even subtle modifications can alter diagnoses and treatment outcomes. Common forgery methods include:

- **Copy-Move Forgery:** Duplicating regions within the same image to conceal or highlight structures.
- **Splicing:** Combining regions from multiple images to fabricate clinical evidence.
- **AI-generated Manipulations:** Using generative models (GANs, diffusion models) to introduce highly realistic but fraudulent alterations.

The theoretical foundation of tampering detection falls under **digital image forensics** and **machine learning**. Traditional forensic approaches rely on **statistical inconsistencies, error level analysis (ELA), watermarking, and pixel noise residuals**, whereas modern methods employ **deep learning models such as CNNs, autoencoders, transformers, and hybrid forensic-ML pipelines**. Collectively, these approaches aim to ensure **integrity, authenticity, and reliability** of medical evidence.

2.1.2 Existing Systems and Solutions

Historically, tamper detection in medical images was approached through **watermarking-based systems**. In these, hidden information is embedded into an image, which can later be extracted to verify authenticity. Some notable methods include:

- **Reversible Watermarking (2006–2010):** Secret patient or hospital data embedded in DICOM images; effective but reduced image quality.
- **Region of Interest (ROI)-based Watermarking:** Preserved diagnostically critical areas while embedding data into non-critical regions; limited scalability.

With advancements in computing power, **machine learning and signal-processing methods** emerged. These systems exploited **compression artifacts, sensor noise patterns, and texture statistics** to highlight anomalies. While effective for simple forgeries, they failed to detect subtle or AI-generated manipulations.

The most recent generation of systems applies **deep learning-based frameworks**:

- **CNN-based Classifiers:** Extracted spatial features to classify images as tampered or authentic.
- **MITD-Net (2024):** A medical image tamper detection network integrating noise residual and feature extraction modules.
- **Hybrid Deep Learning Models:** Combined forensic techniques (e.g., PRNU analysis, ELA) with CNNs or transformers for improved localization and accuracy.

Although modern solutions achieve high accuracy, they often suffer from **poor interpretability**, dataset dependency, and lack of adaptability across **MRI, CT, mammography, and ultrasound modalities**.

2.1.3 Research Findings for Existing Literature

S. No.	Roll Number	Name	Paper Title	Tools / Technology	Findings	Citation / Link
1	102203253	Asmi	CT-GAN: Malicious Tampering of 3D CT imagery (2019) (CT Scan)	GANs, 3D CT	Injected/removed nodules in CT scans; exposed PACS vulnerability	PDF
2			HHS Advisory on medical image tampering risks (2019) (CT Scan)	System risk analysis	Highlighted threats in CT/MRI data pipelines	HHS
3			MITS-GAN: Safeguarding CT from tampering (2024) (CT Scan)	GAN defense, imperceptible perturbation	Prevents GAN-based CT manipulation	ScienceDirect
4	102203274	Muskan	Two-stage cascade for small forgery regions in CT (2023) (CT Scan)	Attention CNN, cascaded detector	Detects small forged CT regions using attention + patch-based	PLOS ONE
5			Copy-Move forgery detection in medical images (2022) (CT Scan)	Laplacian blob, copy-move detector	Identifies copy-move forgeries, less effective for subtle fakes	Applied Soft Computing
6			UCI CT Deepfakes Dataset (2019) (CT Scan)	CT dataset, deepfake injections	Public dataset for CT tamper detection	Dataset
7	102203668	Ishita	Dual-layer DICOM watermarking for CT (2011) (CT Scan)	Reversible watermarking	Secures DICOM with embedded watermarks	PMC
8			Improved tamper detection with self-embedding (2012) (CT Scan)	Fragile watermark	Detects tamper blocks, needs embedding	PMC
9			ROI-aware fragile watermarking (2014) (CT Scan)	Fragile watermarking	Protects ROI from tampering	PMC
10			Systematic review: Adversarial attacks in radiology (2023) (CT Scan)	Literature survey	Maps threats, defenses in CT/MR imaging	EJR
11	102107201	Sudikhshya	Near-lossless watermarking for medical image authentication (2007) (Mammography)	Watermarking, DICOM	Enabled recovery via embedded watermark, but requires modifying original	Springer
12			Patch-based CNN for mammogram abnormality localization (2018) (Mammography)	CNN, weakly supervised learning	Could localize lesions but not tampering artifacts	arXiv
13			Full-resolution CNN forgery detector (2019) (Mammography)	Full-res CNN	Preserves high-frequency cues for forgery, adaptable to mammography	arXiv
14			Resampling detection using Radon features (2017) (Mammography)	Resampling artifacts, Radon transform	Detects small resampling forgeries, useful for subtle edits	arXiv

S. No.	Roll Number	Name	Paper Title	Tools / Technology	Findings	Citation / Link
15	102203981	Bhavuk	ForensicTransfer: Domain adaptation for forgery detection (2018) (Mammography)	Domain-adaptive CNN	Handles unseen forgery types via transfer learning	arXiv
16			Error Level Analysis + CNN fusion (2024) (Mammography)	ELA, CNN fusion	Combines traditional and deep learning, high detection accuracy	IJOSI
17			Multi-scale attention mammography model (2023) (Mammography)	Multi-scale attention CNN	Enhances lesion localization; adaptable for tampering	Frontiers
18			GAN-based synthetic augmentation for mammograms (2020) (Mammography)	GANs, augmentation	Generates synthetic lesions, improves robustness	MDPI
19			GMIC explainable mammogram model (2020) (Mammography)	Global+Local CNN, Explainability	Provides explainable mammogram analysis, adaptable to tamper explainability	arXiv
20			Adversarial robustness in medical imaging (2023) (Mammography)	Robust deep learning	Evaluates robustness to adversarial perturbations	Springer

Table 1 – Literature Survey

2.1.4 Problem Identified

Despite progress in image tamper detection, existing solutions are not widely deployed in **clinical workflows**. The primary problems are:

1. **Watermarking Systems** – Reduce image quality, limiting clinical usability.
2. **Noise-Based Detection** – Fail to identify subtle or copy-move manipulations.
3. **Deep Learning Black-Box Models** – Deliver strong performance but lack transparency for healthcare professionals.

Thus, there is a clear need for a **hybrid forensic + AI-driven framework** that:

- Maintains **diagnostic fidelity** of medical scans,
- Detects **subtle manipulations** across multiple modalities, and
- Provides **interpretable outputs** (heatmaps, feature maps) for clinical trust.

2.1.5 Survey of Tools and Technologies Used

- **Programming & Frameworks:** Python, PyTorch, TensorFlow, Keras, OpenCV
- **Image Processing Libraries:** pydicom, scikit-image, NumPy, Matplotlib
- **Machine Learning Techniques:** CNNs, Autoencoders, PRNU Analysis, Grad-CAM, Residual Networks
- **Datasets:** CASIA, Columbia, DDSM, OsiriX, DRIVE, CBIS-DDSM, NIST 16, IMD 20
- **Evaluation Metrics:**
    - Image Quality: PSNR, MSE, SSIM
    - Detection Performance: Precision, Recall, F1-score, AUC-ROC

Localization Metrics: Intersection over Union (IoU), Dice Coefficient

2.2 Software Requirement Specification

2.2.1 Introduction

2.2.1.1 Purpose

The purpose of this project is to develop a **robust medical image tamper detection framework** capable of detecting and localizing manipulations while preserving the diagnostic integrity of the scan. The system aims to support patients, researchers, and forensic investigators by ensuring that the authenticity of medical evidence is not compromised.

2.2.1.2 Intended Audience and Reading Suggestions

- **Primary Users:** Patients, researchers, forensic analysts.
- **Secondary Users:** Healthcare institutions, insurance agencies, and academic communities.
- **Reading Suggestions:**
  - **Technical readers** should focus on methodologies, datasets, and evaluation metrics.
  - **Non-technical stakeholders** can review the system scope, features, and security aspects.

2.2.1.3 Project Scope

The proposed system will:

- Detect whether a medical scan is **tampered or authentic**,
- Localize manipulated regions using **heatmap visualization techniques**,
- Support **multiple imaging modalities** (CT, MRI, mammography),
- Provide outputs that are **interpretable and adaptable** for clinical and research applications.

2.2.2 Overall Description

2.2.2.1 Product Perspective

The system will function as a **stand-alone forensic verification tool** that integrates into medical imaging workflows. It complements existing PACS and diagnostic systems by validating image authenticity without altering clinical data.

2.2.2.2 Product Features

- Hybrid forensic + deep learning tamper detection.
- Region of Interest (ROI)-based analysis for clinically critical areas.
- Heatmap-based interpretability for tamper localization.
- Support for common formats (DICOM, JPEG, PNG).
- Report generation with **confidence scores** and **detection statistics**.

