Author/s:
Eze, Peter Uchenna

Title:
A Robust and Reliable Tele-medical data Security and Authentication System using Spread Spectrum Steganography

Date:
2020

Persistent Link:
https://hdl.handle.net/11343/253877

A Robust and Reliable Tele-medical data Security and Authentication System using Spread Spectrum Steganography

Peter Uchenna, Eze (793495)
ORCID: 0000-0003-0244-3668

Submitted in total fulfilment for the degree of

Doctor of Philosophy in Engineering

Principal Supervisor
Prof. Udaya Parampalli

Co-Supervisor
Prof. Robin Evans

Data61 Supervisor
Dr. Dongxi Liu

School of Computing and Information Systems

MELBOURNE SCHOOL OF ENGINEERING

July 2, 2020

Declaration

This is to certify that

1. the thesis comprises only my original work towards the PhD,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Peter Uchenna, Eze

Peter Uchenna, Eze (793495) , July 2, 2020

Acknowledgements

My great thanks go to my supervisors: Professors Udaya Parampalli and Rob Evans, Dr. Dongxi Liu and the Chair of the advisory team, Professor Andrew Turpin.

I am also grateful to Melbourne University and CSIRO Data61 for their research scholarships and top-up scholarships, respectively.

My research conferences and publications would not have been possible without the support of my first supervisor Prof. Udaya Parampalli and that of Google Travel scholarship. I am grateful for your support, which enabled me to get my work out there and get feedback for further refinement.

I also appreciate the emotional and psychological support from my wife, Eze Linda Ezinwanne, and the beautiful smile of our twins, Gerald Chinemerem and George Chinweike.

My greatest thanks go to God almighty for keeping me alive and resourceful till date

Preface

Peer-reviewed Publications from this Thesis:

1. **P. U. Eze**, U Parampalli., U.C Iwuchukwu and N.Onuekwusi ***Challenges and Prospects of Blind Spread Spectrum Medical Image Watermarking*** In Proceedings of 3rd IEEE International Conference on Electro-Technology for National Development, Federal University of Technology, Owerri, Nigeria, pp. 10 - 18 , Nov 7 -10, 2017. (Chapter 2).

2. **P. U. Eze**, U. Parampalli, R.J Evans and D. Liu, ***Integrity Verification in Medical Image Retrieval Systems using Spread Spectrum Steganography***. In Proceedings of the 2019 International Conference on Multimedia Retrieval (ICMR '19). ACM, Ottawa, ON, Canada, 10 – 13 June, pp. 53-57, 2019. (Chapter 3).

3. **P. U. Eze**, Udaya, P., Evans, R. ***Medical Image Watermark and Tamper Detection Using Constant Correlation Spread Spectrum Watermarking***. World Academy of Science, Engineering and Technology, International Science Index 135, International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 12, no. 3 , pp. 107 - 114, 2018. (Chapters 3 and 4).

4. **P. U. Eze**, U. Parampalli, R.J Evans and D. Liu, ***Spread Spectrum Steganographic Capacity Improvement for Medical Image Security in Teleradiology***. 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Honolulu, Hawaii, USA, pp. 766-769, 17- 21 July, 2018. (Chapter 4).

5. **P. U. Eze**, U. Parampalli, R.J Evans and D. Liu, ***Evaluation of the Effect of Steganography on Medical Image Classification Accuracy*** In Journal of Applied Bioinformatics & Computational Biology,9:4, 2020. (Chapter 5)

6. **P. U. Eze**, U. Parampalli, R.J Evans and D. Liu, ***A New Evaluation Method for Medical Image Information Hiding Techniques***. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, July 20 -24, 2020, pp. 6119-6122, doi: 10.1109/EMBC44109.2020.9176066. (Chapter 6)

Ethics Approval for use of Human Data (CSIRO): 2019˙061˙LR. See Appendix A for approval document and details.

Abstract

An emerging area with unique security challenges is the area of automated diagnosis (autodiagnosis) in teleradiology. In teleradiology, patients' scans and associated electronic medical records(EMR) are transmitted to a remote location (rural-urban or urban-urban) for image analysis, classification, and diagnosis. The major challenge with this approach is that these scans and EMR are often fragmented and sent out to different users, such as requesting hospitals, independent specialists, patients, external artificial intelligence (AI) systems, and image archives. This occurrence makes it difficult to control the security and privacy of these health information. Therefore, new methods for tamper detection on the image and secrecy preservation of patient's health records are now necessary in this new setting.

Steganography and digital watermarking, collectively known as information hiding (IH) techniques, are among the methods of providing robust security for multimedia (image, video, audio, and text) data. In particular, Spread Spectrum (SS) Steganography and watermarking are hiding techniques that provide secret and robust information hiding, respectively, by using secret keys that are known only to the authorised parties. However, due to the non-standardisation of IH techniques, coupled with the issues of diagnostic quality after data hiding in medical images, the adoption of IH methods in medical practice is currently low. Hence, we are also faced with the challenges of validation and adoption of IH-based algorithms for practical use.

Therefore, we are faced with two major challenges in this thesis: (i) how to improve tamper detection and data hiding capacity of spread spectrum steganography while

retaining its robustness and secrecy and (ii) how to increase the adoption of data hiding security techniques in teleradiology for autodiagnosis. **The ultimate goal of solving these challenges is to improve global healthcare with maximum security but at a low cost.** The quest to achieve this objective led to the following contributions in this thesis:

1. Firstly, we design a new algorithm known as the Spread Spectrum-based Constant Correlation Compression Coding Scheme (C_4S) for cover data Integrity and zero Bit Error Rate (BER) covert message detection. The goal is to allow both accurate and robust detection of secret message in the form of EMR, and content integrity verification by a third-party remote application.

2. Secondly, by leveraging the method developed above and the amplitude modulation techniques, we improved SS Steganographic data hiding capacity. We increased the number of bits that can be embedded in each 8x8 image sub-block from the classical 1 bit to **12 bits** for 16-bit DICOM and **9 bits** for 8-bit natural images. This steganographic capacity was achieved by both increasing the number of unique sequences and the number of frequency channels used for transmission.

3. The predictors and features (known as image biomarkers in medicine) used for remote autodiagnosis, are not usually considered while evaluating medical image IH algorithms. Thus, in this contribution, the effect of IH in computer-aided diagnosis is evaluated based on statistical significance testing of the feature changes, and Machine Learning classification (Support Vector Machine) of Chest X-ray scans of Normal and Pneumonia patients. The results imply that attention should be paid to the specific biomarkers that are sensitive to embedded information but are also relevant in autodiagnosis.

4. Finally, to bring together several algorithms, evaluation mechanisms, and medical image watermarking into practical use, a unified software framework was designed. This unified framework intends to standardise the validation and adoption all IH algorithms for medical image security applications.

In conclusion, this thesis has developed and evaluated new spread spectrum steganography security algorithms for both EMR extraction in the face of attacks and semi-fragile medical image tamper detection, thereby achieving both accuracy and integrity checks, unlike in the basic SS steganography. It also allows higher capacity, especially in the region of non-interest (RONI) of medical images. To enable the adoption of medical image IH techniques in autodiagnosis, a new software framework for unifying algorithms' testing and validation is designed. These contributions are believed to have advanced knowledge in the area of IH and informed practice in the area of medical data security.

Keywords— Security, Privacy, Information Hiding, Digital watermarking, steganography, Medical image, Tamper detection, Integrity, Embedding strength, Watermark payload, Pseuorandom sequence, Biomarkers, Software framework, Teleradiology, Autodiagnosis, Constant correlation, Image classification, Machine learning, AI, BCV, BER, C4S, EMR, PSNR, ROI, RONI, SVM.

Contents

List of Figures

List of Tables

Table 1: List of Common Symbols and Abbreviations

α **(alpha)**	Watermark embedding strength
ϵ **(epsilon)**	Error or fault tolerance.
ρ **(rho)**	Chosen fundamental correlation value by sender and receiver.
AI	Artificial Intelligence
ADNI	Alzheimer's Disease Neuroimaging Initiative
ANOVA	Analysis of Variance
ASCII	American Standard Code for Information Interchange
BCV	Base Correlation Value. The same as ρ
BER	Bit Error Rate
bpc	bits per channel
bps	bits per sample
bpseq	bits per sequence
bpp	bits per pixel
C_4S	Constant Correlation Compression Coding Scheme
CDMA	Code Division Multiple Access
Cr	Compression ratio
DICOM	Digital Image Communication in Medicine
DCT	Discrete Cosine Transform
DSP	Digital Signal Processing
DWT	Discrete Wavelet Transform
ECC	Error-Correcting Codes
EMR	Electronic Medical Records
HSI	Host Signal Interference
HVS	Human Visual System
IH	Information Hiding
ISIC	International Skin Imaging Collaboration
KLD	Kullback-Leibler Divergence
LSB	Least Significant Bit
MIW	Medical Image Watermarking
MIS	Medical Image Steganography
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
PACS	Pictures Archiving and Communication System
PN	Pseudo-noise
PSNR	Peak Signal-to-Noise Ratio
ROI	Region of Interest
RONI	Region of non-interest
SS	Spread Spectrum
SSIM	Structural Similarity Index Measure
SVM	Support Vector Machine

Chapter 1
Introduction

1.1 Thesis Overview

Teleradiology is an aspect of electronic healthcare (e-Health) in which diagnostic data such as medical image scans, biosignals, and patient's EMR are transmitted from the originating healthcare facility to a remote location where they are used for medical diagnosis. In the remote medical practice, medical image scans and related data could be stored locally, transmitted electronically, or uploaded to the cloud archive. Some types of medical image scans involved in teleradiology include X-ray scans, Magnetic Resonance Imaging (MRI), ultrasound (US) images, computerised tomography (CT) scans, among others. The existence of these medical data in digital form has, in turn, enabled their easy manipulation while in transit, storage, or use. These manipulations may be undertaken by authorised multiple physicians, third-parties at different locations, or by malicious (insider or outsider) attackers [127].

The actions of the above users of medical data bring about many challenges for e-health including (i) ensuring the privacy of patients' data, (ii) verifying the integrity and authenticity of the stored or transmitted images, (iii) proving the ownership of archived or transmitted data (iv) ensuring the non-denial of action or inaction on the patient's data, (v) ensuring that enough annotation data are securely transmitted as part of the image, and (vi) providing accurate diagnosis at these remote locations. Solutions are required for these problems.

General information security frameworks exist to solve some of the above issues.

Recent studies [113, 149] have shown that cryptography, a technique in which information is first transformed into a different and unrecognised format using a secret key, is a significant part of most of these security frameworks. However, these studies quickly asserted that all the above challenges could not be solved effectively by cryptographic means. For example, in an open, insecure network prone to noise attacks, encrypted data will get corrupted even with the slightest modification. Cryptography alone provides confidentiality but requires much computational power to be very secure. Besides, encrypted data automatically generates suspicion of secrecy and may force the attacker to destroy the information if s/he cannot decipher it. Hence, other solutions involve hiding security information, robustly, into the medical image itself, or modifying some of its properties to achieve the required kind of solution. These classes of solutions known as ***information hiding (IH) techniques*** have been currently used independently or in conjunction with cryptography to achieve comprehensive security for internet-based solutions. In a broad sense, this thesis is concerned with this aspect of information security.

Despite the challenges mentioned above, the application of IH security techniques for multimedia security is very robust to different types of attacks [40, 38]. It also has continuous security on the multimedia data while they are in use [127, 132]. In particular, **Spread Spectrum IH techniques** are essential in noisy channels, corruptible storage environments, and wireless mobile devices with low-security capabilities. Spread Spectrum (SS) communication technique tends to utilise a transmission bandwidth that is much more than that required by the modulating information. This spreading occurs via the reduction in the signal amplitude and by the spreading signal energy across various channels in the entire carrier bandwidth [42]. This spreading employs a noisy pseudorandom sequence (key-based encryption equivalent to the one used in cryptography). The primary aim of this technique is to achieve a reduction in the possibility of an attacker intercepting or jamming the information transmitted in the entire band. Because this technique provides robust information hiding as well as reliable security through secret keys, it is the specific domain of interest in this thesis. Further descrip-

tion and analysis of SS steganography and watermarking (classifications of IH techniques) will be provided in Section 2.3.1.

The two major shortcomings of SS steganography, currently, are **capacity** and **integrity** verification. Integrity interrogates both the unauthorised modification of the image itself and that of the EMR embedded into it. The amount of information (in bits) that could be encoded into the medical image and transmitted securely without significantly degrading the cover image itself is low. As a patient's health history increases in volume, and information becomes crucial for arriving at an accurate diagnosis, the problem of capacity will require a solution. These two security properties are still inadequately provided for through SS steganography and watermarking.

Having identified the IH technique as the domain of this thesis, there is a further challenge that is specific to this domain, especially when applied to medical images. IH raises some ethical, legal, and diagnostic integrity concerns, which need to be adequately addressed. A reliable framework for evaluating these concerns and establishing how they may eventually affect accurate diagnosis has not yet been established and adopted. The need for such an evaluation framework is made more significant in this era of Artificial Intelligence(AI), where systems are built and operated without human intervention.

This thesis, therefore, first addresses the two challenges that are specific to SS steganography. Next, it develops an evaluation framework for determining if the diagnostic integrity of a medical image scan has been affected by an arbitrary IH method. Finally, this work presents a software framework that enables any kind of data hiding security algorithm to be properly evaluated when used for autodiagnosis in teleradiology.

1.2 Motivation of Research and Problem Statement

This research's motivation stems from current intentional and unintentional security breaches of medical records, often leading to severe issues including panic, loss of money, loss of reputation, fraudulent insurance claims, and political blackmail [110].

Cyberattacks have increased in the local and international health sectors, just as in the financial sector. This rise is due to the increased benefits derivable from health data. According to the *UK Independent News* [1] of 2017, medical information is now becoming more valuable than financial data and could be worth more than ten times the value of credit cards sold on the Deep Web (websites not indexed by search engines). This trend corroborates the statements of Humer and Finkle [2] in *Reuters News* in 2014: *Stolen health records are sold about 10 dollars on the dark web, which is 10 to 20 times what a credit card number is sold in the dark web*. Both occurrences suggest that despite existing security measures, attackers will continue to find means of harvesting health data via the internet because of high pay-off for having such information.

For radiological data breaches, Greenbone Networks Germany [124] recently [3] reported a massive health data breach shows that most internet-based Picture Archiving and Communications Systems (PACS) servers used for radiological image storage are highly insecure.

They recorded the leakage of 24 million data records from patients from across 52 countries. From Australian servers alone, about 50,000 sets of data made up of nearly 2.5 million accounts, including imaging data were leaked. These accounts contained personally identifiable information (PII) along with other medical and billing information. No programming or special software was required to read these data. This leakage shows that the method of archival is flawed, and the researchers in [124] have demonstrated how a modification of the image scans and cloning of the PII could cause significant damage to the patients, the health organisation and also governments. Hence, a more dangerous scenario is a security breach that modifies patient scans and/or electronic records, especially diagnostic data.

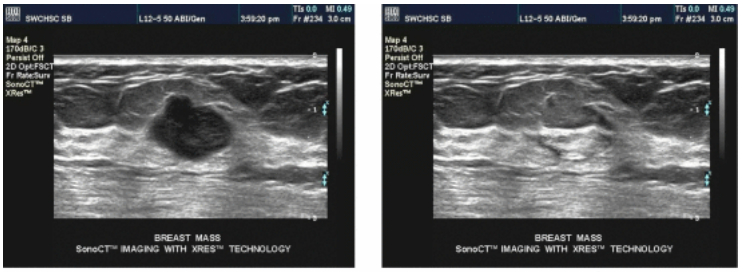
To further illustrate the dangers of medical image tampering, let us consider the

[1] https://www.independent.co.uk/life-style/gadgets-and-tech/news/nhs-cyber-attack-medical-data-records-stolen-why-so-valuable-to-sell-financial-a7733171.html, May 12, 2017

[2] https://www.reuters.com/article/us-cybersecurity-hospitals/your-medical-record-is-worth-more-to-hackers-than-your-credit-card-idUSKCN0HJ21I20140924 , September 25, 2014

[3] September 2019: https://www.itnews.com.au/news/millions-of-australians-sensitive-medical-images-data-left-openly-accessible-531248

case of medical image manipulation of the ultrasound shown in Figure 1.1. A malicious modification of the medical image scans has occurred through lesion insertion and smoothing, thereby changing the medical diagnosis interpretation by either humans or machines. The implication is that a lethal infection could be interpreted as normal or benign.



(a) (b)

Figure 1.1: Cyst lesion replacement. [164]. *How can one determine which of the images is the original one from the ultrasound machine? There is a need to encode a tamper-detectyion security data from the acquisition point before the image is given to a patient, sent to an archive or transmitted for third party diagnosis.*

This thesis is further motivated by the security challenges emanating from unavoidable EMR fragmentation for remote health or third-party health provision. Consider a rural setting that is not connected to a PACS, a central database and network of medical images, nor has an elaborate health information system (HIS), used for storing patient's record. Much information could be captured and transmitted in ad-hoc mode to a remote location with the image for autodiagnosis. Examples of such information include metadata about the scans, patient's health history, and other laboratory test reports employed for diagnosis. There is a need to securely transmit these data with an image without affecting diagnostic quality.

The critical nature of medical data has made medical image watermarking and Steganography (MIW/S) security solutions a challenging research area. This difficulty is due to strict security requirements, high transmission bandwidth requirement (as high compression rate is not allowed), and larger hiding capacity relating to the authenticity and integrity information [143]. Again, a separate transmission of this information increases the chance of mismatching Electronic Medical Record(EMR) with its

source and its owner.

Apart from the above challenges, diagnostic quality degradation may arise, which is a sensitive topic in medical practice as it has both ethical and legal implications[35]. All these challenges are depicted by the famous magic triangle shown in Figure 1.2. One needs to maximise the properties that are most relevant to an application.

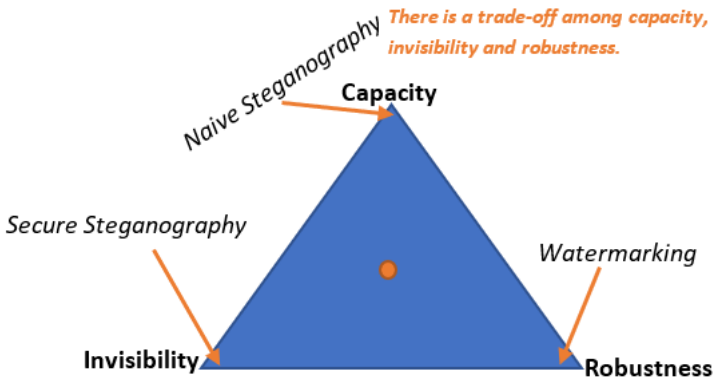


Figure 1.2: Trade-off among data hiding primitives: *Depending on the requirement of an application, the trade-off point may be moved towards any of the edges and vertices to maximise invisibility, capacity or robustness or any of the two but not all.*

A well-designed SS data hiding technique could provide a solution for ensuring secure communication. However, its use of wide-band limits the capacity of actual EMR data embedded. Hence, although it currently provides robustness, it cannot provide integrity checks, which is one of the properties of a reliable security system. So the problem of maximising both security(integrity) and capacity remains unsolved.

Hence, the existing requirements and challenges that motivate this research include:

1. the need to provide integrity checks on images that are utilised in remote healthcare and auto-diagnosis system to ensure accurate diagnosis, but by leveraging a technology that provides robust communication channel security and privacy of health data with or without encryption.

2. the need to securely transmit more annotation and EMR data for correct auto-diagnosis without compromising a patient's privacy and diagnostic image qual-

ity.

3. providing MIW/S evaluation techniques that are more relevant to medical diagnosis, especially autodiagnosis, to closely evaluate the ethical, legal, and diagnostic implication of an IH algorithm, while accepting the security services it provides.

