Democratic and Popular Republic of Algeria

Ministry of Higher Education and Scientific Research

University Mohamed Khider of Biskra

Faculty of Exact Sciences and Science of Nature and Life

Department of Computer Science

Ref: ....

Thesis Presented to Obtain the Degree of

Doctorate in Computer Science

Option: Image and Artificial Life

Entitled:

Deep Learning-Based Data Hiding Technique for Medical Images

Presented by:

Hadjer Saidi

Publicly defended on: 26/06/ 2025

In front of the Jury Committee composed of:

Foudil Cherif	Professeur	President	University of Biskra
Okba Tibermacine	MCA	Supervisor	University of Biskra
Abdelhakim Cheriet	MCA	Examiner	ENSIA, Alger
Bilal Mokhtari	MCA	Examiner	University of Biskra
Abd El Mouméne Zerari	MCA	Examiner	University of Biskra
Ahmed Elhadad	MCA	Co-supervisor	South Valley University, Egypt

Acknowledgement

First and foremost, I express my sincere gratitude to **ALLAH** for endowing me with the strength, courage, and well-being to complete this work. I am profoundly thankful for his blessings, which have sustained me from the inception of this endeavor to its fruition, empowering me to overcome obstacles and achieve this modest success.

I wish to express my gratitude to my thesis supervisor, Dr **Okba Tibermacine**, for his invaluable guidance, motivation, expert knowledge, and constant availability, which enabled me to overcome the challenges throughout this academic journey. It has been a true honor to work under his supervision.

I would also like to express my sincere gratitude to my co-supervisor, Dr **Ahmed Elhadad**, for his outstanding guidance and essential support. His deep expertise has significantly enhanced my work by offering vital insights and broadening the scope of my research.

I also wish to express my sincere gratitude to the members of the jury committee — **Pr Foudil Cherif**, **Dr Abdelhakim Cheriet**, **Dr Bilal Mokhtari**, and **Dr Abd El Mouméne Zerari** — for honoring me with their valued presence.

I wish to express my deep gratitude to my father, **Lahcen**, and my mother, **Karima**, who have always supported, encouraged, and assisted me throughout my studies. They provided every opportunity for my success, firmly believing in my abilities. No words can fully capture the extent of my gratitude.

I would also like to extend my heartfelt thanks and gratitude to my rock **Houssam** , whose steadfast encouragement and assistance have made preparing this work much easier.

I would also like to convey my sincere gratitude to my family, my husband's family, my close friends, and all my colleagues at the LESIA laboratory.

Hadjer Saidi

Dedication

I dedicate this thesis to the memory of those who have been taken from us by death since the inception of this work: my uncle **Riad**, my aunt **Rachida**, and my aunt's husband **Hussein**. May God have mercy on them.

To the source of love, my dear parents **Lahcen** and **Karima**.
To my wonderful brother and sisters: **Mossab**, **Hiba**, **Houda**, and **Hadil**.
To my little family: My husband **Houssam**, my leader **Zakaria**, and my two moons, **Maisson** and **Nadine**.
To my close friends: **Amina Tindert**, **Wafa Oudainia** , **Imen Khaldi**, **Hadjer Tahraoui**, and **Meriem Mezerdi** .
To all who have shared my joy.

Hadjer Saidi

Abstract

This thesis introduces novel deep learning-based frameworks for secure medical image steganography, addressing the critical challenge of protecting sensitive patient information while preserving diagnostic quality. Medical images present unique security requirements due to their dual nature: containing both critical visual data for diagnosis and sensitive patient metadata, which conventional steganographic techniques do not adequately address.

Our work presents two complementary approaches to this challenge. First, we develop a Mask-RCNN based framework that intelligently identifies diagnostically insignificant regions within medical images for strategic data embedding. By combining this region-aware detection with Discrete Cosine Transform (DCT) embedding in the frequency domain, our method achieves remarkable imperceptibility with Peak Signal-to-Noise Ratio (PSNR) values exceeding 115 dB while maintaining high payload capacity.

Second, we propose a clinical quality-aware convolutional neural network architecture that leverages an encoder-decoder framework for end-to-end steganography. This approach employs parallel processing paths with scaled residual learning to embed secret medical images within cover images while preserving diagnostic features. Extensive experimentation across multiple medical imaging modalities (CT, MRI) from datasets including MIDRC-RICORD-1B, and IQ-OTH/NCCD, demonstrates that our method achieves a good trade-off between payload and imperceptibility.

The frameworks developed in this thesis achieve an optimal balance between embedding capacity, imperceptibility, and robust secret recovery, providing healthcare institutions with effective tools for safeguarding patient privacy while maintaining the integrity of medical diagnostics. This work significantly advances the field of medical image security and establishes a foundation for future innovations in secure healthcare information systems.

Keywords: Medical Image Steganography, Deep Learning, DICOM, Mask-RCNN, DCT, Encoder-Decoder, Data Security.

Résumé

Cette thèse introduit de nouveaux Architectures basés sur l'apprentissage profond pour la stéganographie sécurisée des images médicales, en répondant au défi crucial de la protection des informations sensibles des patients tout en préservant la qualité diagnostique. Les images médicales présentent des exigences de sécurité uniques en raison de leur double nature : elles contiennent à la fois des données visuelles essentielles au diagnostic et des métadonnées sensibles relatives aux patients, que les techniques stéganographiques conventionnelles ne prennent pas en charge de manière adéquate.

Notre travail propose deux approches complémentaires pour relever ce défi. Tout d'abord, nous développons une architectures basé sur Mask R-CNN permettant d'identifier intelligemment les régions diagnostiquement non significatives des images médicales afin d'y intégrer stratégiquement des données. En combinant cette détection orientée région avec l'insertion des données dans le domaine fréquentiel à l'aide de la Transformée en Cosinus Discrète (DCT), notre méthode atteint une imperceptibilité remarquable avec des valeurs de Peak Signal-to-Noise Ratio (PSNR) dépassant 115 dB, tout en conservant une capacité d'insertion élevée.

Ensuite, nous proposons une architecture de réseau neuronal convolutif tenant compte de la qualité clinique, exploitant un cadre encodeur-décodeur pour une stéganographie de bout en bout. Cette approche met en œuvre des chemins de traitement parallèles avec un apprentissage résiduel à échelle adaptée, permettant d'intégrer des images médicales secrètes au sein d'images de couverture tout en préservant les caractéristiques diagnostiques. Des expérimentations approfondies sur plusieurs modalités d'imagerie médicale (CT, IRM) à partir de bases de données telles que MIDRC-RICORD-1B et IQ-OTH/NCCD démontre que notre méthode offre un bon compromis entre la capacité de charge utile et l'imperceptibilité.

Les cadres développés dans cette thèse atteignent un équilibre optimal entre capacité d'insertion, imperceptibilité et récupération robuste des données secrètes, offrant ainsi aux établissements de santé des outils efficaces pour la protection de la confidentialité des patients tout en maintenant l'intégrité du diagnostic médical. Ce travail constitue une avancée significative dans le domaine de la sécurité des images médicales et pose les bases de futures innovations dans les systèmes d'information sécurisés pour le secteur de la santé.

Mots-clés: Stéganographie d'images médicales, Apprentissage profond, DICOM, Mask-RCNN, DCT, Encodeur-Décodeur, Sécurité des données.

ملخص

قدمت هذه الأطروحة نهجًا جديدًا قائمًا على التعلم العميق للإخفاء السري للصور الطبية، مما يعالج التحدي المتمثل في حماية معلومات المرضى الحساسة مع الحفاظ على جودة التشخيص. تقدم الصور الطبية متطلبات أمنية فريدة بسبب طبيعتها المزدوجة—التي تحتوي على بيانات مرئية ضرورية للتشخيص وبيانات وصفية حساسة للمرضى—والتي تفشل تقنيات الإخفاء التقليدية في معالجتها بشكل كافٍ.

يقدم بحثنا نهجين متكاملين لهذا التحدي. أولاً، طورنا إطارًا يعتمد على Mask-RCNN يحدد بذكاء المناطق غير المهمة تشخيصياً داخل الصور الطبية لإدماج البيانات بشكل استراتيجي. من خلال الجمع بين هذا الكشف القائم على المناطق مع تضمين تحويل جيب التمام المتقطع (DCT) في مجال التردد، تحقق طريقتنا إخفاءً ملحوظًا مع قيم نسبة الإشارة إلى الضوضاء (PSNR) تتجاوز 511 ديسيبل مع الحفاظ على سعة حمولة عالية.

ثانيًا، نقترح بنية شبكة عصبية تلافيفية تراعي الجودة السريرية تستفيد من إطار المشفر-فك التشفير للإخفاء من البداية إلى النهاية. يستخدم هذا النهج مسارات معالجة متوازية مع تعلم متبقي مقياس لتضمين الصور الطبية السرية داخل صور الغلاف مع الحفاظ على ميزات التشخيص. تُظهر التجارب المكثفة عبر طرق تصوير طبية متعددة (التصوير المقطعي، التصوير بالرنين المغناطيسي) من مجموعات بيانات تشمل MIDRC-RICORD-1B و IQ-OTH/NCCD أن طريقتنا تحقق توازناً جيداً بين سعة الإخفاء وعدم القابلية للاكتشاف.

تحقق الأطر المطورة في هذه الأطروحة توازنًا مثاليًا بين سعة التضمين، والإخفاء، واسترجاع السر بقوة، مما يوفر للمؤسسات الصحية أدوات فعالة لحماية خصوصية المرضى مع الحفاظ على سلامة التشخيص الطبي. يعزز هذا العمل بشكل كبير مجال أمن الصور الطبية ويضع أساسًا للابتكارات المستقبلية في أنظمة معلومات الرعاية الصحية الآمنة.

الكلمات المفتاحية: إخفاء البيانات في الصور الطبية، التعلم العميق، DICOM، Mask-RCNN، DCT، الترميز-الفك، أمان البيانات.

Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

ACGAN:	Auxiliary Classifier Generative Adversarial Network
CNN:	Convolutional Neural Network
DCGAN:	Deep Convolutional Generative Adversarial Network
DL:	Deep Learning
DNN:	Deep Neural Network
DCT:	Discrete Cosine Transform
DFT:	Discrete Fourier Transform
DICOM:	Digital Imaging and Communications in Medicine
DWT:	Discrete Wavelet Transform
EfficientNet:	A family of CNN architectures optimized for efficiency
GANs:	Generative Adversarial Networks
HIPAA:	Health Insurance Portability and Accountability Act
HiDDeN:	High-Capacity Deep Neural Networks for Image Steganography
IWT:	Integer Wavelet Transform
IDCT2:	Inverse Two-Dimensional Discrete Cosine Transform

LSB:	Least Significant Bit
MAE:	Mean Absolute Error
NCC:	Normalized Cross-Correlation
NROI:	Non-Region of Interest
OPAP:	Optimum Pixel Adjustment Process
PSNR:	Peak Signal-to-Noise Ratio
QR Code:	Quick Response Code
ResNet:	Residual Neural Network
RGB:	Red, Green, Blue
ROI:	Region of Interest
STC:	Syndrome-Trellis Codes
VGG16:	16-layer Convolutional Neural Network
WGAN:	Wasserstein Generative Adversarial Network

List of Publications

International Journal Publication

1. Saidi, H., Tibermacine, O., & Elhadad, A. (2024). High-capacity data hiding for medical images based on the mask-RCNN model. Scientific Reports, 14, 7166. https://doi.org/10.1038/s41598-024-55639-9

International conference paper

2. Saidi, H., Tibermacine, O., & Elhadad, A. (2025). Privacy-Preserving Medical Image Steganography: A Clinical Quality-Aware Deep Learn- ing Framework. The 7th International Conference on Pattern Anal- ysis and Intelligent Systems.

Workshop

3. Saidi, H., Tibermacine, O., & Elhadad, A. (2024). Data Hiding in Medical Image using Deep Learning and LSB Steganography. LESIA Laboratory, University of Biskra.

General Introduction

Chapter 1

Introduction

1 Context and Motivation

The digitization of healthcare has transformed medical imaging into an essential component of modern diagnostics, treatment planning, and research. Medical images, such as MRI, CT, and X-rays, contain not only visual diagnostic information, but also sensitive patient metadata embedded within the DICOM (Digital Imaging and Communications in Medicine) format. These images constitute a substantial component of electronic health records and are frequently transmitted across healthcare networks, creating complex data assets that require comprehensive protection against data breaches, unauthorized access, and cybersecurity attacks.

The healthcare sector faces unique security challenges due to the sensitive nature of patient information and the strict regulatory frameworks governing its protection. Current regulatory frameworks, such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), impose strict requirements for protecting medical data. However, conventional security measures such as cryptography primarily focus on securing data in transmission or storage, leaving it vulnerable once decrypted. In contrast, steganography—the practice of hiding information within digital media—offers a promising solution by embedding patient data directly within medical images, ensuring that security remains intact throughout the data lifecycle.

DICOM, the standard format for medical imaging widely used in hospitals, clinics, and research institutions, presents unique challenges for steganography that conventional approaches fail to address:

- **Sensitive Metadata Protection:** Unlike conventional image formats (JPEG, PNG), DICOM files store both pixel data and patient metadata. Securing only the image portion is insufficient—a robust approach must also safeguard metadata.

- **High Diagnostic Sensitivity:** Any alteration in a DICOM image must not compromise diagnostic accuracy. Even minor distortions can lead to misinterpretations by radiologists or medical professionals.

- **Compression and Format Constraints:** DICOM images may undergo lossy or lossless compression, requiring a steganographic method that remains robust across different storage and transmission scenarios.

Traditional steganographic techniques like Least Significant Bit (LSB) substitution and transform domain methods face significant limitations when applied to medical images, particularly in balancing three critical factors: embedding capacity, imperceptibility, and robustness. These methods typically embed data uniformly across the image, potentially modifying critical diagnostic regions and affecting clinical decision-making.

Recent advancements in deep learning have revolutionized medical imaging, creating unprecedented opportunities to enhance steganographic techniques. Deep learning offers several advantages over conventional approaches:

- **Feature-Aware Data Embedding:** Deep neural networks, particularly Convolutional Neural Networks (CNNs), can automatically learn the most suitable regions for embedding hidden data while minimizing perceptual distortions.

- **Context-Aware Steganography:** Mask Region-based Convolutional Neural Networks (Mask-RCNN) can intelligently segment non-diagnostic regions within medical images, ensuring that hidden data does not interfere with critical areas.

- **End-to-End Optimization:** Deep learning architectures such as encoder-decoder networks can dynamically optimize the balance between imperceptibility, robustness, and embedding capacity without manual parameter tuning.

By leveraging these deep learning capabilities, we aim to develop advanced steganographic techniques that address the challenges of medical image security while preserving diagnostic integrity.

2 Problem Statement

Despite the promising potential of medical image steganography, several critical challenges remain unresolved in current implementations. Traditional steganographic techniques often suffer from a fundamental tradeoff between embedding capacity, imperceptibility, and robustness. In medical contexts, this tradeoff becomes even more critical, as any degradation in image quality could potentially affect diagnostic accuracy, with severe consequences for patient care.

Traditional techniques like Least Significant Bit (LSB) substitution and transform domain methods offer limited embedding capacity and often introduce perceptible distortions when attempting to embed substantial amounts of data. These distortions, while possibly acceptable in general image steganography, are problematic in medical imaging, where even subtle artifacts could lead to misdiagnosis. Furthermore, many existing approaches fail to account for the structural heterogeneity of medical images, treating diagnostically critical and non-critical regions equally during the embedding process.

The problem is further compounded by the diverse range of medical imaging modalities, each with unique characteristics, resolutions, and diagnostic requirements. A steganographic approach that works well for one modality might be unsuitable for another, highlighting the need for adaptive, context-aware embedding techniques that can accommodate this diversity while maintaining consistent security and quality standards.

In this work, we focus primarily on developing deep learning-based data hiding techniques for patients' medical information. We address two main deep learning-based image steganography paradigms. For the first type—deep learning for optimal embedding locations—we developed a system that utilizes a deep learning architecture to pinpoint the most suitable regions within medical images for data concealment, ensuring the hidden information remains undetectable by both human observation and steganalysis techniques. For the second type—deep learning for end-to-end stego generation—we propose an innovative framework that employs convolutional neural networks (CNNs) to directly produce stego-images from cover images and secret messages, thereby eliminating the need for manually designed embedding rules.

Our proposed framework consists of embedding patients' information by strategically modifying selected coefficients within multi-resolution medical images, particularly targeting the Digital Imaging and Communications in Medicine (DICOM) data format. This approach aims to overcome the limitations of traditional methods by preserving diagnostic quality while ensuring robust security.

3 Research Questions

The problem to be addressed in this thesis are expressed by the following research questions:

1. **How can deep learning architectures, notably encoder-decoder frameworks, be designed to balance embedding imperceptibility, capacity, and recovery accuracy in medical image steganography?**

   This question explores the architectural considerations necessary to develop neural networks that can effectively hide information within medical images while ensuring that the resulting stego images maintain visual fidelity and that the embedded information can be accurately recovered.

2. **How can diagnostically insignificant regions in medical images be effectively identified and utilized for data embedding**

without compromising clinical value?

This question investigates methods for intelligently identifying areas within medical images that have minimal diagnostic significance, allowing for strategic embedding that preserves essential clinical information while maximizing data hiding capacity.

3. **How can advanced steganographic techniques be optimized to achieve high embedding capacity while maintaining image quality in medical images?**

   This question addresses the technical challenge of increasing the amount of information that can be embedded within medical images without introducing perceptible artifacts or compromising diagnostic utility.

4 Contributions

This thesis makes several significant contributions to the field of medical image steganography.

1. **Development of a deep learning-based framework for medical image steganography:** We propose a new approach that leverages the Mask Region-based Convolutional Neural Network (Mask-RCNN) architecture to identify diagnostically insignificant regions in medical images. This allows for strategic embedding of patient information in areas that do not compromise the image's diagnostic value,

significantly advancing the state-of-the-art in context-aware medical image steganography.

2. **Integration of QR codes for efficient patient information representation:** Our framework incorporates QR code generation to efficiently encode and embed patient information within medical images. This approach provides a structured representation of patient data that can be reliably embedded and extracted, enhancing the practical utility of the steganographic system in clinical workflows.

3. **Implementation of DCT-based embedding in insignificant regions:** We introduce a Discrete Cosine Transform (DCT) based embedding technique specifically designed for medical images. This approach enables high-capacity data hiding in the frequency domain while minimizing perceptual distortion, Attain high PSNR levels, reaching as much as 115 dB.

4. **Design of a clinical quality-aware deep learning framework:** We develop a CNN-based encoder-decoder architecture specifically optimized for medical images that preserves diagnostic features through multi-scale feature extraction, maintains embedding imperceptibility via constrained residual learning, and achieves robust secret recovery through progressive feature reconstruction.

5. **Experimental validation on diverse medical imaging modal-**

ities: Through extensive experimentation on medical datasets like CHAOS, MIDRC-RICORD-1B, and IQ-OTH/NCCD, we demonstrate the versatility and effectiveness of our approach across different medical imaging modalities. The framework achieves consistent performance with superior image quality compared to existing methods, showing up to 9.76 dB improvement in PSNR while maintaining full embedding capacity.

Our work collectively address the fundamental challenges in medical image steganography by providing a comprehensive framework that balances security, capacity, and clinical utility. The practical implications of this work include enhanced protection of patient privacy, streamlined secure data sharing in clinical settings, and improved integration with existing healthcare information systems.

5 Thesis Structure

This thesis is organized into the following chapters:

Chapter 1: General Introduction

This chapter provides an overview of the research context, introduces the problem statement, outlines the research questions, summarizes the main contributions, and presents the structure of the thesis.

Chapter 2: Healthcare Data Security and Data Hiding Tech-

niques

This chapter delivers an overview of fundamental concepts in image steganography, covering traditional techniques such as spatial domain and transform domain methods. It also provides a presentation of medical image characteristics and security requirements.

Chapter 3: Literature Review

This chapter surveys the state-of-the-art in medical image steganography, focusing on both traditional approaches and emerging deep learning-based methods. It critically analyzes the strengths and limitations of existing techniques, identifying research gaps that this thesis aims to address.

Chapter 4: High-Capacity Data Hiding for Medical Images Based on the Mask-RCNN Model

This chapter elaborates on our first major contribution, extending our framework with advanced DCT-based embedding techniques and comprehensive evaluation on the CHAOS dataset. We demonstrate significant improvements in embedding capacity and image quality, with PSNR values exceeding 115 dB while maintaining diagnostic integrity.

Chapter 5: Privacy-Preserving Medical Image Steganography: A Clinical Quality-Aware Deep Learning Framework

This chapter details our second major contribution, presenting our most advanced approach using a CNN-based encoder-decoder architecture specifically designed for medical images. This framework achieves a remarkable balance between embedding capacity and image quality, outperforming existing methods by 9.76 dB in PSNR while maintaining high capacity.

Chapter 6: Conclusion and Perspectives

This chapter summarizes the key findings and contributions of the thesis, discusses the limitations of the current approaches, and outlines promising directions for future research in medical image steganography.