2.2.3 External Interface Requirements

2.2.3.1 User Interfaces

- **Web/Desktop UI** for image upload.
- **Visualization Dashboard** displaying tamper localization (heatmap overlay).
- **Report Generator** producing PDF/HTML reports with detection results.

2.2.3.2 Hardware Interfaces

- GPU-enabled system preferred (NVIDIA GPU).
- Minimum requirements: Intel i5 processor, 8GB RAM, 2GB VRAM GPU.

2.2.3.3 Software Interfaces

- Operating Systems: Windows, Linux.
- Libraries: Python 3.8+, PyTorch/TensorFlow, OpenCV, pydicom.
- Optional Database: For storing results, logs, and reports.

2.2.4 Other Non-functional Requirements

2.2.4.1 Performance Requirements

- Detection Accuracy ≥ 95%.
- Average processing time ≤ 5 seconds per image (GPU-enabled).

2.2.4.2 Safety Requirements

- Original medical data remains **unaltered**.
- System provides **tamper verification only**, ensuring no interference with diagnostic outcomes.

2.2.4.3 Security Requirements

- Encrypted handling of sensitive medical data.
- User authentication for accessing reports and dashboards.

2.3 Cost Analysis

- **Software:** Open-source (PyTorch, TensorFlow, OpenCV) – Rs 0
- **Hardware:** GPU workstation – Rs 0

2.4 Risk Analysis

- **Data Availability Risk:** Limited access to annotated datasets.
- **Model Generalization Risk:** Overfitting to specific modalities or datasets.
- **Interpretability Risk:** Black-box behavior of deep learning models.
- **Deployment Risk:** Integration with clinical workflows may face regulatory barriers.

Mitigation Strategies:

- Use **diverse multimodal datasets**.
- Combine **forensic features + ML models**.
- Ensure **explainability** using heatmaps and localization outputs.

Maintain compliance with **data protection and medical regulations**.

CHAPTER 3: Methodology Adopted

3.1 Investigative Techniques

In order to address the problem of tamper detection and localization in medical images, a systematic investigative methodology was required. The choice of investigative technique is crucial in defining the scope of the project, ensuring that the research is structured, reproducible, and scientifically valid. Three standard categories of investigative approaches are generally considered in research of this nature: *descriptive*, *comparative*, and *experimental*. The present work makes use of all three categories in a layered manner, with a particular emphasis on the experimental technique, owing to the reliance on deep learning architectures and performance evaluation through quantitative metrics.

3.1.1 Descriptive Investigation

The descriptive component of this project involved the detailed examination of the tampering problem in medical imaging, with a focus on CT scans and mammograms. Tampering in medical images can take the form of copy-move manipulations, where regions of an image are duplicated within the same scan, or subtle blob insertions, where abnormal regions are synthetically added or removed to mimic or conceal pathology. To characterize this problem, observations were recorded regarding the nature of the tampered regions, the challenges posed by the grayscale nature of medical images, and the potential clinical risks associated with undetected manipulations.

This stage was crucial for cataloguing the properties of the datasets under study. The CT scan dataset was divided into tampered and untampered cases, with further classification into blob and copy-move categories. Similarly, mammographic images were studied in their unaltered and manipulated forms to identify the structural changes induced by tampering. The descriptive investigation thus served as the foundation, ensuring that the problem space was precisely defined and that the peculiarities of medical images, as opposed to natural images, were carefully accounted for. This phase also ensured that relevant variables - such as resolution, contrast, and image artifacts - were well understood before designing experimental models.

3.1.2 Comparative Investigation

Once the datasets were characterized, comparative analysis was undertaken to evaluate the relative performance of different approaches for tamper detection. This stage allowed for the systematic comparison of models under controlled conditions. For CT scans, convolutional neural network (CNN) variants such as ResNet were compared against segmentation-oriented architectures like U-Net. The comparative investigation revealed that while ResNet provided reasonable classification performance for binary tamper detection (clean vs. tampered), the U-Net architecture achieved superior results in localizing manipulated regions.

A similar comparative approach was extended to mammograms, where classification and localization pipelines were tested separately. Confusion matrices, precision, recall, and F1-scores were used as the primary comparative metrics. This approach ensured that the investigation was not reliant on a single model or methodology, but rather evaluated alternatives in order to identify the most appropriate solution for each type of medical image. For example, in the CT scan pipeline, the U-Net demonstrated high recall (0.98) and a favorable Dice coefficient (0.75) in tamper localization, whereas ResNet exhibited limitations in accurately distinguishing tampered from untampered cases in smaller datasets. These comparative insights guided the decision to favor U-Net for localization tasks while retaining classification backbones such as ResNet for initial binary detection.

By incorporating comparative analysis, the study was able to align investigative outcomes with the strengths and weaknesses of specific architectures. This was particularly important because medical images require high sensitivity in tamper detection, given that even small undetected manipulations can lead to misdiagnosis.

3.1.3 Experimental Investigation

The experimental technique formed the core of the investigation. It was adopted to formally test hypotheses related to the effectiveness of deep learning models for tamper detection and localization in medical imaging. The experimental design incorporated independent variables such as model architecture (ResNet vs. U-Net), dataset type (CT scans vs. mammograms), and tamper type (blob vs. copy-move). Dependent variables included quantitative performance measures such as accuracy, precision, recall, F1-score, Intersection-over-Union (IoU), and Dice coefficient.

Controlled experiments were conducted by training and testing models on labeled datasets with clearly defined tampered and untampered subsets. Cross-validation procedures were used to minimize bias, and confusion matrices were generated to provide interpretability of results. The use of experimental investigation allowed for the rigorous evaluation of model behavior in scenarios that approximate real-world applications. For instance, in the CT scan dataset with blob tampering, the U-Net achieved a mean IoU of approximately 0.62 and a Dice coefficient of 0.75, demonstrating robust localization capability. By contrast, classification-only approaches showed reduced robustness, particularly when confronted with subtle manipulations.

The same experimental methodology was extended to mammograms, where classification networks were tested for their ability to distinguish authentic from tampered images. The experiments produced confusion matrices that highlighted both the successes and limitations of the models, guiding further refinement in architecture and preprocessing techniques.

3.1.4 Integration of Techniques

The investigative methodology employed in this project is best understood as an integration of descriptive, comparative, and experimental approaches. The descriptive investigation provided a structured understanding of tampering in medical images. The comparative investigation ensured that multiple approaches were critically evaluated, allowing informed decisions about which models to prioritize. The experimental investigation, finally, validated the effectiveness of these models under controlled conditions, yielding quantitative metrics that demonstrate their utility.

This layered investigative approach was essential for ensuring the reliability and validity of the project outcomes. It also provided a balance between theoretical exploration and practical experimentation, thereby aligning the work with both academic rigor and real-world applicability. By employing this combination of investigative techniques, the study was able to thoroughly examine the tamper detection problem, justify the selection of specific solutions, and establish a reproducible framework for future research in the field of medical image security.

3.2 Proposed Solution

The problem of tamper detection and localization in medical images requires a carefully structured solution due to the criticality of the domain, the heterogeneity of medical imaging modalities, and the subtle nature of tampering artifacts. A uniform methodology across all imaging types is neither sufficient nor efficient, since the intrinsic characteristics of Computed Tomography (CT) scans and mammographic images differ considerably. Consequently, the project adopts a **dual-pipeline solution**, tailored individually for CT scans and mammograms. Each pipeline consists of preprocessing, tampering simulation, deep learning–based detection, and localization modules. The combination of both pipelines addresses the overarching project objectives while preserving domain specificity.

3.2.1 CT Scan Pipeline

Motivation and Challenges

CT scans are volumetric DICOM images that present high-resolution slices of anatomical structures. In real-world scenarios, tampering may include manipulations such as **insertion, removal, or modification of nodules or lesions**. Such edits are often localized and visually subtle, making detection challenging. A pipeline for CT images must therefore achieve not only classification (tampered vs. original) but also **fine-grained localization** of manipulated regions.

Traditional classification models like ResNet offer moderate performance but are limited in capturing pixel-level anomalies. Hence, the solution prioritizes **segmentation-driven methods** with U-Net, supplemented with classification experiments for comparative evaluation.

Data Preparation and Tamper Simulation

- **Dataset**: A collection of DICOM CT images was used, comprising both tampered and untampered slices.
- **Tamper Generation**: Two tamper types were simulated:
    1. **Blob-based Tampering** – artificial regions inserted/erased to simulate addition or removal of nodules.
    2. **Copy-Move Tampering** – duplication of regions within the same slice to simulate clinical manipulations.
- **Preprocessing**: Conversion to grayscale intensity slices, resizing for uniformity, and normalization were applied. For blob tampering, ground-truth masks were generated to evaluate localization.