4. the need to practically increase the adoption of watermarking and Steganography in teleradiology to reduce the shortcomings of existing security approaches. This challenge was identified by [127], who solved only part of the problem.

In the next section, we describe the scope in which we investigated the above challenges and needs.

1.3 Scope of Study

This thesis focuses on Spread Spectrum (SS) methods (mainly spatial domain) and not on other methods. Thus, we focused on methods that hide one or more bits of information in n-samples of the cover image. Specifically, we are concerned with improving SS methods that use secret keys to hide one bit or more bits in at least 32 ($n \geq 32$) samples of data (pixel or coefficients). We are aware of other methods such as Difference Expansion (DE) methods [100, 132], which uses a few numbers of pixels ($n \leq 8$) to hide one or more bits of information. Although these methods may accommodate higher payload than SS methods, they may also be more vulnerable to noise-related attacks as most of the hidden information could be easily removed by attacks.

The fundamental SS data hiding technique attributed to Cox *et al* [40, 38, 39] forms the basis of the algorithms developed in Chapters 3 and 4. However, whereas these works focused on tamper-resistant solutions, this thesis focused on adding tamper-detection capability and increased hiding capacity to the traditional SS methods. Also, we focused on blind Steganography, where the original image will not be available at the receiver.

We focused our evaluation on medical images: Chest X-ray scans for pneumonia diagnosis, skin cancer images for melanoma, and MRI brain images for Dementia diagnosis. The developed algorithms, however, are also applicable to other non-medical images. The medical images ranged from 8 to 16 bits per pixel. Therefore, we considered medical images from visible light, magnetic resonance fields, and X-ray radiation. Other forms of images, such as ultrasound, captured at different spectra of the electromagnetic wave, were not evaluated. However, these other medical image modalities were reviewed, and the developed methods in this thesis are generalisable to them.

Finally, the domain of the research presented in this thesis is largely focused on practice rather than the evolution of new theories. Mathematical and statistical methods were utilised for analysis of the soundness of a solution. The target goal, however, is to develop a roadmap towards a teleradiology system that could be implemented in practice.

1.4 Research Objectives

The objective of this research is to utilise the spread spectrum (SS) steganography technique to provide dynamic, secure, but low-cost (in terms of human and material resources) access to teleradiological data. The target users are health experts and autodiagnosis Artificial Intelligence (AI) systems.

To achieve this objective, we shall:

1. investigate and analyse the current limitations of SS steganography.
2. design and evaluate an integrity verification and capacity-optimized steganographic algorithm based on SS techniques,
3. develop an evaluation mechanism for IH algorithms based on image biomarkers used in auto-diagnosis instead of generic signal processing parameters
4. Propose a general software framework for implementing and standardising Medical Image information hiding security to enhance its adoption for remote autodi-

agnosis in teleradiology.

To achieve the above goal and sub-objectives, we will need to answer the questions and sub-questions that follow in the next section.

1.5 Research Questions

This thesis addresses the following research questions:

1. ***Can SS Steganography offer a solution that combines medical image integrity verification and zero-error watermark detection for text embedding while preserving its robust characteristics for authentication?*** Since SS watermarking techniques are historically tamper-resistant rather than tamper-detective, this question involves a search into new parameter settings and algorithmic designs that allow tamper detection. In our case, it is not only about the tampering with the inserted watermark and secret EMR data but also about the cover image itself. The spatial location of this modification will be determined, as well.

2. ***How can one practically estimate and improve current capacity-distortion performance of SS medical image Watermarking and steganography (MIW/S) for applications in teleradiology?*** Because of the wide-band nature of SS methods, larger space in the scan is used to transmit a lower number of bits. Attempt t embed more data leads to higher distortion on the medical image. Hence, having talked about detecting external misuse of images in the first research question, one needs to avoid internal distortion while improving the data-carrying capacity of SS steganography. Hence, to answer this question, one must first estimate the inherent embedding size of a medical image using a SS model. After this estimation, we will explore how to increase the amount of patient and image authentication data hidden in a medical image without significantly degrading the image's diagnostic quality.

3. ***How can one evaluate the suitability of a steganographic system for use in teleradiology based autodiagnosis?*** The notion of distortion in watermarking and Steganography has been largely limited to the human visual system (HVS). The advent of computer-aided diagnosis and machine learning (ML) algorithms has brought a different perspective to what a distorted image is. This new perspective led to new questions that relate to the adequate evaluation of Steganography for automated remote diagnosis. Hence, we need to answer more specific questions, such as:

   (a) *On what conditions could IH become adversarial?*

   (b) *Are all medical image classification features equally affected by the same strength of watermark or secret message?*

   The answers to these questions would lead to a deeper understanding of the relationship between security watermarks, covert messages and computer-aided diagnosis.

4. ***How can a conceptual and software framework help to unify and accelerate the design, implementation, and adoption of data hiding security techniques in teleradiology?*** The peculiarity of the medical image space requires a systematic framework that will unify and standardise the implementation, evaluation, and adoption of developed algorithms. This requirement is because human health is critical and can not utilise any arbitrary algorithm for medical image security. This non-standard nature is one of the research problems that we are tackling in this thesis. To answer this question, we first answer other questions such as:

   (a) *What aspects of IH algorithms need to be unified?*

   (b) *How can the confidence of medical experts be obtained through this framework?*

In answering the above questions and in a quest to fulfill the main objective stated

in Section 1.4, we arrived at several contributions in this thesis. We summarise these four contributions in the following section.

1.6 Research Contributions and Significance

This thesis has contributed to existing knowledge, developed new solutions to existing problems, and proposed a software framework that would lead to the adoption of medical image information hiding security techniques in teleradiology practice. The infographic in Figure 1.3 summarises these four contributions.

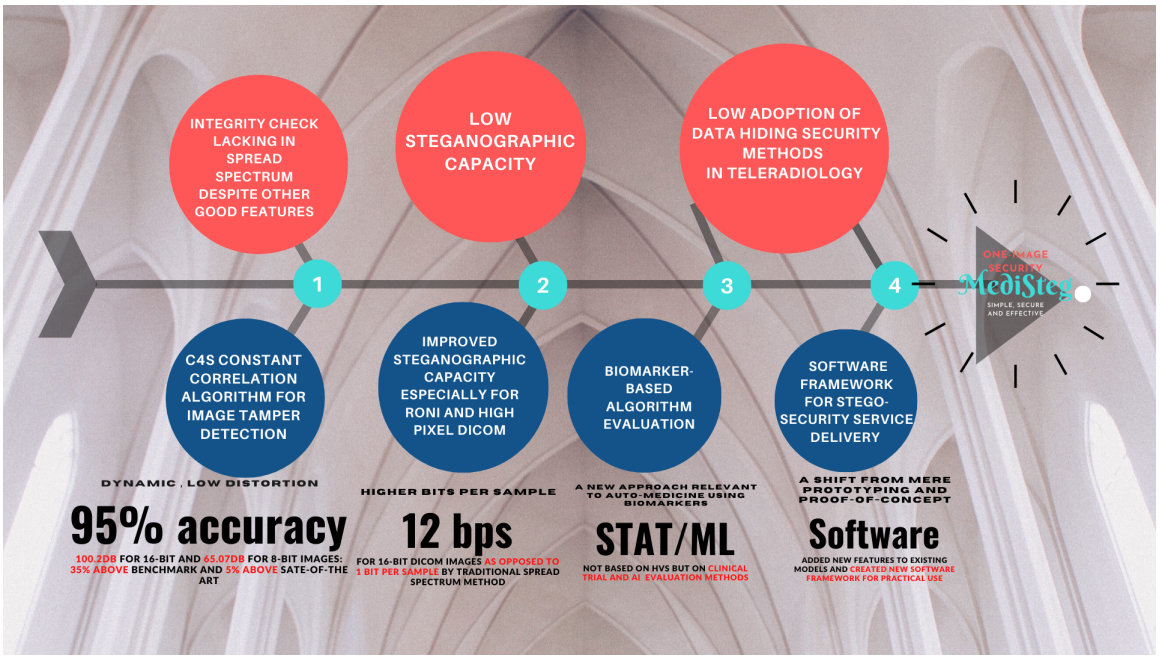


Figure 1.3: Summary of contributions in relation to the research problems

1. A new SS-based image tamper detection method was developed to improve the reliability of teleradiological image scans and associated health record utilised in automated diagnosis while retaining the robust nature of SS techniques (Chapter 3). Specifically:

    - the Spread Spectrum Constant Correlation Compression Coding Scheme (C_4S)

for cover data integrity and zero Bit Error Rate (BER) for extracted watermarks and secret messages is developed. This contribution ensures the sustainability of the SS technique as a solution that is already providing two out of the three security requirements: availability and confidentiality. The third security feature required is integrity, and that is what is provided in this thesis. This feature is also necessary for ensuring correct diagnosis over a network that is vulnerable to malicious data modification attacks.

- a detailed theoretical analysis and comparison between the C_4S algorithm and the existing traditional SS method was provided. The analysis established a novel means through which the new algorithm can be simultaneously employed for both tamper detection and robust authentication. This analysis is important for sound evaluation, criticism, and future improvement.

- an empirical evaluation to detect tampering for the distortion-based attack models was performed. The significance of this contribution is that even though the robust method can be used to extract embedded information, one can still detect tampering on the image and determine if further evaluation as described in Chapter 5 should be performed or not. This rigorous evaluation process helps to ensure an accurate outcome for patient diagnosis. This cover integrity verification capability was not previously possible with SS as it usually focused on retrieving the embedded information but not concerned about the cover image itself.

2. An algorithm to improve the SS Steganographic data hiding capacity of medical images was developed so that more meaningful annotation data can be transmitted for both authentication and accurate remote diagnosis (Chapter 4). The specific contributions here include:

   - the establishment of the parameters that affect capacity improvement, especially for SS steganography and watermarking. This knowledge enables one

to understand the exact trade-offs required to maximise information capacity.

- the design and analysis of a new capacity improvement algorithm by extending the C_4S method. Also, new features that can improve capacity, specifically for medical images, were identified. These will enable one to use both the theoretically proven spread spectrum methods and existing heuristics specific to medical images to increase the amount of data that can be added and successfully extracted from medical images without distorting diagnostic information. Without heuristics, this thesis achieves up to **12 bits per sample** for 16-bit DICOM images as opposed to 1 bit per sample or simple identification code signature in classic ([40, 38]) SS data hiding techniques.

3. An evaluation of the effect of Steganography on diagnostic biomarkers was performed. This thesis particularly evaluated the effect of Steganography on the accuracy of machine learning (ML) based on medical image classification algorithms. The significance is to gain direct insight into any adversarial threat that MIW/S may pose for remote autodiagnosis using ML models (Chapter 5). Specifically:

   - this thesis established that the quality parameters that are based on the human visual system (HVS) are not appropriate for ML algorithms used in autodiagnosis. The significance of this outcome is that data hiding engineers should not rely just on traditional quality parameters such as Peak Signal to noise ratio (PSNR), Mean Square Error (MSE), and the likes. The use of image biomarkers upon which diagnosis is made should be incorporated as an evaluation parameter.
   - to convince the medical community to adopt Steganography as a security tool in teleradiology, the use of evaluation techniques similar to those employed by clinicians during clinical trials is necessary. Hence, a statistical significance evaluation of the effect of Steganography on the statistical distri-

bution of textural image biomarkers for pneumonia chest X-ray was carried out. A support vector machine (SVM) evaluation for medical image data hiding algorithms was also carried out in addition to statistical significance analysis. These evaluations established to what extent, in terms of data capacity and robustness parameter, that medical image IH techniques could start being adversarial. The significance of this contribution is that algorithm designers can evaluate their algorithms in the form of clinical trials as well as use-case scenarios. This approach will help to establish the payload threshold for images that will be used in practical autodiagnosis.

4. A new unified software framework for integration, evaluation, and deployment of medical image watermarking and Steganographic (MIW/S) algorithms were designed and developed. The significance of this contribution is to bridge the gap between data hiding information theories and their use in practice. With a practical framework, it is easy to establish what works and what needs to be redesigned (Chapter 6). Specifically:

    - this work progressed from the unified model presented by Nyeem [127] and an integrated conceptual framework in Qasim *et al* [131], to identify and document the shortcomings of the existing models and conceptual frameworks. We then improved and proposed a model and framework towards increased adoption of MIW/S by the medical experts for whom these systems are being designed. A common gap is that most of these conceptual frameworks have a high focus on the special algorithm developed by the author. This specialisation makes it difficult for the standard implementation of the various algorithms. Secondly, no known existing unified software framework exists. The significance of this contribution is that a unified and complete abstract model is now available for MIW algorithm design and evaluation towards providing the specified security services required in teleradiology. With this, standards can begin to emerge in MIW.

- we designed and analysed a novel generalised conceptual framework with a corresponding new software implementation framework. This framework provides a unified interface to allow any MIW algorithm to be tested without a required knowledge of its detailed implementation. The significance is that both medical and image processing experts can continuously add, update, and refine their algorithms but evaluate them with the same benchmark relevant to disease and diagnosis. This standardised evaluation framework allows for the adoption of the best algorithms that meet ethical and medical specifications. This contribution is a significant improvement from a mere conceptual framework proposed by Qasim *et al* in [131].

The next section provides the roadmap and navigable links for the rest of the chapters in this thesis.

1.7 Thesis Organisation

Table 1.1 summarises the organisation of the rest of the chapters in this thesis.
Chapters 3, 4, 5 and 6 are the major contribution chapters.

Table 1.1: Thesis Organisation

Chapter	Title	Description
2	Spread Spectrum Medical Image Steganography	Background, research conceptual and theoretical framework, review of existing works, research gap and evaluation parameters
3	**Spread Spectrum Image Tamper Detection**	Detection of arbitrary modification of Medical Image
4	**Spread Spectrum Steganographic Capacity Improvement**	Increasing the amount of data embedded per sample image.
5	**Effect of Information Hiding on Image Biomarkers**	Using specific disease biomarkers to evaluate the effect of adding watermarks to medical images.
6	**Medical Image Information Hiding Software Framework**	Software Design Models and Analysis, Partial implementation and justification of designs based on both Software Design principles and Medical Image IH applications.
7	Conclusion and Future Works	Concludes the research and points out directions fo future work

Chapter 2

Spread Spectrum Medical Image Steganography

This chapter establishes the background concepts, the theoretical and the conceptual frameworks underlying this thesis. Section 2.2 is a high-level background that establishes the broad domain of this research. The specific research domain is discussed further in Section 2.3, which introduces, discusses and analyses Spread Spectrum (SS) techniques, spreading sequences and how they are utilised in SS Steganography and digital watermarking. Section 2.4 goes further to discuss biomedical signals, and then Section 2.4.1 discusses their digital processing requirements and techniques, enabling one to understand the kind of images and processing techniques that are applicable to this work. Also, existing related works based on SS medical image information hiding (IH) techniques are reviewed and compared in Section 2.4.3. This review led to the identified research gaps and existing challenges in the area of SS Steganography as presented in Section 2.5. Finally, the design principles for medical image security via Information hiding and the related evaluation parameters are reviewed in Section 2.6. The concepts, challenges, principles and parameters established in this chapter are applicable to the remaining part of this thesis.

Go To Chapter:1, 3, 4, 5, 6, 7

2.1 Introduction

Early information hiding security techniques relied on robust (the ability to withstand adverse conditions) watermark embedding and perfect secrecy of the covert message [144]. It was also assumed that the embedding algorithm itself is unknown to oth-

ers. Whereas robustness may remain a valid assumption for robust watermarking techniques, perfect secrecy of hidden data and that of the algorithms are inconsistent with established security principles such as the popular Kerckhoff's principle [82]. Kerckhoff's principle states that *cryptographic algorithms are public, and only the keys used for encryption provide the required security in the system*. Hence, steganographic algorithms should not rely only on assumed perfect secrecy provided by cover data or host. It should also employ secret keys to secure data even when the algorithm is private. This secret key requirement is even more important for sensitive data such as patients' health records transmitted in teleradiology.

Again, for an open network or storage system that is prone to various noise and channel attacks, the transmitted data need to withstand this noise and also be recoverable even if part of the transmitted signal is lost. Also, there is a requirement to determine the integrity of the received data when error or tampering is unavoidable, and then localise these modifications to a particular region of the transmitted data. Hence, the form of the secret key system required in this environment is not the same as that of the cryptographic key, which is completely random. Cryptographic keys provide end-to-end confidentiality but not the robustness to counter errors and ensure data availability without several re-transmissions. It can establish but cannot localise image data tampering [126].

The above requirements (privacy, localised tamper detection, and data availability in noisy environments) prompts the need to explore Spread Spectrum (SS) data hiding techniques [40, 38, 104, 102], which have some well-established security coding schemes and noise rejection capabilities. The purpose of this thesis is to improve, evaluate and apply the security features in SS steganography (with some aspect of watermarking) to secure the privacy of health data and verify the integrity of medical images and patients' records utilised in teleradiology and remote autodiagnosis. The focus of this chapter, in particular, is to establish a background understanding of SS data hiding techniques, the existing challenges in applying SS to medical image security, establishing research gaps, and defining the evaluation parameters.

2.2 Information Hiding Security Techniques

The security of data, information and knowledge has become increasingly important since the advent of Information and Communication Technologies (ICT). It has taken both new and extended dimensions, as seen in security requirements for smart cities, Internet of things (IoT), cloud computing, Wireless Sensor Networks (WSN), telemedicine, among others [85, 61]. Information Security also includes social and organisational security [28]. The importance of social and organisational information security has made its way into stronger legislation such as the newest General Data Protection Regulation, GDPR law. It was described as the most important change in data privacy regulation in the last 20 years [1].

Some different techniques and technologies could be used to achieve information security. These include Demilitarized zones having firewalls and Intrusion Detection Systems, formal Cryptography, Information Hiding (IH) techniques, and other tools such as anti-virus software that are used to battle unavailability attacks.

Unlike the other methods and tools highlighted in the last paragraph, both Cryptography and IH techniques can protect information irrespective of their physical location. These two techniques have become increasingly important in this era of information ubiquity. The Internet has made all kinds of information to be transmitted even via open networks and to electronic devices that have little or no information security capability. However, the limitation of cryptographic processes in determining image integrity was mentioned in [22] and elaborated in [127]. For example, cryptographic hashes can detect global image tampering (when hash values do not match), but it does not localise the tampering. This shortcoming of the cryptographic method is one of the motivations behind the study of steganographic technique as it has the potential to localise tampering in images.

Information Hiding (IH) is divided into Steganography and Digital Watermarking. Fragile (meant to break easily and allow the detection of tampering) watermarking

[1] https://eugdpr.org/

technique will make it easier to localise medical image and video tampering. This capability is an advantage of IH, especially for medical image forensics in legal proceedings [35].

Steganography completely hides the data from suspicion and tries to maximise capacity and fidelity. In this IH technique, the embedded message is of interest but not necessarily encoded in a robust manner. In digital watermarking, the cover (image, audio or video into which a message is hidden) is of interest to the receiver, and the embedded message further adds value to the cover either for the sender or for the receiver. Also, the message is often encoded in a robust manner [104, 144]. In practical applications, both Steganography and digital watermarking are combined to achieve detection of the modification of either the cover or the embedded secret message. In terms of the inserted data, we use 'secret message' or 'covert message' to represent the embedded data during steganography while 'watermark' remains the inserted data for the purpose of digital watermarking. Both concepts are used in this thesis due to the developed semi-fragile algorithms that bear the properties of both Steganography and Digital watermarking.

Steganography differs from cryptography in that the occurrence of communication is completely concealed, unlike in cryptography, where the communication of information is evident but the content of the information is obscured [104]. For a Steganographic system to be useful, it must ensure that the hidden data is invisible to the eye (imperceptibility), ensure easy extraction of data (extractability), can embed reasonably high data payload (capacity) and can resist attempts to remove the watermark (robustness) [16, 40].