Chapter 2

Healthcare Data Security and Data Hiding Techniques

2.1 Introduction

This chapter delves into the critical subject of healthcare data security by exploring a range of data hiding techniques that play a pivotal role in protecting sensitive information. It begins with a detailed examination of security concerns in the medical field and then transitions to the field of data hiding, explaining its principles and frameworks for covert communication. By discussing both classical approaches—such as LSB substitution, transform domain methods, statistical techniques, and spread spectrum methods—and modern deep learning-based techniques like CNN-based models and GANs, the chapter sets the stage for understanding the evolution and integration of these methods to secure sensitive healthcare data.

2.2 Healthcare Data Protection

Digital imaging is advancing quickly in the modern day due to innovations in image-capturing technology and the development of communication networks that have made life more easier for people. In the medical area, where the sharing of medical pictures across healthcare facilities has become crucial for precise diagnosis and all-encompassing patient care, photos have emerged as one of the most effective means of information transmission. The smooth exchange of medical images is essential for enabling prompt and efficient medical choices, which in turn improves the standard of healthcare services. Therefore, it is even more important to highlight the need to preserve the privacy of medical images, which presents particular concerns in comparison with ordinary images, because any distortion might negatively impact the accuracy of diagnosis. To protect patient privacy, many legal and regulated frameworks for healthcare data security have been developed. The fundamental principles of security and privacy for healthcare data will be discussed in this part, along with the main standard frameworks and regulatory that are intended to safeguard patient data.

2.2.1 Essential Security and Privacy Pillars for Protecting Healthcare Data

For the protection of digital information, various security techniques are employed [1], primarily addressing the following aspects:

Confidentiality

Confidentiality is the concept of ensuring that information is accessible for reading, listening, recording or physical removal only to subjects entitled to it, and that subjects only read or listens to the information to the extent permitted. A subject may be a person, a process or an organization [2]. The confidentiality in the healthcare system ensuring that information is accessed only by authorized subjects and only to the extent permitted—translates into concrete role-based permissions, secure login procedures, and audit trails that protect patient data. This ensures that each subject (whether a person, process, or department) interacts with health information strictly within their scope of authorization

Integrity

Integrity, in the context of information flow, means that reliable outputs must remain unaffected by any untrusted inputs. This concept serves as the counterpart to traditional confidentiality models, where public outputs must not be influenced by sensitive inputs. When viewed through the lens of access control, integrity is about preventing unauthorized or improper modification of data—specifically, stopping any changes by principals who do not hold the necessary rights [3]. In a healthcare context, integrity means ensuring that reliable patient information is not compromised by untrusted sources or unauthorized inputs. This mirrors how, in confidentiality, sensitive clinical details must remain separate from public or untrusted outputs. From an access control perspective, integrity centers on preventing any unauthorized alteration of patient data—only authorized clinicians and staff should be able to modify a patient's records, while all others must be blocked from making changes.

Authentication

Digital authentication is the process of confirming that the individual requesting access to a digital service really owns the credentials linked to their identity [4] . In simpler terms, it checks that the person trying to log in is who they claim to be by verifying they possess the valid "authenticator" (such as a password, token, or fingerprint) associated with that identity [5].

In a healthcare system, authentication ensures that anyone accessing electronic patient records or other clinical information truly is who they claim to be, by verifying they possess the valid credentials linked to their authorized identity. This helps maintain patient privacy, uphold data integrity, and ensure that only approved personnel or patients can view or update sensitive medical information.

Availability

Availability is the ability to make information and related physical and logical resources accessible as needed, when they are needed, and where they are needed [6]. In a healthcare context, availability means ensuring that patient information, medical records, and supporting resources—both physical and digital—are readily accessible to authorized healthcare providers whenever and wherever they are needed. This timely access is crucial for delivering accurate and efficient care, maintaining patient safety, and supporting critical clinical decisions.

Authorization

Authorization is the process of deciding whether a verified individual or system is allowed to carry out certain operations or retrieve specific information. In a healthcare context, authorization is the process of determining whether a verified healthcare provider or system is allowed to carry out certain actions—such as updating medical records or viewing specific patient data—based on their assigned roles and permissions.

Privacy

Privacy is the right of the individual to decide about himself/herself [7]. In a healthcare setting, privacy refers to each patient's right to determine how, when, and with whom their personal health information is collected, used, or shared.

2.2.2 Key Regulatory and Standard Frameworks for Healthcare Data Security

There are numerous established standards and tools for safeguarding personal medical information; here, we will focus on HIPAA [8] and ISO/IEC 27799 [9].

- The Health Insurance Portability and Accountability Act (HIPAA): enacted in 1996, is a federal law that aims to reduce healthcare fraud and abuse while creating nationwide safeguards to protect patient health information. It also ensures that people can keep their health insurance coverage when they change or lose their jobs. Beyond these protections, HIPAA encourages initiatives such as medical savings accounts and expanded long-term care options. In addition, it establishes standards to maintain the confidentiality and security of electronic health data.The primary objectives of Public Law 104-191, the Health Insurance Portability and Accountability Act (HIPAA) of 1996, are to improve the portability and continuity of health insurance coverage, combat health care fraud and abuse, promote medical savings accounts, enhance access to long-term care coverage, and simplify the administration of health insurance. A key part of this law also establishes nationwide standards for the privacy and security of electronic health information, ensuring the confidentiality and integrity of individuals' health data.

- ISO/IEC 27799: is an international standard designed especially for the medical field. In order to safeguard the availability, confidentiality, and integrity of personal health information, it interprets and modifies the security measures of the more general ISO/IEC 27002 standard. It acknowledges that the healthcare industry is a special one, with numerous providers, a wide range of third-party services, and sensitive patient data necessitating a less generic approach to information security management. The main aims of ISO/IEC 27799 is to provide guidance to healthcare organizations and other health information custodians on how to adopt a minimal degree of security suitable for their situation in order to maintain the confidentiality, accuracy, and ease of access to personal health information for patient care. In essence, it offers a comprehensive action plan for the setup and maintenance of an Information Security Management System (ISMS) in the healthcare industry, guaranteeing that particular risks and demands of the field are consistently met, such as patient privacy and legal compliance.

2.3 Medical Images and DICOM File Structure

Since medical imaging is crucial for diagnosing diseases and planning treatments, it is essential that both the images and their associated data are consistently maintained and exchanged. DICOM (Digital Imaging and Communications in Medicine) is one of the most commonly adopted standards in medical imaging [10]. DICOM is a standard used for handling, storing, printing, and transmitting information in medical imaging. It encompasses both a file format and a network communications protocol, ensuring seamless interoperability among various medical devices and software.

2.3.1 DICOM File Structure

A DICOM file consists of two main components:

1. Header (Metadata Section): Contains patient information, acquisition details, and imaging parameters.
2. Image Data Section: Stores the actual pixel data of the medical image.

DICOM files use a tag-based structure, where each piece of information is represented as a data element identified by a unique tag (Group, Element).

2.3.2 Patient Data Information Tags in DICOM

Patient-related metadata is essential for identifying, managing, and ensuring the integrity of medical records. DICOM provides specific tags to store patient demographics and medical history. Table 2.1 displays the key tags associated with patient information.

Table 2.1: Patient Data Information Tags in DICOM

Tag (Group, Element)	**Attribute Name**	**Description**
(0010,0010)	Patient's Name	Stores the patient's full name.
(0010,0020)	Patient ID	A unique identifier for the patient.
(0010,0030)	Patient's Birth Date	Date of birth (YYYYMMDD).
(0010,0040)	Patient's Sex	Gender (M = Male, F = Female, O = Other).
(0010,1010)	Patient's Age	Patient's age in years.
(0010,1020)	Patient's Weight	Weight of the patient in kilograms.
(0010,21B0)	Additional Patient History	Relevant medical history or diagnosis.

2.3.3 Image Data Information Tags in DICOM

DICOM also contains image-specific metadata that provides critical details about the imaging process and the properties of the acquired image. Table 2.2 presents the primary tags corresponding to image information.

Table 2.2: Image Data Information Tags in DICOM

Tag (Group, Element)	**Attribute Name**	**Description**
(0020,000D)	Study Instance UID	A unique identifier for the imaging study.
(0020,000E)	Series Instance UID	A unique identifier for a series within a study.
(0020,0013)	Instance Number	The position of this image in the series.
(0008,0060)	Modality	Imaging type (e.g. CT, MR, X-ray, Ultrasound).
(0028,0010)	Rows	Number of rows (image height, in pixels).
(0028,0011)	Columns	Number of columns (image width, in pixels).
(0028,0030)	Pixel Spacing	Physical spacing between pixels (in mm).
(0028,0100)	Bits Allocated	Number of bits allocated per pixel.
(0028,1050)	Window Center	Center of the grayscale display range.
(0028,1051)	Window Width	Width of the grayscale display range.

2.4 Data hiding

Data hiding has historically referred to the act of concealing communications or information so that only the intended recipients are aware of its existence. Although modern data hiding techniques often rely on digital technology, information hiding is the process of embedding secret data within various forms of redundant cover media (e.g., images, audio files, videos, or text documents) in such a way that the presence of the data remains undetectable to unauthorized parties [11]. Data hiding includes two sections: Steganography and Watermarking. Steganography is the process of secretly encoding data into a carrier media , such as an image or audio file, so that its existence is not detected. In contrast, watermarking usually entails adding ownership or identifying information to a piece of media without substantially changing the original content, usually for copyright protection or authentication.

The complete process of concealing information relies on two operations: insertion, which involves embedding the information into the medium (the original document in watermarking, or the cover medium in steganography), and extraction, which retrieves this information. Additionally, the term detection is used when verifying the presence of information in the watermarked medium, without necessarily intending to extract it. In data hiding, three fundamental concepts serve as essential performance indicators: Imperceptibility, capacity, and robustness.

- Imperceptibility : Imperceptibility is the characteristic of a data-hiding technique that allows information to be embedded without causing visible or audible distortions in the cover medium, ensuring that the hidden data remains unnoticeable to human perception [12]. Imperceptibility in steganography ensures hidden data does not noticeably alter the original medium, keeping it indistinguishable from an unmodified version. In digital watermarking, it maintains the host signal's quality while embedding a watermark without perceptual degradation.

- Capacity: Capacity in data hiding refers to the highest amount of information that can be embedded in a cover medium while preserving its quality and keeping the hidden data undetectable [12]. Capacity in steganography is the maximum data that can be embedded while ensuring imperceptibility and resistance to detection. In watermarking, it refers to the amount of embedded data that preserves the host signal's quality and robustness.

- Robustness: Robustness in data hiding is the capability of embedded information to remain intact despite intentional or unintentional alterations to the cover medium, including compression, noise, filtering, or geometric transformations [13]. Robustness in steganography ensures hidden data remains intact despite processing or attacks like compression or format conversion. In digital watermarking, it signifies the watermark's resistance to distortions while remaining detectable and recoverable under transformations such as cropping and filtering.

2.4.1 Degital watermarking

In the following section, we introduce the fundamental concepts of digital watermarking and explore the different types of watermarking schemes.

What is watermarking?

Digital watermarking is the process of imperceptibly embedding auxiliary information (a "watermark") into a host signal—such as an image, audio track, video stream, or software binary—so that the added data remain invisible or inaudible to human perception [14]. The embedded watermark (which may be a bit-sequence, text string, logo, or cryptographic hash) can later be extracted or detected to prove ownership, trace distribution, authenticate content, or signal tampering.

Watermarking is classified into visible and invisible types. Visible watermarking alters

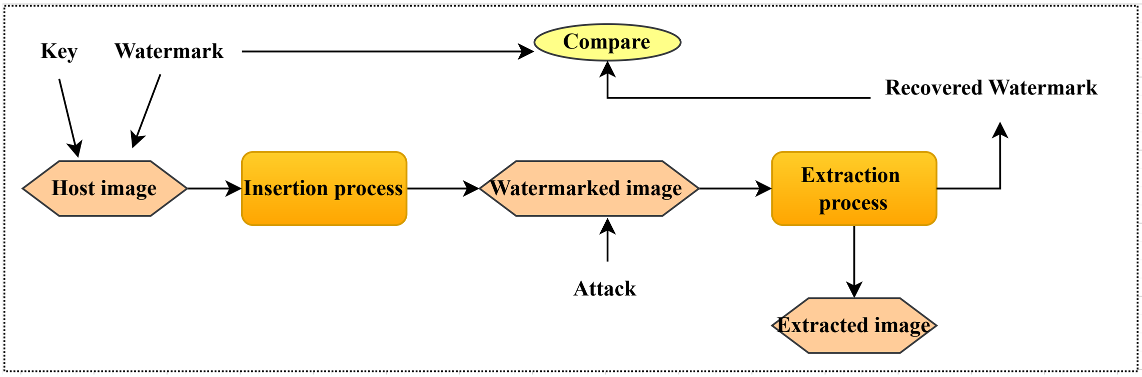


Figure 2.1: Digital watermarking framwork

a file by adding a noticeable mark, such as a copyright symbol on images. Invisible watermarking subtly modifies the document without detection, often by altering the least significant bits, making it a form of steganography.

2.4.2 Watermarking Framework

A watermarking framework is a methodical procedure for adding a watermark on a host image and then extracting it. The following is a description of the main elements and procedures that make up this framework:

- Host image: The host image is the original digital image that serves as the medium for embedding the watermark. It remains visually unchanged or minimally affected after watermark insertion.
- The key: In order to guarantee security and permitted access, the watermarking procedure uses a secret parameter called the key. It aids in the extraction and embedding procedures by preventing the watermark from being removed or altered without authorization.
- Watermark: The distinctive information included into the host image is known as the watermark, and it may take the shape of a binary pattern, text, serial number,

or logo.

- Insertion process: By employing a specialized algorithm, the watermark is embedded into the host image during the insertion phase, ensuring that it remains unobtrusive and resistant to attacks.
- Watermarked image: The image that results from embedding the watermark is known as the watermarked image. It should safely include the concealed watermark and look like the original host image.
- Extraction process: Using the appropriate key and extraction technique, the watermark is extracted from the watermarked picture in the extraction phase, which is the opposite of the insertion process.
- Recovered watermark: It is the extracted watermark obtained from the watermarked image. Ideally, it should match the original watermark embedded in the host image.
- Compare process (Original vs. Extracted Watermark): The resemblance between the recovered watermark and the original watermark is assessed in this comparison stage. While differences might be a sign of manipulation or integrity loss, a high similarity ensures legitimacy by confirming effective watermark recovery.
- Extracted image: The Image that remains after the extraction or watermark removal operation is known as the extracted image. This image may still show signs of watermarking or it may be quite similar to the original host image, depending on the watermarking method employed.

2.4.3 Watermarking classification

Digital watermarking can be classified based on various criteria, including visibility, robustness, domain of embedding, and application [15].

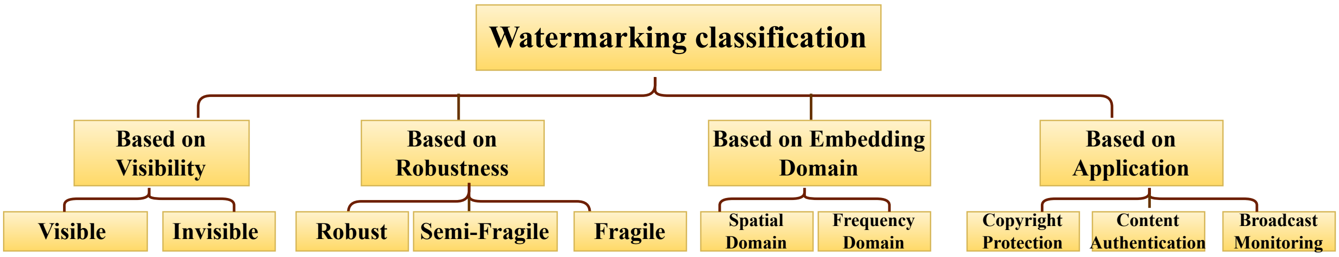


Figure 2.2: Classification criteria of digital watermarking

- Based on visibility: In visible watermarking, the watermark is clearly noticeable to users, appearing as logos or text superimposed on images or videos, primarily for copyright protection. In invisible watermarking, the watermark is discreetly embedded within the content, making it imperceptible to the human eye or ear, and is commonly utilized for authentication and content tracking.

- Based on robustness: Robust watermarking designed to withstand common signal processing operations, such as compression, filtering, cropping, and format conversion. Used for copyright protection. Fragile watermarking easily destroyed or altered if the host media is modified, making it useful for integrity verification and tamper detection. And for the semi-fragile watermarking can tolerate some level of modification, such as compression, but is sensitive to malicious tampering, it used for content authentication.

- Based on embedding domain: In spatial domain watermarking, the watermark is embedded directly by modifying pixel or sample values in images, audio, or video. It is simple but less robust against attacks. As for frequency domain watermarking, the watermark is embedded in transformed coefficients, such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), or Discrete Fourier Transform (DFT), making it more robust.

- Based on application: Used for copyright protection to establish ownership and prevent unauthorized use, content authentication to detect tampering and verify integrity, and broadcast monitoring to track media distribution.

Digital watermarking plays a key role in securing digital content through embedding and authentication techniques. With various classifications and applications, it ensures protection while maintaining media integrity. In the next section, we explore the other aspect of data hiding—steganography.

2.4.4 Image Steganography

What is Steganography?

Steganography is the practice and science of covert communication, achieved by embedding information within other data to conceal its existence [16]. The term originates from the Greek words ”stegos” (meaning ”cover”) and ”grafia” (meaning ”writing”), which together translate to ”covered writing.”

2.4.5 Image steganography Framework

The fundamental aim of image steganography is to securely transmit an image (cover) that maintains its original appearance while covertly carrying concealed information (secret message) from sender Alice to recipient Bob. Despite being targeted by Eve's attacks in the communication channel. While maintaining secrecy amidst the steganalysis process, which endeavours to reveal concealed information within digital files. Fig. 2.4 displays the fundamental framework of image steganography, and the several key elements and processes are described below:

- Secret: Refers to the hidden information that needs to be securely transmitted. It can be text, an image, audio, or any other form of digital data.

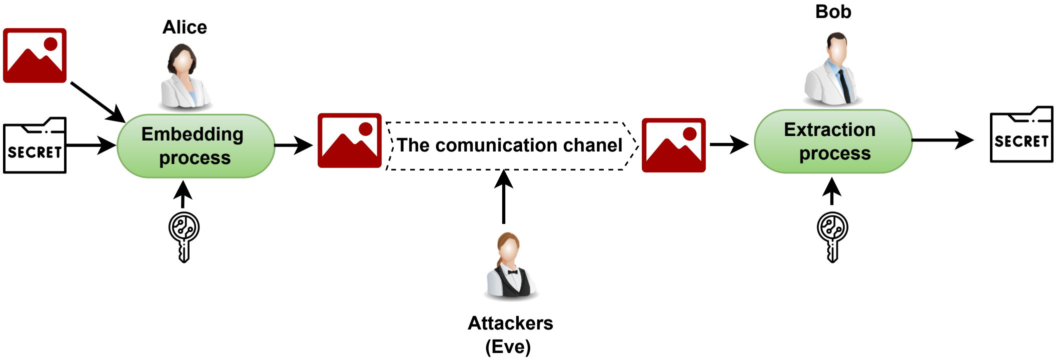


Figure 2.3: Steganography Framework

- Cover: Is the original medium in which the secret is embedded. It can be an image, audio, video, or any digital file used to hide the secret data without raising suspicion.
- Key: Is a secret parameter used in the steganographic process to enhance security. It ensures that only the intended recipient with the correct key can extract the hidden information.
- Alice: Represents the sender in the steganographic system. She embeds the secret information into the cover using an embedding process before transmitting it through a communication channel.
- Embedding Process: Is the method used to hide the secret within the cover. This can be done using different techniques, such as modifying pixel values in images, altering frequency components in audio, or appending data in text files. The result of this process is a stego object.
- Communication Channel: Is the medium through which the stego object is transmitted from Alice to Bob. It can be the internet, email, cloud storage, or any other means of data transfer.

- Stego: Is the final output after the secret has been embedded into the cover. It should resemble the original cover as closely as possible to avoid detection.

- Extraction Process: Is the reverse of the embedding process. It involves retrieving the hidden secret from the stego object using the correct extraction method and key.

- Bob: Represents the receiver in the steganographic system. He uses the extraction process and the correct key (if required) to retrieve the hidden secret from the stego object.

- Extracted Secret: Is the recovered hidden information that Bob successfully retrieves from the stego object. If the process is executed correctly, the extracted secret should match the original secret embedded by Alice.

The steganography framework provides a structured approach to concealing and transmitting secret information securely. By embedding data within a cover medium and ensuring proper extraction, it enables covert communication between Alice and Bob while minimizing the risk of detection. In the following section, we explore the methods and approaches that demonstrate the digital steganography framework.