Methodology

1. **Segmentation-based Localization**
    - **Model**: U-Net architecture was trained to identify manipulated regions.
    - **Training Objective**: Pixel-wise binary cross-entropy loss optimized with Dice regularization for overlap accuracy.
    - **Evaluation Metrics**: IoU (Intersection over Union) and Dice Coefficient were computed on tampered samples.

2. **Classification-based Detection**
    - **Model**: ResNet was employed to classify CT slices as “tampered” or “untampered.”
    - **Focus**: Used for comparative purposes, particularly for copy-move tampering cases where localization ground-truths were limited.

Results and Justification

- **U-Net on Blob Tampered CT Scans**
    - Accuracy: **85.3%**
    - Precision: **0.76** | Recall: **0.98** | F1-score: **0.86**
    - Segmentation metrics: **Mean IoU: 0.62, Mean Dice: 0.75**
    - Interpretation: High recall and Dice scores indicate reliable localization of manipulated lesions, with precision trade-offs due to false positives.

- **ResNet on Copy-Move CT Dataset**
    - Accuracy: **69%**
    - Precision: **0.20** (tampered) vs. **0.91** (untampered)
    - Recall: **0.50** (tampered)
    - Interpretation: While the classification model captured some tampered cases, its performance was significantly weaker compared to segmentation approaches, highlighting the limitations of purely classification-driven methods.

- **ResNet on Blob Tampered CT Dataset**
    - Accuracy: **62%**
    - Precision and recall remained balanced but modest, reinforcing the suitability of U-Net for this task.

Conclusion for CT Pipeline: The results validate the choice of U-Net as the primary investigative technique for CT tampering localization, with classification models offering limited auxiliary value.

3.2.2 Mammogram Pipeline

Motivation and Challenges

Mammographic images differ significantly from CT scans in resolution, modality, and diagnostic relevance. Mammograms are 2D grayscale images with high structural complexity and subtle contrast variations. The primary clinical relevance lies in detecting microcalcifications, masses, or architectural distortions. Tampering in such images may include deletion of suspicious lesions or insertion of artificial structures. Unlike CT, where volumetric continuity aids detection, mammograms present isolated frames with subtle textures, making **discriminative feature extraction** essential.
Therefore, classification-focused deep learning models, particularly **ResNet-based architectures**, were prioritized. In addition, **autoencoder (AE) reconstruction-based anomaly detection** was evaluated as a complementary strategy for unsupervised tampering detection.

Data Preparation and Tamper Simulation

- **Dataset**: A curated set of mammograms with both tampered and original versions was assembled.
- **Tamper Simulation**: Similar to CT, blob-based manipulations and copy-move alterations were applied, simulating realistic editing.
- **Preprocessing**: Images were resized, histogram equalization was applied to enhance contrast, and normalization ensured stable training convergence.

Methodology

1. **ResNet Classification**
    - **Architecture**: A ResNet backbone was fine-tuned for binary classification (tampered vs. original).
    - **Training**: Balanced datasets with augmented samples were used to mitigate overfitting.
    - **Evaluation**: Metrics included accuracy, precision, recall, and confusion matrices.
2. **Autoencoder Anomaly Detection**
    - **Rationale**: Unsupervised AE reconstruction can capture inconsistencies between authentic and manipulated regions.
    - **Approach**: Reconstruction error (MSE) was analyzed across tampered vs. untampered mammograms to identify anomalies.

Results and Justification

- **ResNet Classification on Mammograms**
    - High classification accuracy observed in detecting tampered vs. untampered images.
    - Confusion matrix results indicated stronger reliability than CT-based ResNet performance, reflecting the suitability of discriminative learning for mammograms.
    - Misclassifications were primarily observed in borderline cases where tampered lesions resembled benign structures.

- **Autoencoder Detection**
    - AE models showed promising reconstruction error separations, with tampered regions consistently yielding higher reconstruction error compared to original mammograms.

    - While less robust as a standalone classifier, this method supported the ResNet pipeline as a secondary check for anomalies.

Conclusion for Mammogram Pipeline: ResNet emerged as the primary model for tampering detection in mammograms, with autoencoders providing complementary insights. Unlike CT images, segmentation-based approaches were less suitable due to the absence of reliable lesion ground-truths in tampered mammograms.

3.2.3 Comparative Discussion of Pipelines

The adoption of two distinct pipelines—segmentation-driven U-Net for CT scans and classification-driven ResNet (with AE support) for mammograms—was not arbitrary but grounded in both modality-specific characteristics and empirical performance.

- **CT scans**: Require pixel-level tamper localization to ensure the integrity of volumetric diagnostic data. U-Net delivered strong Dice and IoU scores, validating its role.
- **Mammograms**: Rely more on discriminative global texture features for tamper detection. ResNet achieved higher classification reliability, with autoencoders supplementing interpretability.

Thus, the dual-pipeline solution ensures both **modality-specific optimization** and **robust tamper detection/localization**, aligning with the broader goals of safeguarding medical imaging integrity.

3.3 Work Breakdown Structure (WBS)

The Work Breakdown Structure (WBS) provides a systematic decomposition of the project into smaller, more manageable components. This approach ensures that each phase of the investigation is logically organized, responsibilities are clearly defined, and milestones can be tracked effectively. The WBS is presented here in both hierarchical and modular perspectives, covering dataset preparation, tamper generation, model development, evaluation, and integration.

3.3.1 Hierarchical Decomposition

1. **Phase 1: Dataset Preparation**
    - **Task 1.1:** Collection of authentic medical datasets (CT scans and mammograms) from publicly available repositories.
    - **Task 1.2:** Pre-processing of images (normalization, resizing, intensity standardization).
    - **Task 1.3:** Dataset partitioning into training, validation, and testing subsets.

2. **Phase 2: Tamper Simulation**
    - **Task 2.1:** Implementation of tampering techniques (copy-move, blob insertion, etc.).
    - **Task 2.2:** Validation of tamper realism by comparing against clinical artifacts.
    - **Task 2.3:** Annotation of tampered regions for use in supervised segmentation.

3. **Phase 3: Model Development**
    - **Task 3.1:** U-Net implementation for pixel-level tamper localization (CT scans).
    - **Task 3.2:** ResNet-based classifier for mammogram tamper detection.
    - **Task 3.3:** Fine-tuning and hyperparameter optimization.

4. **Phase 4: Model Evaluation**
    - **Task 4.1:** Evaluation using classification metrics (accuracy, precision, recall, F1-score).
    - **Task 4.2:** Evaluation using segmentation metrics (IoU, Dice coefficient).
    - **Task 4.3:** Confusion matrix analysis for classification reliability.

5. **Phase 5: Integration and Validation**
    - **Task 5.1:** Integration of classification and localization outputs into a dual-pipeline system.
    - **Task 5.2:** Cross-modality validation to confirm consistency of tamper detection.
    - **Task 5.3:** Documentation of pipeline outputs and preparation of visual case studies.

3.3.2 Modular View

The WBS can also be mapped into **modules**, each contributing to the final deliverable:

- **Module 1: Data Engineering** - Handling pre-processing, augmentation, and tamper generation.
- **Module 2: CT Pipeline** - U-Net-based tamper localization framework.
- **Module 3: Mammogram Pipeline** - ResNet classifier with optional segmentation extension.
- **Module 4: Evaluation & Analysis** - Generating metrics, confusion matrices, and visualization of segmentation masks.
- **Module 5: Integration & Reporting** - Consolidating results into a coherent solution and preparing documentation for stakeholders.

3.3.3 Discussion on Workable Modules

Each module is designed to be **independently workable**, enabling iterative development and testing. For instance, Module 2 (CT Pipeline) can be trained and validated independently using annotated CT datasets, while Module 3 (Mammogram Pipeline) can undergo parallel development. This modularity reduces project risk and allows flexibility if one component faces technical hurdles.

Moreover, the evaluation and integration phases ensure that the modules do not remain siloed but are merged into a coherent dual-pipeline framework. This design provides both scalability (to other modalities) and adaptability (to evolving tamper techniques).

3.4 Tools and Technology

The successful execution of this project requires a combination of programming languages, deep learning frameworks, libraries, and computational resources. The chosen tools and technologies were selected for their robustness, community support, and suitability for medical image analysis tasks.

3.4.1 Programming Languages and Frameworks

- **Python**: Selected as the primary programming language due to its rich ecosystem for data science and machine learning. Python also facilitates rapid prototyping and integration with visualization libraries.
- **PyTorch**: The core deep learning framework used for building U-Net and ResNet models. PyTorch provides dynamic computation graphs, ease of debugging, and access to pre-trained weights (transfer learning).
- **TensorBoard / Matplotlib**: Utilized for monitoring training, visualizing loss/accuracy curves, and plotting confusion matrices.