The classification of Steganographic/Watermarking methods is shown in Figure 2.1. Even though Steganography and Watermarking do not strictly have the same meaning, they have almost the same taxonomy as shown. The difference between the two comes down to what properties of information hiding, such as robustness, imperceptibility and hiding capacity, would be maximised and whether the cover or the hidden message is of utmost importance [43, 118]. Also, we have adopted the classification in Figure 2.1

because SS Steganography, as applied to medical image security, is concerned with one or more of the sub-classes as described below.

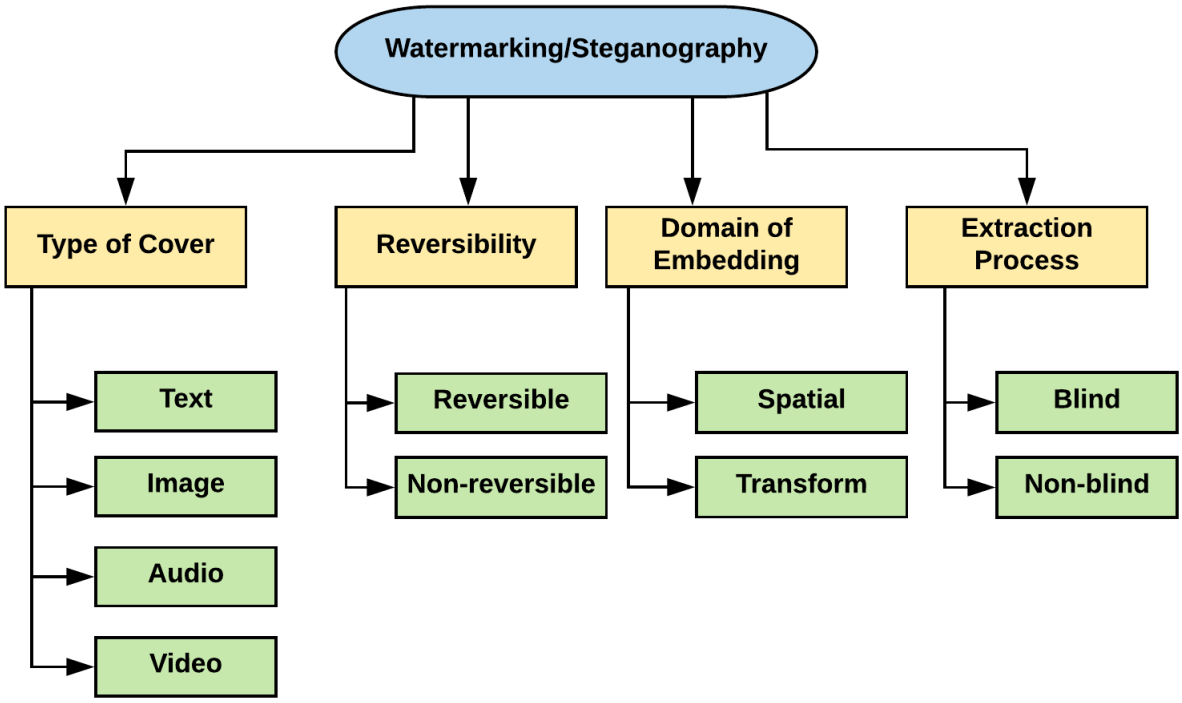


Figure 2.1: Steganographic Methods: *A system can combine these methods. For example, medical image cover with non-reversible blind watermark embedded in the spatial domain*

The type of cover used in Steganography could be Text, Image, Audio or Video. Any kind of data (Text, Image, Audio or Video) could be hidden in the cover provided it can be encoded in some form. Because this research focuses on medical images, we are exploring Image Steganography in this thesis. However, due to the volumetric or 3-D nature of certain DICOM images, ideas from video Steganography and watermarking are incorporated.

In military and medical applications, reversible watermarking techniques have been strongly advocated. In reversible watermarking, one can restore both the cover and hidden message to their respective original forms. It enables zero distortion when the wa-

termark or hidden message is extracted. This method is also used for tamper-proofing and authentication, while the image exists within an insecure environment. However, reversible watermarking does not offer continued protection of the cover or the secrecy of the hidden data after the watermark is removed. On the other hand, irreversible watermarking ensures continuous protection of data but introduces permanent distortion to the original cover. The important factor here is determining the maximum allowable distortion for the given application.

Irrespective of the type of cover or whether the watermark is intended to be reversible or not, the embedding could be in the Spatial or Transform domain. In spatial embedding, there is a direct modification of image pixels or information bits to embed data. The time-domain of the signal is used. This domain of embedding is simpler and tends to provide higher capacity than frequency or transform domain methods. However, it is prone to watermark removal by signal processing algorithms. Robust embedding can be achieved in the frequency domain, although capacity may be lower and distortion higher [118]. The common transform domains of embedding include the Discrete Wavelet Transform (DWT), the Discrete Fourier Transform (DFT), the Discrete Cosine Transform, (DCT), the Non-subsampled contourlet transform (NSCT), among others. Both the Spatial and Transform domain of embedding could be utilised in medical image steganography.

Whether one requires access to the original cover before extracting the hidden information determines if an IH method is blind or non-blind. In blind steganography, the original cover is not required for retrieving the hidden information. With non-blind or *private* watermarking/steganography, all participants have the original cover, and they can subtract this from the watermarked image to retrieve the hidden data. This method is not suitable for medical security in Teleradiology as the receiving hospital will not have the original cover *a priori*. Because of this reason, we are using only blind steganographic techniques in this research.

2.3 Spread Spectrum Information Hiding Techniques

This section introduces the theories of spread spectrum technology in general and SS steganography in particular, including spreading sequences and Code division multiple access (CDMA) techniques.

2.3.1 Spread Spectrum Communication theory

Spread Spectrum (SS) communication originated in the 1950s and is defined as the form of information transmission in which the bandwidth used for the transmission is well above the minimum bandwidth required by such information [130]. A spreading code is used at the encoder, and the same is used as a despreading code at the receiver through a synchronised code generator. These codes should be independent of the original data to be transmitted. The known advantages of this method of communication over the existing ones like Frequency modulation (FM) and Pulse Code Modulation modulation PCM) include interference immunity, anti-jamming capability, Multiple users simultaneous access, low probability of intercept, among others.

In a simple mathematical approach, one can begin to explain SS principles using the Shannon-Hartley channel capacity theorem stated in [162] as:

$$C = B\log_2(1 + \frac{S}{N}), \tag{2.1}$$

where C is the maximum channel capacity in bits per second (bps) at a given bit error rate(BER), B is the required channel bandwidth in Hertz while $\frac{S}{N}$ is the signal power to noise ratio. Now the goal of SS is to maintain a high data rate (C) even when there is higher noise, N at the the receiver. Hence, (2.1) can be re-written as:

$$\frac{C}{B} = \log_2(1 + \frac{S}{N}). \tag{2.2}$$

By very rough estimation, one can say that:

$$\frac{C}{B} \approx \frac{S}{N}. \tag{2.3}$$

Hence, at decreased $\frac{S}{N}$, one can only keep the channel capacity, C constant, by increasing B, which is the transmission bandwidth. This approach is the simplest justification of the spread spectrum technique, and it can be restated thus: *if data is transmitted at a rate C over a channel occupying a bandwidth much greater than C, then a reliable communication can still be achieved even at a reduced Signal-to-noise ratio (SNR).*

Spectrum spreading can be achieved by **direct sequence**, where fast pseudo-random generated sequence is used to change carrier phase; **frequency hopping**, in which the carrier pseudo-randomly changes its frequency; by **time hopping**, in which modulated information is sent in pseudo-random time-slots. A hybrid method uses a combination of these methods. Although the use of SS communication was mainly in the area of military application, in recent times, it has extended to radio communications, mobile networks, wireless communications, and many more. In this thesis, its principles were adapted for digital watermarking and steganography. The direct sequence and frequency hopping SS methods will be discussed further because of their widespread use and relevance to this research.

2.3.2 Typical Spread Spectrum Steganography System

Spread Spectrum Steganography and watermarking have been identified to be among the best hiding methods as they are resilient to various adversary conditions, including noise and other signal processing attacks [104]. The general block diagram of an SS watermarking system is shown in Figure 2.2. The watermark[2], **b**, is spread into the original Image, **X**, using a pseudo-random sequence, **W**, generated through a secret key. The output is the watermarked or stego image, **Y**, which is then transmitted to

[2]Whether an algorithm is a watermarking or steganographic algorithm, once the information to be embedded is converted into binary, we generally call it the generated watermark.

the decoder side or receiver. The PN sequence (**W**) generators, for spreading (during embedding) and despreading (during extraction) are required to be synchronised.

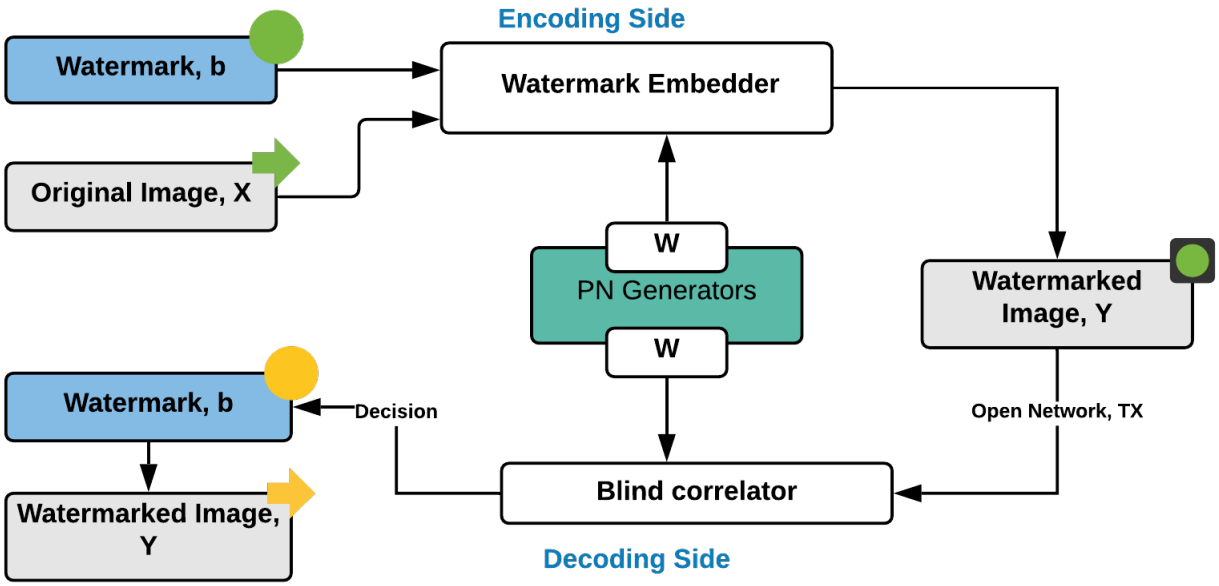


Figure 2.2: General Blind SS Steganography Scheme: *For blind methods, the original cover image, X, is not needed at the receiver. The yellow marker at the reciever means that the original message (watermark), image X and stego image Y may have been modified by attacks while on transit.*

In order to ensure that the message is successfully hidden into the cover, a perceptual and statistical analysis should prove that the hidden message is undetectable both in transit and at the reception of the cover [120, 1], without the required key to decode the hidden message. This implies that only the watermarked object and a secret key will be transmitted. The original cover image or secret message file should not be required to recover the message, reverse the watermarking or both. This is called **blind watermarking/Steganography** [127].

2.3.3 Theory of Direct Sequence SS Steganography

Direct Sequence Spread Spectrum (DSSS) uses a Pseudo-noise sequence to spread a bit of information across wider frequency bands. if a single user's data source is to be hidden using SS technology, then only the robustness and interference against noise are leveraged. Equation (2.4) gives the general additive embedding function for additive SS watermarking of this kind, which is based on DSSS.

$$Y_{ij} = X_{ij} + \alpha W_{ij}(-1)^{S_k}, \tag{2.4}$$

for $1 \leq$ i $\leq$ m and $1 \leq$ j $\leq$ n, $\boldsymbol{Y}$ is the watermarked (or stego) image. $\boldsymbol{X}$ is the original cover sub-block image of size m x n, W is the Pseudorandom Noise (PN) and S_k is the k^{th} watermark bit to be embedded into X. $\boldsymbol{W}$ is a typically a vector of length, l , from an m-sequence or other PN codes such as gold code, Walsh code etc.

An example data structure for $\boldsymbol{X}$(In Spatial domain) and $\boldsymbol{W}$ is given below.

$$\mathbf{Y}= \begin{bmatrix} 91 & 70 & 47 & 24 & 24 & 13 & 3 & 3 \\ 74 & 47 & 48 & 37 & 25 & 12 & 3 & 3 \\ 75 & 44 & 41 & 48 & 27 & 9 & 3 & 3 \\ 66 & 56 & 38 & 46 & 28 & 11 & 3 & 3 \\ 47 & 64 & 54 & 38 & 26 & 14 & 3 & 3 \\ 57 & 57 & 53 & 35 & 26 & 16 & 3 & 3 \\ 48 & 45 & 38 & 29 & 28 & 17 & 3 & 3 \\ 42 & 46 & 37 & 21 & 32 & 17 & 3 & 3 \end{bmatrix} \pm \alpha \begin{bmatrix} -1 & 1 & 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ -1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 \end{bmatrix}$$

Equation (2.4) can be simplified into Equation (2.5) as follows.

$$Y_{ij} = \begin{cases} X_{ij} + \alpha W_{ij}, & \text{if } s = 0 \\ X_{ij} - \alpha W_{ij}, & \text{if } s = 1 \end{cases}, \tag{2.5}$$

where α is the ***embedding strength*** that determines the robustness and imperceptibility of the embedded watermark in the cover image.

The general method of retrieving a watermark [3], $\bar{s}$, blindly using SS technique, is to perform a linear correlation between the secret code W and the stego image, Y (or an attacked version of it with some noise). This correlation ($\langle . \rangle$) retrieval criterion is shown in Equation (2.6).

$$\bar{s} = \begin{cases} 0, & \text{if } \langle Y, W \rangle > 0 \\ 1, & \text{if } \langle Y, W \rangle < 0 \end{cases}. \tag{2.6}$$

Experimental results [125] have shown that that equation (2.6) can lead to false positives - a situation where an image without an embedded watermark indicates that it has one. Figure 2.3 below illustrates this problem. Even though most of the sub-blocks had their correlation value distribution with all of mean, mode and median almost at zero, it is obvious that a reasonable number of the sub-blocks do not have zero correlation value, even when it contains no watermark. Equation (2.6) is the classical SS watermark retrieval method and this problem is called ***Host Signal Interference (HSI)***. It is a fundamental problem of most SS IH methods, which this research would address by exploring the concepts of embedding with side information at the encoder and reverse engineering of the decoder.

A general approach to reducing the HSI problem is to design watermarking systems subject to a given permissible error. Hence, a threshold value, T_h, is often applied instead of zero to cater for this error. This has led to the modification of Equation (2.6) into Equation (2.7).

$$\bar{s} = \begin{cases} 0, & \text{if } \langle Y, W \rangle > T_h \\ 1, & \text{if } \langle Y, W \rangle < -T_h \end{cases}. \tag{2.7}$$

According to Nguen and Tuan [125], T_h should be approximately equal to $\alpha/2$. Equation (2.7) improves detection accuracy, but the error rate can still be high even for un-attacked watermarked images depending on the texture of the image. For example, if $\alpha = 3$, then $T_h = 1.5$. Looking at Figure 2.3, any block whose un-watermarked corre-

[3] Again, we refer to the binary form of the embedded data.

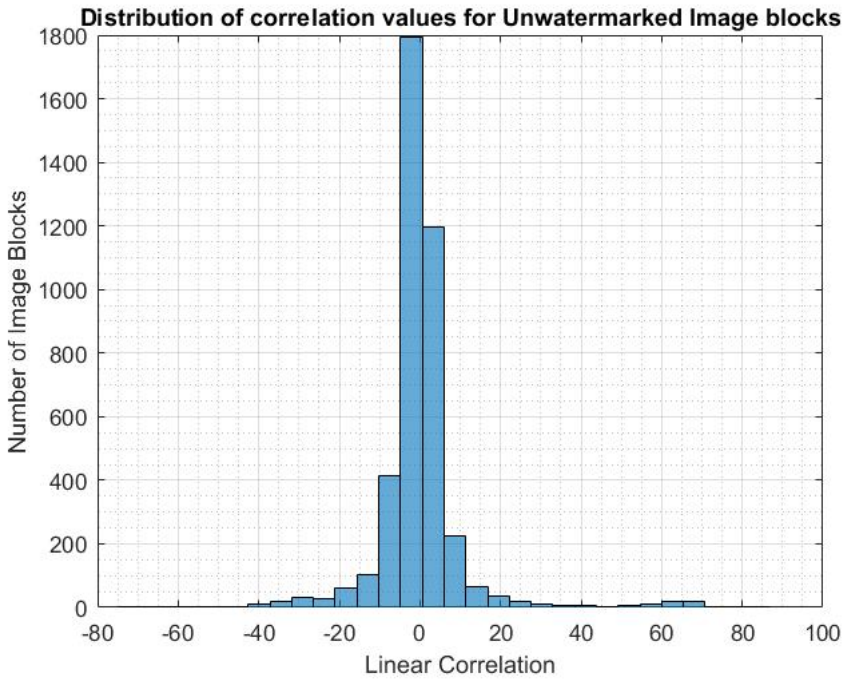


Figure 2.3: Correlation values for image blocks with no watermark.

lation value is greater than 1.5 will be recognised as having been watermarked. This is called ***False Positive (FP)***. All watermarked blocks with correlation values between 0 and 1.5 will not be recognised as having been watermarked. This is called ***False Negative (FN)***. Apart from quantisation errors caused by underflow and overflow for spatial domain IH, FP and FN are the major causes of bit errors during watermark extraction.

The result of this error rate would affect the accuracy of a text-based hidden message, unlike the higher tolerance for image-based watermarks. This prompts the need to design new SS algorithms that can achieve the level of extraction accuracy for text-embedding and extraction. This high level of accuracy is required in remote autodiagnosis as the embedded information may include values representing biomarkers used for disease classification and diagnosis. This requirement has been incorporated into the newly designed data hiding algorithms in Chapters 3 and 4.

2.3.4 Analysis of Basic Spread Spectrum Steganography

The additive spread spectrum watermark embedding originally proposed by [40] was given in Equations (2.4) and (2.5) in various forms. However, in order for it to conform with notations in Section 3.2, we replace the embedded bit as $s \in \{\pm 1\}$ and hence we have Equation 2.8:

$$Y = X + \alpha s W. \tag{2.8}$$

There could be additive noise or other forms of attack introduced by the communication channel as the watermarked image is being sent to the destination. This is modelled by an additive Gaussian noise, $\boldsymbol{n}$. According to Gkizeli *et. al* [56], additive Gaussian model is widely accepted as being appropriate for modelling quantization errors, channel transmission disturbances, and image processing attacks. Thus, what is received at the destination is:

$$Z = Y + \boldsymbol{n}. \tag{2.9}$$

It should be noted that all of $\boldsymbol{W}$, $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ and $\boldsymbol{n}$ are of same length (N) or dimensions (mxn).