2.5 The Steganographic Approaches for Data Concealment

Over the years, steganography has been extensively studied through various research efforts, employing different techniques. These approaches can be broadly categorized into two main types based on their feature selection and classification algorithms. The first category, known as classical steganography, relies on traditional embedding methods such as Least Significant Bit (LSB) to conceal information within a cover medium. The

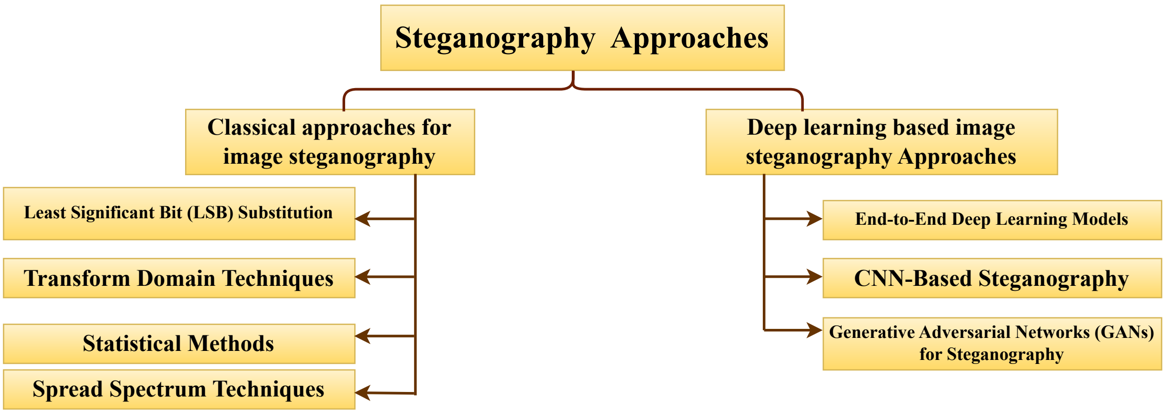

Steganography Approaches
Classical approaches for image steganography
Deep learning based image steganography Approaches
Least Significant Bit (LSB) Substitution
Transform Domain Techniques
Statistical Methods
Spread Spectrum Techniques
End-to-End Deep Learning Models
CNN-Based Steganography
Generative Adversarial Networks (GANs) for Steganography

Figure 2.4: Steganography Framework

second category leverages deep learning techniques to enhance the concealment process, offering more advanced and adaptive methods for data hiding. The figure 2.4 illustrates the general classification of this method.

2.5.1 Classical approaches for image steganography

Traditional image steganography conceals data by leveraging regular patterns in digital pictures, and the field is chiefly organized around four prominent technique families: Least Significant Bit (LSB) substitution, transform-domain embedding, statistical steganography, and spread-spectrum techniques.

Least Significant Bit (LSB)

Least Significant Bit (LSB) insertion is a widely used and straightforward technique for embedding information in an image. It is one of the most recognized methods for concealing secret text within an image. This approach replaces the LSBs of pixel values in the cover image with bits from the hidden message. Since only the LSBs are modified, the resulting stego-image remains nearly identical to the original, as the alterations have little visible impact [17].

- Advantages of LSB-based techniques:

  - Simplicity and Efficiency – LSB substitution is straightforward to implement and requires minimal computational resources.

  - Minimal Visual Distortion – Since only the least significant bits are modified, the changes are imperceptible to the human eye, preserving the visual quality of the image.

  - High Data Hiding Capacity – In 24-bit images (RGB), three bits per pixel can be used, allowing a reasonable amount of data to be embedded.

  - Easy to Extract Hidden Data – The embedded message can be easily retrieved if the method and key (if used) are known.

  - No Need for Complex Algorithms – Unlike other steganographic techniques, LSB embedding does not require sophisticated mathematical models.

- Weaknesses of LSB-based techniques:

  - Low Resistance to Image Processing – Common operations like compression, resizing, filtering, or format conversion can alter the LSBs, potentially corrupting the hidden message.

  - Vulnerability to Detection – Statistical and steganalysis tools can detect patterns in modified LSBs, making the method less secure against attacks.

  - Limited Robustness – If an image undergoes slight modifications, such as noise addition or cropping, the hidden data can be lost or altered.

  - Lower Security Compared to Advanced Techniques – Since LSB substitution is well-known, it is more susceptible to steganalysis, making it less secure for highly confidential data.

  - Capacity vs. Quality Trade-off – Embedding too much data can create noticeable distortions in the image, making the steganography detectable.

Transform Domain Techniques

Transform domain techniques are advanced methods used in steganography, where secret information is embedded in the frequency domain of an image instead of directly altering pixel values in the spatial domain (as in LSB-based techniques) . These methods utilize mathematical transformations such as the Discrete Cosine Transform (DCT) [18], Discrete Wavelet Transform (DWT) [19] , and Discrete Fourier Transform (DFT) to embed data into less noticeable regions of the image, making it more resistant to common image processing operations like compression, filtering, and noise addition [20]

- Advantages of frequency domain techniques:
  - Higher Robustness: Frequency domain techniques are more resistant to common image processing operations such as compression, filtering, and noise addition compared to spatial domain methods.
  - Better Security: Since the secret information is embedded in transformed coefficients rather than raw pixel values, it is harder for attackers to detect and extract the hidden data.
  - Decreased Perceptual Distortion: To maintain the overall quality of the stego-image, data is frequently placed in high-frequency components where human eyesight is less sensitive.
  - Effective for Compressed images: JPEG pictures frequently employ techniques like Discrete Cosine Transform (DCT), which makes them appropriate for concealing data in compressed forms without appreciably compromising visual quality.
  - More Data Hiding Capacity: Depending on the transformation technique, frequency domain methods can provide a larger embedding capacity without making noticeable changes to the image.
  - Resistant to Statistical Attacks: Unlike LSB substitution, which can be detected through histogram analysis, transform domain techniques distribute hidden data

more evenly, making steganalysis more challenging.

- Weaknesses of frequency domain techniques:
  - Higher Computational Complexity – Transform domain methods involve mathematical transformations like DCT, DWT, or DFT, which require more processing power and time compared to simpler spatial domain techniques.
  - More Difficult Implementation: Unlike LSB-based approaches, frequency domain techniques require specialized knowledge in signal processing and transformation algorithms.
  - Limited Applicability to Certain Image types: These methods perform best with compressed forms like JPEG, but they might not work as well with uncompressed formats like BMP or PNG.
  - Possibility Loss of Hidden Data: Some embedded information may be lost or damaged if an image is subjected to substantial modifications, including significant compression or scaling.
  - In our research presented in the chapter 4 , we prefer to use the 2-D DCT because it concentrates most visual energy into a few low-frequency coefficients within small blocks, allowing data to be embedded in the mid-frequencies with minimal distortion while maintaining block-level control. DCT coefficients are real and supported by fast hardware instructions, giving simpler quantization and lower computational cost than the multiscale DWT or the global, complex-valued DFT—both of which spread embedding changes more widely and are less resilient to routine post-processing.

Statistical Methods

Statistical methods in image steganography focus on embedding information by modifying the statistical properties of an image while ensuring that global statistical characteristics remain unchanged. These methods are designed to counteract steganalysis techniques that detect anomalies in image structures caused by hidden data [21] [22]. Unlike spatial domain techniques (such as LSB substitution) or frequency domain techniques (such as DCT-based steganography), statistical steganography does not directly alter pixel values or frequency coefficients in an obvious manner. Instead, it manipulates statistical distributions to ensure that embedding does not introduce detectable irregularities. Changing the histogram of a picture is one of the most basic statistical steganographic methods [23]. Since the frequency distribution of pixel values is represented by an image's histogram, data may be embedded by subtly altering pixel intensities while preserving the form of the histogram to guarantee that the changes are imperceptible. For instance, an algorithm may: Group pixels into pairs (e.g., 120 and 121). If a message bit is 0, keep the pair unchanged. If a message bit is 1, slightly increase or decrease one of the pixel values to encode data while maintaining the histogram's overall structure. This method is effective against basic histogram analysis attacks, which attempt to detect anomalies in pixel distributions.

- Advantages of statistical Methods

  - More Secure than LSB Substitution: Since these methods avoid directly modifying LSBs in a predictable way, they are less susceptible to detection by chi-square steganalysis or histogram-based attacks.

  - Higher Robustness: These techniques offer better resilience against image processing operations like compression, filtering, and noise addition.

  - Difficult to Detect Using Basic Statistical Tests: By preserving first-order and second-order statistics, these methods can bypass simple steganalysis techniques that look for anomalies in pixel distributions.

- Weaknesses of statistical Methods

  - Lower Embedding Capacity: Since these methods must maintain statistical con-

sistency, they often allow less data to be hidden compared to LSB-based techniques.
- More Computationally Complex: Adjusting pixel distributions while preserving statistical properties requires more processing power than simple bit replacement methods.
- Still Vulnerable to Advanced Steganalysis: While they evade basic detection methods, more advanced machine learning-based steganalysis techniques can still detect statistical irregularities.

Spread Spectrum Techniques

Spread Spectrum (SS) techniques are inspired by wireless communication systems [24], where a signal is spread over a wide frequency spectrum to make it resistant to noise and interference. In the context of image steganography, these techniques distribute the hidden data across multiple frequency components of the image, making detection and removal more difficult [25]. Unlike LSB or transform domain methods, Spread Spectrum Steganography (SSS) ensures that the hidden message is embedded in a way that closely resembles natural image noise, increasing its robustness against attacks. Spread spectrum techniques introduce a small amount of modification across a large number of pixels or frequency coefficients, rather than concentrating the changes in a specific region. This is achieved through pseudorandom sequences, which help distribute the hidden message throughout the image in a noise-like manner.

- Advantages of Spread Spectrum Techniques
  - High Robustness: The embedded data is spread across multiple frequencies or spatial locations, making it resilient to compression, filtering, and noise.
  - Strong Security: Because the message resembles natural image noise, it is difficult for attackers to detect using statistical or steganalysis methods.
  - Low Perceptibility: The changes made to the image are subtle and spread over a

wide area, making them hard to notice visually.
- Resistance to Steganalysis: Traditional detection methods struggle to differentiate the hidden message from normal variations in the image.

- Weaknesses of Spread Spectrum Techniques
  - Lower Embedding Capacity: Since the message is distributed across a large part of the image, the total amount of data that can be hidden is lower compared to LSB-based methods.
  - Higher Computational Complexity: Encoding and decoding require additional processing, such as correlation techniques and pseudorandom sequence generation.
  - Susceptibility to Desynchronization: If the original image undergoes significant modifications (e.g., cropping, rotation, or excessive compression), retrieving the embedded message may become difficult without proper synchronization.

2.5.2 Deep Learning-Based Image Steganography Approaches

The integration of deep learning techniques into steganography has brought substantial advancements and improvements [26]. This is achieved by enhancing the concealment quality, allowing confidential information to be seamlessly embedded within the cover media in a way that makes it difficult to differentiate between the original and the stego. Deep learning-based steganography, in contrast to traditional techniques, may automatically discover the best embedding techniques, boosting security, resilience, and imperceptibility. Deep learning-based image steganography approaches are generally divided into two main categories:

- Deep Learning for Optimal Embedding Locations: In these techniques, the cover image is processed using deep learning models to identify the best places for data embedding, with the goal of embedding information in areas where steganalysis or human eye are less likely to notice changes.

- Deep Learning for End-to-End Stego Generation: These approaches use deep neural networks to directly generate stego-images from the original cover images and secret messages, eliminating the need for handcrafted embedding rules.
- GAN-Based Steganography: leverages generative adversarial networks, where the generator creates stego images that contain hidden data, while the discriminator pushes the generator to produce outputs that are visually indistinguishable from normal images.

End-to-End Deep Learning Models

End-to-end deep learning models automate the entire steganographic process, from embedding to extraction, without relying on predefined rules. These models typically use encoder-decoder architectures, where:

The encoder takes a cover image and a secret message as input and learns to embed the message while maintaining image quality. The decoder is trained to extract the hidden message from the stego-image. A well-known example of this approach is HiDDeN (High-Capacity Deep Neural Networks for Image Steganography) [27], which uses CNN-based architectures to embed secret messages while optimizing for robustness and imperceptibility.

CNN-Based Steganography

Convolutional Neural Networks (CNNs) are widely used to improve the embedding process in image steganography. These methods leverage CNNs for two main tasks:

Feature-based embedding: CNNs analyze image features (e.g., textures, edges, noise levels) to determine the best regions for embedding data, ensuring that modifications blend naturally with the cover image. Steganalysis-resistant embedding: CNNs can be trained to hide information in ways that are difficult for traditional steganalysis methods to detect.

Generative Adversarial Networks (GANs) for Steganography

GAN-based steganography is one of the most advanced deep learning approaches for hiding information in images. GANs consist of two networks:

A Generator that learns to embed secret data while making the stego-image indistinguishable from real images. A Discriminator that tries to distinguish between real images and stego-images, forcing the generator to improve its ability to hide data effectively. A well-known model in this category is StegGAN [28], which uses adversarial training to ensure that stego-images closely resemble natural images. By continuously improving through adversarial learning, these models can create stego-images that are highly resistant to detection.

2.6 Conclusion

In summary, this chapter underscores the importance of robust data hiding techniques in protecting healthcare data. The comprehensive analysis of traditional watermarking and steganographic methods alongside emerging deep learning approaches demonstrates how these technologies contribute to enhanced security, resilience, and imperceptibility. By integrating these techniques, the chapter provides valuable insights into building secure systems that effectively guard against unauthorized access and tampering, ensuring the integrity of sensitive healthcare information.

Chapter 3

Literature Review on Data Hiding in Medical Images

3.1 Introduction

In this chapter, we provide an overview of the latest research and methodologies in the field of data hiding, with a particular focus on image steganography. We explore both traditional steganographic techniques and modern deep learning-based approaches. Additionally, we review studies that leverage steganography to enhance the confidentiality of medical data. Furthermore, we detail the evaluation metrics employed in this study to assess the performance of the proposed methods, ensuring a thorough analysis of their effectiveness.

3.2 Image Steganography Metrics

In this introductory section, we outline the essential metrics employed in image steganography. Table 3.1 encapsulates these metrics, each addressing a distinct aspect of image quality or the embedding process. Collectively, they serve to evaluate the imperceptibility, robustness, and capacity of the data-hiding method.

Metric	Description	When is it better ?
Peak Signal-to-Noise Ratio (PSNR)	It calculates the discrepancy between two images by comparing the maximum signal power to the power of noise that disrupts the signal. It is mesured in decibels (db)	A higher value (Exceeds 30 dB)
Structural Similarity Index (SSIM)	It compare luminance, contrast and the structural of the cover and the stego .	Close to 1
Mean Square Error (MSE)	It measures the average squared difference between the pixel values of the cover image and the stego image.	A lower value
Universal Image Quality Index (UIQI)	It computes the similarity between the cover and the stego based on similarities in luminance, contrast, and structure.	close to 1
Embedding Capacity (EC)	It is the number of secret bits that are embedded per pixels, it is calculated in Bpp (Bit Per Pixel)	A higher value
Payload capacity	It is the quantity of information that can be concealed within the cover media. It is represented as a specific number of bits or the percentage designated for concealing data relative to the total size of the cover.	A higher value
Bit Error Rate (BER)	Determining the proportion of wrongly recovered bits relative to the concealed data's original value.	close to 0

Table 3.1: Most common image steganography matrices

- PSNR: A higher PSNR value indicates that the stego image maintains a quality closer to the original cover image, suggesting minimal perceptible differences [29].

- SSIM: A value near 1 implies that the structural information in the stego image closely matches that of the cover image, ensuring strong visual similarity [30].
- MSE: A lower MSE signifies fewer errors between the cover and stego images, indicating higher fidelity in the stego image [31].
- UIQI: A UIQI near 1 reflects a high level of similarity in luminance, contrast, and structure between the cover and stego images [32].
- Embedding Capacity (EC): A higher EC allows more secret data to be embedded without significantly affecting the cover image's quality [33].
- Payload Capacity: A greater payload capacity enables the concealment of larger amounts of information within the cover media.
- Bit Error Rate (BER): A BER close to 0 emphasizes that the extracted secret data closely matches the original, ensuring data integrity [34].

3.3 Classical Approaches to Medical-Image Steganography

Numerous studies have applied steganography in medical imaging. Notably, Bozhidar et al. [35] introduced an innovative method known as BOOST, designed to conceal user data within medical images. Their approach unfolded in two distinct stages: Initially, the confidential patient data was encrypted using a novel ”pseudo-random generator based on the nuclear spin generator” technique, resulting in encrypted data. This encrypted output was subsequently transcribed into a binary sequence using an ASCII table. In the subsequent step, this binary sequence found its place within the least significant bit of the non-black pixels in the image. In particular, their method achieved remarkable results, boasting PSNR values exceeding 113 dB, all while accommodating a payload capacity of

0.74 bits per pixel. The substantial payload capacity emphasizes the potential for real-world applications. However, it is important to consider the computational overhead of these encryption and embedding processes, especially when dealing with large medical image datasets.

In [36], Romany et al. introduced an encompassing steganography method that amalgamates several techniques for robust data hiding within medical images. They proposed the application of RSA encryption for safeguarding sensitive information, the Ripplet Transform for image manipulation, and LSB substitution for embedding secret data. An adaptive genetic-algorithm-based Optimum Pixel Adjustment Process (OPAP) was implemented to enhance imperceptibility by fine-tuning the stego image. This comprehensive approach demonstrated resilience against RS attacks and established that Discrete Ripplet Transform (DRT) yielded superior results in comparison to Integer Wavelet Transform (IWT). Notably, the achieved PSNR values ranged from 49 to 56 dB, indicating a trade-off between visual quality and payload capacity.

In [37], Songul K and Engin A presented an innovative steganography technique termed ”Genetic Algorithm-Optimum Pixel Similarity.” This approach leverages pixel similarity and LSB embedding to seamlessly integrate a substantial amount of data, specifically 10,000 characters, into 256 × 256 medical images. What sets this method apart is its ability to achieve embedding without resorting to data compression techniques. The fitness function for the genetic algorithm is adopted from PSNR, with random selection as the key method. Impressively, the average PSNR achieved was recorded at 47.41 dB, highlighting the delicate balance between imperceptibility and embedding capacity.

Partha et al. explored patient data protection in [38] through a novel steganographic method, employing Support Vector Machine (SVM) and Discrete Wavelet Transform (DWT). The SVM was utilized for the recognition of Regions of Interest (ROI) and Non-ROI (NROI) within medical images. RGB components were subjected to IWT, and a circular array technique facilitated the integration of confidential information within NROI pixels. Impressively, this approach yielded an average PSNR value of 64 dB, showcasing its potential for robust and secure patient data embedding. In another study [39], a robust and reversible data hiding scheme was proposed, involving a Support Vector Neural Network (SVNN) classifier and the Contourlet Transform method. The SVNN was trained to identify suitable pixels for concealment, with the HL band of the CT coefficient serving as the container for hidden data. The method was rigorously analyzed with and without noise, demonstrating exceptional results with a PSNR value of up to 89.3253 dB, outperforming the SVNN-wavelet approach from [40].

In [41], an innovative approach was introduced that encoded patient data using Enhanced Huffman compression coding for enhanced payload capacity and security. This encoded data was then concealed within medical images using Pixels Contrast (PC) and the Henon map algorithm. The study evaluated outcomes based on histogram analysis, PSNR, and SIMM metrics, with achieved PSNR values ranging between 70 dB and 71 dB. A novel steganography technique was proposed in [42], utilizing a combination of a 3-D chaotic system, one-particle Quantum Walk (QW), and Particle Swarm Optimization (PSO). This intricate methodology ensured the privacy of medical data by generating sequences for PSO through chaotic systems and QW, which were then utilized to replace confidential medical images with concealed data. Despite its high visual quality, this technique achieved an average PSNR of 44.1 dB, reflecting inherent limitations in data capacity.

In [43], Quantum Walks were employed for a pixel-based steganography technique with added security. The approach aimed to identify two LSB pixels in the carrier image for the concealed two bits of confidential data. The technique leveraged quantum walk states as a private key, promising robust security and resilience against data loss attacks. With a PSNR close to 44.40 dB, the method achieved remarkable visual quality.

The authors in [44] proposed an alternative method using the Rabin public key cryptosystem to safeguard patient medical records. This concealed information within medical cover images using diagonal queue replacements and the LSB plane. The technique showcased robustness and yielded high PSNR values, reaching 76 dB. However, careful consideration is needed when using significant areas, as they might impact diagnostic accuracy. The steganography technique presented by Hashim et al. in [45] targeted data security during transmission within an IoT framework. Encrypted patient data was divided into blocks and concealed within medical images using the Henon map parameters for random pixel selection. This technique demonstrates effective use of steganography mechanisms for IoT data security.

Prasanth et al. introduced an invisible watermarking scheme in [46] for embedding patient information into EGG signals for telemedicine applications. A QR code of patient data was decomposed and utilized for watermarking EGG signals. This intricate approach provided a unique approach to securing medical information within telemedicine applications. Thiyagarajan et al.'s approach in [47] centered on reversible steganography, using LSB for embedding. Canny edge detection distinguished between Regions of Interest (ROI) and Non-ROI (NROI) in medical images, with NROI serving as payload pixels. Graph 3 coloring was employed to generate a private dynamic key, adding an innovative layer of security.