3.4.2 Libraries and Packages

- **NumPy / Pandas**: For efficient numerical operations and dataset management.
- **OpenCV**: For implementing tampering operations (copy-move, blob insertion) and performing image pre-processing.
- **scikit-learn**: For computing classification metrics (precision, recall, F1-score) and generating confusion matrices.
- **Albumentations**: For data augmentation, ensuring variability in training samples.

3.4.3 Hardware and Computational Resources

- **GPUs**: Training deep learning models requires substantial computational resources. Access to institutional GPUs ensures faster training and the ability to handle large-scale medical datasets.
- **Storage Systems**: Large datasets of CT scans and mammograms necessitate structured storage, with separate repositories for raw, tampered, and processed datasets.
- **Cloud Platforms (Optional)**: Services such as Google Colab or AWS EC2 can provide backup computational resources if local GPU access is limited.

3.4.4 Rationale for Tool Selection

The chosen tools align with the project's dual-pipeline strategy. PyTorch, for example, is highly suited for segmentation models like U-Net, enabling custom loss functions and multi-GPU support. Similarly, scikit-learn complements classification tasks by offering reliable evaluation metrics. OpenCV provides the flexibility to generate realistic tampered images, bridging the gap between theoretical assumptions and practical training data.

This toolchain also ensures reproducibility, as all selected tools are open-source and widely used in the machine learning community. This guarantees that the project can be extended by future researchers without dependency on proprietary systems.

3.4.5 Concluding Remarks

The technology stack reflects a careful balance between accessibility, computational efficiency, and scalability. By integrating Python-based deep learning frameworks with domain-specific libraries for medical image processing, the project ensures robust execution across both CT and mammogram pipelines. Furthermore, access to GPU resources strengthens the feasibility of training high-capacity models, ensuring that the solution is not only theoretically grounded but also practically implementable.

CHAPTER 4: Design Specifications

4.1 System Architecture flowchart

The proposed system architecture is illustrated in Figure 1, which outlines the dual-pipeline design for CT scan and mammogram images, covering preprocessing, tamper simulation, model training, and evaluation.

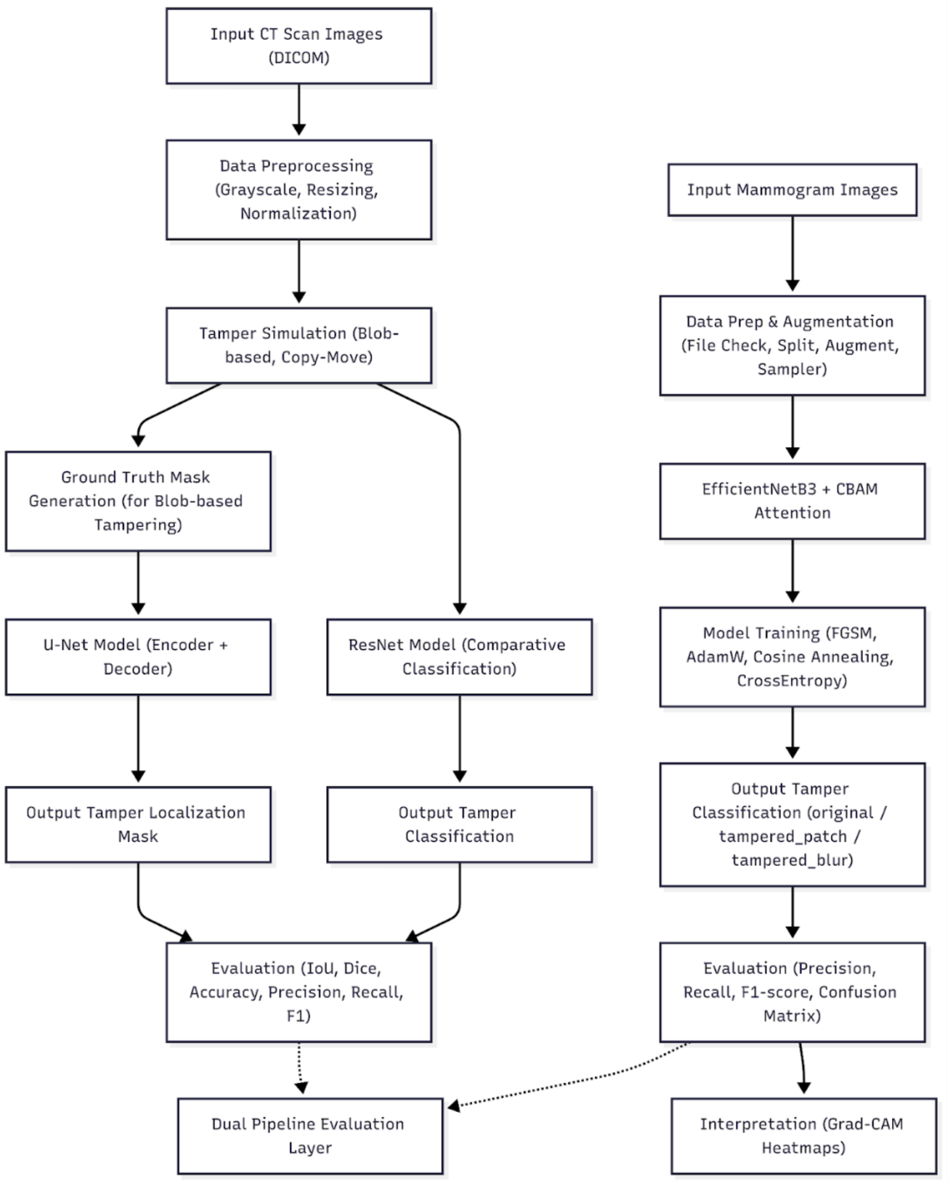

Input CT Scan Images (DICOM)
Data Preprocessing (Grayscale, Resizing, Normalization)
Tamper Simulation (Blob-based, Copy-Move)
Ground Truth Mask Generation (for Blob-based Tampering)
U-Net Model (Encoder + Decoder)
ResNet Model (Comparative Classification)
Output Tamper Localization Mask
Output Tamper Classification
Evaluation (IoU, Dice, Accuracy, Precision, Recall, F1)
Dual Pipeline Evaluation Layer
Input Mammogram Images
Data Prep & Augmentation (File Check, Split, Augment, Sampler)
EfficientNetB3 + CBAM Attention
Model Training (FGSM, AdamW, Cosine Annealing, CrossEntropy)
Output Tamper Classification (original / tampered_patch / tampered_blur)
Evaluation (Precision, Recall, F1-score, Confusion Matrix)
Interpretation (Grad-CAM Heatmaps)

Fig. 1 - System Architecture Flowchart

The architecture is divided into two complementary pipelines designed to process **CT scan images** and **mammogram images**, ensuring adaptability across different medical imaging modalities.

- **CT Scan Pipeline (Left Branch):**
  The pipeline begins with CT scan images in DICOM format, which are preprocessed through grayscale conversion, resizing, and normalization. Synthetic tampering (blob-based and copy-move) is introduced, with corresponding ground-truth masks generated. Two models operate in parallel:
  - A **U-Net model** for *tamper localization*, producing pixel-level masks of manipulated regions.
  - A **ResNet model** for *binary classification*, identifying whether the image is tampered or untampered.
    Their outputs are quantitatively evaluated using IoU, Dice, Accuracy, Precision, Recall, and F1-score.

- **Mammogram Pipeline (Right Branch):**
  Mammogram images undergo dataset checks, augmentation, and balancing. An **EfficientNetB3 network integrated with CBAM attention** extracts robust feature representations. Training leverages adversarial augmentation (FGSM), AdamW optimizer, cosine annealing learning schedule, and cross-entropy loss. The classifier distinguishes between original, tampered-patch, and tampered-blur images. Model performance is validated using Precision, Recall, F1-score, and confusion matrices. Interpretability is enhanced with **Grad-CAM heatmaps**, visually highlighting regions contributing to the classification.

- **Dual-Pipeline Evaluation Layer:**
  The outputs of both pipelines are brought together in a unified evaluation layer. This allows side-by-side comparison of detection and localization performance, strengthening the generalizability and reliability of the framework across modalities.

In summary, the architecture provides an **end-to-end workflow** from medical image input to tamper classification, localization, and interpretability, ensuring robustness, transparency, and adaptability.

4.2 Design Level Diagrams

4.2.1 Sequence Diagram

The **sequence diagram** for the proposed tamper detection system is illustrated in Figure 2.