At the receiver, the watermark detection is achieved by computing the correlation statistics, $\boldsymbol{r}$, defined by:

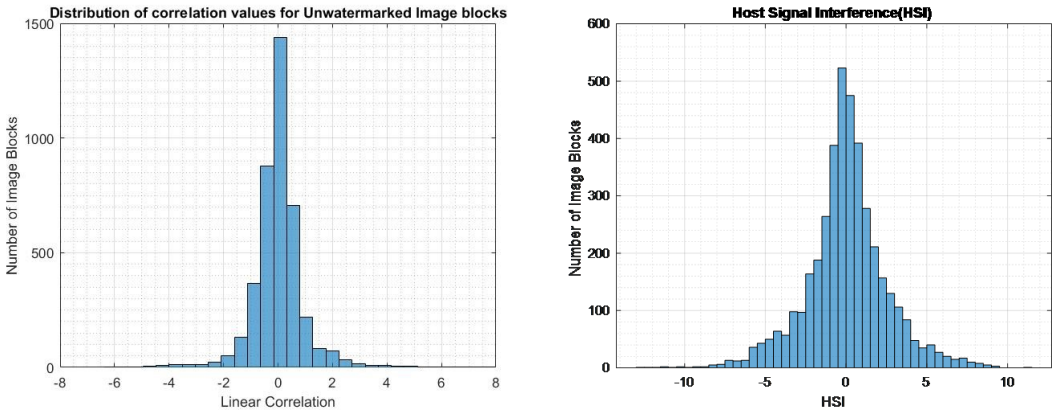
$$r = \frac{\langle Z, W \rangle}{(|W|)} = \frac{\langle \alpha s W + X + n, W \rangle}{(\sigma_w^2)} = \langle \alpha s W, W \rangle + \langle X, W \rangle + \langle n, W \rangle. \tag{2.10}$$

Equation (2.10) can be simplified into:

$$r = \alpha s + x + n \tag{2.11}$$

where $\boldsymbol{x}$ and $\boldsymbol{n}$ are the linear correlation coefficients of the cover and noise signal respectively and each is of the form:

$$\langle Z, W \rangle = \frac{1}{m \cdot n} \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij} W_{ij} \tag{2.12}$$



(a) Low HSI – Cancer Lesion (JPEG) (b) High HSI – Mandrill Image (Tiff)

Figure 2.4: Correlation values for image blocks with no watermark

with **Z** replaced with **X** and $\boldsymbol{n}$ respectively. Also, $m \times n = N$, the length of the pseudo-random sequence. The actual embedded bit is estimated by computing:

$$\bar{s} = sign(r). \tag{2.13}$$

A value of +1 represents a 1 – bit while that of -1 represents a 0-bit. In some improved SS watermark detection, thresholding value, T_h, is often applied instead of *sign(r)* to cater for error due to x in (2.11). This leads to the equation already stated in 2.7 and according to Nguen and Tuan in [125], T_h should be approximately equal to $\alpha/2$. Equation (2.7) improves detection accuracy but error rate can still be high even for un-attacked watermarked images depending on the nature of the image. For example, if $\alpha = 3$, then $T_h = 1.5$. Looking at Figure 2.4, any block whose un-watermarked correlation value is greater than 1.5 will be recognised as having been watermarked. This is called *False Positive (FP)*, already defined earlier.

We now formally analyse the distortion and error performance (BER) of traditional SS. These are the major parameters used to measure all of the imperceptibility, security, and robustness (fragility) in the field of information hiding (Steganography and digital watermarking).

In terms of Distortion Analysis, without the addition of any noise, we first derive

the distortion, **D**, due to the watermark.

$$\mathbf{D} = \mathbb{E}[\| Y - X \|] \implies \mathbb{E}[\| \alpha s W + X - X \|^2]. \tag{2.14}$$

This implies:

$$\mathbf{D} = ||W||\alpha^2 = \alpha^2 \sigma_w^2. \tag{2.15}$$

The meaning of this is that distortion of the cover image depends on the square of the embedding strength, α. The larger the α, the larger the distortion.

2.3.5 Spreading Sequences

A sequence is an ordered set of real or complex numbers with finite or infinite elements. In spread spectrum (SS) technologies, each user is assigned a spreading sequence (code), $\boldsymbol{W}$, which has unique properties. The *spreading* effect comes from the fact that these sequences, which *appear random* though deterministic, is a wideband signal that increases the bandwidth occupancy within the transmission medium or frequency spectrum. The deterministic nature ensures that different signals can be differentiated while keeping its random nature for security. As shown in 2.7, the Pseudo-noise (PN) generator generates sequences that get mixed with each bit or group of bits of EMR recorded and then spreads it across several pixels of the medical image.

One of the most important properties of spreading sequences is that there should be little or no interference with other users existing within the same spectrum. Secondly, the code should be secure enough that it should not be easily forged by unauthorised users. Thirdly, it should be easily reproduced at the receiver so that it could be used to extract the transmitted message from the carrier [109]. The desirable properties of spreading codes include:

- being periodic with a constant length, L.
- easy to distinguish each code sequence with its time-shifted version.
- being easily distinguished from other code sequences.

- being deterministically reproducible but appear random to a listener.
- that the correlation between two different codes (Cross-correlation) should be small.
- the correlation of a sequence (Length, L) with a time-delayed version of itself (autocorrelation) must be small (tends to zero). This autocorrelation for period, τ is given as:

$$r_\tau = \frac{1}{L}\sum_{k=1}^{L} W_k W_{k+\tau}, \tag{2.16}$$

  where the elements of $\boldsymbol{W}$ are in the domain [-1, 1].
- any unintended receiver should find it difficult to obtain and reproduce the spreading sequence [158].

The measure of the level of interference by the sets of codes belonging to the same code set is achieved by determining the **autocorrelation** and **cross-correlation** between the individual code sequences within and across code sets, respectively. Different types of PN sequences and other spreading sequences exist today, each having different autocorrelation and cross-correlation properties. These sequences have been grouped into three [80]:

- **independent and identically distributed (i.i.d) random sequences** - This includes the Gaussian i.i.d real-valued sequences used by Cox *et al* [40].
- **Sequences used in SS Communication** - These are noise-like signals with good correlation properties that allow the spreading and recovery of co-existing signals. The most popular examples are m-sequences, Gold sequences, Kasami sequences, and Walsh sequences. We are concerned with this class of sequences as they have been successfully used in SS Communication systems include Global Positioning Systems, radar tracking, Wideband CDMA, and Very small aperture terminal systems (VSATS) [9].

- **Other PN-like Sequences** - The most popular example in this category is the chaos sequence. It uses an initial condition to generate a scrambled message which can be deterministically recovered if the initial condition is provided.

The type of sequence, length, and the chirp rate determine the characteristics of a particular SS system.

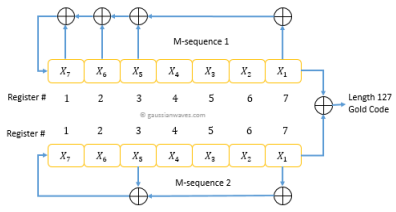
For two sequences $\boldsymbol{V}$ and $\boldsymbol{W}$ each of length $\boldsymbol{L}$, the correlation between them is given as:

$$r_i = \frac{1}{L} \sum_{k=1}^{L} V_k W_{k+i}, \tag{2.17}$$

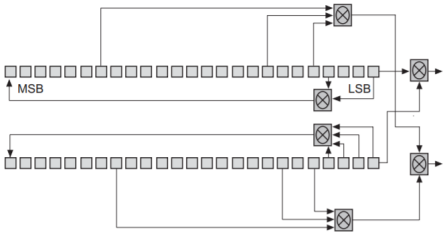
where $\boldsymbol{i}$ is a function of time delay.

2.3.5.1 Generation of Gold Sequences

The SS sequence utilised in running experiments in this thesis is the Gold Code. This is because it has the best discriminating and decoding properties for CDMA applications as well as resistance to collusion attacks, with wide usage for radar and GPS applications. Gold codes have good periodic cross-correlation properties that allow many signals to co-exist within a channel at a minimal interference among one another [9, 30]. The automatic generation of Gold codes of different lengths using PN sequence pairs is shown in Figure 2.5.



(a) 127 bit Gold code ($2^7 - 1$)



(b) 33554431-bit Gold Code ($2^{25} - 1$)

Figure 2.5: Gold sequence generation using Linear Feedback Shift Registers (LFSR)

By formal definition, a Gold code sequence, $G(u, v)$, is obtained by through a set of

operations on two special PN sequences, $\boldsymbol{u}$ and $\boldsymbol{v}$, called the ***preferred pairs***. The final operation is XORing ($\oplus$)(modulo-2 addition) the output together. Preferred pairs are a set of specific primitive polynomials over Galois Field 2 (GF[2]) which will give the ideal and desired properties of a spreading sequence. They have period: $N = 2^n - 1$. The required properties of a preferred pair, $\boldsymbol{u}$ and $\boldsymbol{v}$, of period $N = 2^n$–1 include [87]:

1. n is not divisible by 4.
2. $v = u[q]$, where:
   - q is odd,
   - $q = 2^k + 1$ or $q = 2^{2k}$–$2^k + 1$,
   - $\boldsymbol{v}$ is obtained by sampling every *qth* symbol of $\boldsymbol{u}$.
3. The greatest common divisor (gcd) - also known as highest common factor (hcf) - of n and k is such that:

$$gcd(n,k) = \begin{cases} 1, & n \equiv 1 \ mod \ 2 \\ 2, & n \equiv 2 \ mod \ 4 \end{cases}. \tag{2.18}$$

Preferred pairs are often represented using a set of polynomials of order, n. The sequence set generated is given by:

$$G(u,v) = \{u, v, u \oplus v, u \oplus Tv, u \oplus T^2v, \ldots, u \oplus T^{N-1}v\}, \tag{2.19}$$

where T is a left cyclic shift vector operator. It shifts the vectors one place at a time.

An example of gold code sequence with length,L=63 is : [-1, 1, 1, -1, 1, -1, 1, 1, 1, -1, -1, 1, 1, 1, 1, -1, 1, 1, 1, 1, 1, -1, -1, -1, -1, -1, -1, 1, -1, 1, -1, 1, -1, -1, 1, 1, -1, -1, 1, -1, -1, -1, 1, -1, -1, 1, -1, 1, 1, -1, 1, 1, -1, -1, -1, 1, 1, 1, -1, 1, -1, -1, -1]. The security features of code sets differ by length and correlation properties. The design challenge is evaluating application-specific requirements and selecting or designing codes of suitable length and correlation properties.

2.3.6 Multiple-User SS Steganography

In spread spectrum (SS) technologies, multiple signals are allowed to co-exist within the same channel or frequency band. There is no need for frequency or time division. This is called CDMA, and it is only possible if the secret PN codes assigned to each user (data source) are mutually independent (orthogonal). Two signals, W_1 and W_2, are mutually independent if they have zero correlation and the joint probability mass function (pmf) or probability density function (pdf) is simply the product of their marginal pmfs or pdfs i.e $p_W = p_{W1} \times p_{W2}$. The practical implication of this is that the data from one user would not interfere with another user's data when each user's code is used to perform a correlation in order to extract the embedded data. Another implication is that the value of W_1 provides no information regarding the value of W_2. The CDMA or multi-user steganographic process is formulated here.

Let there exist x number of users, each with a message, ba, to be embedded into the same sub-block, C, using each users' orthogonal code, W_a, where $1 \leq a \leq$ x. Hence, $b = (b_1, b_2, \ldots, b_{x-1}, b_x)$ and $W = (W_1, W_2, \ldots, W_{x-1}, W_x)$.

Originally, $b_a \in \{0, 1\}$. We need to transform it into a polar binary domain, $P_a \in \{1, -1\}$, using Equation (2.20).

$$P_a = 1 - 2b_a. \tag{2.20}$$

In order to generate and multiplex (combine) the watermarks, each P_a is used to multiply the corresponding W_a. The combined watermark for **x** users, Q, is given in Equation (2.21).

$$Q = \sum_{a=1}^{x} P_a W_a. \tag{2.21}$$

The polarisation of the bits is important in order to ensure that the correlation values during extraction are polarised as well. It should be noted that Equations (3.1) and (3.2) are both used for watermark generation and embedding for zero-bit SS watermarking while equation (2.20) and (2.21) are used for watermark generation only for multiple-bit

SS steganography. To embed Q into X, Equation (2.22) below is applied.

$$Y = X + \alpha Q. \tag{2.22}$$

Watermark extraction for each of the x users follows closely from Equations (2.20) and (2.21). However, there is a need for some modifications as well as the proof for independent watermark extraction for each user. The extraction of watermark bit b_a for user a from Y requires only the secret PN code W_a of the particular user among the **x** users. This is given in equation (2.23).

$$b_a = \begin{cases} 0, & \text{if } \langle Y, W_a \rangle > T_h \\ 1, & \text{if } \langle Y, W_a \rangle < -T_h \end{cases}. \tag{2.23}$$

If $\forall \quad W_a, W_i \in Wx$ are mutually orthogonal , then for all $i \neq a$, $\langle W_i, W_a \rangle = 0$. Only $\langle W_a, W_a \rangle = 1$, when $i = a$. Hence, we can show that:

$$\langle Y, W_a \rangle = \langle X, W_a \rangle + \alpha P_a. \tag{2.24}$$

If Equation (2.24) holds true for x users, then it is clear that a problem called ***Multiple Access Interference (MAI)*** has been eliminated. From studies in traditional SS technology, it has been shown that this can only hold if all users'' watermark (PN sequences) are perfectly synchronised and for an upper bound on the value of x. The validity of these assumptions in medical image steganography applications in teleradiology shall be investigated.

The first term in the right-hand side of equation (2.9) is the ***Host Signal Interference (HSI)*** and is desirable to be eliminated in order to ensure accurate detection of watermark and decrease Bit Error Rate (BER). The second term is the ***watermark amplitude***, and it determines imperceptibility, robustness, and the BER of the steganographic system.

If W_a is arranged into a 2-D matrix of size $m \times n$, which is the same dimension as

the cover, **X**, then for this research, Equation (2.25) below will be used as the correlation equation for watermark extraction for both zero-embedding and multi-user SS information hiding.

$$Corr(Y, W) = \frac{1}{m \cdot n} \sum_{i=1}^{m} \sum_{j=1}^{n} Y_{ij} W_{ij}. \tag{2.25}$$

Equations (2.21), (2.24) and (2.25) serve as the basis for more specific embedding and extraction strategies that will be used in this research in Section 3.4.4. We use these equations to derive the Constant Correlation functions for watermark embedding and extraction.

2.3.7 Error Correcting Codes

Error Correcting Codes (ECC) is designed to help in the recovery of error bits caused by channel noise (HSI), MAI, and other malicious attacks during data transmission, processing and storage. The design of these codes through the use of code words enables a code word to represent a bit or group of bits depending on if ***binary*** or ***non-binary*** encoding is being used. In some sense, most ECC has a spreading effect, just as spreading codes mentioned in Section 2.3.5. It could be worth considering the use of either spreading codes or ECC or both, depending on the type of application, the envisaged channel error or other malicious attacks.

Brando *et al* in [20] performed both theoretical and experimental analysis of the performance of common ECC in Spread Spectrum Image steganography. The use of multilevel signalling and ECC could increase the data-carrying capacity and robustness of spatial SS steganographic systems. In [20], binary convolutional codes and non-binary block codes were identified to have the best performance. However, the advantage of using the Reed-Solomon (RS) code is not evident as Bit error rate (BER) did not improve beyond a certain limit (10^{-4}).

ECC, similar to spreading codes, reduces the data-carrying capacity of the encoded data. For this reason, we have employed only spreading codes and not both spreading codes and ECC. Alternative methods of ensuring low BER but higher embedded bits at

low distortion are explored.

2.4 Medical Images and Security Requirements

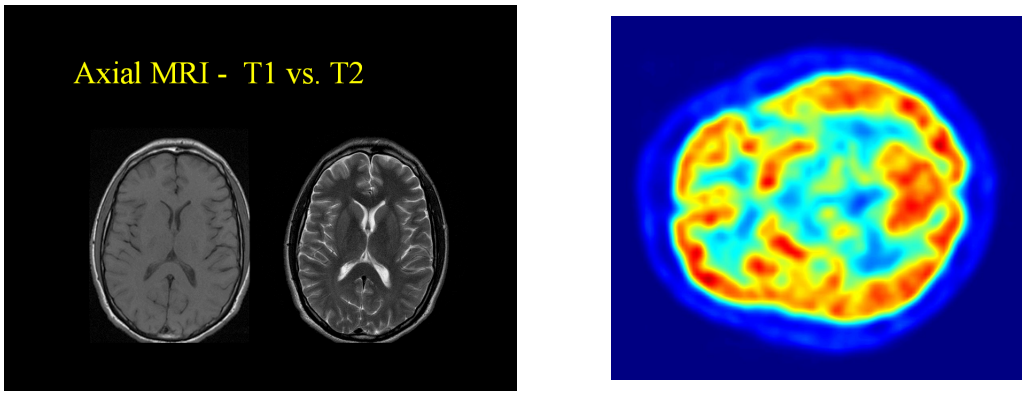
Medical Imaging is an inverse mathematical problem that uses various techniques and processes to create a visual representation and functionality of the human body's internal parts for clinical analysis and medical diagnosis [84]. The 2-D, 3-D or higher dimensional data created is in this process is known as medical image. Medical imaging is an essential part of a branch of medicine called ***Radiology***. As mention in Chapter 1, when ICT is incorporated in rendering radiological services, it is known as Teleradiology.

The major medical images used in modern medicine for medical diagnosis and clinical research are examined below.

- ***Computed Tomography (CT)*** - CT scan uses X-ray technology to create a 3-D image of the body part. Unlike the conventional X-ray scan, it takes not just one picture but several pictures of the body section. Patients may be injected with an X-ray die for better visualisation of the internal organs. CT scan involves exposure to a small amount of x-ray radiation. The amount of total radiation that one is exposed to is proportional to the number of cross-sectional pictures taken.

- ***Magnetic Resonance Imaging (MRI)*** - MRI is used to diagnose soft tissue problems, unlike X-ray, which is used for hard tissue problems such as a bone fracture. MRI scans can be used to diagnose brain tumors, multiple sclerosis, spinal infections, torn ligaments, early stroke, among others [19]. A typical T1 vs T2 axial brain image is shown in Figure 2.6a.

  Our body tissues are made up of water and fats, which contain a hydrogen nucleus. These hydrogen nuclei are mini-bar magnets. When a magnetic field is applied, they tend to align with or against the magnetic field. When additional energy is applied in the form of radio-frequency (RF) signals, the alignment or deflection of the spins of this hydrogen nuclei increases due to the absorption of



(a) T1 vs T2 Brain Image [14]

(b) PET Brain Image

Figure 2.6: Medical Image Samples

this RF energy. When the RF energy is removed, they release this energy and go back to their previous state. The time is taken by different materials such as water or fat to release their excess energy and go back to the relaxation state differs. These release time and energy generate data that is used to plot the tissue characteristics to which the RF signal is exposed. This data is generated in the K-space frequency domain. The Fourier transform of this K-space data results in the MRI image. The popular T1 and T2-weighted MRI images relate to the two different relaxation times of the hydrogen nucleus known as ***Spin-Lattice relaxation time***, (T1) and ***Spin-Spin relaxation time***, (T2) [19].

- ***UltraSound (US)*** - Ultrasound refers to waves whose frequency is beyond the audible frequency (20KHz). The echoes received by the transducer as US signals get reflected or back-scattered by tissues of different types and contents for the diagnostic information that gets displayed as a 2-D US scan picture. Hence, US imaging is based on the principle of echoing. An A-Scan, B-Scan, M-Scan or Doppler-Scan could be produced depending on which signal (amplitude, brightness, motion or frequency difference) [95].

- ***Positron Emission Tomography (PET)*** - PET scans belong to the field of nuclear medicine and molecular imaging. It involves using the isotopes of positron-

emitting radionuclide elements such as Fluorine or Nitrogen to quantitatively measure biochemical and physiological processes *in vivo* or non-evasively [128]. The major advantage of PET scans is its ability to detect changes at the cellular level and thus enable early detection of disease occurrence. This feature gives it an advantage over other image scans previously discussed. The combination of PET and CT scans known as PET/CT scans have been used to diagnose diseases like cancer, neurological disorders, heart diseases, among others. This combination provides better diagnostic information than each done separately [135]. Figure 2.6b shows example of a PET brain image.