S. Jeevitha et al.'s technique in [48] utilized the ElGamal encryption algorithm for patient data security, concealing data within Non-ROI (NROI) areas to ensure patient information remains secure. The approach relied on the comparison between DWT segmentation and Canny Edge Detection results to determine pixel locations for embedding. The method showcased the advantages of DWT-based approaches and achieved strong embedding rates. In [49], R. Bala Krishnan introduced a unique approach where patient data was not encrypted, but the image itself was encrypted using a Sudoku-based mechanism. The encrypted image was then employed for data hiding using Queen Traversal Based Secret Code Substitution based on LSB. This intricate process added layers of protection to the concealed information, yielding high-quality stego images.

Arunkumar S et al. proposed a novel technique in [50] for secure medical image transmission. The medical image was encrypted using the logistic chaotic map, followed by embedding using an Embedding Distortion Measure based co-accurate matrix. The method prioritized security and yielded high visual quality while ensuring secure image transmission. The secure steganography method outlined in [51] incorporated a shell matrix and LSB for enhanced data security. While the method exhibited high payload capacity, it required substantial computational resources for high-resolution images, thereby impacting complexity. The authors, in [52], introduced a novel approach involving a genetic algorithm to enhance PSNR levels in Stego images. The technique aimed to cover a medical image with a natural image, utilizing a combination of mechanisms such as one-point crossover, random resetting mutation, and tournament selection. While the method achieved infinite PSNR and SSIM values without causing distortion, its complexity remained a significant consideration.

Early works present in [53] by Lina Zhang et al. propose a reversible selective encryption scheme tailored for medical images. Their approach extracts the region of interest (ROI) using advanced edge detection and boundary tracking techniques, encrypts the ROI through coupled chaotic maps, and embeds key information via a steganographic method. This method not only secures sensitive medical data but also ensures lossless decryption and improved computational efficiency.In tests on 342 medical images, the average Shannon entropy of the encrypted images was 7.9979, indicating near-optimal randomness. Additionally, the scheme reached NPCR and UACI values of approximately 99.61% and 33.51% respectively, demonstrating strong resistance against differential attacks while maintaining lossless decryption.

Gulfam A. et al. [54] propose a secure medical data transmission method that embeds sensitive patient information into DICOM images using steganography. Their approach first encrypts an RGB patient image using Adversarial Neural Cryptography combined with SHA-256 (ANC-SHA-256) and then hides it in the Region of Non-Interest (RONI) via an adaptive LSB replacement technique. A digital signature is generated from the DICOM metadata using SHA-256 to ensure the file's integrity and authenticity. Evaluations on various medical image datasets—including MRI, CT, X-ray, and ultrasound—demonstrate high visual quality, with average metrics of 67.55 dB for PSNR, 0.9959 for NCC, 0.9887 for SSIM, 0.9859 for UQI, and 3.83 for APE, while maintaining strong robustness against geometric and physical attacks.

These various techniques highlight the diversity of approaches employed for medical image steganography, each balancing imperceptibility, security, and capacity in distinct ways.

3.4 Deep Learning-Based Image Steganography Approaches

This section explores deep learning-based image steganography approaches that leverage the power of neural networks to embed and extract information with enhanced security and robustness. These innovative methods often employ convolutional neural networks, autoencoders, and generative adversarial networks to improve the imperceptibility and capacity of steganographic techniques. In the following discussion, we review key research contributions and evaluate their methodologies.

3.4.1 Encoder-Decoder Image Steganography

Encoder-decoder image steganography leverages deep neural network architectures(often in the form of autoencoders) to embed secret data into cover images while preserving visual quality. These methods learn to extract meaningful features from both cover and secret images, optimize the embedding process jointly, and then decode the hidden information with high fidelity, striking a balance between payload capacity and imperceptibility.

In [55], Buljia introduced a pioneering architecture for image steganography that employs neural networks to conceal a color image within another image of identical dimensions. This technique adopts the structure of auto-encoding networks. The first network is responsible for resizing the secret image to match the cover's size, while the second network serves as the hiding network. The encoder extracts cover characteristics, and the decoder conceals the image within latent space bits randomly. The third network handles extraction. All networks are jointly trained to minimize differences between cover-stego and secret-reconstructed images. This method guarantees imperceptibility and withstands detection by StegExpose, albeit solely designed for lossless images.

In [56], Atique et al. proposed an automated steganography approach using an encoder-decoder architecture to embed gray images into color images. The encoder generates stego images, and the decoder retrieves hidden images. Dual CNN branches in the encoder extract features from cover and secret images, merged to create stego images. This method supports substantial payload (8 bpp) but introduces noticeable distortion detectable by human visual perception. Constraints on hidden image types and sizes exist.

Abhishek Das et al. [57] built upon methods from [55] and [58], aiming to hide three secret images within one cover. Their encoder, called "Prep Networks," comprises three CNNs that upscale secret images to match the cover size. Another CNN, the "hiding network," concatenates pre-processed secret images with the cover, yielding stego images. Decoding employs multiple reveal networks, each trained independently for message recovery. This approach faces challenges, including a significant loss rate for recovered secrets and stego images. In [59], Nandhini et al. introduced a steganography technique concealing color images within other color images, leveraging auto-encoders. Pre-processing networks extract features from both cover and secret images through convolutional layers. Extracted features merge in a central layer before feeding into a decoder network for stego image reconstruction. Payload capacity reaches 1, with a PSNR of 34.55 dB as an imperceptibility metric.

Biswarup et al.'s [60] method employs CNNs to differentiate edge pixels from others. Edge pixels are prioritized in the embedding process based on their resistance to attacks. A supervised CNN identifies edge pixels from a secret image transformed into a 1D array of bitstreams. Edge images undergo masking and threshold testing before embedding secret data in the 5 least significant bits of edge maps. This approach outperforms some spatial domain methods, with PSNR ranging from 36.16 dB to 50.12 dB.

In [61], Toan et al. proposed CNN-based steganography utilizing two networks. The first, employing a U-net architecture, crafts container images. The second network, comprising a six-layer CNN, retrieves the hidden image. This approach reduces training time by half, yielding a lower mean square error of pixel differences than [55]. Akshay et al.'s [62] steganography method, akin to U-net and based on the Dilated Inception Net Module, conceals one image within another. The network, comprised of convolutional layers, concatenation layers, and Dilated Inception Net Blocks, learns to extract hidden image features and incorporate them within cover features. However, this approach advances [56] in terms of Stego quality.

3.4.2 GAN-Based Image Steganography

GAN-based image steganography employs a generator to embed secret data within a cover image while a discriminator learns to distinguish between cover and stego images [63]. Through adversarial training, the generator is optimized to minimize visual distortions, ensuring the stego image remains indistinguishable from the original while maintaining a high payload capacity; some frameworks also include an extractor network to accurately retrieve the hidden data.

In this context, Volkhonskiy et al. introduced the first GAN-driven image steganography model called SGAN [64]. This approach utilizes a DCGAN [65] to convert random noise into cover images, followed by conventional embedding techniques to encode confidential data into these generated covers, thereby producing Stego images. However, the use of DCGAN in SGAN resulted in training instability, leading to suboptimal transparency conforming to steganography criteria [66]. To overcome this shortcoming, Shi et al. [67] replaced DCGAN with WGAN [68], giving rise to SSGAN, which exhibits visually more realistic cover images compared to SGAN and demonstrates partial resilience against training instability.

Lui and al. [69] proposed ACGAN, a variant of a GAN based steganography technique that takes as input noise and label classes to generate new images by linking the secret data to the corresponding class/category label. The receiver trains the ACGAN with the same noise, label categories, dataset to get the same discriminator responsible for extracting the secret information.

Tang et al. [70] combined GAN with Syndrome-Trellis Codes (STC) [71] encoding to propose ASDL-GAN, although it displayed comparatively inferior performance versus conventional adaptive steganography techniques. Nevertheless, the introduction of the Ternary Embedding Simulator activation function elongated the training cycle, in contrast to traditional steganography methods. Thus, Yang et al. [72] replaced the Ternary Embedding Simulator in ASDL-GAN by a Tanh resulting in the UT-SCA-GAN model, that enhances performance and minimizes training time.

Wang et al. [73] proposed the Stego-WGAN framework. Unlike SGAN and SSGAN, Stego-WGAN incorporates both the stego image and the original image as inputs for the discriminative network. This technique ensures not only the concealed image's suitability for secret information embedding but also maintains visual fidelity between the Stego-image and the original one.

Yu et al. [74] proposed to enhance the GAN architecture by including an attention mechanism on top of the architecture, which increased the accuracy of the recovery process.

Similarly, Zhang et al. [66] proposed a model that fuses a GAN, attention mechanisms, and image interpolation techniques to generates images containing confidential information without using cover images and used GANs for information embedding, thus having better anti-detection capability. Moreover, the IDGAN uses an attention mechanism to improve the image details and clarity and optimizes the steganography effect through an image interpolation algorithm.
in [75] Ambika V et al. propose an attention vector-guided GAN (AVG-GAN) for coverless image steganography on medical images that preserves diagnostic features by applying transformations only in non-critical regions. Unlike traditional GAN-based steganography which globally alters the image and can distort important features used for disease diagnosis, the AVG-GAN employs an attention vector to selectively modify only non-discriminative areas, thereby maintaining high image quality. Evaluated on brain tumor, glaucoma, and ovarian cancer datasets, the proposed method achieved a PSNR of approximately 40 dB, RS-BPP around 6.3, WPSNR nearly 38.5 dB, and an SSIM of about 0.98. Additionally, it maintained classification accuracy with only a 1–2% difference between the original and transformed images, and reduced embedding capacity by less than 2% confirming its effectiveness in preserving the clinical utility of medical images while securely embedding secret information.

The generative nature of these techniques presents significant challenges in medical image steganography, as these images often contain intricate details crucial for accurate diagnosis, and any loss of detail can compromise their diagnostic value. In such cases, encoder-decoder architectures are preferable because they better preserve the fine details necessary for effective medical analysis.

3.5 Summary of literature review for medical image steganography

Table 3.2 offers an overview of various image steganography studies. The table is organized into the following columns:

- Approach/Method: Lists the technique along with its associated reference.
- Key Techniques: Highlights the primary methods employed (e.g., encryption, transform domain techniques, neural networks).
- Performance Metrics: Summarizes the reported performance (e.g., PSNR values, payload capacity).
- Remarks: Provides brief insights on strengths, limitations, or notable trade-offs.

Table 3.2: Summary of data-hiding approaches for medical-image steganography

Approach / Method	**Key Techniques**	**Performance Metrics**	**Remarks**
BOOST [35]	Nuclear-spin generator for encryption; LSB embedding in non-black pixels	PSNR $>$ 113 dB; Payload 0.74 bpp	Very high visual quality and capacity; computational overhead may be a concern
Romany et al. [36]	RSA encryption; Ripplet transform; LSB substitution; adaptive OPAP	PSNR 49–56 dB	Robust against RS attacks; quality–payload trade-off
Genetic Algorithm / Optimum Pixel Similarity [37]	GA-based pixel selection; LSB embedding	PSNR 47.4 dB; embeds 10 000 chars in 256×256 images	Balances imperceptibility and capacity without compression
Chowdhuri et al. [38]	SVM for ROI/NROI detection; IWT; circular-array embedding	PSNR 64 dB	Secure embedding within medically relevant regions
SVNN approach [39]	Support Vector Neural Network; contourlet transform; HL-band embedding	PSNR up to 89.3 dB	Reversible; robust even with noise
Quantum Walks [43]	Quantum walks for pixel selection; chaotic LSB embedding	PSNR 44.1–44.4 dB	Extra security via quantum keys; moderate visual quality
Deep-Learning-Based Approaches			
Encoder–Decoder [55]	Autoencoders for resizing, hiding, extraction	Imperceptible (near-lossless) stego images	Joint training minimises differences
Encoder–Decoder [56]	Dual CNN branches; automated embed/extract	Supports payload 8 bpp; some visible distortion	Higher capacity introduces artifacts
GAN models (SGAN, SSGAN, ACGAN) [64, 67, 69]	GANs with adversarial training; attention mechanisms	Mixed PSNR; better training stability (WGAN)	Preserving fine medical details remains hard

3.6 Critical Analysis of Medical Image Steganography Techniques

The literature on medical image steganography demonstrates a broad array of methods, each offering distinct benefits and presenting specific challenges.

3.6.1 Traditional Approaches

Techniques such as BOOST [35] exhibit exceptionally high PSNR values (exceeding 113 dB) and robust payload capacities, indicating that the stego images maintain high visual fidelity. However, these methods often rely on complex multi-stage processes—such as pseudo-random generators based on nuclear spin generators for encryption followed by LSB embedding—which can incur significant computational overhead. This may limit their scalability when processing large medical datasets, a critical factor in real-world clinical applications.

Romany et al. [36] provide a comprehensive approach that combines RSA encryption, Ripplet Transform, and adaptive pixel adjustment. While their method demonstrates robustness against RS attacks and effective trade-offs between quality and capacity (PSNR ranging from 49 to 56 dB), the reliance on multiple sequential processes can introduce latency and complicate implementation. Similarly, the Genetic Algorithm-Optimum Pixel Similarity approach [37] efficiently embeds large amounts of data without compression. Despite its innovation, it achieves moderate PSNR (around 47.41 dB), highlighting the inherent trade-off between payload capacity and imperceptibility.

Other methods, such as those employing SVM for ROI/NROI detection [38] and SVNN with Contourlet Transform [39], focus on preserving diagnostically significant regions. While they offer higher PSNR values (up to 89.33 dB) and targeted embedding, these techniques depend heavily on accurate segmentation and can be sensitive to noise, potentially affecting their robustness in varied clinical conditions.

3.6.2 Deep Learning-Based Approaches

Encoder-decoder architectures [55, 56] leverage the power of deep neural networks to automate feature extraction and optimize the embedding process. These models, when jointly trained, can produce nearly lossless stego images and offer significant advantages in preserving image quality. However, they are typically designed for specific image types (often lossless) and require large, annotated datasets, which may not be readily available in medical imaging. GAN-based steganography methods, including SGAN and its derivatives [64, 67], introduce adversarial training to improve anti-detection capabilities. Although these models can generate realistic cover images and effectively hide data, they frequently suffer from training instability and may struggle to preserve fine details crucial for diagnostic accuracy. The incorporation of attention mechanisms [66, 74] has shown promise in addressing some of these issues, yet the challenge of ensuring consistent performance across diverse medical images remains unresolved.

3.7 Conclusion

The literature review outlined in this chapter demonstrated that traditional steganographic techniques excel in security and high visual fidelity but often come with the drawback of high computational complexity and scalability concerns. In contrast, deep learning-based methods offer automated, high-quality embedding but require substantial computational resources and large datasets, and they may face issues related to training stability and detail preservation.

In the following chapter, we delve into our primary contribution, which employs advanced deep learning techniques in the realm of medical image steganography. This method is designed to optimize the selection of embedding areas, ensuring that hidden information is placed in appropriate regions while utilizing a secure concealment technique based on the DCT transform. This approach maximizes both security and integrity while minimizing the risk of influencing medical diagnosis.

Chapter 4

Mask-RCNN Based High-Capacity Data Hiding Model for Medical Images

4.1 Introduction

This study's contribution lies in the creation of DICOM files that seamlessly integrate patient information into medical images with an exceedingly minimal impact—almost inconsequential—in order to safeguard against misdiagnosis, all achieved through the application of steganography principles. The devised approach involves concealing patient data within areas of the medical image that hold marginal relevance. Here, ”insignificant areas” refer to regions devoid of crucial medical data, such as the black segments found in grayscale DICOM images. Non-essential image regions are first located with Mask R-CNN- [76] a two-stage deep-learning model that detects, classifies, and produces pixel-accurate masks for every object in the scene. The model trains end-to-end with a combined loss over these outputs, achieving state-of-the-art accuracy for tasks like autonomous driving, medical imaging, and augmented reality. Sensitive medical information is then covertly embedded into these low-salience areas using DCT-based steganography.

4.2 The proposed Method

A comprehensive visual representation of the proposed methodology is depicted in Figure 4.1, outlining three fundamental stages: Neural network training, Embedding, and Extraction.

4.2.1 Neural network training

The key-concept in the proposed method is the detection of insignificant areas in medical DICOM images which will be exploited to conceal sensitive information. We assume that the best way to detect these regions is by correctly detecting the main objects in the image. In the literature, CNN-based methods outperform traditional techniques in the detection and segmentation of objects inside images. Thus, we adopt Mask-RCNN architecture [77], as one of the efficient techniques especially in the field of medical images. This architecture is proposed to detect the main objects which represent the significant area that should be kept safe dur-

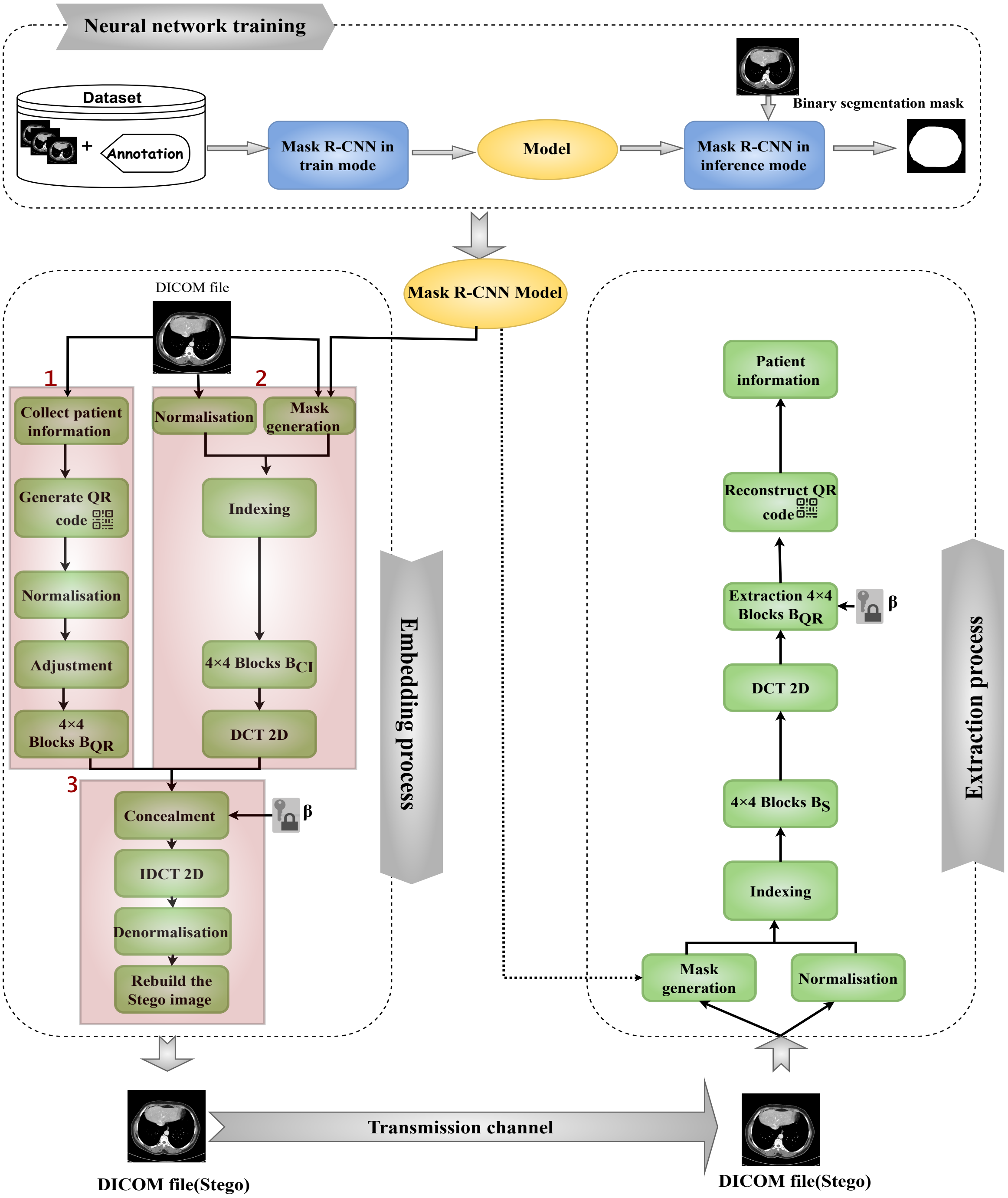


Figure 4.1: Overview of the proposed steganography method

ing information embedding. To train the Mask-RCNN model to obtain binary segmentation masks, we use the architecture depicted in Figure 4.2 on various DICOM files datasets. This architecture is divided into two stages :

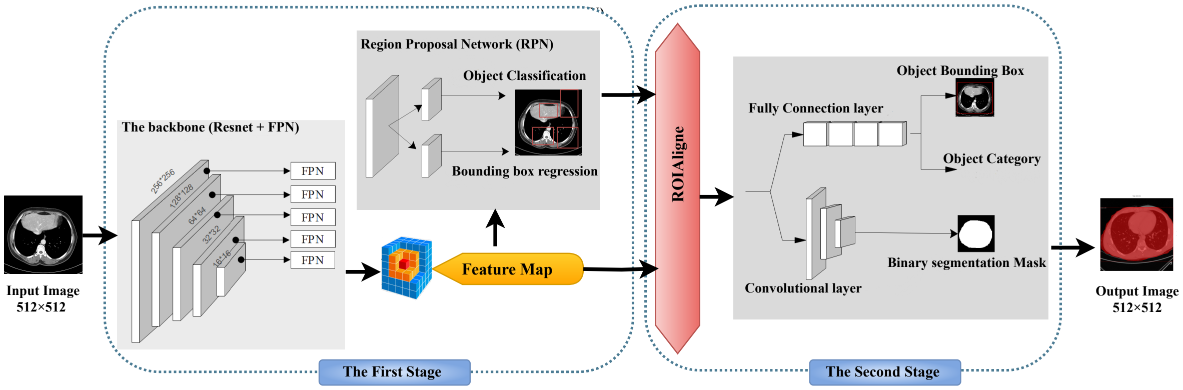


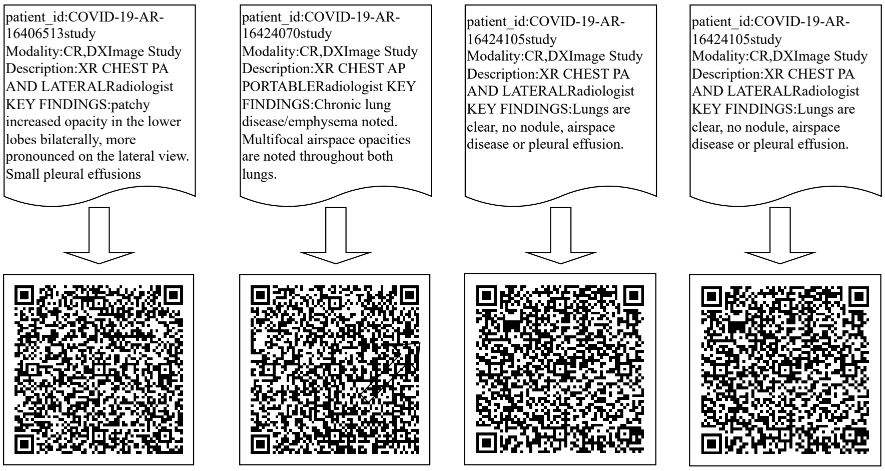
Figure 4.2: Mask R-CNN architecture

1. **First Stage**: It takes an image and produces a feature map and regions proposals. The feature map is obtained through a series of operations conducted on the original image by a backbone constructed from CNN layers (ResNet and feature pyramid. Region proposal Network [77] takes the feature map and produces regions that may contain objects.