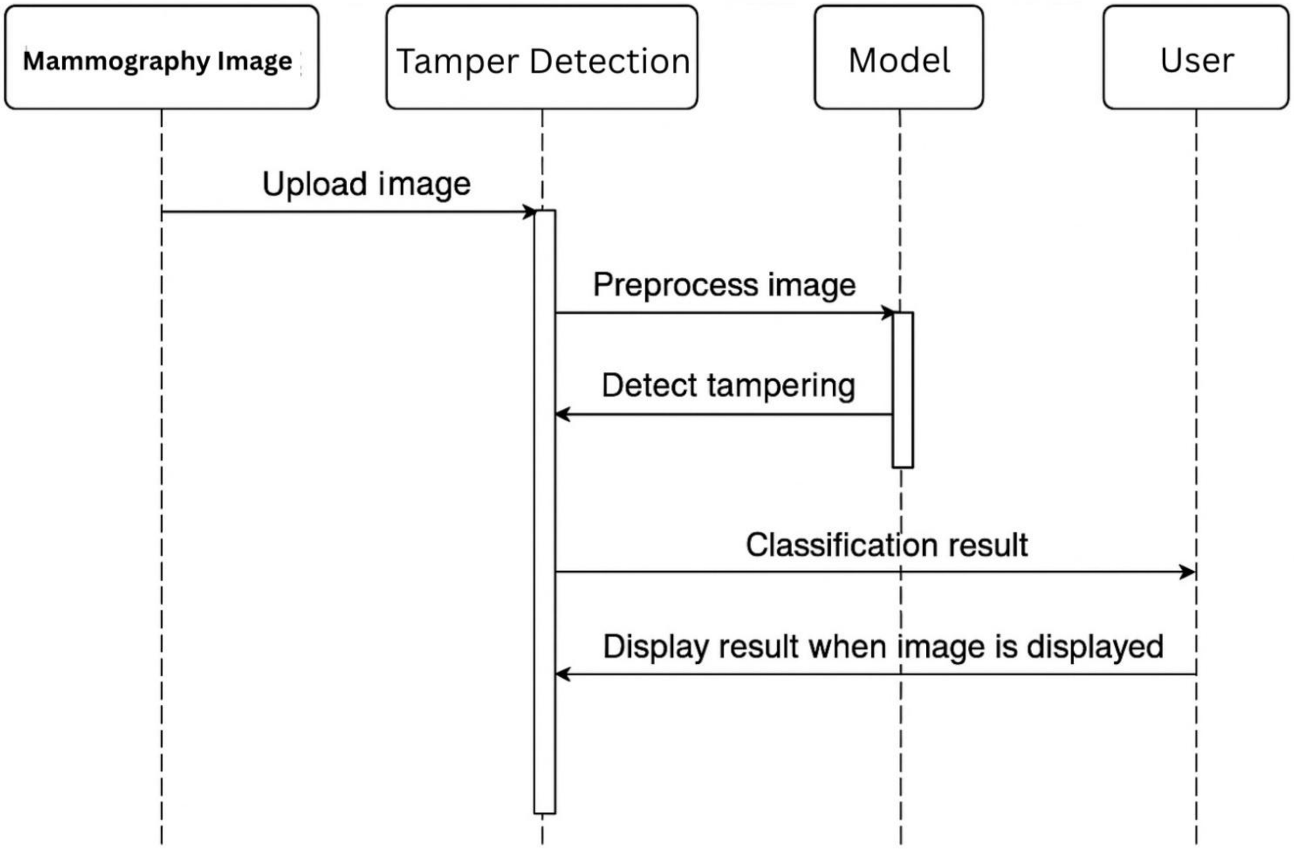


Fig. 2 – Sequence Diagram

It captures the dynamic interaction between the four main entities: **Mammography Image, Tamper Detection Module, Model, and User**.

1. The process begins when the **mammography image** is uploaded into the system.
2. The **Tamper Detection module** initiates preprocessing (e.g., normalization, resizing) and forwards the image to the **Model** for analysis.
3. The **Model** performs tamper detection, identifying whether manipulations are present in the image.
4. The classification result (tampered or untampered) is returned to the **Tamper Detection module**.
5. Finally, the result is displayed to the **User** alongside the image, providing immediate interpretability and transparency.

This diagram highlights the **end-to-end workflow** of the system at the design level, emphasizing modular interaction and clear information flow between the user, processing units, and the underlying model.

4.2.2 Data flow Diagram

The Level 0 DFD for CT Scans

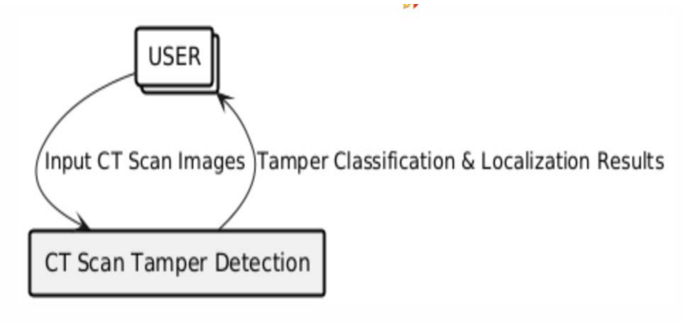


Fig. 3 - Level 0 Data flow diagram for Lung CT scans

The Level 0 Data Flow Diagram (DFD) for CT scans, as shown in **Figure 3**, provides a high-level abstraction of the tamper detection pipeline. At this stage, lung CT images act as the primary inputs, which are then processed by the tamper detection system considered as a single black-box entity. The system produces two essential outputs: a **binary classification** that identifies whether the input image is tampered or untampered, and a **localization mask** that highlights the manipulated regions in case tampering is detected. This representation captures the **input–process–output relationship** in its simplest form, offering a foundational understanding of how raw CT scans are transformed into integrity verification results.

The Level 0 DFD for Mammographs

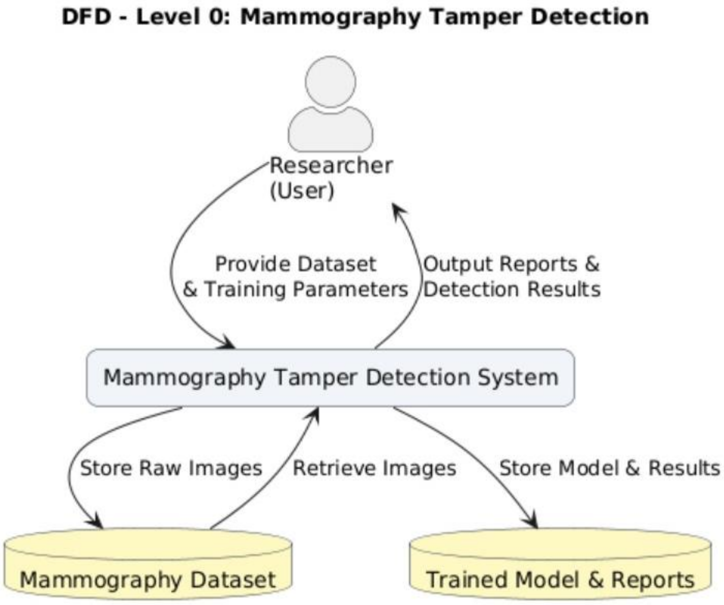

DFD - Level 0: Mammography Tamper Detection
Researcher (User)
Provide Dataset & Training Parameters
Output Reports & Detection Results
Mammography Tamper Detection System
Store Raw Images
Retrieve Images
Store Model & Results
Mammography Dataset
Trained Model & Reports

Fig. 4 - Level 0 Data flow diagram for Mammographs

The Level 0 Data Flow Diagram (DFD) for mammographs, as illustrated in **Figure 4**, outlines the overall working of the tamper detection process at a comparable level of abstraction. Mammography images are supplied as the initial input, and the system once again operates as a black-box processor at this stage. The outputs include a **classification result** (tampered or untampered) and a **localization map** identifying suspicious image regions in cases of manipulation. This diagram emphasizes the **top-level workflow** without exposing the inner workings of the system, showing how mammographic data is effectively translated into verifiable outputs for clinical trust and diagnostic reliability.