2.4.1 Medical Image Processing

As medical images are commonly stored in digital formats in recent times, they benefit from digital signal processing (DSP) techniques for image acquisition, reconstruction, processing and storage. DSP is the set of mathematical tools, algorithms and techniques used to manipulate and extract information and insights from the digital version of sensory data acquired from the real world [145]. During biomedical signal processing, certain digital signal processing techniques and statistical modelling methods are used to analyse and extract relevant information from the signal for further use. Signal processing could be used to remove noise and pre-condition the signal for data and feature extraction. In general, the importance of digital signal processing in biomedicine includes Signal denoising, correct signal model prediction and quantification, feature extraction for diagnostic purposes and prediction of future occurrences and characteristics of the physiological or pathological process [119, 145].

Even though acquisition, reconstruction, processing and storage of the artifacts that generate medical images utilise most of the capabilities of conventional DSP, the concern of Biomedical Engineers and Computer Scientists in the use of DSP tools is different from that of an electronic engineer. Whereas electronic engineers will be expected to improve and expand existing DSP tools, the concern with DSP tools is to understand

the existing ones, their underlying principles, applications and limitations.

Both the conventional and special DSP techniques than can apply to medical images have been described in [59, 119, 139, 145, 84]. Certain transforms such as Radon Transform, Hankel Transform, and K-Space transforms are not as common as Fourier, Wavelet and Contourlet transform. Such transforms, among other polar transformations, are commonly used in creating and reconstructing medical images such as CT and MRI images [59]. The ones considered relevant to this research are discussed further below.

- ***Discrete Fourier Transforms (DFT)*** - Fourier transforms are generally used for design and analysis of Linear Time-invariant (LTI) systems. For a discrete-time period signal F(n), the Fourier Transform, F(ω), which is the frequency component of F(n), is given by:

$$F(\omega) = \sum_{n=-\infty}^{\infty} (F(n) \exp(-j\omega n). \tag{2.26}$$

  Transforms allow one to see the spectral or frequency components of a signal. This spectral decomposition frequently simplifies the analysis of the existing signal or design of a new signal or system that would use those signals. The power of Fourier transform comes from the fact that any periodic function can be broken down into its sine and cosine frequency components using the fact that:

$$e^{\theta} = \cos(\theta) + j\sin(\theta). \tag{2.27}$$

  However, the Fourier transform's limitation is that it cannot enable one to view both time and frequency localisation of a signal. It is not useful for analysing the frequency behaviour of a signal at a particular point in time - that is, time-variant or non-stationary signals. Biosignals such as ECG can be non-stationary. It is also not suitable for representing discontinuities (such as edges in images).

- ***Discrete Cosine Transform (DCT)***

  There are so many well-known probability distributions that a sample dataset

or random variable could follow. A goodness-of-fit test is usually carried out in statistics to determine the extent to which a dataset follows a distribution. In most cases the *Kolmogrov-Smirnov test (KS test)* or the $\tilde{\chi}^2$ *test* is used to compare a sample with a reference probability distribution. The KS test can also be used to compare two samples to know if they have the same distribution.

We first examine the widely acclaimed distribution of the DCT coefficient and use the KS test to select the one that will be used in this research.

The DCT coeffiecient has been purported to follow either Gaussian, Generalised normal, Laplacian or Gamma distribution [17, 122]. However, the analysis by Lam and Goodman in [89] convincingly puts forward the Laplacian and Generalised Gaussian Distribution to be the best fit for the AC components of image DCT coefficients. They also took into account the variance of the 8x8 image blocks whose DCT are taken for JPEG image compression. The type II Discrete Cosine Transform used in JPEG compression is given as:

$$X_{m,n} = \frac{C(m)}{2}\frac{C(n)}{2}\sum_{p=0}^{7}\sum_{q=0}^{7} i_{p,q}.Cos\left(\frac{\pi m(2p+1}{16}\right).Cos\left(\frac{\pi n(2q+1}{16}\right) \tag{2.28}$$

where:

$$C(m) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } m = 0 \\ 1, & \text{if } m > 0 \end{cases} \tag{2.29}$$

and

$$C(n) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } n = 0 \\ 1, & \text{if } n > 0 \end{cases}. \tag{2.30}$$

If p(.) represents probability density function (pdf), then the Laplacian pdf for

$X_{m,n}$ is:

$$p(X_{m,n}) = \frac{\sqrt{2/s}}{2} exp(-\sqrt{2/s}|X_{m,n}|), \tag{2.31}$$

where s is a parameter that controls the variance of the 8x8 blocks across the chosen image. In relation to normal Laplacian parameters $s = \frac{2}{\mu^2}$.

The mathematical definition of other distributions purported for both DCT and DWT are given below:

Gaussian:

$$p(X_{m,n}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_{m,n}-\mu)^2 / 2\sigma^2}. \tag{2.32}$$

Raleigh:

$$p(X_{m,n}) = \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}. \tag{2.33}$$

- ***Discrete Wavelet Transform (DWT)*** - A wavelet is a mathematical function that could be used to divide a time-domain signal into different scale component. DWT uses wavelets of different kinds to create the required signal transformation into the frequency domain for better analysis and feature selection.

  DWT is one of the most popular transform techniques used to post-process, extract features and classify medical images as either normal or abnormal [59]. It is also used for image compression because most of the mother wavelets decompose a signal or image into coarse components and fine details. The fine details could be removed. A wavelet function, $\psi(t)$ has the property that :

$$\int_{-\infty}^{0} \psi(t)dt = 0, \tag{2.34}$$

  which shows that it oscillates equally above and below the x-axis. Also, most of its energy is confined within some finite duration[32]. Hence:

$$\int_{-\infty}^{0} |\psi(t)|^2 dt < \infty. \tag{2.35}$$

Thus, the region with maximum energy could be located, and other less significant parts attenuated. Unlike Fourier transform, wavelet transforms do not only show frequency components of a signal but at what time the set of frequencies existed in the signal. Time-frequency analysis is important in biomedical signal analysis.

A wavelet is produced by combining a chosen mother wavelet with a scaling function $\phi(t)$ or **s**. Thus in simple terms, a wavelet transform sums a signal over its duration and multiplies the sum with a scaled and shifted version of the wavelet function to generate insight into its behaviour at the given time scale.

DWT g(t), is given by:

$$g(t) = \sum_{m=0}^{M} \sum_{n=0}^{N} < x, \psi_{m,n} > .\psi_{m,n}[t] \tag{2.36}$$

where:

$$\psi_{m,n}[t] = s_0^{-m/2} \psi(s_0^{-m} t - n\tau_0). \tag{2.37}$$

The parameter ψ is the Mother wavelet, s is the scale parameter and τ is the translation parameter [138].

Though a generic model for DWT coefficients could be established, accurate modeling for a given application will have to consider the wavelet bases and the level of decomposition being considered.

- ***Image Pixel modeling*** Image modeling in the pixel domain has been widely treated as a discrete time Markovian process with Gaussian increases [98]. The probability that a pixel along the row of an image is in the state f_i at a given time i depends depends on the previous state of a pixel at the instant $i-1$. Hence, there is a transition probability given by: $p(f_i|f_{i-1}$. Assuming that the Markovian process is non-stationary with a Gaussian increase, the transition probability is

given as:

$$p(f_i|f_{i-1}) = \frac{1}{\sigma_s\sqrt{2\pi}} exp\left[-\frac{s^2}{2\sigma_s^2}\right], \tag{2.38}$$

where $s = f_i - f_{i-1}$.

In practice, $p(f_i|f_{i-1})$ is represented by a *stochastic matrix, M,* of size *GxG*, where *G* is the number of gray levels in the image. Hence:

$$p(f_i) = Mp(f_{i-1}). \tag{2.39}$$

The limitation of the above model is that it does not consider the two-dimensional nature of an image. The vertical spatial relationship among pixels does not seem to have been captured. Only the horizontal relationship between pixels is represented by this Markov model.

2.4.2 Teleradiology Data Security Requirements

Even though some medical images are in many ways similar to other natural images, there are certain additional requirements for their handling due to their sensitivity and importance to human health. According to [127, 73, 101, 123, 88, 118, 126, 34] the required parameters and requirements for Medical image watermarking include: *Imperceptibility, Robustness in Region of Non-interest (RONI), Fragility in Region of Interest (ROI), Reversibility, Blindness, Capacity, Integrity, Authenticity, Applicability to different Modalities and anatomies, Computational Complexity and Medical Expert Subjective rating*.

Both Rohini *et al.* [137] and Stallings [147] mentioned the security services that must be provided by a system if it is to be considered secure. These include: ***User Authentication***, ***Access Control***, ***Non-repudiation***, ***Data Integrity*** and ***Confidentiality***. During user authentication, people using the telemedical system should be verified against the one they claim to be. Spoofing of identity should not be allowed. With Access Control, users authenticated users can only use the resources they are authorized to use. Viewing and modification of Medical information should be controlled by access privileges,

and this should be as fine-grained as necessary. For non-repudiation, a secure system should ensure that none of the communication parties will deny their action or inaction during the communication process. The action/inaction of medical personnel involved in a teleradiology diagnosis should not be deniable. This ensures accountability. These three services related to the user and the system in use. This research is concerned with those requirements and services that relate to the data being secured. Hence we are concerned with the following:

- ***Data Integrity*** – This refers to data tamper prevention, detection and recovery. Medical data should be received without any form of unauthorized modifications. Any actual modification should be detected and possibly recovered. The motives for data tampering and why medical image integrity will need to be implemented have been discussed in Section 1.2.

- ***Confidentiality*** – This is related to access control. Whereas access control is more within the organization, confidentiality means ensuring the privacy of the patient within and outside the organisation. Confidentiality ensures that patients' data are protected over the network, in storage and in the ability for others to derive meaning from the flow of the data traffic. Because infinite information can not be hidden using data hiding techniques, it is important to determine and then improve the amount of information that can be hidden while maintaining the confidentiality of such information. This is the reason for the contribution of Chapter 4. The amount of information that can be hidden in multimedia while maintaining confidentiality and integrity is called the ***Steganographic capacity*** of the data hiding technique.

Hence, any system deployed for use in teleradiology should implement security mechanisms to deliver the above security services. Different ways of proving the existence of such services for a developed system will need to be provided during the system evaluation process. How efficient and effective a system provides these services is also an issue for further research. Existing attempts to provide these requirements

(using spread spectrum methods) by current researchers are described and compared further in Section 2.4.3.

2.4.3 Medical Image Data Hiding Techniques

Depending on the major security feature required by an application, various methods have been proposed for achieving confidentiality, integrity and availability through IH techniques. These methods can be classified according to embedding domain, extraction method, reversibility or tamper detection capability. However, because of the high dependence of medical image steganography on the distortion level, the following categorisation is the most relevant to medical images:

1. **RONI-only watermarking** - only the region of non-interest (RONI) has watermark inserted. There is often a challenge of separating the image into RONI and ROI. In the absence of medical experts, a reasonable assumption is made, like in the work of Nyeem [127].

2. **Reversible watermarking** - The watermark can be embedded anywhere in the image but must also be removed before any form of diagnosis could be carried out. For the algorithms in this category, attacks may cause the reversible process not to be successful, and once the watermark is removed, the security of the image is discontinued.

3. **Imperceptible watermarking/steganography** - This embeds watermarks anywhere in the image but subject to predefined distortion constraint. This method helps to achieve both continued protection, larger watermark capacity and authentication. The particular features required for diagnosis must not be affected, and the tolerance level needs to be predefined. This is more of steganography than watermarking.

4. **Zero-watermarking** - No watermark is embedded directly into the image. Features are extracted and used to generate a watermark that is transmitted differ-

ently from the image. There is a possibility of mismatching the actual patient data with the scanned image at the receiver. This technique often only works for image watermarks and not text watermarks.

In the next section, we will focus on various implementations of the above categories but with special attention to the spread spectrum (SS) approach, as SS methodology is the concern of this research.

2.4.4 Spread Spectrum Medical Image Steganography

In the last decade, researchers began to explore the theories and practices of utilising SS IH techniques for Medical image security. This is as a result of the success recorded by SS in military communication systems in terms of security and robustness [77]. A typical SS Medical Image Watermarking (MIW) or steganography (MIS) system is shown in Figure 2.7. It is an adaptation of the generalised SS steganography system in Figure 2.2.

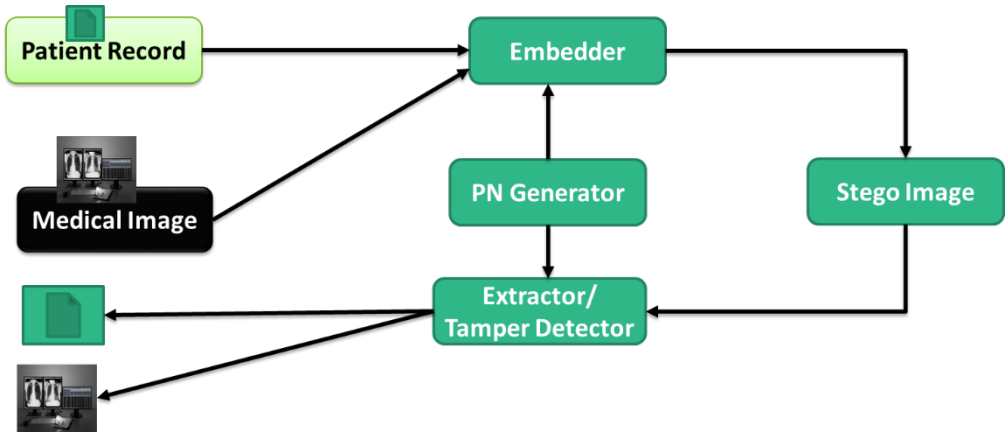


Figure 2.7: Medical Image Steganography: *The cover image is a medical image. It is embedded with the patient record through a secret key provided by a PN generator. At the receiving hospital or facility, the same sequence is used to retrieve the embedded patient record from the medical image cover.*

Maity and Maity in [101] applied SS for the robust watermarking of X-ray, brain MRI and head CT scan. They used a polygon ROI and embedded logos of 32x32 into medical

images of size 512x512. They combined reversible and robust watermarking through reversible contrast mapping and Integer Wavelet Transform Spread Spectrum (IWT SS) watermarking. They got the best performance using X-ray cover, where they achieved PSNR of 44.34 at an embedding rate of 0.491bpp. However, the watermark extraction from the RONI was non-blind. The correlation equation had the original cover image as one of the parameters for extraction. Whereas non-blind extraction is more efficient as it eliminates the host cover energy interference with the watermark [105], it is very unlikely that the receiving hospital would have the cover medical images beforehand in order to enable non-blind extraction.

A Spread Spectrum invertible watermarking system utilising Residue Number System (RNS) and Chaos was implemented by Naseem *et al* in [123]. They used an ROI/RONI mechanism to achieve authentication and data hiding schemes. There test cover image was a 194x259 US image, and the watermarks were a 50x50 image and a 256-bit image hash. Robust watermarking of the logo was done in the RONI, while fragile insertion of the hash was done in the ROI. They also performed various types of attacks on the watermarked image. Their study, however, focused on robustness rather than perceptibility. No data on PNSR was given.

An attempt on multi-access SS for medical image watermarking was implemented by Kumar *et al* in [88] on MRI and US image covers of 512x512. They studied the effect of variation in gain factor, level of decomposition of DWT sub-bands, type of wavelet filter used and the type of medical image modality on the performance of their algorithm. They achieved a very low embedding rate of 0.0039bpp at a PSNR of 41.412dB. Their highest embedding rate was 0.0244, which was achieved at PSNR of 30.138 dB. Though the embedding rates and PSNR are relatively low, their algorithm was highly robust against various attacks that they simulated on the watermarked image. They utilised the Gaussian N (0, 1) pseudo-noise watermark distribution for multi-user embedding. That is a watermark distribution with zero mean and variance of one. Their SS watermarking method was fully blind and thus is a better model for teleradiology.

Table 2.1 summarises the different steganographic methods applied by different re-

searchers for medical image watermarking and Steganography. Not all the methods in this table are SS-based. All of the non-SS methods, such as Memon *et. al* [108] and Zain *et. al* [164] have more data hiding capacity than the SS-based methods. However, as they are based on LSB replacement alone, they are very vulnerable to the slightest attacks. Hence, SS steganography does not have much data-carrying capacity and other security features such as tamper-detection application, and therefore, there is a need to improve SS steganography for medical images in such areas that it is lacking.

Least Significant Bit, DE = Difference Expansion, DWT =Discrete Wavelet Transform, RCM = Reversible Contrast Mapping, IWT = Integer Wavelet Transform, SS = Spread Spectrum.

Author(s)	Modality	Purpose	Region	Method	Capacity	Characters
Memon *et. al* [108]	CT	Authentication	RONI	LSB	64Kb	8,192
Zain *et. al* [164]	US	Integrity Authentication	RONI	LSB	510Kb	65,280
Wakatani [155]	-	Data Hiding	RONI	-	5.9Kb	755
AlQuershi *et. al* [134]	MRI,US, CT,CR	Data Hiding, Authentication	ROI,RONI	DE, DWT	413.8Kb	52,966
Maity *et. al* [101]	X-ray, CT, MRI	Data Hiding Authentication	ROI, RONI	RCM,IWT,SS	126Kb	16,235
Pan *et. al* [129]	X-ray, MRI US, PET	Data Hiding	Whole	DWT	45Kb(0.2bpp)	5,760
Kumar *et. al* [88]	MRI, US	Data Hiding	Whole	SS	1Kb	128
Naseem *et. al* [123]	US	Data Hiding, Authentication	ROI,RONI	SS,RNS Chaos	2.69Kb	344

2.5 Identified Research Gaps

Concerning Spread Spectrum Steganography, in particular, and medical image data hiding security techniques, in general, the following research gaps and challenges still exists.

1. ***Spread Spectrum Stgnography is not yet optimised for hiding capacity*** - The original robust watermarking method by Cox *et al* [38, 40] and the subsequent robust watermarking and Steganographic methods [104, 86, 168] are never optimised for message hiding capacity but for detecting an inserted watermark fingerprint. This robust method solves only copyright problems but not for secure metadata and EMR data, which are needed for interpreting the medical images for manual or autodiagnois, as is the case in teleradiology.

2. ***Practical Estimation of SS Steganographic Capacity*** - The amount of information (in bits) that could be practically embedded, transmitted and reliably extracted with acceptable imperceptibility and message fidelity is yet to be established. Existing methods for estimating this value, such as those proposed by [65] is overly theoretical and draw directly from information theory while the model by [79] is specific to DCT and uses a complex steganlyser. There is still a need to determine the maximum value of information that can be embedded for practical purposes subject to a given distortion parameter that is relevant to a given application.

3. ***Balancing security with medical diagnostic accuracy*** - Even though there is existing research on Medical image IH security, there has not been adequate evaluation on how these watermarks affect diagnostic information. The parameters used are related more to digital signal processing than to diseases indicators or biomarkers. This research gap was explicitly stated by Finlayson *et al* in [53], but has not been solved yet for many diseases. An attempt was made by Garcia in [54] but in a manner limited to deep learning. There is a need to study this for

different algorithms, diseases, image modalities while using different statistical techniques.

4. ***Mitigating the ever-increasing cybersecurity threats and adversarial examples in teleradiology*** - The popular NHS hack of 2017 had health records stolen, and A. Sulleyman had described in [148], how medical information could be ten times more valuable than credit card information in the dark web. Also, Finlayson *et al* in [53] while Mirsky *et al* [110] have demonstrated different adversarial examples in medical image deep learning as well as different incentives that could encourage adversarial attacks in medical image and e-health.