2. **Second Stage**: it consists of aligning regions of Interests (RoIalign). It takes as input feature map and region proposals and generates as output the fixed size regions of interest from region proposals, and three parallel branches for predicting: object category, instance bounding-box, and binary segmentation masks [77].

This architecture is trained on various datasets to determine the binary segmentation mask where 1's represents object pixels (significant region) and 0's represents the background of the image (i.e. the insignificant area). The loss function used to train this model is defined by Equation (4.1) [77].

$$L = L_{class} + L_{box} + L_{mask} \tag{4.1}$$



patient_id:COVID-19-AR-16406513study Modality:CR,DXImage Study Description:XR CHEST PA AND LATERALRadiologist KEY FINDINGS:patchy increased opacity in the lower lobes bilaterally, more pronounced on the lateral view. Small pleural effusions

patient_id:COVID-19-AR-16424070study Modality:CR,DXImage Study Description:XR CHEST AP PORTABLERadiologist KEY FINDINGS:Chronic lung disease/emphysema noted. Multifocal airspace opacities are noted throughout both lungs.

patient_id:COVID-19-AR-16424105study Modality:CR,DXImage Study Description:XR CHEST PA AND LATERALRadiologist KEY FINDINGS:Lungs are clear, no nodule, airspace disease or pleural effusion.

patient_id:COVID-19-AR-16424105study Modality:CR,DXImage Study Description:XR CHEST PA AND LATERALRadiologist KEY FINDINGS:Lungs are clear, no nodule, airspace disease or pleural effusion.

Figure 4.3: Examples of patient data with their corresponding QR codes

where L_{class} is the Classification Loss, L_{box} is the Bounding Box Regression Loss, and L_{mask} is the mask Loss.

4.2.2 Embedding Process

The process of embedding, also known as the concealment process, encompasses a series of sequential steps aimed at seamlessly integrating confidential patient-related data into DICOM images, ultimately giving rise to a Stego image. To elaborate, the focal point of this embedding is the inconspicuous area previously identified using the Mask-RCNN model. This entire procedure can be delineated into three distinct phases, outlined in the subsequent subsections.

Sensitive information preprocessing

In this first phase, the sensitive patient information is retrieved from the DICOM file and transformed into a QR code image via QR Code generator. Figure 4.3 depicts some examples of patient information with their corresponding QR codes generated using the Zxing library [78].

The Generated QR code image is then normalized using Equation 4.2 to convert the values within the image to a range between 0 and 1. After that, the normalized image is adjusted using Equation 4.3, where the value of α is fixed experimentally (α=0,02). The adjusted image is then divided into 4×4 mini matrices called Blocks (B_{Qr}) that are hidden one by one during the concealment phase.

$$\text{Norm}(\text{Img}) = \frac{\text{Img} - \min(\text{Img})}{\max(\text{Img}) - \min(\text{Img})}\,. \tag{4.2}$$

Equation (4.2) applies min–max normalisation (feature scaling) to the image matrix Img:

- min(Img) is the smallest grey-level (or colour-channel) value in the image.
- max(Img) is the largest value in the image.
- Subtracting the minimum shifts every pixel so the new minimum becomes 0.
- Dividing by the dynamic range $\big(\max - \min\big)$ stretches the data so the new maximum becomes 1.
- After scaling, every pixel intensity lies in the interval $[0, 1]$; contrast is preserved but rescaled, which is convenient for many algorithms and neural-network inputs.

$$Adjustment(Msg) = \begin{cases} 1-\alpha, Msg \geq 1 \\ \alpha, Msg = 0 \end{cases} \tag{4.3}$$

Equation 4.3 applies a simple smoothing / clipping transform to a binary message symbol (B_{Qr}) :

- $\alpha \in (0, 1)$ is a small, tunable constant.
- If the symbol is logic "1" (or any value ≥ 1), its adjusted value becomes $1 - \alpha$ (slightly less than one).
- If the symbol is logic "0", its adjusted value becomes α (slightly greater than zero).
- This prevents the adjusted data from taking the extreme values exactly 0 or 1, which can help to conceal the B_{Qr} within the cover image's insignificant DCT coefficients.

Cover preprocessing

The cover, derived from the DICOM file, is subjected to a dual stage preprocessing procedure. Initially, the mask-RCNN model is employed in inference mode to pinpoint the inconsequential regions within the cover, which yields a binary segmentation mask. Subsequently, given that DICOM images are frequently encoded in 16-bit grayscale, the cover is normalized using Equation 4.2 to effectively remap its values within the [0, 1] interval.

The resulting mask obtained from the first preprocessing step is utilized to determine the indices corresponding to the insignificant areas within the normalized cover image, referred to us as Cover insignificant (C_{ins}) image. These insignificant areas are subsequently divided into blocks with dimensions of 4×4, denoted as (B_{Cins}). The transformation coefficients of (B_{Cins}) are then computed using the two-dimensional discrete cosine transform function (2D DCT) [79], resulting in the generation of DCT coefficient of the cover insignificant (B_{DCins}) blocks. The 2D DCT for a matrix I (with dimensions M×N) is calculated using the formula specified in Equation 4.4.

$$C(u,v) = \alpha(u).\alpha(v) \times \left[\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(m,n) \times cos\frac{(2m+1)u\pi}{2M} cos\frac{(2n+1)v\pi}{2N} \right] \begin{array}{l} 0 \leq u \leq M-1 \\ 0 \leq v \leq N-1 \end{array} \tag{4.4}$$

Where: $a(u) = \begin{cases} 0 \leq u \leq M-1 \\ 0 \leq v \leq N-1 \end{cases}$

(m,n) and $I(m,n)$ correspond to the position values and the pixel value at position (m,n) in the spatial domain respectively. $C(u,v)$ isthe corresponding position value and the frequency coefficient at position (u,v) in the transform domain.

Information Concealment

During this stage, the 4×4 blocks originating from the cover, denoted as (B_{DCins}), and the 4×4 blocks representing the secret message, referred to as (B_{Qr}) and generated in the prior phase, are merged together—specifically, (B_{Qr}) is concealed within (B_{DCins})—resulting in the formation of the corresponding block (B_s) within the Stego image. This concealment process is executed through the application of Equation 4.5.

$$B_S = IDCT2(Qun(B_{DCins}, B_{Qr})) \tag{4.5}$$

Where:

- B_s: Represents the block of the Stego image resulting from the concealment process.
- IDCT2 is the inverse two-dimensional discrete cosine transform function (2D-IDCT). This function is explained by Equation 4.6.

$$I(m,n) = \alpha(u).\alpha(v) \times \left[\sum_{u=0}^{M-1} \sum_{v=0}^{N-1} C(u,v) \times cos\frac{(2m+1)u\pi}{2M} cos\frac{(2n+1)v\pi}{2N} \right] \tag{4.6}$$

- Qun: Denotes the function that integrates the B_{Qr} block into B_{DCins} and generates the pre-Stego blockB_{QS}. This block is computed using the formula depicted in (Eq.4.7).

$$B_{QS}^{i,j} = k + (\frac{4}{\beta} \times B_{Qr}^{i,j}) \quad ; \qquad (\frac{4k}{\beta}) <\mid B_{DCins}^{i,j} \mid< \frac{(4k+1)}{\beta} \tag{4.7}$$

  where:

  - i,j: Are respectively the i-th and j-th ligne and column in the block.
  - β: Is the number of intervals that satisfy the cover coefficients on the interval of [0, 4].
  - $k \in 1, 2, 3, ..., \beta - 1$.

Subsequently, the inverse two-dimensional discrete cosine transform function (2D-IDCT) is employed on B_{QS}, yielding the ultimate B_S that characterizes the Stego blocks. Following this, the Stego blocks are amalgamated to forge the Stego image, which subsequently undergoes de-normalization to confine values within the [0, 65535] range. This critical step guarantees that the Stego image adheres to the 16-bit DICOM file format, conserving the initial encoding scheme. An illustrative demonstration of the concealment process can be found in Figure 4.4. The main idea illustrated in Fig. 4.4 is to divide the DCT-coefficient range $[0, 4]$ into β equal sub-intervals, since most B_{DCins} values fall within this span.

Embedding example To embed the secret value $B_{QR} = 0.7$ into the cover coefficient $B_{\text{DCins}} = 1.2$:

1. **Locate the sub-interval.** Since $1.2 \in [1.0,\ 1.5[$, the left-hand bound is $k = 1.0$.

2. **Compute the stego coefficient.** Using the quantisation rule

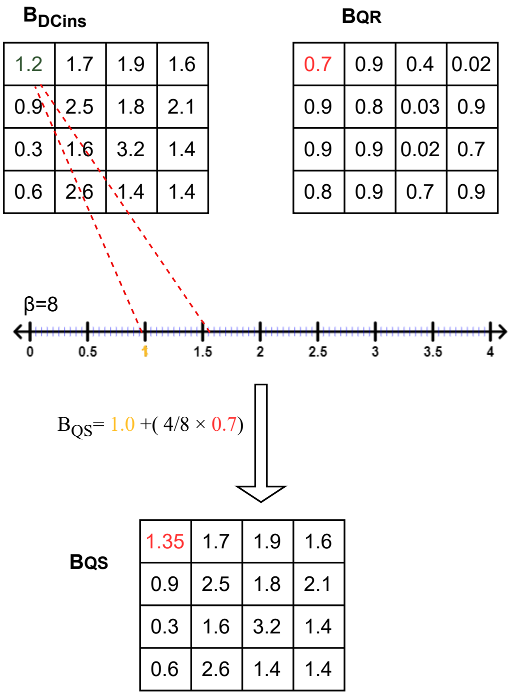
   $$B_S = k + (\frac{4}{8} \times B_{QR})$$

   we obtain

   $$B_S = 1.0 + (\frac{4}{8} \times 0.7) = 1.35$$

   Note that B_S also lies within the same interval $[1.0,\ 1.5[$.

Algorithm 4.1 prescribes the sequential steps that constitute the embedding process.



1.2	1.7	1.9	1.6
0.9	2.5	1.8	2.1
0.3	1.6	3.2	1.4
0.6	2.6	1.4	1.4

0.7	0.9	0.4	0.02
0.9	0.8	0.03	0.9
0.9	0.9	0.02	0.7
0.8	0.9	0.7	0.9

1.35	1.7	1.9	1.6
0.9	2.5	1.8	2.1
0.3	1.6	3.2	1.4
0.6	2.6	1.4	1.4

Figure 4.4: A visual illustration of the hiding technique when $\beta = 8$

Algorithm 4.1 Embedding Algorithm

```
Input : Medical image with size m × n, the QR code image of patient information
Output :Steganography medical image with size m × n
Generate the binary segmentation mask
Normalize the cover using equation 4.2
Normalize and adjust the QR code image using equations 4.2, 4.3.
Get the index of 0 in the binary mask and generate the insignificant area of the Cover
Split the insignificant cover and the QR code into 4×4 blocks
Calculate the DCT transformation coefficients to cover insignificant block using 4.4
Conceal the QR code block in the DCT coefficients with equation 4.7
Calculate the IDCT of the result blocks of step 7, to transform the image into the spatial domain
Reconstruct the Stego image
```

4.2.3 Extraction Process

The extraction process serves as the reverse of the concealment process, with the aim of recovering the patient data hidden within Stego images. This retrieval is exclusively authorized for users possessing the requisite key. The extraction unfolds across two distinct phases: Mask Generation and the Extraction Process.

In the Mask Generation phase, the Stego image is subjected to the Mask R-CNN model operating in inference mode. This operation generates a binary segmentation mask tailored to pinpoint insignificant regions. This mask, a critical tool, facilitates the identification and indexing of these areas—precisely where the secret data has been concealed.

Moving to the Extraction Process, an essential preliminary step involves normalizing the Stego image to ensure pixel values are confined within the [0, 1] range. Leveraging the binary mask associated with the Stego image, the positioning of pixels utilized for concealment is discerned. This determination is pivotal in the creation of the insignificant Stego matrix. Subsequently, the matrix is partitioned into 4x4 blocks and subjected to the Discrete Cosine Transform (DCT) function, transitioning it into the frequency domain. The outcome is a set of transformation coefficients. By employing the inverse merge operation ($IQun$), these coefficients facilitate the extraction of the mini block corresponding to the QR code, denoted as B_QR. The extraction equation is formally expressed as detailed in Equation 4.8.

$$B_{QR} = IQun(DCT2(B_S)) \tag{4.8}$$

where the inverse merge equation (IQun) is calculated by Equation 4.9.

$$B_{QR} = (B_S - k) \times \beta/4; \qquad (\frac{4k}{\beta}) <\mid B_S^{i,j} \mid< \frac{(4k+1)}{\beta} \tag{4.9}$$

Finally, the resulting B_Q blocks are then concatenated to produce the QR code of the patient information. The reverse process is summarized by Algorithm 4.2

Algorithm 4.2 Extraction Algorithm

```
Input :Stego medical image with size m × n
Output :Private patient information.
1: Generate the binary segmentation mask with size m × n
2: Normalize the Stego
3: Get the index of 0 in the binary mask and generate the insignificant Stego area
4: Split the insignificant Stego into 4×4 blocks
5: Calculate the 2DCT transformation coefficients for Stego insignificant block using
   Eq. 4.4
6: Extract the QR code block from the DCT coefficients using the formula 4.9
7: Recreate the private patient information QR code
```

4.3 Experiments

4.3.1 CHAOS Dataset

CHAOS dataset aims to segment abdominal organs (liver, kidneys, and Spleen) using CT and MRI data [80]. It consists of two datasets, Each one corresponding to a series of DICOM images. The first comprises CT images of 40 different patients with a healthy liver. The patient orientation and alignment are the same for all the data sets. The data consists of 16-bit DICOM images with a resolution of 512x512, an x-y spacing of 0.7-0.8 mm, and an inter-slice distance

(ISD) of 3 to 3.2 mm. The second database includes 120 DICOM data sets from two different MRI sequences, each of which is being routinely performed to scan the abdomen using different radiofrequency pulse and gradient combinations. The data sets are acquired by a 1.5T Philips MRI, which produces 12 bit DICOM images having a resolution of 256 x 256. The ISDs vary between 5.5-9 mm (average 7.84 mm), x-y spacing is between 1.36 - 1.89 mm (average 1.61 mm) and the number of slices is between 26 and 50. we have randomly selected 1200 DICOM images from the CHAOS dataset. These images are divided into 1023 images (11 patient images) for training, 199 images for validation (3 patient images) 10 images for testing. We manually created the annotation of these images using VIA Annotation Software [81], with the intention to make a semantic segmentation to separate the foreground that we consider as Significant area from the background that we consider as insignificant area.

4.3.2 Mask RCNN model training and evaluation

We leveraged transfer learning to prepare a Mask-RCNN model that detects insignificant areas in DICOM images. We started by fine-tuning the pre-trained weights of the MS COCO model [82], which is a large object detection and instance segmentation dataset that comprises 328k images with 91 labeled categories. To fine-tune this model, we used the implementation of MASK-RCNN proposed by Matterport in [83] and we started the training on the CHAOS dataset with the MS COCO weights to produce a variation of the network that targets our detection goals.

Table 4.1 presents the configuration details for training our variant of the Mask-RCNN model [84]. The parameters listed in the table include *the backbone architecture*, which is the ResNet101 architecture in this case. The *batch size*, which is the number of images used in each training iteration, is set to 4. The *Feature Pyramid Network (FPN)* used for classification is a fully connected layer with a size of 1024. The *learning rate* and *momentum parameters* are set to 0.001 and 0.9 respectively. And to prevent overfitting, we set the *weight decay* parameter to 0.0001. The *RPN Train Anchors per Image* parameter, which refers to the number of anchors used in the region proposal network (RPN) during training, is set to 256. and the *Images*

Parameter	**Value**
Backbone	resnet101
BATCH_SIZE	4
FPN_CLASSIF_FC_LAYERS_SIZE	1024
PU_COUNT	1
IMAGES_PER_GPU	4
LEARNING_MOMENTUM	0.9
LEARNING_RATE	0.001
RPN_TRAIN_ANCHORS_PER_IMAG	256
STEPS_PER_EPOCH	10
VALIDATION_STEPS	50
WEIGHT_DECAY	0.0001

Table 4.1: Mask R-CNN configuration

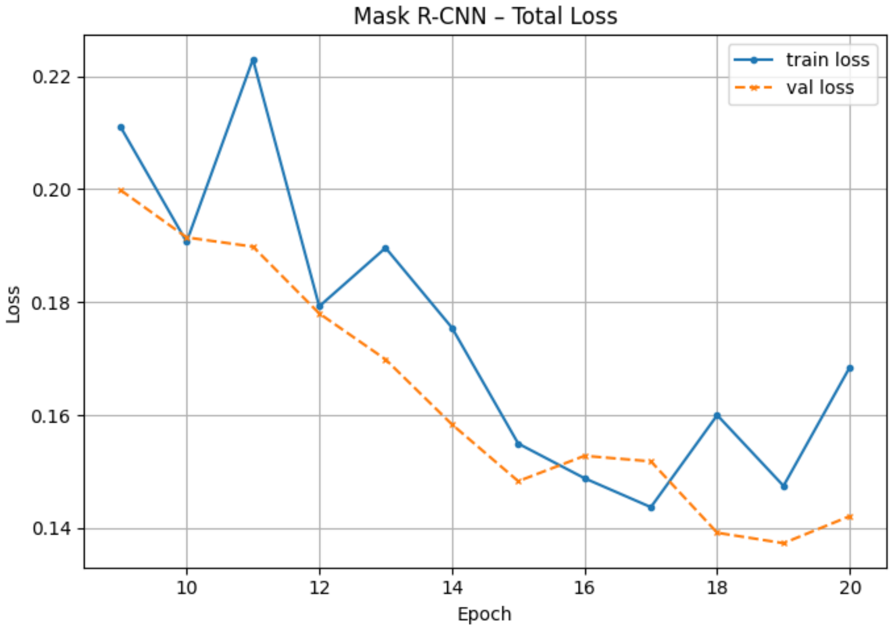
Per GPU parameter is set to 4, indicating that each GPU processes 4 images at a time. The *Steps Per Epoch* parameter is set to 10, and the *Validation Steps* parameter is set to 50. These parameters control the training process and the number of training and validation iterations.