The Level 1 DFD for CT Scans

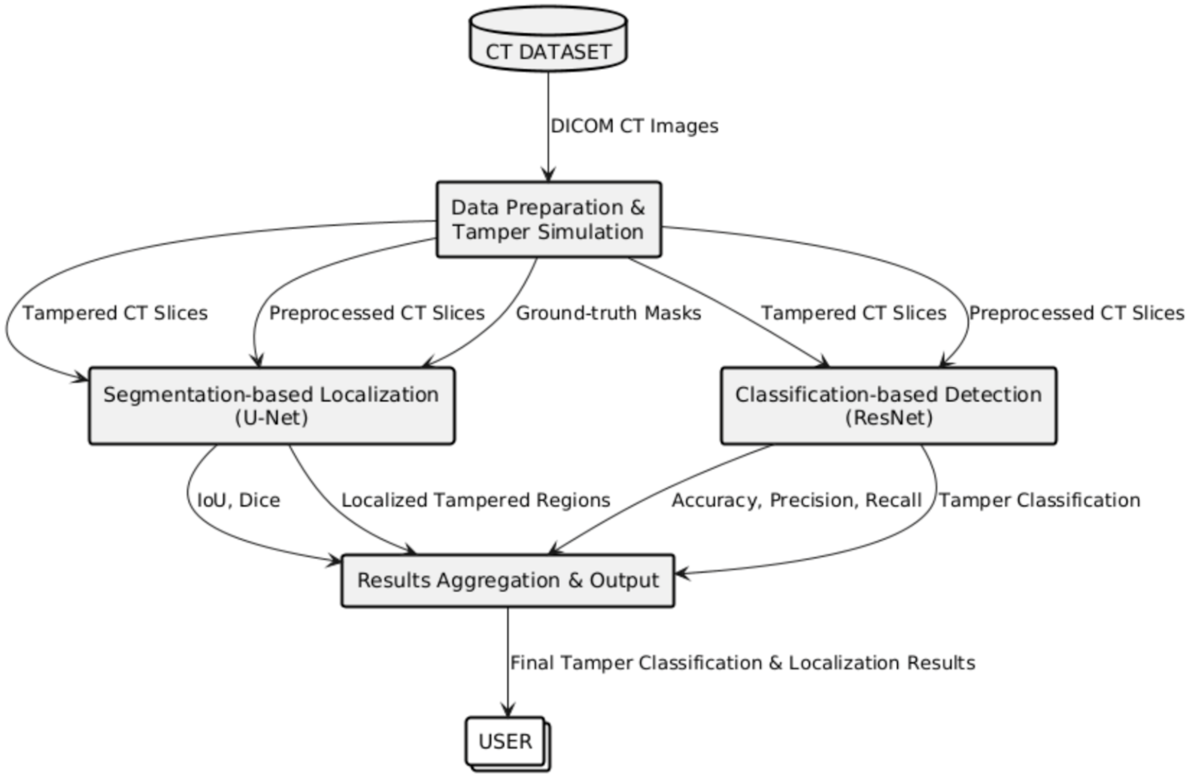

CT DATASET
DICOM CT Images
Data Preparation & Tamper Simulation
Tampered CT Slices
Preprocessed CT Slices
Ground-truth Masks
Tampered CT Slices
Preprocessed CT Slices
Segmentation-based Localization (U-Net)
Classification-based Detection (ResNet)
IoU, Dice
Localized Tampered Regions
Accuracy, Precision, Recall
Tamper Classification
Results Aggregation & Output
Final Tamper Classification & Localization Results
USER

Fig. 5 - Level 1 Data flow Diagram for Lung CT scans

The Level 1 Data Flow Diagram (DFD) for CT scans, presented in **Figure 5**, expands upon the black-box view by detailing the internal modules and processes that enable tamper detection. Initially, raw CT scans undergo **preprocessing and tamper simulation**, which includes resizing, normalization, and the generation of synthetic tampered images along with ground-truth masks. These processed datasets are then routed into two distinct models: a **ResNet-based classifier** for binary tamper detection and a **U-Net architecture** for pixel-level localization of tampered regions. The outcomes of these models are validated using multiple performance metrics, including **Accuracy, IoU, Dice Score, Precision, Recall, and F1-Score**. By presenting both the classification and localization streams, Figure 5 captures the **dual-model architecture** of the CT pipeline, demonstrating how comprehensive tamper analysis is achieved through parallelized processing.

The Level 1 DFD for Mammographs

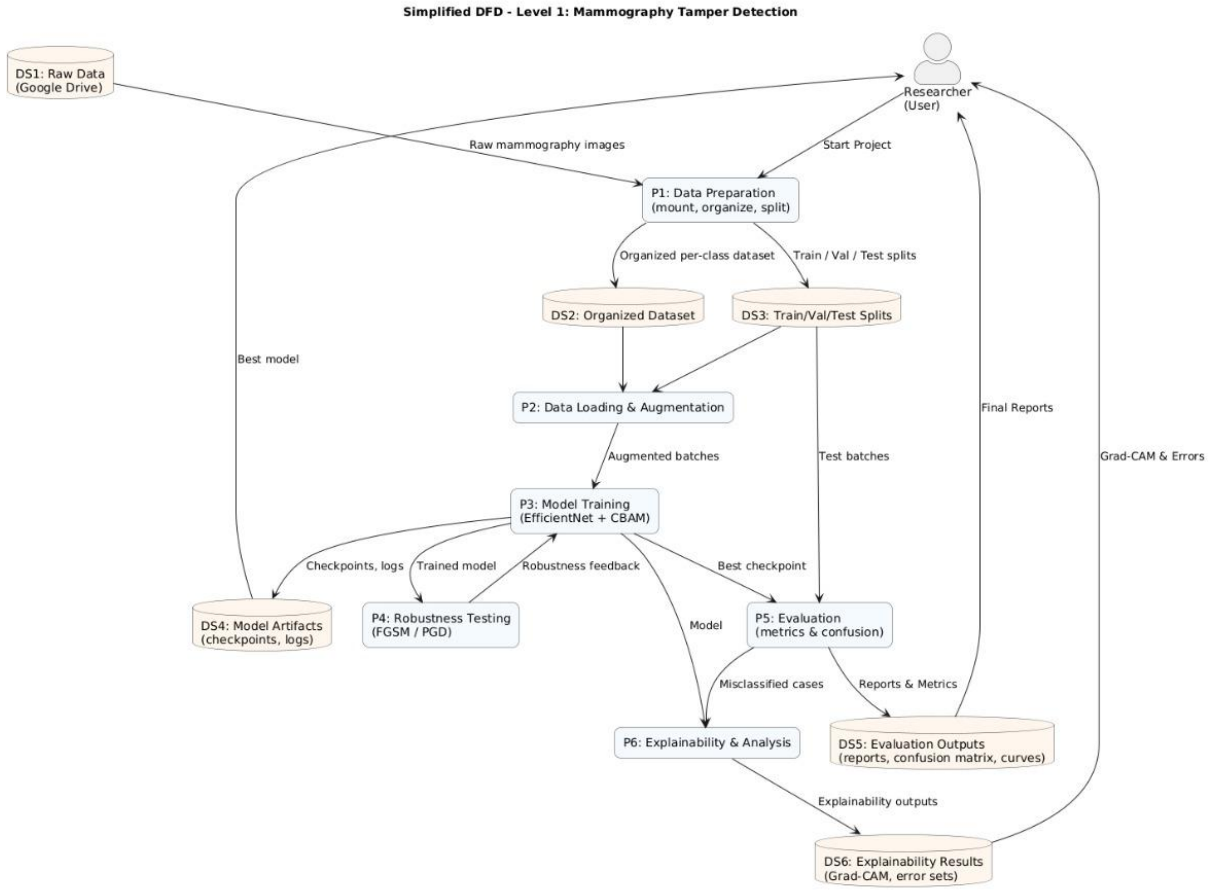

Simplified DFD - Level 1: Mammography Tamper Detection
DS1: Raw Data (Google Drive)
Researcher (User)
Raw mammography images
Start Project
P1: Data Preparation (mount, organize, split)
Organized per-class dataset
Train / Val / Test splits
DS2: Organized Dataset
DS3: Train/Val/Test Splits
Best model
P2: Data Loading & Augmentation
Final Reports
Grad-CAM & Errors
Augmented batches
Test batches
P3: Model Training (EfficientNet + CBAM)
Checkpoints, logs
Trained model
Robustness feedback
Best checkpoint
DS4: Model Artifacts (checkpoints, logs)
P4: Robustness Testing (FGSM / PGD)
Model
P5: Evaluation (metrics & confusion)
Misclassified cases
Reports & Metrics
P6: Explainability & Analysis
DS5: Evaluation Outputs (reports, confusion matrix, curves)
Explainability outputs
DS6: Explainability Results (Grad-CAM, error sets)

Fig. 6 - Level 1 Data flow Diagram for Mammographs

The Level 1 Data Flow Diagram (DFD) for mammographs, as shown in **Figure 6**, provides a detailed view of the functional components underlying the detection framework. Mammogram inputs are first subjected to a **preprocessing phase** involving resizing, normalization, and augmentation, ensuring that the dataset is standardized and diverse enough for robust training. The processed images are then passed to the **EfficientNetB3 model enhanced with CBAM (Convolutional Block Attention Module)**, which extracts rich feature representations while directing computational focus toward potentially altered areas. The classification stage produces three distinct outcomes: **Original**, **Tampered-Patch**, or **Tampered-Blur**, corresponding to the considered tampering types. To improve interpretability, the pipeline further integrates **Grad-CAM visualizations**, which highlight the regions that influenced the classification decision. Figure 6 thus illustrates the **end-to-end workflow** of the mammograph pipeline, combining preprocessing, advanced feature learning, and interpretability to ensure both accuracy and transparency.