5. ***Standardisation and adoption of Watermarking/steganography in Teleradiology*** - Existing benchmarks only exist in research papers but have not been properly evaluated and widely adopted in practice. There is a constant race towards creating algorithms that beat state-of-the-art, but more practical evaluation and adoption framework have not emerged to put this framework in practice. This challenge and gap were clearly stated by Mirsky *et al* [110], who recommended watermarking and steganography as means of solving their medical image tampering attacks but stated that fear of missing diagnosis might have hampered their adoption.

In this research, various methods are explored in order to solve these existing problems. Also more critical literature reviews will be carried out in each of the contribution chapters (3, 4, 5 and 6) of this thesis.

2.6 Design Principles and Evaluation Parameters

In this section, we present the general design principles that guided the research presented in this thesis. We also state the major evaluation parameters that were employed to evaluate, validate and compare the algorithms and solutions developed in this work.

2.6.1 Design Principles

From the analysis so far, the accuracy of the detected watermark, detected tampering, the capacity (in bits), allowable distortion level to the image and the robustness of the embedded watermark are jointly dependent on: The choice of detection threshold T_h, the HSI of the particular cover image, length of W called ***Process Gain***, number of users, $\mathbf{x}$ and the magnitude of embedding strength, α. Because of these, the following design principles, supported by Guo and Zhuang's recommendations in [62], are adopted in this research to optimise the above constraints towards achieving SS-based tamper detection and steganographic capacity improvement for security in teleradiology:

1. Defining an acceptable range of distortion for Medical Image steganography and watermarking.

2. For higher watermark detection accuracy or zero BER, the HSI cancellation method will be used to dynamically estimate the best value of α in order to ensure accurate retrieval of the embedded watermark.

3. Separating an Image into ROI and RONI for the purpose of tamper detection and high capacity EMR insertion, respectively. Dividing a Medical Image into Region of Interest (ROI) and Region of Non-Interest (RONI), as shown in Figure 2.8.

4. Following Kerckhoff's principle of cryptography during steganographic algorithm designs. Hence, the secret key is always required to retrieve watermarks.

5. Giving higher priority to high-pixel depth and 3-D medical images. They could provide more security for medical image integrity and higher capacity for EMR embedding [93].

2.6.2 Evaluation Parameters

Suitable Technical evaluation methods are used to determine the success of this research work. These evaluations will determine the contribution to knowledge as well

Figure 2.8: ROI-RONI segmentation Example of Medical Images: *More controlled embedding will be exerted in the ROI*

as the relevance of the research to the health industry. The variables of interest in this research include *Steganographic capacity, detection accuracy, imperceptibility, robustness and diagnostic quality*. The parameters that follow will be used to determine these variables. Whereas some benchmarks for Medical images are established for some of these variables, the aim of this research includes establishing new benchmarks using medical-related parameters such as image biomarkers (see Section 5.2).

2.6.2.1 Peak Signal-to-Noise Ratio (PSNR)

- this is the ratio of the original cover over the noise (standard error) introduced by the embedded data.

$$PNSR = 20 * log_{10}(\frac{B}{\sqrt{MSE}}). \tag{2.40}$$

B is the largest value of signal or the dynamic range for the pixel values ($2^n - 1$, where n is pixel depth) and **MSE** is the Mean Square Error per pixel. PSNR is a statistical degradation measure. It is a measure of statistical imperceptibility and its value ranges from 0 to ∞. However, according to [29, 72, 49] , an effective medical image IH algorithm is **effective if its PSNR is greater than or equal to 40 dB**.

2.6.2.2 Structural Similarity Index (SSIM)

This parameter is known to consider the Human Visual System (HVS) in its computation. Hence, it is a measure of the visual imperceptibility of the watermark. It is also used to measure the degree of reversibility of an IH algorithm in order to determine if the watermark has been removed and the stego image restored to its original state. Reversibility is sometimes a desirable property of a medical image IH algorithm.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{2.41}$$

where μ_x, μ_y and σ_x^2, σ_y^2 are the corresponding mean and variance of the images x and y respectively. The parameter σ_{xy} is covariance of x and y. The value of SSIM ranges from 0 to 1, where 1 indicates a perfect similarity between two images.

2.6.2.3 Bit Error Rate

Let B_{error} = Number Of wrongly extracted bits and B_{total} = Total number of embedded bits. Then Bit Error Rate (BER) is defined as:

$$BER = \frac{B_{error}}{B_{total}} \tag{2.42}$$

BER is a very important parameter in defining Steganographic capacity. It is also important in determining the accuracy of the retrieved patient's record. A single-bit error in a character changes its meaning. For instance, the ASCII code for character 'a' is 01100001 (97 decimal). If the Least Significant Bit (LSB) of this ASCII code is retrieved as 0 instead of 1, then we have 01100000 (96 decimal), which returns the non-alphabetic character "'. Hence, error bits and their distribution is an important measure in this research.

2.6.2.4 Clinical Evaluation using Image Bio-markers

Image biomarkers are objective measures obtained from radiological images or scans. These measures are combined with other biomarkers such as those obtained from blood tests in order to diagnose a disease.

In this research evaluation, we focus on textural image biomarkers that help in diagnosing pediatric pneumonia disease. Specifically, we used Xray images. The purpose of this evaluation is to determine if the embedded message or influences the quantification result. The quantification is performed before and after embedding, and various statistics is used to determine how watermarking affects the possible diagnostic outcome.

The evaluation of the algorithms developed in this research will also be performed on public image databases of medical images. In this evaluation technique, trained Support Vector Machine (SVM) are used to classify Xray images as *Normal* or *Pneumonia*. After this, the images are then embedded with EMR data for remote diagnosis and treatment. Then, a new model is built and tested using the watermarked images. Different levels of degradation are tested to determine at what level the machine model starts to behave differently from the original baseline model. A similar approach, like the one applied by Chaplot *et al* [27], will be used. This approach will help us establish a new benchmark and verify the claim by Chen [29] concerning PSNR values for medical images.

2.7 Summary

In this chapter, we have reviewed the meaning of SS information hiding techniques. We mentioned that the use of spreading sequence, which serves as a symmetric key and has some desirable correlation properties, is at the heart of this technique. We have also reviewed medical images as being signals from the human body and utilised for diagnosis. The integrity of medical images and the privacy of associated EMR in teleradiology are the major concern of this thesis. However, we also want to achieve

this through the SS data hiding technique in particular. We will also develop robust evaluation mechanisms for this and other existing algorithms in order to encourage their adoption for use in practice.

In the next chapter, we will dive into the design of the SS-based tamper detection and accurate watermark detection algorithm. This design will culminate in the fundamental stage of our new Constant Correlation Compression Coding Scheme (C_4S) algorithm. This fundamental stage is only for tamper detection. The extended stage defined in Chapter 4 includes the capability for Steganographic capacity improvement.

Chapter 3

Spread Spectrum Medical Image Tamper Detection

This chapter aims to propose and test a Spread Spectrum (SS)-based steganographic technique for medical image tamper detection and zero Bit Error Rate (BER) watermark retrieval. These security and accuracy features are accomplished by modifying the basic SS Steganography technique to design semi-fragile steganographic method algorithms. This modification involves placing a limit on the variation of the correlation values used to detect watermark signals at the receiver. The new algorithms would enable accurate detection of text watermark in the form of Electronic Medical Records (EMR), and the verification of medical image content integrity. In Section 3.2, we introduce the notations and symbols that we will use in this and the following chapters. The motivation for attacking medical images are provided in Section 3.3 together with attack models and methods. We then propose our new method and the algorithms in Section 3.4. The theoretical and design analysis are then presented in Section 3.5 before empirical evaluation and results are presented in Sections 3.6 and 3.7, respectively. The discussion of these results follows in Sections 3.9. The chapter concludes with a summary in Section 3.10.

Go To Chapter: 1,2, 4,5, 6, 7

3.1 Introduction

In Section 2.3.2, we introduced the fundamental Spread Spectrum (SS) Steganography technique. As a recap, an SS steganography technique involves the use of Pseudo-

random Noise (PN) sequences of length, L to embed bits (0 or 1) of information by spreading it across L cover data elements. The PN code and the entire embedding/extraction strategies have the property such that performing a linear correlation between the PN code and the watermarked image will have a positive or negative correlation value. Negative correlation extracts a 0-bit while a positive correlation extracts a 1-bit (or vice versa).

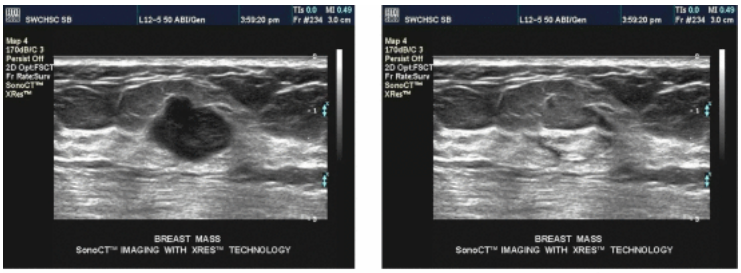
Currently, detecting tampering with SS Steganography is generally difficult because it has the inherent characteristics of robustness, where the effect of external noise and attacks are intended to be decimated and resisted as much as possible. For instance, correlation values can be allowed to vary from $-\infty$ to ∞, thereby removing the chance of localising the cause of any possible shift in correlation values. Secondly, most researchers have focused on using only image data, which has higher error tolerance, as the watermark for authentication of cover data. This trend is opposed to the use of text watermarks and integrity-checking hashes as metadata(such as Electronic health record (EMR)) or integrity checks, respectively. Here, error tolerance is approximately zero.

Text watermarks (not encoded as an image) involve the conversion of the text to the American Standard Code for Information Interchange (ASCII) and embedding the binary equivalent or its compressed version in an image. With this encoding method, any error in the extracted watermark would change the resulting ASCII character. Hence, there is a need for 100% accuracy in the retrieved hidden data. This leads to the ***question***:

How can SS Steganography offer a solution that combines medical image integrity verification and zero-error watermark detection for text embedding while preserving some robust characteristics for authentication?

As medical images are increasingly being stored, transmitted and processed digitally, there is also an increased potential of intentionally and unintentionally modifying the contents of medical images for various reasons. One of the popular malicious modification of medical image scans is through the insertion or deletion of lesions in the

medical image [74, 164]. This concept is illustrated in Figure 3.1 for a lesion replacement in an ultrasound scan.



(a) (b)

Figure 3.1: Cyst lesion replacement [164]: *Lesions can be added to falsely cause panic or removed to conceal evidence.*

Mousavi *et al* in [118] identified two classes of attacks that could be performed on images, including medical image scans. These include *signal processing* and *geometric* attacks. These are the two groups of attacks that respectively maintain and change the spatial position of pixel elements. Some geometric attacks, such as scaling and rotation, even change the global image dimensions.

Various signal processing methods are applied to medical images before and after their acquisition through a scanning machine. A typical signal processing operation performed before transmission of medical images over the internet is lossless compression, where image content is largely preserved while reducing its size [114]. For watermarking and Steganography, compression is considered an attack, as it may lead to the loss of embedded information. However, this process of size reduction is often necessary for large-sized images transmitted in Teleradiology. Medical images also include coloured images in JPG format.

The ***objective*** of this chapter is to modify the fundamental SS method such that it can be used to retrieve the embedded information at 100% accuracy while detecting and localising different kinds of tampering on the medical image. We assume that the hidden message is embedded either immediately after acquisition from the machine or when all processing must have occurred, but before diagnosis, archival, storage or transmission to a third-party. Moreover, this third-party (or a patient) is not required to

modify the image in any form. The intended outcome in practice is to provide complimentary security for lightweight cryptography by using simpler algorithms that could be implemented in mobile health Teleradiology applications such mobile Telediagnosis and wearables (including video cameras) collecting, analysing and transmitting biosignals from the human body.

We show both theoretically and empirically, that based on the **distortion model of attacks**, our algorithm can yield zero error probability when there are no attacks and that very high image fidelity of up to *65dB* can be obtained for even 8-bit pixel depth of images and up to *100.89dB* for 16-bit DICOM images thereby preserving the diagnostic integrity of medical images. These values are obtained at a hiding capacity of about 90% of the maximum image bandwidth due to the controlled embedding technique.

In the following sections, we will present the notations applied in this and the following chapters. We will also introduce general concepts and works done in integrity detection before we propose, analyse and evaluate our new method.

3.2 Symbols and Notations

Unless where otherwise stated the following notations and corresponding interpretations shall be assumed throughout this thesis.

1. $\mathbf{X} \in \mathbf{M}^{m \times n}$ is a host (or cover) image, where $m \times n$ is the image size and $\mathbf{M}$ is the image alphabet (in pixel or transform domain). The steganographer intends to embed a set of l watermark bits, $s_i = \{s_1, s_2, s_3, \ldots, s_l\}$ into $\mathbf{X}$.

2. The image, $\mathbf{X}$ is broken into $\mathbf{L}$ sub-blocks of size $m_1 \times n_1$ such that $\mathbf{L} = \frac{m \times n}{m_1 \times n_1}$. Hence, $\mathbf{X} = \{x_1, x_2, \ldots, x_{l \leq \mathbf{L}}\}$. We assume that l, the number of watermark bits to be inserted is less than or equal to the number of blocks in the image, L. That is, $l \leq L$.

3. Each sub-block, x_1 carries one hidden information bit, $s_i \in \{\pm 1\}$ for the purpose of this chapter but can carry more as will be seen in Chapter 4.

4. Embedding can be performed in either pixel domain or transform (DCT, DWT, SVD or NSCT) domain, $\boldsymbol{T}$.

5. $\mathbf{Y}$ is the stego or watermarked image. It is the result of watermarking $\mathbf{X}$ with s.

6. $\mathbf{W}$ is a single spreading (pseudo-random) sequence, which is a member of a key-generated set of sequences. We have used the Gold sequence in this thesis.

7. α is the embedding strength. In our methods, α is not constant but is dynamically computed by a host-cancellation method to meet the requirements of integrity checks and zero error during watermark retrieval.

8. *base correlation value (BCV)* denoted as ρ is a pre-chosen value. It defines the channels in an image into which watermarks are embedded.

9. $\boldsymbol{n}$ refers to different distortion attacks for a given mean and standard deviation (where applicable) could be added to $\mathbf{Y}$ in transit or storage.

10. $\mathbf{Z}$ is an attacked version of $\mathbf{Y}$.

11. μ, σ and σ^2 are mean, standard deviation and variance, respectively.

The parameter ρ, as well as the key(s) used to generate $\mathbf{W}$, are shared between the sender and receiver and are used as information security key and tamper detection threshold, respectively. The variables $\boldsymbol{n}$, $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are empirical values from a distribution $\mathcal{X}$. When the variables are assumed to have a distribution, a probability density function (PDF) or Probability mass function (PMF) will be defined and used for analysis; if not, statistical parameters from the empirical data will be used. For convenience, the subscript for variables will be omitted where the context is clear.

3.3 Attack Motivations and Detection Models

In this section, we explore the reason for integrity degradation attacks for medical images, different types of active attacks and current tamper detection methods.

3.3.1 Motivation for Medical Image Tampering

It is pertinent to understand the various reasons why attackers can choose to tamper with a medical image. According to insights on these reasons as put forward by Mirsky *et al* in [110] and summarised in Table 3.1 there are different motivations for an attacker. These include personal ideology, political reasons, financial gains, revenge and craving for attention and recognition. The attacker may not have the ability to carry arms but has access, as an insider or outsider, to the health record of the targeted individual or group. They can alter their medical scans by adding or removing lesion to cause misdiagnosis, leading to injury, death, fraudulent insurance claims and mental trauma. As an example of insurance fraud, someone can electronically alter his/her scan to include evidence of brain hemorrhage after a faked accident to make huge financial claims from the applicable insurance company. This and other motives, including how to use man-in-the-middle to execute these attacks, have been recently demonstrated by Mirsky *et al* in [110].

For the above reasons, both data security and medical forensics sometimes mandate that certification is obtained about the integrity of transmitted medical images used for diagnosis. In the next section, we will introduce the model of attacks that we consider in this thesis, which are only the attacks that distort the cover image **X**.

3.3.2 Distortion Attacks on Medical Images

Our attack model is based on the fact that any form of manipulation introduces distortion on the original medical image. In this chapter, we focus on distortion attacks as defined by [127], where an attacker processes the image to modify the content in some ways.

By formal definition:

Definition 3.1. (Distortion Attack). Let there be a watermarked Image, $\overline{X}$ containing the spreading sequence, W. A distortion attack is defined as $q(\overline{X})$ such that $D(q(\overline{X}), W) = failure$ due to a degradation of magnitude ω from the watermarked image, $\overline{X}$. q(.) is

Table 3.1: Motivations, goals and effects of medical image tampering

Motivations	**Actions/Goals of Attacker**	**Effects on Victim**
Ideological	• Remove a Leader • Terrorize people	Life course, injury, death, trauma and loss of money
Political	• Remove a Leader • Alter election • kill opponent	Life course, death, trauma, loss of money
Financial gain	• Steal job position • Remove Leader • Falsify Research • Insurance Fraud • Ransom Payment	Loss of money, life course, trauma, injury
Attention	• Steal job position • Remove Leader • Falsify Research	Loss of money, life course, injury, trauma
Revenge	• Steal job position • Remove Leader • Murder an individual • Terrorize people	Loss of money, injury, death, trauma, life course.

an applicable image processing technique. D(.) is a decoding function

In Figure 3.2 the distortion, D_1 is caused by embedding one or more bits in a sub-block. The distortion, D_2, is external to the watermarking system, and it is introduced by either an adversary, a communication channel or a post-processing operation after watermarking. D_2 is what we consider as an attack in this research even though it is agreed that watermarking introduced allowable distortion to the original image. This assumption is valid necessary in practical steganography because perfect security is nearly impossible in this case due to a lack of accurate models to describe non-stationary sources such as images [6, 65].

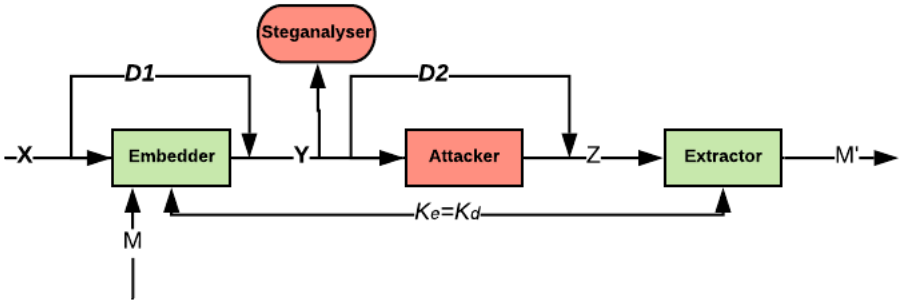


Figure 3.2: Attack Model: *There are two major distortion points, D_1 and D_2. In this chapter, we consider only D_2 distortion attack as it is external to the encoder system and thus probably malicious.*

Distortion attacks to Steganography and watermarking are broadly classified into ***signal processing*** attacks and ***Geometric*** attacks [127, 126]. The signal processing attacks can be intuitively divided into those attacks that add noise to an image and those who intend to remove these noise and re-condition the image for further use (pre-processing operations). Noise-based attacks are often modelled as additive noise of a given mean (μ_n) and standard deviation (σ_n). For Example, a generalise Gaussian noise attack can be represented by $N(\mu_n, \sigma_n^2)$, with white Gaussian noise given as $N(0, \sigma_n^2)$. The specific signal processing attacks considered in this work are discussed further next.