The training was conducted on a machine empowered by Core I7 and 10th generation processor, Intel UHD graphics, and 16 GB of RAM. The training is done in 20 epochs (8 epochs for the head and 12 epochs to fine-tune all layers). The curves of the training and validation losses are presented in Figure 4.5. Sub-figure (a) displays the general training and validation losses, while sub-figure (b) displays the losses of the MR-CNN mask training and validation. The MR-CNN general loss is recorded as 0.1291 at the end of the training, and the MR-CNN mask loss is noted as 0.0450.

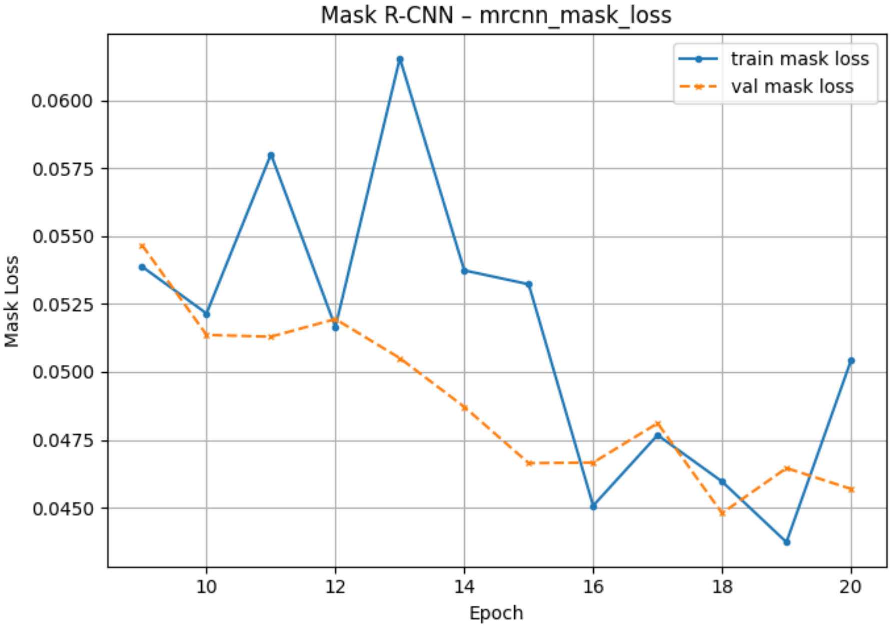
We evaluated the overlap between the annotated and the generated masks of the validation dataset using the Intersection Over Union (IoU) metric [85]. IoU is calculated using Equation 4.10.

$$(IOU) = \frac{Area\,of\,intersection\,of\,two\,masks}{Area\,of\,Union\,of\,two\,masks} \tag{4.10}$$

A lower value of IoU indicates inadequate prediction (i.e. poor prediction) [86], whereas a value of 1 represents an entirely accurate prediction. The validation process yielded an Average IoU of 0.9146, signifying that the model can be safely used.



(a) MR-CNN Loss



(b) MR-CNN Mask Loss

Figure 4.5: Model training and validation losses

4.3.3 Imperceptibility measurement

The second part of the experiment was dedicated to the embedded process that uses the trained MR-CNN model. We tested the process on 10 images. After extracting patient information from the DICOM files and transforming them into QR codes, we applied the embedding method described above to conceal the QR codes (Message) in the Cover images to obtain Stego images as a first step. Then, we applied the extraction process to retrieve secret messages from the Stego images as a second step.

Table 4.2 shows the obtained results (the cover, message, Stego, and the Retrieved message) for a sample of 4 DICOM files. Based on visual inspection, it appears that there is no discernible disparity between the original (Cover and Message) and generated images (Stego and retrieved Qr code). However, we used the PSNR (Peak Signal to Noise Ratio) metric to evaluate the visual quality of the generated images and The CNN (Normalized Correlation Coefficient) metric to check the similarity between the cover mask and the stego mask obtained by the MR-CNN model, which affects the correctness of the extraction process.

Furthermore, PSNR is calculated in decibels between two images using Equation 4.12:

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N} \tag{4.11}$$

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \tag{4.12}$$

Where M and N are the numbers of rows and columns in the input images. R is the maximum fluctuation in the input image data type.

Moreover, NCC is used in our case to measure the robustness of the model and estimate the difference between the cover mask and the Stego mask, NCC value adjacent to 1 implies that the two masks are similar. NCC formula is given in Equation 4.13 :

$$NCC = \frac{\sum_{i=0}^{m} \sum_{j=0}^{n} (MC - \mu MC)(MS - \mu MS)}{\left(\sqrt{\sum_{i=0}^{m} (MC - \mu MC)^2}\right) \left(\sqrt{\sum_{i=0}^{m} (MS - \mu MS)^2}\right)} \tag{4.13}$$

Where, μMC and μMS are the mean pixel values of the cover mask and the Stego mask,

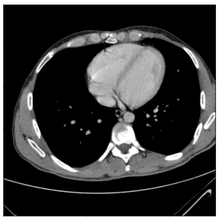
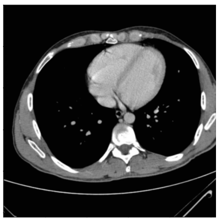
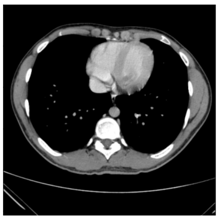
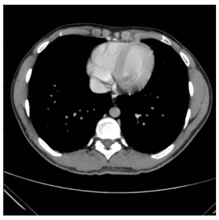
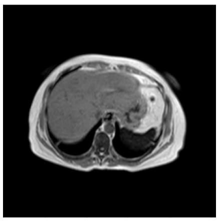
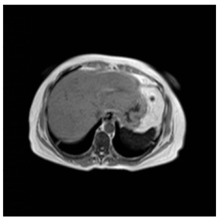
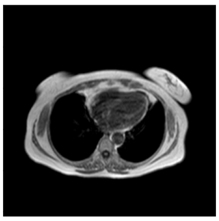
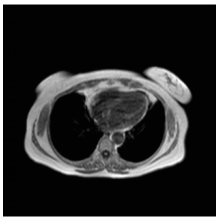
	The cover	The message	The stego	The Retrieved Message
CT 1				
CT 2				
MRI 1				
MRI 2				

Table 4.2: The obtained results (the cover, message, Stego, and the Retrieved message) for a sample of 4 DICOM files

respectively. Table 4.3 presents the NCC and PSNR values for the tested DICOM images, along with their image sizes and scan types. The NCC values listed in the table fall within the range of 0.83 to 1 for all images, signifying that our MR-CNN models can predict identical masks from both the Cover and Stego images. This successful prediction enables the accurate detection of the insignificant area where concealment operations take place. The PNSR values depicted in the table 4.3, in case of the embedding parameters β and α are set to 1000 and 0.02 respectively, are ranging between 107.47 and 116.57, indicating that our method effectively conceals sensitive patient information with a high level of imperceptibility.

Patient ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Scan Type	CT	CT	CT	CT	CT	MRI	MRI	MRI	MRI	MRI
Image size (Pixel)	512 x 512	512 x 512	512 x 512	512 x 512	512 x 512	256 x 256	256 x 256	512 x 512	256 x 256	256 x 256
NCC	1.00	1.00	0.83	1.00	0.99	1.00	0.99	1.00	0.99	0.99
PSNR (dB)	**116.07**	**113.69**	**115.53**	**116.57**	**116.02**	**107.47**	**110.3**	**117.4**	**110.3**	**110.33**

Table 4.3: The NCC and PSNR values for the tested DICOM images.

To examine the effect of the β coefficient on the final quality of the Stego, we measured the PSNR by variating the embedding parameter β. Table 4.4 shows visual examples for various values of β.

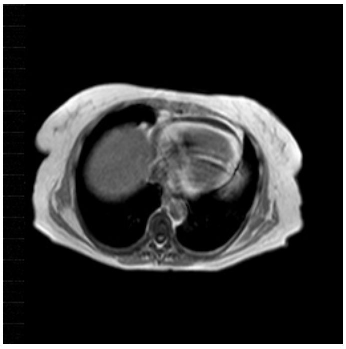
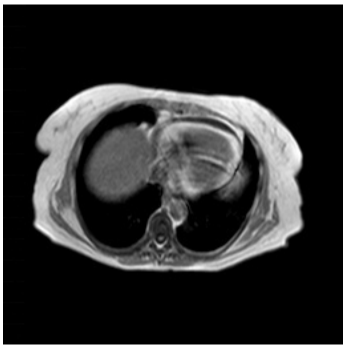
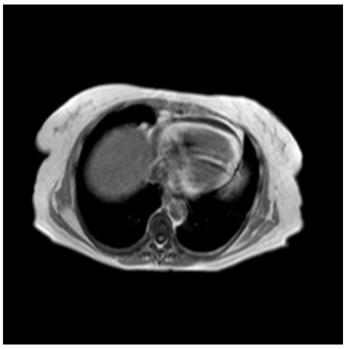
β	100	500	1000
PSNR (dB)	90.222	104.303	110.333
The stego			

Table 4.4: Visual examples when βtakes different values

Table 4.5 shows the obtained average results of the PSNR values between the Cover and Stego DICOM images of 10 patients for different β values. The value of PSNR was between 101.447 and 111.386 for $\beta = 500$. The minimum and maximum value of PSNR are 66.494 dB (β =10, patient MRI 1) and 117.400 dB (β = 1000, patient MRI 3) demonstrating that the highest value of β produces a high-quality Stego image and low β value produces low stego quality . Generally, PSNR higher values refer to the invisibility of higher quality.

Scan type		β										
		10	100	200	300	400	500	600	700	800	900	1000
CT	1	75.051	95.955	102.01	105.557	108.094	110.039	111.578	112.993	114.129	115.225	116.074
	2	72.661	93.551	99.64	103.166	105.729	107.661	109.222	110.618	111.748	112.888	113.695
	3	74.555	95.424	72.483	101.500	105.029	107.563	109.494	112.45	113.585	114.696	115.532
	4	75.136	96.313	102.442	105.998	108.55	110.517	112.07	113.444	114.644	115.723	116.571
	5	75.076	95.914	101.994	105.526	108.067	109.995	111.557	112.945	114.076	115.191	116.022
Averag		**74.4958**	**95.4314**	**95.7138**	**104.3494**	**107.0938**	**109.155**	**110.7842**	**112.49**	**113.6364**	**114.7446**	**115.5788**
MRI	1	66.494	87.347	93.425	96.949	99.500	101.447	102.984	104.394	105.533	106.643	107.473
	2	69.389	90.217	96.281	99.803	102.345	104.283	105.823	107.227	108.365	109.476	110.304
	3	76.472	97.313	103.385	106.904	109.447	111.386	112.925	114.321	115.472	116.581	117.400
	4	69.389	90.217	96.281	99.803	102.345	104.283	105.823	107.227	108.365	109.476	110.304
	5	69.39	90.222	96.29	99.816	102.362	104.303	105.847	107.253	108.393	109.505	110.333
Avrag		**70.2268**	**91.0632**	**97.1324**	**100.655**	**103.1998**	**105.1404**	**106.6804**	**108.0844**	**109.2256**	**110.3362**	**111.1628**

Table 4.5: The average results of the MSE values between the Cover and Stego DICOM.

4.3.4 Capacity and payload

	CT images					MRI images					
Patient id	1	2	3	4	5	6	7	8	9	10	**Average**
Capacity	0.5	0.54	0.50	0.48	0.49	0.63	0.61	0.54	0.72	0.58	**0.54**
Payload	0.248	0.296	0.248	0.256	0.248	0.72	0.8	0.16	0.64	0.8	**0.43**

Table 4.6: The resultant Capacity and payload.

Table 4.6 presents a detailed analysis of capacity and payload values for both CT (Computed Tomography) and MRI (Magnetic Resonance Imaging) images across ten different patient IDs. These values provide crucial insights into the performance and efficiency of a data hiding technique when applied to medical images. It's essential to consider these capacity and payload values when designing and implementing data hiding techniques for medical images, as they provide insights into the trade-off between data capacity and image quality in the context of medical data security. Capacity refers to the amount of secret data that can be embedded within the medical image while maintaining the image's visual quality and integrity. The capacity is calculated using the equation 4.14.

$$\rho = \frac{N_{\text{sel}}}{N_{\text{total}}}\,\beta, \tag{4.14}$$

where

- N_{sel} — total number of DCT coefficients
- N_{total} — total number of DCT coefficients in the entire image
- β — bits embedded per selected coefficient ($\beta = 1$).

Equation 4.14 yields the embedding capacityρ in *bits per pixel*. For CT images, the capacity ranges from 0.24 to 0.29, with an average capacity of 0.25. This indicates that, on average, approximately 25 % of the image can be utilized to hide secret data without significant degradation in image quality. For MRI images, the capacity varies from 0.16 to 0.72, with an average capacity of 0.62. MRI images show a slightly lower but still substantial capacity, with approximately 62 % of the image available for data embedding. Payload refers to the amount of secret data that is successfully embedded within the image. The payload is calculated using the equation 4.15 Equation 4.15 yields the Payload in *bits per pixel.* .

$$Payload = \frac{H_s\, W_s\, b}{N_{sel}}\, \beta \tag{4.15}$$

where

H_s, W_s Height × width of the secret image.

b Bits per secret pixel (8 for grayscale, 24 for RGB).

N_{sel} Number of cover coefficients you actually modify.

β Bits embedded in each selected coefficient (often 1).

It is a critical metric as it indicates how much data can be reliably hidden within the image. For CT images, the payload values range from 0.24 to 0.29, with an average payload of 0.50. This suggests that, on average, 50 % of the image can be effectively used to conceal secret data. For MRI images, the payload varies between 0.16 and 0.8, with an average payload of 0.5, mirroring the payload results of CT images.

4.3.5 Robustness analysis

A critical aspect of our proposed data hiding technique for medical images is its robustness against various types of noise attacks. To evaluate the resilience of our method, we subjected the stego images to three common noise attacks: Gaussian noise, uniform noise, and salt and pepper noise. Te robustness was assessed by retrieving the embedded QR code from the noise-afected stego images and measuring the similarity and visual quality through normalized cross-correlation (NCC) and peak signal-to-noise ratio (PSNR), respectively.

In 4.7, the NCC values obtained under Gaussian noise, uniform noise, and salt and pepper noise attacks were 0.3188, 0.3150, and 0.3193, respectively. Tese results indicate a moderate level of correlation between the original and retrieved QR codes post-attack, demonstrating the method's capability to withstand noise perturbations to a certain extent. Moreover, the PSNR values remained above 53 dB across all noise types, suggesting that the visual quality of the

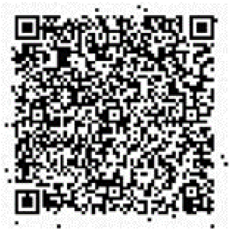
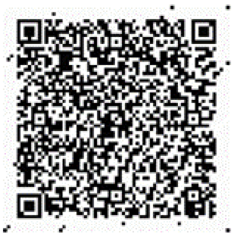
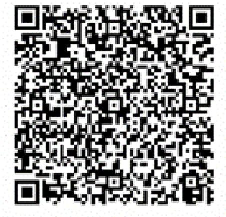
Stego attack	Gaussian noise	Uniform noise	Salt and pepper noise
The retrieved QR code			
NCC	0.3188	0.3150	0.3193
PSNR (dB)	53.36	53.33	53.39

Table 4.7: The robustness evaluation results under the noise attacks.

stego images is preserved well above acceptable thresholds, even in the presence of noise. This is significant as it ensures that the diagnostic value of medical images is not compromised due to embedding and subsequent noise attacks. In developing our data hiding approach for medical images based on the Mask-RCNN model, we meticulously balanced the trade-off between robustness, visual quality, and payload capacity. This delicate equilibrium ensures that while concealing data within the medical images, the method maintains resilience against various attacks, preserves high visual fidelity, and accommodates a significant payload for information embedding. By optimizing the embedding process and incorporating error correction coding techniques, we mitigate the risk of information loss and maintain the integrity of the stego images, even in the presence of noise or other forms of interference. Furthermore, careful selection of embedding parameters and compression algorithms allows us to strike an optimal balance between concealing capacity and visual imperceptibility, ensuring that the embedded data remains imperceptible to the human eye while maximizing the amount of information that can be securely hidden within the images. This careful consideration of trade-offs empowers our method to deliver robust and high-quality stego images suitable for secure transmission and storage of sensitive medical data.

Method	Image Size	Technique	Embedding Capacity(Bpp)	Payload (bits)	Best PSNR
Bozhidar S and Borislav S. [35]	336×336	Nuclear Spin Generator	0.75	83883	113.50
Subhadip M and al [51]	256×256	LSB and shell matrix	3	786432	48.42
Akshay K and al [62]	128×128	Deep learning	24	-	37.55
Atique R and al [56]	300×300	Deep learning	8	89910	36.58
Proposed Method	512×512	Deep learning	0.50	131524	**115.53**

Table 4.8: Compare the proposed method with other methods.

4.3.6 Comparison

Table 4.8 provides a comprehensive comparison between the proposed data hiding method and several other existing techniques, highlighting key parameters such as image size, embedding capacity, payload, and best PSNR (Peak Signal-to-Noise Ratio). Bozhidar S and Borislav S.'s [35] method employs a Nuclear Spin Generator on 336x336 images, achieving a relatively high embedding capacity of 0.75 Bpp. It successfully hides 83,883 bits of data while maintaining a remarkable PSNR of 113.50, indicating good image quality preservation. Subhadip M [51] and his team utilize a combination of LSB (Least Significant Bit) and a shell matrix technique on 256x256 images, resulting in a much higher embedding capacity of 3 Bpp. This approach allows the concealment of a substantial 786,432 bits of data. However, the PSNR of 48.42 indicates some loss in image quality compared to the previous method. Akshay K [62] and his collaborators employ deep learning on 128x128 images, achieving an impressive embedding capacity of 24 Bpp. While the exact payload is not specified, this approach prioritizes data capacity over PSNR, which is lower at 37.55, indicating some visual quality degradation. Atique R [56] and his team also apply deep learning, but on 300x300 images, resulting in an embedding capacity of 8 Bpp. They manage to conceal 89,910 bits of data with a PSNR of 36.58, indicating some trade-off between capacity and image quality. In contrast, the Proposed Method operates on 512x512 images using deep learning and achieves a reasonable embedding capacity of 0.50 Bpp, which allows the concealment of 131,524 bits of data. Notably, it outperforms the other methods [51] [62] [56] in terms of PSNR, attaining an impressive 115.53, signifying exceptional image quality preservation. In summary, the proposed method strikes a balance between embedding capacity and image quality, achieving a competitive capacity while maintaining outstanding PSNR, making it a promising choice for data hiding in medical images.

4.4 Conclusion

In this chapter, we presented our first contribution to the field of high-capacity data hiding in medical images. We introduced a novel approach that integrates Mask R-CNN segmentation with DCT steganography.

By fine-tuning and pretraining the Mask R-CNN model, our proposed method ensures consistency in identifying embedding regions across both original and stego images, thereby guaranteeing the accurate recovery of medical information. Experimental results demonstrate the effectiveness of this approach in achieving high-capacity data embedding with minimal distortion, as reflected by PSNR values ranging from 70 to 115, ensuring imperceptibility to the human eye. Additionally, comparative analysis highlights the superior performance of our method over existing alternatives.

In the next chapter, we present our second contribution, which explores another aspect of deep learning-based image steganography, where a CNN-based architecture is solely responsible for both concealment and extraction.

Chapter 5

Quality-Aware DNN-based Framework for Medical-Image Steganography

5.1 Introduction

This chapter presents the description of our second research contribution to the field of medical image steganography and its experimental results. Our second contribution focuses on developing a deep learning-based approach where a CNN architecture is fully responsible for both the concealment and extraction of hidden information within medical images. Unlike traditional steganographic methods that require manual optimization for embedding security and quality, our proposed approach leverages deep neural networks to learn an optimal embedding strategy while preserving medical image integrity.

This study specifically aims to conceal a medical image containing sensitive diagnostic information within another medical image that does not contain such information, utilizing the structure of a single DICOM file, which is inherently designed to store a series of images (or ”frames”) together. This approach ensures that confidential medical data is securely embedded within a diagnostically neutral image.

To evaluate our method, we conducted extensive experiments on medical imaging datasets, assessing the performance of our CNN-based framework in terms of embedding capacity, imperceptibility, and secret image recovery accuracy. We also compare our approach with existing steganographic techniques to highlight its advantages in terms of robustness, security, and reconstruction fidelity.

This chapter is structured into four sections, starting with an introduction and a review of background information in Section 5.2. Section 5.3 introduces the proposed deep learning-based steganography method and provides details on its implementation, including dataset information, network architecture, and training procedure. Section 5.4 presents the analysis of experimental results, covering both quantitative evaluation and visual assessment, along with a comparative analysis of competitive methods. Finally, Section 5.5 concludes the chapter with a summary of key findings and potential directions for future research.