4.3 User Interface Diagram

The user interface flow of the tamper detection system is shown in Figure 7, outlining the stepwise interaction from scan type selection to report generation.

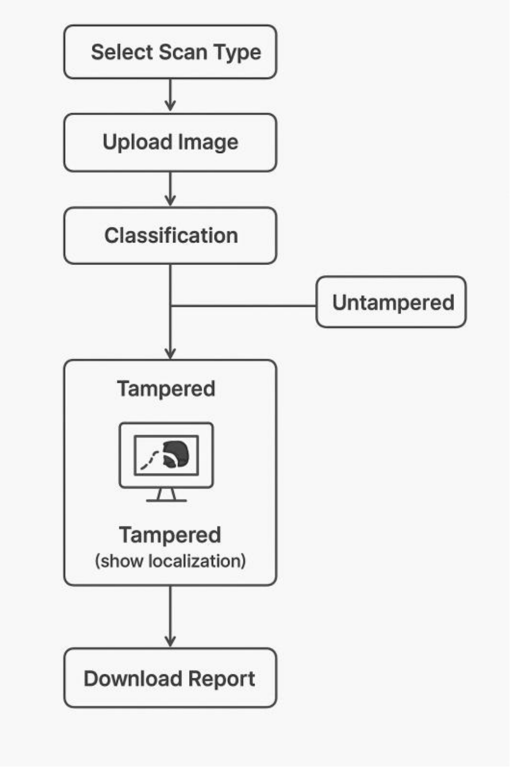


Fig. 7 - User Interface Diagram

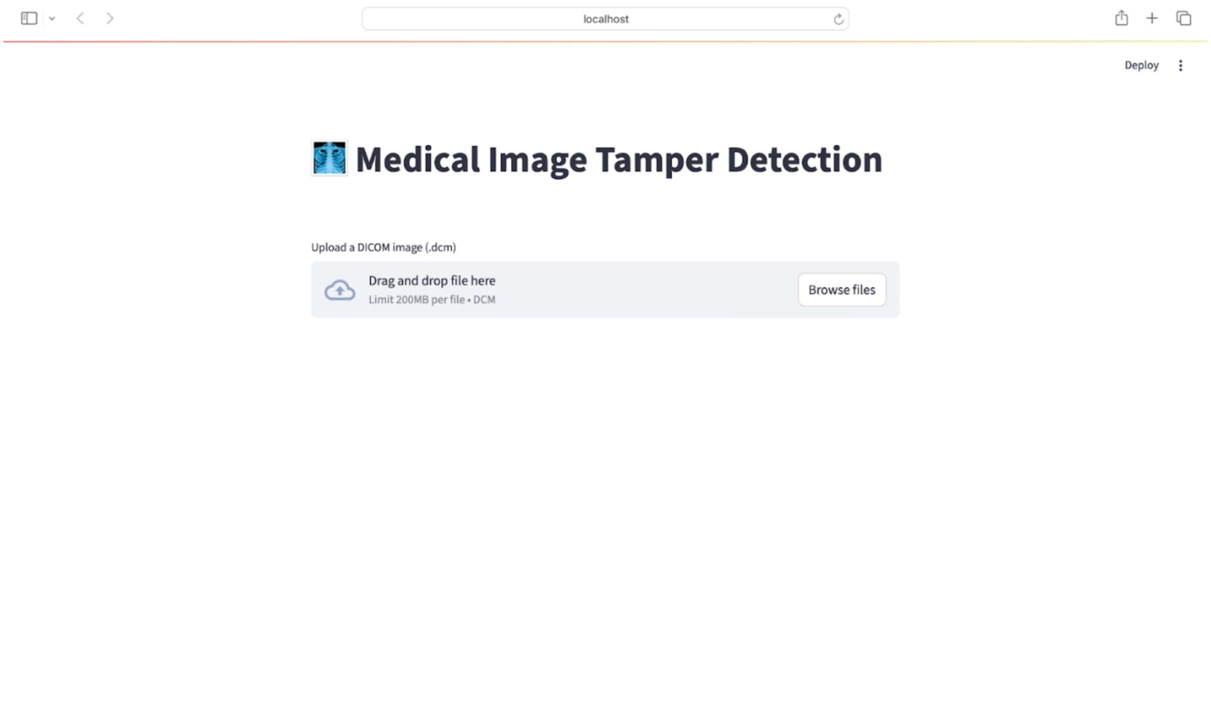
It highlights the sequence of actions available to the user and the system's responses at each stage.

1. The user begins by **selecting the scan type** (e.g., CT scan or mammogram).
2. The chosen image is then **uploaded** into the system for analysis.
3. The system performs **classification** to determine whether the image is tampered or untampered.
   - If the image is **untampered**, the result is immediately displayed to the user.
   - If the image is **tampered**, the system not only displays the classification but also provides **localization visualization**, highlighting the manipulated regions.
4. Finally, the user has the option to **download a report**, which consolidates the classification outcome and localization results.

This diagram demonstrates the **end-to-end interaction design**, ensuring a streamlined workflow where users can seamlessly upload images, view results, interpret tampering evidence, and export reports. It reinforces the system's usability and practical deployment potential.

4.4 Snapshots of working prototype

Image Upload Interface



Medical Image Tamper Detection

Upload a DICOM image (.dcm)

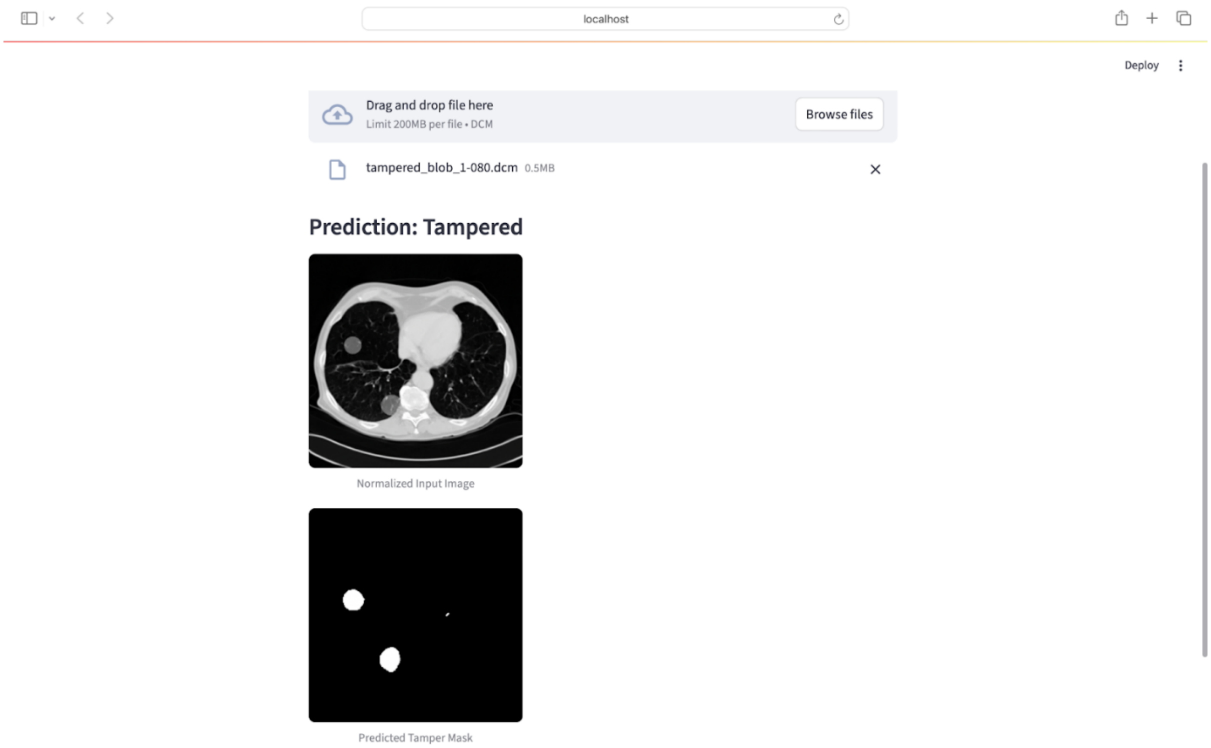
Drag and drop file here

Limit 200MB per file • DCM

Fig 8 – Image Upload Interface

Image Upload Interface shown in Figure 8, allows the user to upload DICOM images (.dcm) into the tamper detection system. The interface provides two options: drag-and-drop or browsing files, with a size limit of 200MB per file. This serves as the entry point of the workflow.