Noise-based Attacks

Noise contaminates the quality of medical images and affects the accuracy and precision of medical image interpretation and feature extraction by humans and machines. Noise can be introduced into a medical image during its acquisition, transmission, or storage [60]. It is important to detect these noise effects in medical images as not being able to do so may lead to diagnostic errors both by humans and auto-diagnostic systems. Therefore, it is obvious that malicious users can deliberately introduce noise into existing high-quality medical images to cause misdiagnosis. Among the noises that significantly affect the interpretability of medical images, Gaussian is the most prevalent, especially the Additive White Gaussian Noise (AWGN) [60]. Gaussian noise uniformly affects medical images and greatly reduces the interpretation accuracy by humans and computers. We briefly summarise other types of noise attacks below:

1. ***Gaussian White noise*** - This refers to uncorrelated Gaussian and additive noise added to image pixels. For example, the amount of noise added to one pixel is unrelated to the amount of noise added to the next pixel.

2. ***Poisson noise*** - This is also called a short noise or impulse noise. It can be modelled as a Poisson process. A Poisson process is a set of events that occurs at known average time intervals, but the exact time for each and next event is unknown [1]. This occurs mostly in electronic equipment.

3. ***Salt and Pepper*** - This is a fixed value impulse noise introduced into an image. It is usually seen as sharp black and white disturbances in an image. They are made up of the minimum and maximum intensity values in an image such as 0 and 255 for an 8-bit image.

4. ***Speckle Noise*** - This is a common type of noise in medical images due to the effect of the environment on the imaging machine [99]. Speckle noise reduces the

[1] https://towardsdatascience.com/the-poisson-distribution-and-poisson-process-explained-4e2cb17d459

quality and accuracy of image interpretation and analysis. Speckle is not always a noise but could be formed as a result of the signal fluctuations itself. It often appears as a granular interference and is common in ultrasound and optical coherence tomography (OCT) images. It can be reduced using the median, mean, and Lee filtering.

Attackers utilise the mathematical and statistical properties of the above noise processes while designing how to introduce them into medical images to achieve any of the motives in Section 3.3.1.

Processing Attacks

This includes the application of filters and other mathematical and statistical functions to modify a medical image's intensity or content. The textures in medical images may seem like noise, but they are the differentiating features for machine learning algorithms used in autodiagnosis. The attempt to remove these features may lead to misclassification of images in terms of disease severity. The most common processing attacks considered in this thesis include:

1. ***Gaussian Filtering*** - This is used to remove noise and details in an image. It is used to blur images. The behaviour of the Gaussian function is highly influenced by the standard deviation. The data points located at $\pm\sigma_n$ represents 68% of the data set. Values within $\pm 3\sigma_n$ from the mean represent 99% of the entire data set. Gaussian filters are low pass filter and for images, is based on the 2-D Gaussian function given as [58, 60]:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{3.1}$$

   Central pixels on the spatial image location has a higher weighting than those at the edges. Gaussian filter is used for smoothening images.

2. ***Median and Mean Filtering*** - These are filtering operations that are often used to

reduce salt and pepper noise in an image. It is also an attack because it is a non-linear filter which filters pixels that are not likely salt and pepper pixels. Most of them are non-adaptive and thus introduce artefacts in an image, thus becoming a form of attack on the image.

3. ***Histogram Equalisation*** - This is a form of contrast adjustment where the values of pixels in an image histogram are redistributed such that all intensities get the almost equal frequency. This processing attack is often used when the ROI of the image is represented by very close intensity values. As this changes the image content and brightness, it is a form of distortion on the original image. It generally provides a better view of bone structures in X-ray films.

4. ***Contrast adjustment*** - This another form of intensity adjustment but not necessarily equalising the number of pixels for each intensity value. Contrast adjustment is one of the most popular signal processing applied to medical images as it usually has poor contrast for computer vision and pattern recognition. Specifically, we used the ADAPTHISTEQ function, which is a Contrast-limited Adaptive Histogram Equalization (CLAHE). It enhances the contrast of images by transforming the values in the intensity image, I. Unlike HISTEQ() function; it operates on small data regions (tiles of [8 8] default value), rather than on the entire image.

5. ***JPEG Compression*** - Source coding is a common practice in digital image processing. In this case, the original data is represented in a form that reduces the amount of transmitted data without significantly changing the perceptual content of an image. The Joint Photographic Experts Group (JPEG) compression standard is probably the most common image format. It is a lossy compression algorithm, though it provides a quality factor to enable a trade-off between image quality and storage/transmission space/bandwidth. Because it normally removes some parts of the image, it is considered an attack.

Geometric attacks

Various types of Geometric attacks are modeled differently. Some of these models are presented in [78], while they were further explained in [72]. The following geometric attacks on an image are considered in this thesis:

1. **Rotation**: This is a form of geometric attack that moves the position of a pixel (x_1, y_1), to the location of another pixel (x_2, y_2) by rotating the first pixel through an angle θ and about an origin (x_0, y_0), according to (3.2a) and (3.2b) respectively.

$$x_2 = x_1 cos\theta + y_1 sin\theta \tag{3.2a}$$

$$y_2 = -x_1 sin\theta + y_1 cos\theta \tag{3.2b}$$

   This attack would have a distortion effect on medical images as some pixels may move out of the originally defined region of interest and thus create a different interpretation for an expert. In terms of auto-diagnosis, this will have a different interpretation for a machine learning model, depending on the dataset used for the training. It is, therefore, important to detect distortions caused by rotations at a block-level within an image. The direction and angle of rotation affect object recognition.

2. **Scaling**: This is the resizing of an image. It includes both reductions in size and enlargement of an image. In this research, we assume that an attacker will not change the final size but could scale up or down and then back to the original size of an image. During scaling up or down, the editing algorithm determines what intensity value to give new pixel positions or what pixels to remove from the original image. Hence, scaling up or down and then back to the original image size is a form of distortion because the algorithm would not produce the same pixel intensity values as the original image of the same size.

3. **Translation**: In a translation attack a pixel position (x_1, y_1) is moved to a new pixel position (x_2, y_2) according to a pre-defined distance (β_0, γ_0). Hence, the translation equation becomes:

$$(x_2, y_2) = (x_1 + \beta_0, y_1 + \gamma_0) \tag{3.3}$$

Translation attack moves pixels around and could lead to a change in image texture.

4. **Block replacement attacks**: This is a kind of attack that involves the replacement of a block of an image with either perceptually similar or dissimilar blocks. This attack scenario has two separate implications in this thesis. If the replacement block and the original block are perceptually similar, then the diagnostic outcome may not change, but watermark detection may still be affected, and then tampering may need to be detected. On the other hand, if the intention is to replace benign lesions with malignant ones or vice versa, both watermark detection and the diagnostic outcome will be affected. Again, tampering will need to be detected.

RST attacks: Rotation, Scaling and Translation (RST) are grouped as affine Transforms.

For all attacks, it should be noted that both authorised users and attackers are subject to distortion constraints and image size changes. Specifically, in this thesis, we assume that image sizes are standardised and both the watermark security provider and the attacker are not allowed to change the image size. Hence, attacks that result in the change of a medical image's original size were not considered further in this thesis.

Finally, whether watermarking itself or an attack on a watermarked image is perceptually significant or not, the implication for auto-diagnosis, where computer vision and machine learning are applied is a different consideration altogether. In this chapter, we are concerned with an individual block or several blocks in a single image. This is opposed to statistical machine learning and big data computational techniques for decision making in auto-diagnosis. In this chapter, we assume the traditional man-in-the-loop medical diagnosis, monitoring, and treatment. In Chapters 5 and 6, we will

look more into automated diagnosis and the implications of image watermarking and attacks on autodiagnosis.

3.3.3 Image Tamper Detection Methods

Image tamper detection is a form of content authentication or content integrity verification. Multimedia content authentication is implemented by cryptographic means, through watermarks [74, 92] or via statistical computation on the retrieved image [5]. These approaches could be either active or passive.

The method of image authentication and tamper detection, which does not involve the introduction of extra data into the multimedia is said to be ***passive*** while the methods which require some data such as watermarks to be introduced is said to be ***active*** [111, 5]. Cryptographic signatures (if embedded) and watermarking methods are active methods, while most statistical methods are considered passive. Though [111] had highlighted the setbacks of active methods, they acknowledged that it remains the most efficient method of image authentication. Figure 3.3 is a summary of this classification.

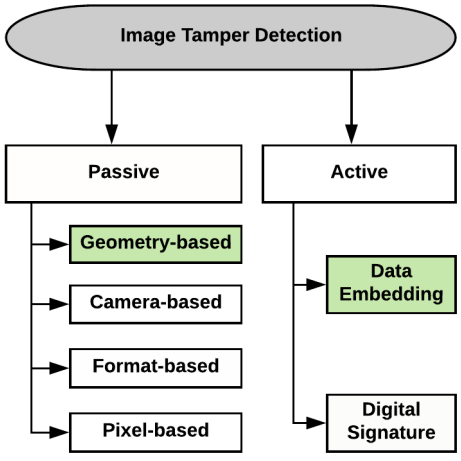


Figure 3.3: Methods of Tamper Detection: *active methods involve the embedding of data which may serve other purposes apart from detecting tampering. Passive methods do not embed any data but uses computational methods to estimate tampering*

Digital signatures have been historically used to identify unauthorised modification of an encrypted document from a cryptographic point of view. In recent times, both digital watermarking and steganography have also contributed to multimedia tamper detection and possible recovery [74, 92]. However, the digital signature uses a cryptographic hash that does not tolerate even a single bit change in the original data. On the other hand, traditional digital watermarking is so robust that it can withstand several manipulations, including some significant compression of data content. Whereas the case of a digital signature is not realistic for medical image processing and storage, the later is also not ideal as some diagnostic information may have been lost at some extreme robustness. This phenomenon leads to the quest for a virtue that lies in the middle.

C.U Lin in [92] was able to differentiate between the concepts: complete authentication and content authentication. Complete authentication considers the entire multimedia as a whole and does not allow any form of modification, including transformations. This technique is appropriate for text (EMR), secret keys or transaction data, where a change in bitstream affects the meaning of the final decoded data. For multimedia data, some bits could be changed without changing the meaning of the image or video. In this sense, the content is still preserved. Examples include lossless compression, contrast adjustment, signal filtering, among others. Digital signatures and fragile watermarking can be used for complete authentication, while robust digital signatures or semi-fragile watermarking/steganography are used for content authentication [92, 118].

Wolfgang and Delp in [157] developed a block-based authentication system with the help of m-sequences in the [-1, 1] domain. The watermark was generated from different bit planes of a pixel apart from the LSB. They assumed that only the LSB could be affected by a lossy compression algorithm. However, it is not true that only LSBs are affected by all compression algorithms. Though a lossy compression is not recommended for medical images, other acceptable post-processing for a medical image that affects other bit planes apart from the LSB will result in a false negative.

Saini *et al* in [5] proposed a non-blind passive method of determining image tampering. They used the mean vector and correlation coefficient methods for detecting forged parts of BMP images. Both the original and possibly forged images are required. For an image of M rows and N columns, the row mean vector(RMV) is given as:

$$RMV = \frac{1}{M}\sum_{i=1}^{M} r_i \tag{3.4}$$

While the column mean vector(CMV) is given as:

$$CMV = \frac{1}{N}\sum_{j=1}^{N} c_j \tag{3.5}$$

r_i is the *ith* row while c_j is the *jth* column. For forged images, Original Image (r1,c1) value will not match forged image's (r2,c2) value.

They also proposed the 2-D correlation between original Image and potentially tampered image. This correlation value is always in the domain [-1,1]. An image is deemed tampered with if the correlation per sub-block at the destination differs by more than 0.025 between the computation at the source and the computation at the destination.

This method requires the availability of both forged and original images to detect tampering, which is non-blind and not suitable for Teleradiology, where the original image is not available at the destination. Furthermore, even the lossless compression method in the ROI may trigger the 0.025 correlation threshold defined by these researchers. This method would lead to an unnecessary false alarm.

It is important to note how methods that apply to other multimedia may be different from those of medical images. For instance, the level of compression acceptable in a non-medical JPEG image may have caused many diagnostic defects in medical images. For this and other reasons such as localisation of tampering in the ROI, this research will focus on techniques implemented using Spread Spectrum (SS) techniques, which we have noticed would give the kind of flexibility needed depending on the domain(spatial or transform) of watermarking.

On a related note, one can easily advocate passive tamper detection methods for

medical images as no distortion of any kind is expected. However, with active embedding, side information is transmitted, and more utility applications could be implemented. The challenge is to minimise distortion, preserve diagnostic information content while improving the capacity of such side information transmitted. The solution to and validation of the severity of these later challenges constitute the subjects of Chapters 4 and 5. In the next section, we present our tamper detection solution based on SS Steganography.

3.4 Proposed Tamper Detection Technique

We opted for blind but active tamper detection methods that detect hidden information with a high probability of accuracy. A blind watermarking method does not require the original cover image at the decoder. Active tampering changes the content of the image attacked. The proposed method starts with a high-level framework shown in Figure 3.4. It then goes further to present a tamper detection model design and analysis using Figure 3.5. Further analysis is presented using some of the equations already stated in section 2.3.3. The pseudo-code of the resultant algorithms are then presented.

3.4.1 High-Level Design Framework

Image Tamper Detection mechanism proposed in this chapter (and thesis) involve encoding fragile tamper detection information in the region of interest (ROI) and other information and other authentication information in the region of non-interest (RONI). This framework is illustrated in Figure 3.4. Thus, the main goal is to design a watermarking method that could be used as both semi-fragile and robust watermarking method but while keeping distortion in ROI as low as possible, but at zero bit error rate during watermark extraction provided no tampering occurred. We have utilised data sets whose ROI has been pre-defined by experts and thus did not develop any new segmentation algorithm. Also, there are existing [38, 40] robust tamper-resistant algorithms that are based on the spread spectrum that can be used in RONI. Hence, this

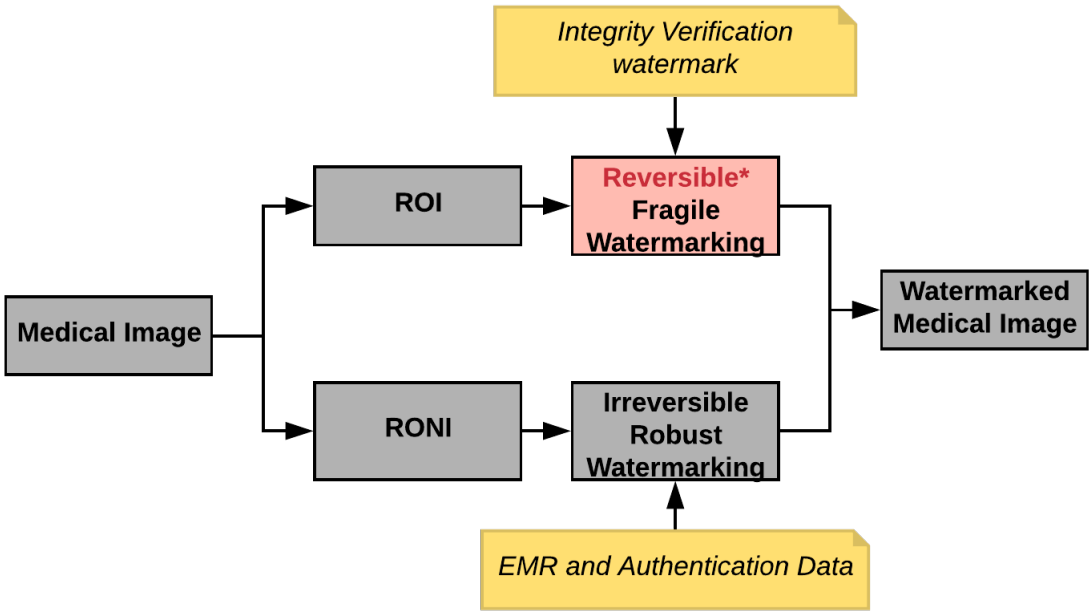


Figure 3.4: Overall Proposed Framework: *Medical images were divided into regions of interest (ROI) and Regions of non-interest (RONI). In this work, the integrity verification data was inserted in the ROI to ensure localisation of tampering. Reversibility is optional.*

chapter's major contribution is the fragile SS tamper detection technique that can also be configured for robust embedding and extraction. Hence, a Semi-fragile watermarking in the spread spectrum spatial/transform domain is proposed to detect tampering in the ROI. Next, we present the design and analysis of the proposed technique.

3.4.2 Design Concepts

In this proposed technique, the general idea is that instead of allowing the correlation value for decoding a 1-bit to be any correlation value greater than zero, we want to ensure that the correlation value is within a pre-defined positive value, ρ. Also, a 0-bit is returned if the computed correlation value is $-\rho$.

As ρ can be any real number (not just integer) bounded - upper and lower - by some values determined by the allowable level of distortion (measured mainly by PSNR in

this thesis) in the cover image, the sender and receiver can agree on the value of ρ beforehand. This parameter serves as a secondary secret key after the key used to generate **W**.

The detailed design and analysis are based on the model shown in Figure 3.5. The value ρ (and its complement, $-\rho$) represents the ***base correlation value (BCV)*** agreed by the sender and the receiver as the correlation value, equivalent to a frequency channel in a frequency hopping spread spectrum communication, to be utilised for current transmission in the image sub-block.

For blind tamper detection at the receiver, the linear correlation, called the detection statistics, $\boldsymbol{r}$, must fulfil the condition:

$$r = \pm\rho \pm \epsilon \tag{3.6}$$

The value of ϵ represents a tolerance on the original value of ρ. For instance, the allowable deviation from BCV above which a sub-block can be flagged as tampered lies in the range $\pm\rho \pm \epsilon$.

Equation (3.6) also defines the preferred location for our security watermark bit, called *Bit Zone (BZ)* as it is close to the origin and thus has the lowest distortion due to lower embedding strength required to move the correlation value from its original position before watermarking (See HSI in Section 2.3.3).

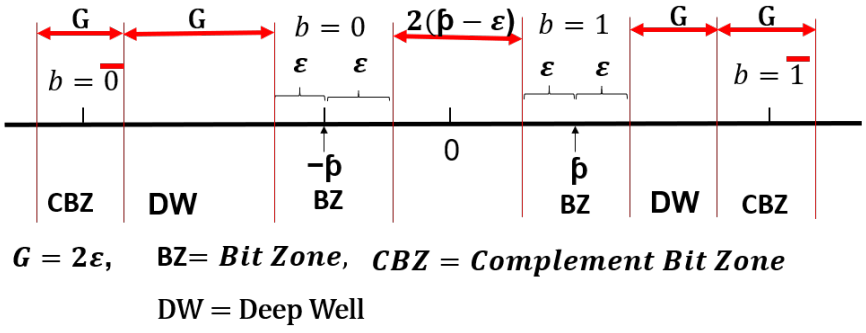


Figure 3.5: Model for C_4S Algorithm Design: *we modified the classical SS embedding strategy such that one is exactly sure of the correlation value, ρ to expect in the BZ and CBZ instead of checking for it at* ± 1 *to* $\pm\infty$ *or at* $\pm T_h$ *to* $\pm\infty$.

A *Complement Bit Zone (CBZ)* was introduced to encode a tamper detection bits as its complement bits if the embedding strength required to encode the original bit was too large. For example, if the computed embedding strength required to change a block from negative to positive correlation (1-bit) will introduce a distortion larger than ϵ, then the 1-bit will be encoded in the defined negative CBZ. Hence, the 1-bit is encoded as $\overline{0}$.

A correlation distance of width $G = 2\epsilon$ is maintained between any two embedding channels to create a deep well, DW that could trap an attacker who does not have the secret key, **W**. From Figure 3.2, the distortion D_1 is equivalent to ϵ and is tolerable based on the application. However, any modification (by an attacker or steganographer) that is greater than ϵ lies within the D_2 distortion and thus must be detected. Hence, the authorised user's advantage is the knowledge of the secret key, which could be used to maximise security and minimise distortion. The attacker will not have this advantage and will either perform a perturbation that remains within the steganographic bounds or cause a noticeable distortion that could be detected by either out-of-bound correlation value or increase in BER at the receiver.