5.2 Background

Convolutional Neural Networks (CNNs): CNNs are designed to automatically extract hierarchical features from images using layers of convolution, pooling, and activation functions. Pre-trained CNN architectures such as VGG16 [87], ResNet [88], and EfficientNet [89]are commonly used for feature extraction and image authentication. Autoencoders: An autoencoder is a method based on neural networks designed to reconstruct input data, capturing meaningful representations of the underlying information. [90]. It is made up of three primary parts:

- **Encoder** – Transforms the input into a lower-dimensional latent feature representation.
- **Latent Representation** – A compact, meaningful representation of the input data.
- **Decoder** – Reconstructs the input from the latent representation.

An autoencoder seeks to strike a compromise between learning a practical latent representation that captures key data features and reducing reconstruction error. Applications such as feature extraction, classification, and clustering can make use of this model. An ideal autoencoder is one that minimizes reconstruction error as much as possible given the constraints of its bottleneck layer size. A well-trained autoencoder with sufficient capacity approximates this ideal state at the end of training, providing a theoretical limit for information preservation (IP) during learning. However, the exact training trajectory depends on the optimization process.

By combining steganography principles with deep learning techniques, advanced secure image transmission frameworks can be developed, ensuring both data confidentiality and authenticity in sensitive applications such as medical image security.

Our method leverages an autoencoder-inspired architecture for image steganography, where the encoder embeds the secret image within the cover image through feature extraction and residual learning, ensuring minimal distortion. The decoder then reconstructs the secret image from the stego image using convolutional layers. The training process optimizes a reconstruction loss, similar to autoencoders, to ensure accurate recovery of the hidden information while preserving the visual integrity of the cover image.

5.3 The Proposed Method

In this work, We present a convolutional auto-encoder steganography framework that conceals a medical image inside a cover image while introducing virtually no visible distortion. The architecture pairs a CNN-based encoder, which learns a subtle pixel-level residual to embed the secret, with a CNN decoder that reconstructs the hidden image from the resulting stego image. Both subnetworks are trained jointly, end to end, with tailored loss functions that enforce two objectives at once: (i) high cover fidelity (PSNR > 40 dB) and (ii) precise secret recovery. This fully convolutional design therefore supports secure, high-capacity embedding without compromising the diagnostic quality of the host image.

The proposed method consists of two principal components:

- **Encoder:** A deep convolutional network that integrates the secret image into the cover image while preserving its visual integrity.
- **Decoder:** A corresponding convolutional network that accurately retrieves the hidden image from the stego image with high precision.

5.3.1 Model Overview

Figure 5.1 provides an overall diagram of the model, showing how the CNN encoder embeds the secret as a low-amplitude residual and how the accompanying CNN decoder extracts it. Together, these two components form a unified architecture trained end-to-end.

Encoder Network

The encoder processes two input images: the *cover image* (a typical medical image) and the *secret image* (another medical image intended for embedding). The encoder consists of two separate convolutional branches:

- **Cover Image Processing Branch:** Extracts spatial features from the cover image using a series of convolutional layers.

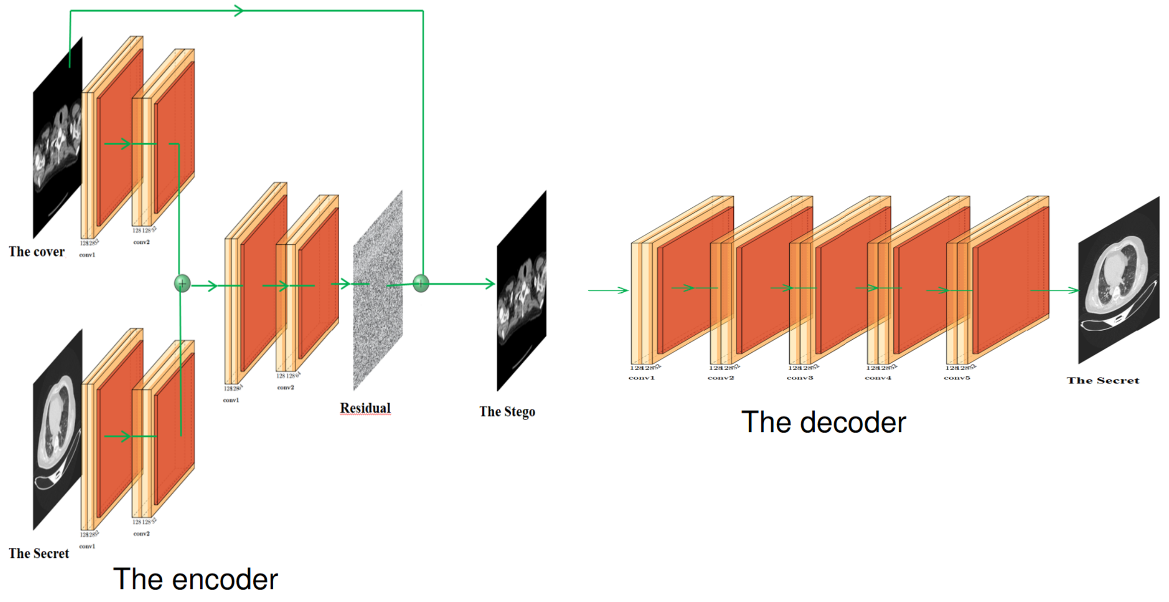


Figure 5.1: The Model architecture

- **Secret Image Processing Branch:** Encodes the secret image using a separate convolutional pathway to ensure effective integration into the cover image.

These extracted features are then concatenated and passed through additional convolutional layers to generate a *residual image*. This residual represents a small perturbation that, when added to the cover image, forms the stego image. The residual is scaled using a factor of 0.1 and clipped to ensure minimal perceptual distortion in the cover image.

Mathematically, the stego image can be represented as:

$$S = C + 0.1\,R \tag{5.1}$$

where S is the stego image, C is the cover image, and R is the learned residual.

Decoder Network

The decoder is designed to recover the hidden secret image from the generated stego image. It consists of multiple convolutional layers that progressively reconstruct the secret image from

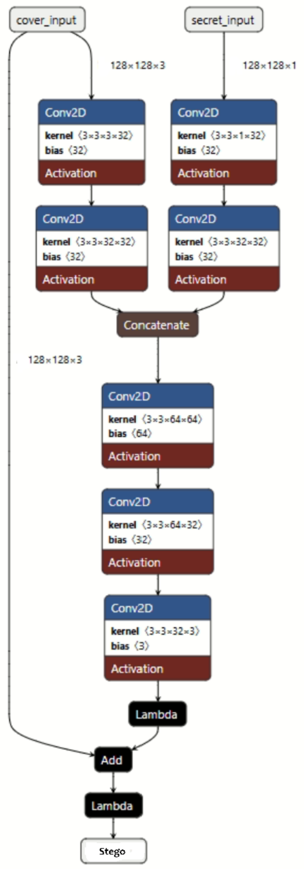


Figure 5.2: The encoder architecture

the stego image input. The final output of the decoder is a single-channel grayscale image that represents the recovered secret.

5.3.2 Training Objective

To ensure effective learning, we define a composite loss function:

- **Cover Loss:** Ensures that the stego image remains visually similar to the cover image, measured using Mean Absolute Error (MAE):

$$\mathcal{L}_{\text{cover}} = \text{MAE}(C, S)$$

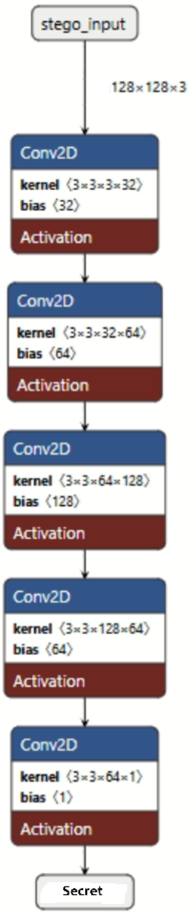


Figure 5.3: The decoder architecture

where C is the cover image and S is the stego image.

- **Secret Loss:** Ensures that the extracted secret image is as close as possible to the original secret image:

$$\mathcal{L}_{\text{secret}} = \text{MAE}(S', S_{\text{orig}})$$

where S' is the recovered secret image and S_{orig} is the original secret image.

The total loss function is defined as:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{cover}}\mathcal{L}_{\text{cover}} + \lambda_{\text{secret}}\mathcal{L}_{\text{secret}}$$

where λ_{cover} and λ_{secret} are weighting factors controlling the trade-off between image concealment and extraction accuracy. We set $\lambda_{\text{cover}} = \lambda_{\text{secret}}$ to balance both objectives equally.

5.3.3 Implementation Details

Dataset description

For training, we utilize two distinct medical image datasets: the MIDRC-RICORD-1B dataset, serving as the source for cover images, and the IQ-OTH/NCCD - Lung Cancer Dataset, providing the secret images. The IQ-OTH/NCCD dataset was gathered from the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases. It contains 1190 CT scan images from 110 patients, categorized into normal (55 cases), benign (15 cases), and malignant (40 cases). The scans, marked by oncologists and radiologists, were acquired using a Siemens SOMATOM scanner with a 120 kV protocol and 1 mm slice thickness. Images were collected in DICOM format, deidentified, and analyzed with institutional review board approval. The dataset represents patients from various backgrounds across central Iraq and provides multiple CT slices (80–200 per scan) capturing different chest angles. The MIDRC-RICORD-1B dataset is part of the RSNA International COVID-19 Open Radiology Database (RICORD), containing 120 de-identified thoracic CT scans from COVID-negative patients. It includes supporting clinical variables such as age, sex, exam date/time (pseudonymous), exam description, symptomatic status, testing result, specimen source, study UID (pseudonymous), image count, and modality. This dataset is designed for research in medical imaging, providing structured data for analyzing thoracic CT characteristics in non-COVID cases.

Training

The Adam optimizer is used to train the model with a batch size of 16 and a learning rate of 0.001 across 70 epochs. The dataset includes paired medical images, where each pair consists of a cover image and a secret image. As part of preprocessing, the images are resized to 128 × 128 pixels, and their pixel values are adjusted through normalization to fit within a defined range. For evaluation, we utilize Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) to assess the similarity between the cover and stego images, as well as between the extracted secret and the original secret images. Higher PSNR values indicate better performance and lower distortion, and for SSIM 1 indicates perfect similarity, 0 indicates no similarity.

5.4 Results

5.4.1 Quantitative Evaluation

The table presents PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) for two processes Cover-Stego and Secret-Recsecret across five test samples. A high PSNR ($\geq$ 40 dB) indicates minimal distortion, while an SSIM value close to 1.0 reflects strong structural similarity. In Samples 1 and 3, the stego images achieve very high PSNR values ($\geq$ 45 dB), suggesting minimal distortion, and the recovered secrets also exhibit impressive quality (PSNR $\geq$ 37 dB, SSIM $\geq$ 0.985).

Samples 2 and 5 maintain excellent fidelity, with stego images around 44 dB PSNR and the recovered secrets remaining around 38 dB PSNR with an SSIM of about 0.99. Sample 4, though it shows the lowest PSNR (41.73 dB), still remains within the high-fidelity range by standard image-processing metrics, and its recovered secret also demonstrates solid performance (35.72 dB, SSIM = 0.9881).

The stego image's average PSNR across all samples is around 44.31 dB, with an SSIM of roughly 0.9626, suggesting that the cover images have been well-preserved after embedding. The recovery of the secrets demonstrates the resilience of the technique, with an average PSNR of around 37.57 dB and an SSIM of approximately 0.9897. Overall, these metrics reflect both minimal impact on the cover image and near-lossless retrieval of the hidden content.

	Cover - Stego		Secret - Recsecret	
	PSNR	SSIM	PSNR	SSIM
Test Sample 1	45.37 dB	0.9731	38.11 dB	0.9889
Test Sample 2	44.46 dB	0.9578	38.83 dB	0.9931
Test Sample 3	45.69 dB	0.9751	37.11 dB	0.9855
Test Sample 4	41.73 dB	0.9531	35.72 dB	0.9881
Test Sample 5	44.31 dB	0.9539	38.09 dB	0.9928

Table 5.1: PSNR and SSIM values for Cover-Stego and Secret-RecSecret comparisons.

5.4.2 Sample Visuals

Table 5.2 presents visual sample results. When comparing the columns for each sample—cover, secret, stego, and recovered secret—it is clear that the stego images (third column) remain visually consistent with their respective cover images (first column). This indicates that the embedding process introduces minimal or nearly imperceptible alterations, preserving key structural elements such as tissue density contrasts and organ boundaries. Meanwhile, the recovered secrets (fourth column) closely resemble the original secrets (second column), with only subtle variations in brightness or artifact visibility. The high degree of similarity between the cover and stego images, coupled with the faithful reproduction of the hidden content, suggests that the steganography method is effective at both preserving the cover images' appearance and accurately reconstructing the embedded secret images.

From a practical standpoint, these results indicate a highly effective steganographic method: the cover images remain visually indistinguishable from the originals, and the secret images can be recovered with minimal loss. Both the quantitative metrics (PSNR and SSIM) and a visual inspection support the conclusion that the embedding and extraction processes exhibit high fidelity.

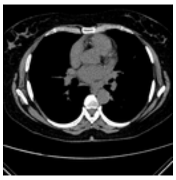
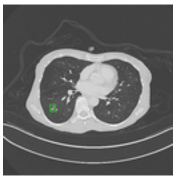
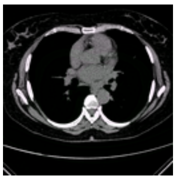
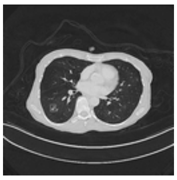
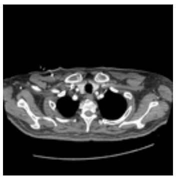
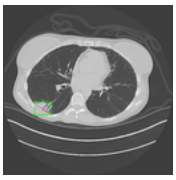
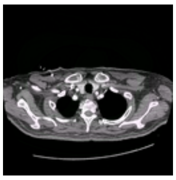
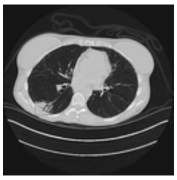
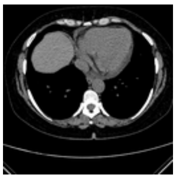
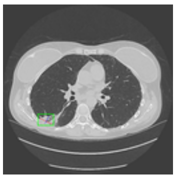
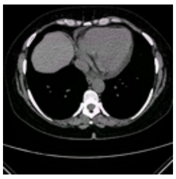
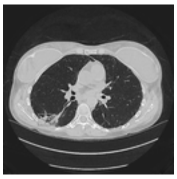
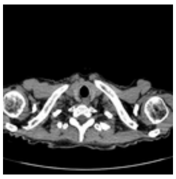
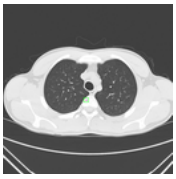
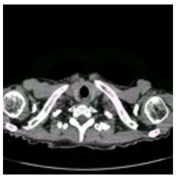
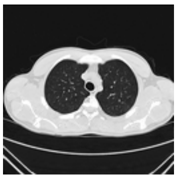
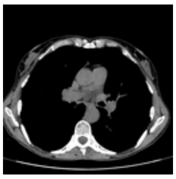
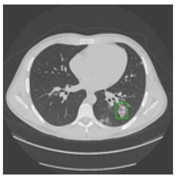
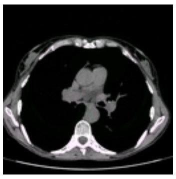
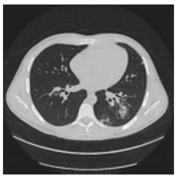
Test Sample	**Cover**	**Secret**	**Stego**	**Recovered Secret**
Test Sample 1				
Test Sample 2				
Test Sample 3				
Test Sample 4				
Test Sample 5				

Table 5.2: Results of steganography encoding and decoding for five test samples.

Approach	**Based on**	**Cover Size**	**Secret Size**	**Payload (%)**	**PSNR (dB)**
Rehman [91]	Encoder–Decoder	32×32×3	32×32×1	33	29.6
Zhang [92]	GAN	256×256×3	256×256×1	33	33.92
Chen [93]	GAN	300×300×3	300×300×1	33	34.07
Subramanian [59]	Encoder–Decoder	256×256×3	256×256×3	100	34.55
Proposed	Encoder–Decoder	128×128×3	128×128×3	100	**44.31**

Table 5.3: Comparison of our method with other deep-learning schemes

5.4.3 Comparison with competitive methods

The proposed method performance was evaluated against several state-of-the-art deep learning-based steganography approaches, as shown in the table 5.3. The comparison reveals significant advantages of our proposed architecture: First, while methods like Rehman's and Zhang's achieve PSNR values of 29.6 dB and 33.92 dB respectively, our approach demonstrates superior imperceptibility with a PSNR of 44.31 dB - an improvement of approximately 9.76 dB over the next best performer. This substantial gain in PSNR indicates that our method introduces significantly less distortion in the cover image while maintaining the same embedding capacity.

Second, unlike competitive methods that often limit payload capacity to 33% of the cover image size, our method achieves 100% capacity, matching only Subramanian's method in this aspect. However, we maintain this full capacity while delivering nearly 10 dB higher PSNR than Subramanian's approach (44.31 dB vs 34.55 dB), demonstrating that our architecture more effectively balances the traditional trade-off between capacity and imperceptibility.

5.5 Conclusion

In this chapter, we presented our second research contribution to the field of medical image steganography, focusing on a deep learning-based approach where an auto-encoder architecture is fully responsible for both the concealment and extraction of hidden information. Unlike traditional steganographic techniques, our method leverages an encoder-decoder architecture that learns an optimal embedding strategy while preserving medical image integrity.

A key aspect of our approach is the secure concealment of a medical image containing sensitive diagnostic information within another diagnostically neutral medical image. By utilizing the structure of a single DICOM file, which is designed to store multiple images (or ”frames”) together, our method ensures seamless integration into existing medical imaging workflows.

Through extensive experiments on medical imaging datasets, we demonstrated the effectiveness of our approach in terms of embedding capacity, imperceptibility, and secret image recovery accuracy. Quantitative evaluations using PSNR and SSIM confirmed that our method achieves high-quality embedding with minimal distortion, while visual assessments further validated the imperceptibility of stego images and the accuracy of extracted secrets.

Additionally, comparative analysis with state-of-the-art steganographic methods highlighted the superiority of our approach in achieving a balance between high payload capacity and imperceptibility. Our method achieves a significantly higher PSNR value while maintaining full embedding capacity, outperforming existing deep learning-based steganographic techniques.

These findings establish a strong foundation for further research in secure medical image steganography. Future work may explore improvements in robustness against steganalysis attacks, adaptation to varying medical imaging modalities, and real-time applicability in clinical settings.

Conclusion and Future works

1 Summary of Contributions

This thesis explores the challenge of securing medical data in digital healthcare systems where medical imaging plays an important role in diagnostics and patient management. We studied how deep learning and steganography can be combined to improve confidentiality of medical images while maintaining their diagnostic integrity.

Our work focuses on three main contributions. First, we developed a region identification method based on Mask-RCNN that reliably distinguishes between diagnostically significant and insignificant regions. This segmentation method achieves an average Intersection over Union (IoU) score of 0.9146 and ensures consistent region identification between original and stego images, as demonstrated by normalized cross-correlation (NCC) values ranging from 0.83 to 1.0, thereby maintaining clinical integrity. Second, integration of a DCT-based frequency-domain embedding approach combines spatial and transform-domain techniques to achieve exceptional image quality, with PSNR values exceeding 115 dB for CT images. The embedding process is finely tunable through a coefficient that controls the quality-capacity trade-off, and the use of QR code representations for patient information standardizes diverse data formats, optimizing embedding efficiency. Finally, we introduce a clinical quality-aware deep learning framework that expertly balances embedding capacity, imperceptibility, and recovery accuracy. This is achieved by a dual-branch encoder that preserves essential diagnostic features while embedding secret data in less critical areas, and a decoder that employs progressive feature reconstruction to improve recovery. This design achieves an average PSNR of 44.31 dB at full embedding capacity. We evaluated our approach using multiple medical imaging datasets, including CT and MRI scans. The results show improvements in image quality, with PSNR gains of 9.76 dB over existing methods, and better embedding capacity, ensuring minimal distortion while securing data.

2 Addressing Research Questions

We reflect on the research questions introduced in this thesis:

RQ 1: How can deep learning architectures, particularly encoder-decoder models, balance imperceptibility, capacity, and recovery accuracy?

The proposed clinical quality-aware CNN has achieved this balance through specific design choices. The dual-branch encoder preserved diagnostic features while embedding secret data, and the decoder's progressive feature reconstruction improved recovery. Using a scaled residual learning approach, we maintained good imperceptibility, achieving an average PSNR of 44.31 dB with 100% embedding capacity, performing 9.76 dB better than other methods.