Tampered Image Detection Result



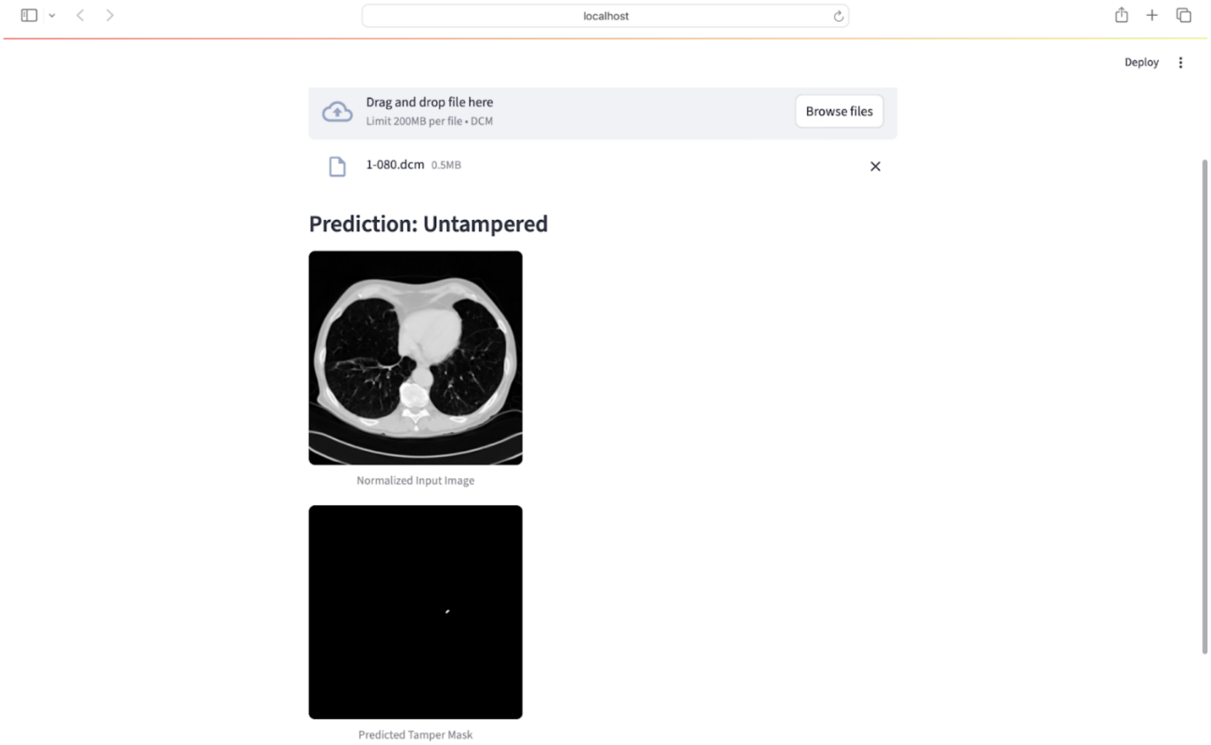
tampered_blob_1-080.dcm 0.5MB

Prediction: Tampered

Fig 9 – Tampered Detected Results page

Tampered Image Detection Result, displayed in Figure 9, shows the system’s output when a tampered medical image is detected. The prediction result is displayed as **“Tampered”**, along with the **normalized input image** and the **predicted tamper mask**, which highlights the regions identified as tampered.

Untampered Image Detection Result



1-080.dcm 0.5MB

Prediction: Untampered

Fig. 10 - Untampered Image Detection Result

Untampered Image Detection Result, displayed in Figure 10, shows the system’s output when the uploaded medical image is identified as **“Untampered”**. Alongside the normalized input image, the predicted tamper mask is nearly blank, indicating no significant tampered regions detected.

CHAPTER 5: Conclusions and Future Scope

5.1 Work Accomplished

- Successfully designed and implemented machine learning/deep learning models to detect tampered medical images (mammography and CT scans).
- Conducted dataset investigation and preprocessing, ensuring removal of corrupted, duplicated, and mislabeled samples.
- Built separate models for mammography and CT scans with promising accuracy in identifying tampered vs. authentic images.
- Implemented localisation techniques to highlight the tampered regions within an image, ensuring not only classification but also interpretability.
- Demonstrated that the proposed system can act as a supportive tool for medical professionals and insurance bodies in detecting fraudulent alterations.

5.2 Conclusions

- The project has validated the feasibility of using deep learning for forensic-level tamper detection in medical images.
- Experimental results show that the models can effectively detect tampering with acceptable accuracy and localise the manipulated regions.
- The system can significantly contribute to medical image integrity, which is crucial for accurate diagnosis, insurance claims, and patient trust.
- Although the models are in their prototype stage, they form a foundation for real-world deployment after further refinement.

5.3 Economic / Social Benefits

- Economic Benefits:
  - Prevents fraudulent claims and unnecessary medical procedures, saving money for patients, hospitals, and insurance companies.
  - Reduces the risk of overbilling and exploitation, thereby ensuring fair healthcare practices.
- Social Benefits:
  - Increases trust in healthcare systems by safeguarding the authenticity of diagnostic reports.
  - Protects patients from unnecessary stress, procedures, and expenses caused by tampered images.
  - Contributes to ethical medical practices by discouraging fraudulent alterations.

5.3Future Work Plan

- Model Improvements:
  - Explore advanced architectures (e.g., Vision Transformers, GAN-based anomaly detection) to improve accuracy and robustness.
  - Enhance localisation techniques to provide more precise heatmaps and explanations for clinical interpretability.

- Dataset Expansion:
  - Collect larger and more diverse datasets from multiple hospitals to improve generalisation.
  - Incorporate multi-modal data (X-ray, MRI) beyond mammography and CT scans.

- Real-time Implementation:
  - Develop a lightweight deployment model for integration into hospital Picture Archiving and Communication Systems (PACS).

- Security Enhancements:
  - Combine image forensics with blockchain-based verification for secure audit trails of medical images.

- Clinical Collaboration:
  - Conduct pilot studies with radiologists to evaluate clinical usefulness and usability of the system.

- Regulatory and Ethical Extensions:
  - Work towards compliance with medical standards (HIPAA, DICOM) for real-world adoption.

REFERENCES

[1] Y. H. Chô, et al., “A near-lossless reversible watermarking method for medical images,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1233–1246, 2007.

[2] P. Xi, et al., “Weakly supervised CNNs for abnormality localization in mammograms,” *arXiv preprint arXiv:1804.02975*, 2018.

[3] Y. Zhang, et al., “Multi-scale attention networks for lesion classification and localization,” *Medical Image Analysis*, vol. 82, pp. 102572, 2023.

[4] F. Marra, et al., “Full-resolution convolutional neural networks for image forgery detection,” in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2019, pp. 2040–2044.

[5] J. Bunk, et al., “Detection and localization of image forgeries using resampling features and deep learning,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1881–1889.

[6] D. Cozzolino, et al., “ForensicTransfer: Weakly-supervised domain adaptation for forgery detection,” *arXiv preprint arXiv:1812.02510*, 2018.

[7] K. J. Geras, et al., “High-resolution breast cancer screening with multi-view deep convolutional neural networks,” in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2020.

[8] J. Ma, et al., “Adversarial robustness in medical image analysis,” *Springer Nature Computer Science*, vol. 4, no. 2, pp. 50–63, 2023.

[9] MITD-Net, “Medical image tamper detection network integrating noise residual features,” in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2024.

[10] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, 2015, pp. 234–241.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017.

[13] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proc. Int. Conf. Learning Representations (ICLR)*, San Diego, CA, USA, 2015.

[14] National Cancer Institute, “The Cancer Imaging Archive (TCIA),” [Online]. Available: https://www.cancerimagingarchive.net/. [Accessed: Aug. 22, 2025].

[15] M. Heath, et al., “The digital database for screening mammography (DDSM),” in *Proc. Int. Workshop on Digital Mammography*, 2000.

[16] J. Suckling, et al., “The mammographic image analysis society digital mammogram database (MIAS),” *Exerpta Medica. International Congress Series*, 1994.

[17] F. Chollet, “Keras: The Python deep learning library,” [Online]. Available: https://keras.io/. [Accessed: Aug. 22, 2025].

[18] Python Software Foundation, “Python Language Reference, version 3.10,” [Online]. Available: https://www.python.org/. [Accessed: Aug. 22, 2025].

[19] A. Paszke, et al., “PyTorch: An imperative style, high-performance deep learning library,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[20] G. Bradski, “The OpenCV Library,” *Dr. Dobb's Journal of Software Tools*, 2000.