RQ 2: How can diagnostically less relevant regions in medical images be identified and used for embedding without affecting clinical value? The Mask-RCNN framework we developed proved remarkably effective at distinguishing between significant and insignificant regions, with an average IoU score of 0.9146. This region-aware approach ensured embedding occurred only in areas that would not affect diagnosis, with NCC values between 0.83 and 1.0 confirming the consistency of region identification between original and stego images. This intelligent segmentation was fundamental to maintaining diagnostic integrity while maximizing embedding capacity.

RQ 3: How can steganographic techniques be improved to increase embedding capacity while maintaining image quality?

Our integration of deep learning for region identification with DCT-based embedding in the frequency domain achieved remarkable results, with PSNR values exceeding 115 dB for CT images. The parameterizable embedding process through the β coefficient provided precise control over the quality-capacity trade-off. Additionally, the QR code representation of patient information optimized the embedding process by standardizing diverse data formats.

3 Limitations

Although the proposed work has introduced promising techniques for medical image security, there are certain limitations that should be acknowledged. The computational cost of deep learning models, particularly Mask-RCNN, may restrict practical implementation in healthcare settings with limited hardware resources. Optimization strategies such as model pruning and quantization could help improve efficiency.

While our method demonstrated robustness against common image processing operations, a more detailed analysis against advanced steganalysis attacks is necessary to assess long-term security. Future research could explore adversarial training or other defensive mechanisms to strengthen resistance against detection.

Finally, this study primarily focuses on 2D medical images. However, modern healthcare increasingly relies on 3D imaging, such as MRI and CT volumes. Applying our steganographic approach to volumetric data presents additional challenges, including the need to maintain spatial consistency across multiple slices. Addressing these issues would enhance the applicability of this method in real-world medical environments.

4 Future Research Directions

To address these limitations and expand this work, we propose the following future research directions:

4.1 Optimizing for Resource-Limited Environments

Deploying deep learning models in clinical settings with limited computational resources requires optimizing model efficiency. Techniques such as model compression, quantization, and knowledge distillation can reduce memory and processing requirements while maintaining performance. Developing lightweight architectures specifically tailored for edge devices and hospital systems could make steganography-based security solutions more accessible in real-world healthcare environments.

4.2 Integrating Steganography with Blockchain for Secure Medical Data Management

Combining steganography with blockchain technology could enhance security, integrity, and authenticity in medical image management. Blockchain offers decentralized and tamper-proof storage, ensuring that any modification to a medical image is traceable. By embedding cryptographic hashes or digital signatures within steganographic content, we can create a secure framework for verifying image authenticity while maintaining confidentiality. This integration could be useful for medical data sharing across institutions while complying with regulatory standards such as HIPAA and GDPR.

4.3 Extending Steganographic Techniques to 3D Medical Imaging

Modern medical imaging increasingly relies on 3D volumetric data, such as MRI and CT scans, which require new steganographic methods. Unlike 2D images, 3D datasets consist of multiple slices with spatial dependencies, making conventional embedding techniques less effective. Future research should explore volumetric steganography approaches that maintain spatial coherence across slices, ensuring that hidden information remains robust while preserving diagnostic accuracy. Techniques such as 3D convolutional neural networks (CNNs) or graph-based embeddings could be investigated for this purpose.

4.4 Enhancing Robustness Against Advanced Steganalysis Techniques

As steganalysis techniques evolve, it is crucial to ensure that steganographic methods remain resistant to detection. Future work should focus on adversarial training strategies, where steganographic models are trained alongside steganalysis networks to improve their ability to evade detection. Additionally, generative adversarial networks (GANs) can be used to create more adaptive and imperceptible embedding strategies, ensuring that hidden data remains secure against increasingly sophisticated detection methods.

These perspectives aim to improve medical image security by building on the findings of this thesis while addressing challenges in digital healthcare data protection. Advancing these areas will contribute to the development of more secure, efficient, and clinically applicable steganographic methods for protecting sensitive medical data.

Bibliography

[1] Poonam Kadian, Shiafali M Arora, and Nidhi Arora. Robust digital watermarking techniques for copyright protection of digital data: A survey. *Wireless Personal Communications*, 118:3225–3249, 2021.

[2] Johs Hansen Hammer and Gerardo Schneider. On the definition and policies of confidentiality. In *Third International Symposium on Information Assurance and Security*, pages 337–342. IEEE, 2007.

[3] Arnar Birgisson, Alejandro Russo, and Andrei Sabelfeld. Unifying facets of information integrity. In *International Conference on Information Systems Security*, pages 48–65. Springer, 2010.

[4] Juanita Blue, Joan Condell, and Tom Lunney. A review of identity, identification and authentication. *International Journal for Information Security Research*, 8(2):794–804, 2018.

[5] Javier Lopez, Rolf Oppliger, and Günther Pernul. Authentication and authorization infrastructures (aais): a comparative survey. *Computers & Security*, 23(7):578–590, 2004.

[6] Suhail Qadir and Syed Mohammad Khurshaid Quadri. Information availability: An insight into the most important attribute of information security. *Journal of Information Security*, 7(3):185–194, 2016.

[7] James Q Whitman. The two western cultures of privacy: Dignity versus liberty. *Yale LJ*, 113:1151, 2003.

[8] Accountability Act. Health insurance portability and accountability act of 1996. *Public law*, 104:191, 1996.

[9] Tembisa G Ngqondi. The iso/iec 27002 and iso/iec 27799 information security management standard: A comparative analysis from a healthcare perspective. *Port Elizabeth: Nelson Mandela Metropolitan University*, 2009.

[10] Mario Mustra, Kresimir Delac, and Mislav Grgic. Overview of the dicom standard. In *2008 50th international symposium ELMAR*, volume 1, pages 39–44. IEEE, 2008.

[11] Neil F Johnson and Sushil Jajodia. Exploring steganography: Seeing the unseen. *Computer*, 31(2):26–34, 1998.

[12] Fabien AP Petitcolas, Ross J Anderson, and Markus G Kuhn. Information hiding-a survey. *Proceedings of the IEEE*, 87(7):1062–1078, 2002.

[13] Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, and Ton Kalker. *Digital watermarking and steganography*. Morgan kaufmann, 2007.

[14] Christian Rey. Tatouage d'image: gain en robustesse et intégrité des images. Avignon, 2003.

[15] Gaurav Chawla, Ravi Saini, Rajkumar Yadav, et al. Classification of watermarking based upon various parameters. *International Journal of Computer Applications & Information Technology*, 1(II), 2012.

[16] Tayana Morkel, Jan HP Eloff, and Martin S Olivier. An overview of image steganography. In *Issa*, volume 1, pages 1–11, 2005.

[17] Mehdi Hussain, Ainuddin Wahid Abdul Wahab, Yamani Idna Bin Idris, Anthony TS Ho, and Ki-Hyun Jung. Image steganography in spatial domain: A survey. *Signal Processing: Image Communication*, 65:46–66, 2018.

[18] Fabien AP Petitcolas and Stefan Katzenbeisser. *Information hiding techniques for steganography and digital watermarking (Artech House Computer Security Series)*. Artech House, 2000.

[19] T Narasimmalou and R Allen Joseph. Discrete wavelet transform based steganography for transmitting images. In *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012)*, pages 370–375. IEEE, 2012.

[20] Kirti D Nagpal and PDS Dabhade. A survey on image steganography and its techniques in spatial and frequency domain. *International Journal on Recent and Innovation Trends in Computing and Communication*, 3(2):776–779, 2015.

[21] Gandharba Swain. Digital image steganography using nine-pixel differencing and modified lsb substitution. *Indian Journal of Science and Technology*, 7(9):1444–1450, 2014.

[22] Jarno Mielikainen. Lsb matching revisited. *IEEE signal processing letters*, 13(5):285–287, 2006.

[23] Adnan Sondas and Harun Kurnaz. H nmh: A new hybrid approach based on near maximum histogram and lsb technique for image steganography. *Wireless Personal Communications*, 126(3):2579–2595, 2022.

[24] Oleksandr Kuznetsov, Emanuele Frontoni, and Kyrylo Chernov. Beyond traditional steganography: enhancing security and performance with spread spectrum image steganography. *Applied Intelligence*, 54(7):5253–5277, 2024.

[25] Lisa M Marvel, Charles G Boncelet, and Charles T Retter. Spread spectrum image steganography. *IEEE Transactions on image processing*, 8(8):1075–1083, 1999.

[26] Nandhini Subramanian, Omar Elharrouss, Somaya Al-Maadeed, and Ahmed Bouridane. Image steganography: A review of the recent advances. *IEEE access*, 9:23409–23423, 2021.

[27] Jiren Zhu, Russell Kaplan, Justin Johnson, and Li Fei-Fei. Hidden: Hiding data with deep networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 657–672, 2018.

[28] Brijesh Singh, Prasen Kumar Sharma, Shashank Anil Huddedar, Arijit Sur, and Pinaki Mitra. Steggan: hiding image within image using conditional generative adversarial networks. *Multimedia Tools and Applications*, 81(28):40511–40533, 2022.

[29] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.

[30] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[31] Kamred Udham Singh. A survey on image steganography techniques. *International Journal of Computer Applications*, 97(18), 2014.

[32] Fauzia Puspa Lestari, Choirul Anam, Yati Hardiyanti, and Freddy Haryanto. Automated universal image quality index measurement vs. automated noise measurement: Which method is better to define ct image quality? *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)*, 9(2):132–139, 2019.

[33] Hiba Abdel-Nabi and Ali Al-Haj. Frequency domain based data hiding for encrypted medical images. In *Intelligent Data Security Solutions for e-Health Applications*, pages 21–56. Elsevier, 2020.

[34] De Rosal Ignatius Moses Setiadi. Improved payload capacity in lsb image steganography uses dilated hybrid edge detection. 2022.

[35] Bozhidar Stoyanov and Borislav Stoyanov. Boost: medical image steganography using nuclear spin generator. *Entropy*, 22(5):501, 2020.

[36] Romany F Mansour and Elsaid M Abdelrahim. An evolutionary computing enriched rs attack resilient medical image steganography model for telemedicine applications. *Multidimensional Systems and Signal Processing*, 30:791–814, 2019.

[37] Songul Karakus and Engin Avci. A new image steganography method with optimum pixel similarity for data hiding in medical images. *Medical Hypotheses*, 139:109691, 2020.

[38] Partha Chowdhuri, Pabitra Pal, and Tapas Si. A novel steganographic technique for medical image using svm and iwt. *Multimedia Tools and Applications*, pages 1–20, 2023.

[39] VK Reshma, RS Vinod Kumar, D Shahi, and MB Shyjith. Optimized support vector neural network and contourlet transform for image steganography. *Evolutionary Intelligence*, pages 1–17, 2020.

[40] Reshma VK and Vinod Kumar RS. Pixel prediction-based image steganography by support vector neural network. *The Computer Journal*, 64(5):731–748, 2021.

[41] Mohammed Mahdi Hashim, Ali A Mahmood, and Mohammed Q Mohammed. A pixel contrast based medical image steganography to ensure and secure patient data. *International Journal of Nonlinear Analysis and Applications*, 12(Special Issue):1885–1904, 2021.

[42] Bassem Abd-El-Atty. A robust medical image steganography approach based on particle swarm optimization algorithm and quantum walks. *Neural Computing and Applications*, pages 1–13, 2022.

[43] Bassem Abd-El-Atty, Abdullah M Iliyasu, Haya Alaskar, and Ahmed A Abd El-Latif. A robust quasi-quantum walks-based steganography protocol for secure transmission of images on cloud-based e-healthcare platforms. *Sensors*, 20(11):3108, 2020.

[44] Mamta Jain and Saroj Kumar Lenka. Diagonal queue medical image steganography with rabin cryptosystem. *Brain informatics*, 3(1):39–51, 2016.

[45] Mohammed Mahdi Hashim, Suhad Hasan Rhaif, Ali A Abdulrazzaq, Adnan Hussein Ali, and Mustafa Sabah Taha. Based on iot healthcare application for medical data authentication: Towards a new secure framework using steganography. In *IOP Conference Series: Materials Science and Engineering*, volume 881, page 012120. IOP Publishing, 2020.

[46] Prasanth Vaidya Sanivarapu, Kandala NVPS Rajesh, NV Rajasekhar Reddy, and N Chandra Sekhar Reddy. Patient data hiding into ecg signal using watermarking in transform domain. *Physical and Engineering Sciences in Medicine*, 43(1):213–226, 2020.

[47] P Thiyagarajan and G Aghila. Reversible dynamic secure steganography for medical image using graph coloring. *Health Policy and Technology*, 2(3):151–161, 2013.

[48] Jeevitha Sankaran and Amutha Prabha Nagarajan. Steganography technique based on wpt and elgamal encryption with confusion for robust medical image. *Journal of Advanced Research in Dynamical and Control Systems*, 11:177–192, 01 2019.

[49] R Bala Krishnan, N Rajesh Kumar, NR Raajan, G Manikandan, A Srinivasan, and D Narasimhan. An approach for attaining content confidentiality on medical images through image encryption with steganography. *Wireless Personal Communications*, pages 1–17, 2022.

[50] S Arunkumar, V Subramaniyaswamy, and R Logesh. Hybrid robust image steganography approach for the secure transmission of biomedical images in cloud. *EAI Endorsed Transactions on Pervasive Health and Technology*, 5(18):e1–e1, 2019.

[51] Subhadip Mukherjee, Somnath Mukhopadhyay, and Sunita Sarkar. A shell-matrix-based image steganography technique for multimedia security and covert communication. *Innovations in Systems and Software Engineering*, pages 1–16, 2022.

[52] Eduardo Vazquez, Stephanie Torres, Giovanny Sanchez, Juan-Gerardo Avalos, Marco Abarca, Thania Frias, Emmanuel Juarez, Carlos Trejo, and Derlis Hernandez. Confidentiality in medical images through a genetic-based steganography algorithm in artificial intelligence. *Frontiers in Robotics and AI*, 9:1031299, 2022.

[53] Lina Zhang, Xianhua Song, Ahmed A Abd El-Latif, Yanfeng Zhao, and Bassem Abd-El-Atty. Reversibly selective encryption for medical images based on coupled chaotic maps and steganography. *Complex & Intelligent Systems*, 10(2):2187–2213, 2024.

[54] Gulfam Ahmad, Fatima Gulzar, and Hina Riaz. Secure medical data transmission: An adversarial neural cryptography-based steganography technique with digital signature and lsb replacement. *Physical Education, Health and Social Sciences*, 3(1):52–85, 2025.

[55] Shumeet Baluja. Hiding images in plain sight: Deep steganography. *Advances in neural information processing systems*, 30, 2017.

[56] Atique ur Rehman, Rafia Rahim, Shahroz Nadeem, and Sibt ul Hussain. End-to-end trained cnn encoder-decoder networks for image steganography. In *Computer Vision–ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings, Part IV 15*, pages 723–729. Springer, 2019.

[57] Abhishek Das, Japsimar Singh Wahi, Mansi Anand, and Yugant Rana. Multi-image steganography using deep neural networks. *arXiv preprint arXiv:2101.00350*, 2021.

[58] Felix Kreuk, Yossi Adi, Bhiksha Raj, Rita Singh, and Joseph Keshet. Hide and speak: Towards deep neural networks for speech steganography. *arXiv preprint arXiv:1902.03083*, 2019.

[59] Nandhini Subramanian, Ismahane Cheheb, Omar Elharrouss, Somaya Al-Maadeed, and Ahmed Bouridane. End-to-end image steganography using deep convolutional autoencoders. *IEEE Access*, 9:135585–135593, 2021.

[60] Biswarup Ray, Souradeep Mukhopadhyay, Sabbir Hossain, Sudipta Kr Ghosal, and Ram Sarkar. Image steganography using deep learning based edge detection. *Multimedia Tools and Applications*, 80(24):33475–33503, 2021.

[61] Toan Pham Van, Thoi Hoang Dinh, and Ta Minh Thanh. Simultaneous convolutional neural network for highly efficient image steganography. In *2019 19Th international symposium on communications and information technologies (ISCIT)*, pages 410–415. IEEE, 2019.

[62] Akshay Kumar, Rajneesh Rani, and Samayveer Singh. Encoder-decoder architecture for image steganography using skip connections. *Procedia Computer Science*, 218:1122–1131, 2023.

[63] Bingbing Song, Ping Wei, Sixing Wu, Yu Lin, and Wei Zhou. A survey on deep-learning-based image steganography. *Expert Systems with Applications*, page 124390, 2024.

[64] Denis Volkhonskiy, Ivan Nazarov, and Evgeny Burnaev. Steganographic generative adversarial networks. In *Twelfth international conference on machine vision (ICMV 2019)*, volume 11433, pages 991–1005. SPIE, 2020.

[65] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[66] Chunying Zhang, Xinkai Gao, Xiaoxiao Liu, Wei Hou, Guanghui Yang, Tao Xue, Liya Wang, and Lu Liu. Idgan: Information-driven generative adversarial network of coverless image steganography. *Electronics*, 12(13):2881, 2023.

[67] Haichao Shi, Jing Dong, Wei Wang, Yinlong Qian, and Xiaoyu Zhang. Ssgan: Secure steganography based on generative adversarial networks. In *Advances in Multimedia Information Processing–PCM 2017: 18th Pacific-Rim Conference on Multimedia, Harbin, China, September 28-29, 2017, Revised Selected Papers, Part I 18*, pages 534–544. Springer, 2018.

[68] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[69] Ming-ming Liu, Min-qing Zhang, Jia Liu, Ying-nan Zhang, and Yan Ke. Coverless information hiding based on generative adversarial networks. *arXiv preprint arXiv:1712.06951*, 2017.

[70] Weixuan Tang, Shunquan Tan, Bin Li, and Jiwu Huang. Automatic steganographic distortion learning using a generative adversarial network. *IEEE Signal Processing Letters*, 24(10):1547–1551, 2017.

[71] Tomáš Filler, Jan Judas, and Jessica Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Transactions on Information Forensics and Security*, 6(3):920–935, 2011.

[72] Jianhua Yang, Kai Liu, Xiangui Kang, Edward K Wong, and Yun-Qing Shi. Spatial image steganography based on generative adversarial network. *arXiv preprint arXiv:1804.07939*, 2018.

[73] Yaojie WANG, Ke NIU, and Xiaoyuan YANG. Information hiding scheme based on generative adversarial network. *Journal of Computer Applications*, 38(10):2923, 2018.

[74] Cong Yu, Donghui Hu, Shuli Zheng, Wenjie Jiang, Meng Li, and Zhong-qiu Zhao. An improved steganography without embedding based on attention gan. *Peer-to-Peer Networking and Applications*, 14:1446–1457, 2021.

[75] Ambika, Virupakshappa, and Deepak S Uplaonkar. Deep learning-based coverless image steganography on medical images shared via cloud. *Engineering Proceedings*, 59(1):176, 2024.

[76] Puja Bharati and Ankita Pramanik. Deep learning techniques—r-cnn to mask r-cnn: a survey. *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, pages 657–668, 2020.

[77] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[78] Jens Richter. Qr code generator 1.1 based on zxing, 06 2023.

[79] Nawal Balaska, Aissa Belmeguenai, Ahcène Goutas, Zahir Ahmida, and Selma Boumerdassi. Securing medical data by combining encryption and robust blind medical image watermarking based on zaslavsky chaotic map and dct coefficients. *SN Computer Science*, 3:1–17, 2022.

[80] A. Emre Kavur, N. Sinem Gezer, Mustafa Barış, Sinem Aslan, Pierre-Henri Conze, Vladimir Groza, Duc Duy Pham, Soumick Chatterjee, Philipp Ernst, Savaş Özkan, Bora Baydar, Dmitry Lachinov, Shuo Han, Josef Pauli, Fabian Isensee, Matthias Perkonigg, Rachana Sathish, Ronnie Rajan, Debdoot Sheet, Gurbandurdy Dovletov, Oliver Speck, Andreas Nürnberger, Klaus H. Maier-Hein, Gözde Bozdağı Akar, Gözde Ünal, Oğuz Dicle,

and M. Alper Selver. CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. *Medical Image Analysis*, 69:101950, April 2021.

[81] Abhishek Dutta and Andrew Zisserman. The via annotation software for images, audio and video. In *Proceedings of the 27th ACM international conference on multimedia*, pages 2276–2279, 2019.

[82] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[83] Waleed Abdulla. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN, 2017.

[84] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[85] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 658–666, 2019.

[86] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[87] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[88] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[89] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[90] Umberto Michelucci. An introduction to autoencoders. *arXiv preprint arXiv:2201.03898*, 2022.

[91] Rafia Rahim, Shahroz Nadeem, et al. End-to-end trained cnn encoder-decoder networks for image steganography. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

[92] Ru Zhang, Shiqi Dong, and Jianyi Liu. Invisible steganography via generative adversarial networks. *Multimedia tools and applications*, 78(7):8559–8575, 2019.

[93] Beijing Chen, Jiaxin Wang, Yingyue Chen, Zilong Jin, Hiuk Jae Shim, and Yun-Qing Shi. High-capacity robust image steganography via adversarial network. *KSII Transactions on Internet and Information Systems (TIIS)*, 14(1):366–381, 2020.