In Summary, there are two major principles upon which the C_4S algorithm depend:

1. Pre-determining the distribution of the detection statistics, r, by initially choosing its mean value ρ at the encoder.

2. Using the encoder's knowledge of the host signal interference(HSI), x to determine optimum embedding strength to cancel this effect [38] and ensure that ρ is effectively the detection statistic at the receiver if the image has not tampered in any way.

To solve overflow and underflow problems, each sub-block within the medical image ROI was dynamically and conditionally pre-processed to reduce HSI and quantisation errors before embedding. Adaptive histogram shifting methods were employed for this purpose.

For practical applications, ***a desired error probability, as well as distortion in the***

form of Signal-to-Noise (SNR), are often specified. These will be used to design or select the appropriate IH method for the application. This assumption shall form the basis for practical evaluation of this or any other algorithm in this thesis, using the 40dB benchmark for medical images.

Before going into analysis and empirical experiments, it is important to describe the embedding function, dynamic embedding strength computation, and the decoding function. These form the fundamental component of any data hiding algorithm as they determine accuracy, distortion, and fragility (or robustness). The corresponding pseudo-codes are also presented alongside the embedding and extraction functions.

3.4.3 Embedding function

The embedding equation remains the same as that of the classical spread spectrum additive embedding:

$$Y_{ij} = \begin{cases} X_{ij} + \alpha W_{ij}, & \text{if } s_k = 1 \\ X_{ij} - \alpha W_{ij}, & \text{if } s_k = 0 \end{cases}, \tag{3.7}$$

where $\mathbf{Y}$ is watermarked image block, $\mathbf{X}$ is the cover image block, α is the embedding strength or watermark amplitude or watermark power and $\mathbf{W}$ is the PN code. $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{W}$ are of the same dimension, $\mathbf{m} \times \mathbf{n}$. i and j refer to individual pixels or transform coefficients in the k_{th} sub-block cut out from the global image $\mathbf{X}$ and $\mathbf{Y}$.

if $s_k = [0, 1] \rightarrow [-1, 1]$, then we can simply re-write (3.7) as:

$$Y_{ij} = X_{ij} + \alpha s_k W_{ij}. \tag{3.8}$$

Equation (3.8) is the embedding equation, and it remains the same as that of traditional SS watermarking. We will ignore the subscript in s and simply write s instead of s_k. This is because only a single sub-block will often be used for specific design and analysis. The detailed algorithm for the embedding function is shown in Algorithm 1.

Table 3.2: Parameters and functions used in the algorithms

Parameter	Definition
Cr	Number of bits embedded in a sub-block
X	Original (host DICOM) Image
X_i	i^{th} sub-block in X
Y	Watermarked Image
Y_i	i^{th} sub-block in Y
ρ	Base Correlation Value
ϵ	Error tolerance and determinant of channel gaps
nearNegZeroCorr	The first embedding on 0^-
numChannels	number of embedding channels required (2 by default)
nearPosZeroCorr	The first embedding on 0^+
d	Local distortion value in a sub-block
Msg	Message to be sent in binary format
$\langle yblock, W \rangle$	Linear Correlation
$L = length(Msg)$	Length in bits of watermark to be embedded
mod(L,Cr)	Modulo Cr division of L

The parameters used in this and other algorithms in this thesis are listed in Table 3.2.

Algorithm 1: Compression Encoding Algorithm (CEA)

Data: X, ρ, W, Cr,ϵ, Msg

Result: Watermarked Image, Y, Quality Parameters, list[]

Divide X into 8x8 sub-blocks, X_i. (m = n = 8);
Compute length, L = length(Msg);
numChannels $= 2^{Cr}$;
$nearNegZeroCorr = -\rho; nearPosZeroCorr = \rho$;
ChannelGap $= 4 * \epsilon$, i=1;
msg = Msg(i);
while *msg in Msg* **do**
    Generate spreading code, W;
    compute α using ρ as in (3.9);
    Embed current message bit into X_i using (3.7);
    Compute block-based Quality Parameters;
    $i = i + 1$;
    msg = Msg(i);
end
Compute Global ROI Parameters;
Plot the required graphs;

The naming of this algorithm as the Compression Encoding Algorithm (CEA) will be explained further in Chapter 4, where it was extended to embed more than a single bit of information in a sub-block.

3.4.4 Embedding Strength

For tamper detection, any image block that does not return a correlation value of $\rho \pm \epsilon$ (or -$\rho \pm \epsilon$) is deemed to have been tampered by an adversary or has been affected by a burst error or noise. This image block will be detected and highlighted. The parameter ϵ is a small allowable error resulting from quantisation of image values (pixel intensity

for spatial domain embedding) or mild processing. The idea of constant correlation method is to ensure the L.H.S of (2.25) is constant and equal to the predetermined value, ρ. That is, $\rho = \langle Y, W \rangle$. This can be achieved by making the embedding strength, α, the subject of the expression in (2.24) and using the computed value of α to perform the embedding in (3.7). This gives the value of α to be substituted in (3.7) as (3.9).

$$\alpha = \frac{\langle Y, W \rangle - \langle X, W \rangle}{s} = \frac{\rho - \langle X, W \rangle}{s}. \tag{3.9}$$

Hence, our technique is a shift from a constant embedding strength to a *dynamic embedding strength* and from a dynamic linear correlation to a *constant linear correlation*.

3.4.5 Extraction function

The desired constant correlation, ρ, is any real number agreed in advance between the sender and receiver. This means that decoding equations of (2.6) or (2.7) can no longer assume just any value on the number line but $\pm\rho \pm \epsilon$. This gives us the watermark decoding function in (3.10).

$$\bar{s} = \begin{cases} 0, & \text{if } r = \langle Y, W \rangle = \rho \pm \epsilon \\ 1, & \text{if } r = \langle Y, W \rangle = -\rho \pm \epsilon \end{cases}, \tag{3.10}$$

where ϵ is the tolerated error deviation from unintentional attacks, noise and quantisation errors. We will assume $\epsilon = 0.5$ in this thesis for pixel domain of embedding (based on Continuity Correction in Statistics). However, in general: $0 < \epsilon \leq \rho$. The detailed extraction and tamper detection techniques are provided in Algorithm 2.

In the next section, a detailed design and analysis will be provided. We call this technique, ***Constant Correlation Compression Coding Scheme*** , C_4S. This naming is more obvious in Chapter 4 than in this chapter. However, we will use this term from this point onward.

Algorithm 2: Decompression Decoding Algorithm (DDA)

Data: Y, numChannels,Y, W, Cr, ρ, ϵ, L
Result: EMR, tamperdetected

Divide it into 8x8 sub-blocks, Y_i (m = n = 8);
polarisedChannels, $pc = numChannels/2$;
nzbitvalue = $pc - 1$;
pzbitvalue = pc;
thisbitgroupvalue =$'$ 1000000$'$;
$i = 1$;
Msg = $''$;
while *i less than or equal to L* **do**
  $P = \langle yblock, W \rangle$;
  $binarybits = null$;
  $blockStatus =' NoWatermark'$;
  if P_j= -$\rho + \epsilon$ **then**
    while *j less than or equal to nzbitvalue* **do**
      if $P = -j * \rho \pm \epsilon$ **then**
        $thisbitgroupvalue = f(j, P)$;
        $binarybits = binary(thisbitgroupvalue)$;
        $Msg = Msg + binarybits$;
        $blockStatus =' WatermarkFound'$;
      end
      $j = j + 1$;
    end
    if $binarybits == null$ **then**
      reportTampering();
      $blockStatus =' Tampered'$;
    end
  end
  if $P >= \rho - \epsilon$ **then**
    while *j less than or equal to pzbitvalue - 1* **do**
      if $P = j * \rho \pm \epsilon$ **then**
        $thisbitgroupvalue = f(j, P)$;
        $binarybits = binary(thisbitgroupvalue)$;
        $Msg = Msg + binarybits$;
        $blockStatus =' WatermarkFound'$;
      end
      $j = j + 1$;
    end
    if $binarybits == null$ **then**
      reportTampering();
      $blockStatus =' Tampered'$;
    end
  end
  $i = i + 1$;
  reportBlockStatus(blockStatus) ;
end
Compute Watermark Quality Parameters;
Convert extracted EMR, *Msg* bits to text;
Flag tampered blocks, if any;
Display EMR and tampered version of medical image;

3.5 Analysis of the Proposed Technique

In this section, we perform a theoretical analysis of the new method introduced in Section 3.4. However, the analysis is connected to empirical evaluation to reduce the gap between theory and practice.

3.5.1 Problem formulation

Regarding the general notations defined in Section 3.2 and algorithmic symbols provided in Table 3.2, we then proceed with the analysis of the new tamper detection method.

To proceed, let us consider the detection statistic in (2.10) and make the embedding strength α the subject of the equation:

$$\alpha = \frac{r - x - n}{s}. \tag{3.11}$$

The variable x can be completely determined at the sending end because it is the host signal interference (HSI) - the correlation between the host signal, $\mathbf{X}$ and the spreading sequence, $\mathbf{W}$. This is popularly called the projection of x on $\mathbf{W}$. However, not all the components of $\boldsymbol{n}$ (attack noise) can be determined at the sending end because they are introduced by either the physical channel or by an attacker (apart from the quantisation errors where applicable).

Moreover, in tamper detection problems, it is the presence of $\boldsymbol{n}$ that we want to detect at the receiver. Hence, without any form of attack the detection statistics, $\boldsymbol{r}$ is simply same as removing $\boldsymbol{n}$ from (2.11). Also, the embedding strength, α required to ensure that the pre-determined $\boldsymbol{r}$ (which is called ρ at the encoder) is correctly retrieved at the receiver is determined by also removing $\boldsymbol{n}$ from (3.11). Hence, at the encoder, the dynamic embedding strength α , for each sub-block is first computed as:

$$\alpha = (r - x)/s. \tag{3.12}$$

Thus, the challenge is to find a modulating function, g, which has the following properties. It:

1. embeds the desired correlation ρ, that enables one to detect tampering at the receiver if $r > |\pm\rho \pm \epsilon|$. This is a *Tamper Detection* Problem,

2. cancels the HSI, x via the dynamic computation of α, to ensure accurate watermark ($s = [-1, 1]$) retrieval in the absence of attacks. This is a *Watermark Detection* Problem.

3. construct a controlled watermark insertion strategy that ensures that the distortion from watermark alone does not exceed a threshold allowable for accurate diagnosis ($PSNR >= 40dB$ in this case). This is a *Distortion Control* Problem.

3.5.2 The C_4S Spread Spectrum Solution

We seek for a modulation function of the form, g(r, x, s) which can maintain the the nature of the original spread spectrum embedding equation . This function would replace αs in (2.8). Hence, the general form of our new embedding equation is:

$$Y = X + g(r, x, s)W. \tag{3.13}$$

We then co-design the encoder and decoder by using the value r that we want to keep constant at both ends. The original detection statistics is given by (2.12) and thus substituting 3.13 for Y (not Z as attack is not considered yet), we have :

$$r = \langle Y, W\rangle = \frac{1}{m \cdot n}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}W_{ij} \implies \frac{1}{m \cdot n}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} + g(r, x, s)W_{ij})W_{ij} \tag{3.14}$$

As g(r, x) is a scalar, (3.14) implies that:

$$r = \langle Y, W\rangle = \frac{1}{m \cdot n}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}W_{ij} + g(r, x, s)\frac{1}{m \cdot n}\sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij}W_{ij} \tag{3.15}$$

W is $\mathcal{N}(0,1)$ distributed and the first term in (3.15) is simply the host signal interference, $\boldsymbol{x}$. Also, in our design and as is the case for traditional SS, it is enforced that $\boldsymbol{r}$ should carry the sign of the polar bit it is representing as should be the case for correct watermark extraction. Hence, the predetermined correlation value retains the principle that sign($\boldsymbol{r}$) = {-1,1}. One can then re-write (3.15) as:

$$rs = x + g(r, x, s)\frac{mn}{mn}. \tag{3.16}$$

reducing (3.16) to the lowest terms and solving for g(r,x), we have:

$$g(r, x, s) = rs - x \tag{3.17}$$

Hence, the traditional SS embedding equation changes to a new one with a dynamic embedding strength, α, determined by x,s and r (ρ), where x is dependent on each sub-block, $s = [-1, 1]$ and ρ is a real number greater than zero. Hence, we have:

$$Y = X + (rs - x)W = X + (\rho s - x)W. \tag{3.18}$$

The detection equation remains

$$\overline{s} = \begin{cases} 0, & \text{if } r = \langle Z, W \rangle = \rho \pm \epsilon \\ 1, & \text{if } r = \langle Z, W \rangle = -\rho \pm \epsilon \end{cases}. \tag{3.19}$$

Equation (3.19) uses **Z** and not **Y**. This shows that the possibility of some noise addition or tampering may have occurred. Beyond the tolerance limits stated in (3.19), the image block is deemed tampered even though the embedded watermark can still be correctly retrieved by relaxing ϵ further. The parameter,ϵ is an error tolerance for acceptable distortion within an application. It determines both the level of fragility and robustness of the C_4S algorithm. This makes it to be compatible with traditional SS by simply removing this limit in the range of detection statistic $\boldsymbol{r}$ but not across the other side zero.

Hence, our detection condition now has upper and lower bounds instead of having just lower bounds as in traditional SS (2.6).

Figure 3.6 shows the intended outcome of the C_4S algorithm at the decoder in the absence of any attacks. After embedding, all correlation values for embedding negative binary bits (-1) should be located on $-\rho$, while all correlation values for embedding positive binary bit should be located on ρ.

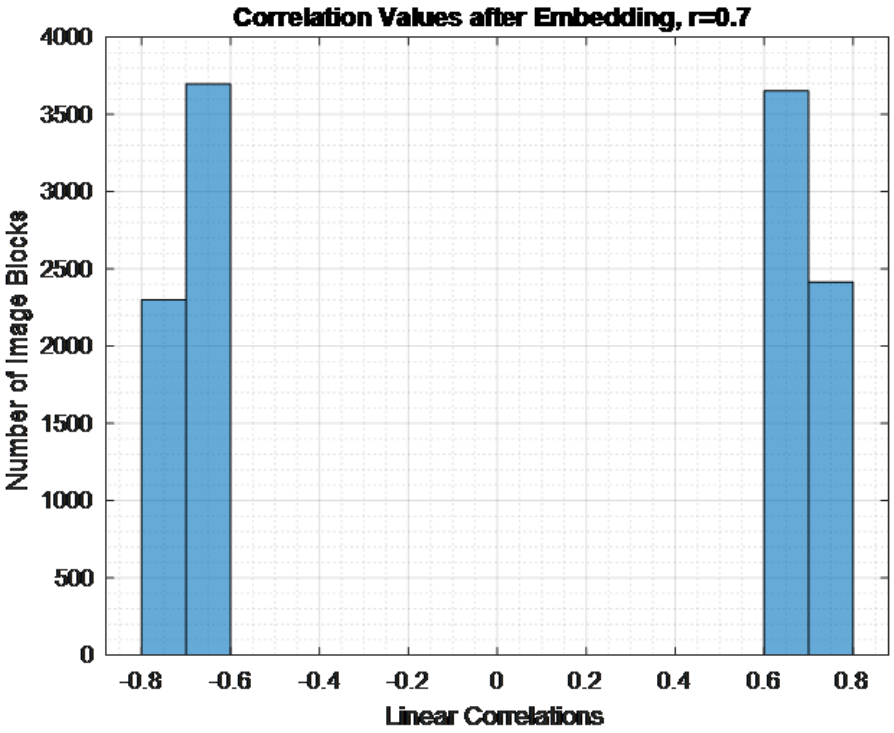


Figure 3.6: Distribution of statistic r without tampering: *In this ideal situation,* $r = \rho = 0.7$. *All correlation values at the decoder were centred on* ± 0.7. *This illustrates the solution to tamper and watermark detection problems.*

If ρ is predetermined and used for embedding, then we assume that it will also be constant at the receiver provided that there is no significant attack on the stego image. A significant deviation from ρ as determined by the distribution of $\boldsymbol{r}$ at the receiver constitutes tampering in the sub-block. Whereas $\boldsymbol{r}$ (or ρ depending on which side of the system is considered) is not recognised as a primary security key in our algorithm, it can act as a secondary security key. By analogy, the security capability of ρ in the

C_4S algorithm can be compared to the *initialisation vectors* (IV) used for block chaining in block-based cryptographic ciphers [147], though a bit weaker as it is message-dependent.

3.5.3 Steganographic Performance Analysis

As we stated in Section 2.3.4, signal distortion and bit error rate (BER) are the major parameters (Capacity will be considered in Chapter 4) for analysing the performance of Steganographic systems. Image distortion is measured using parameters such as Mean Squared Error (MSE), Signal-to-interference-plus-noise ratio (SINR), Peak Signal-to-noise ratio (PSNR), among others. In this chapter, an image Distortion, $\boldsymbol{D}$, is represented by the PSNR value of the image after data embedding or attack. This value must be greater than or equal to a minimum value for the given application. For 8-bit medical images, the minimum PSNR is 40 dB [49, 29]. On the other hand, the BER is a measure of how accurate the watermark was retrieved. Accuracy reduces if more blocks are effectively modified.

3.5.3.1 Image Distortion Analysis

Distortion can be modeled either at the encoder or at the decoder. This approach is important because we are concerned with blind image steganography - the one in which there is no original image at the decoder. Hence, at the encoding side, we are modeling distortion as the difference between the original cover image and the watermarked image. At the extractor, the distortion is that of the ratio of the embedded watermark to that of the noise introduced by the additive noise components. This is measured on the detection statistic, $\mathbf{r}$.

The distortion is the deviation of each pixel or transform coefficients from its original value before watermarking. The simplest approach is to take the expectation of the absolute differences (*norm*) among the elements of the watermarked entities, $\mathbf{Y}$, and

the original host signal, **X**. This can be expressed as:

$$\mathbb{E}[D] = \mathbb{E}[\| Y - X \|], \tag{3.20}$$

where $\mathbb{E}$ is *mathematical expectation* and **D** is a chosen *distortion parameter*. With respect to our embedding new embedding equation (3.18), distortion at the encoder can be expressed as:

$$\mathbb{E}[D] = \mathbb{E}[|\rho s - x|^2 \sigma_w^2|]. \tag{3.21}$$

It can be seen that minimum distortion can be achieved when $\rho s = x$. The chosen base correlation value multiplied by the polar watermark bit to be embedded equals the host signal interference. This condition produces zero distortion as the embedding strength, α, becomes equal to zero. This approach is not feasible for all sub-blocks, as we would lose the tamper detection capability of the algorithm. Hence, we want to choose the ρ that minimised α across the entire image.

$$\begin{aligned} \underset{\alpha_{sum}}{\text{minimize}} \quad & \alpha_{sum} = \sum_{i=1}^{N} |\rho s_i - x_i| \\ \text{subject to} \quad & 0 < \rho \leq 1.0, \end{aligned} \tag{3.22}$$

where N is the number of sub-blocks in an image, s_i is the bit of information to be embedded into sub-block i whose HSI is x_i.

A comparison of our method and traditional SS finds a good analogy between α and ***g(r,x,s)*** . Our goal is to design α such that it allows tamper detection while still being compatible with existing traditional SS watermarking and steganography while improving or extending its usefulness. Next, we consider the BER analysis of C_4S.

3.5.3.2 BER Analysis

Without any form of noise, the ideal detection statistic remains **r =** ρ and BER computation is not required. However, even at the encoder, some distortions (D_1) - such

as rounding off, quantisation and file format compression mechanism - could be introduced by the embedding process . We have used the error tolerance value, ϵ in the detection equation (3.10) to account for these. The comparative part of (3.10) for detecting $s = 1$ can be re-written as an inequality:

$$a \leq r \leq b, \tag{3.23}$$

where $a = \rho - \epsilon$ and $b = \rho + \epsilon$ [2].

The probability of correctly detecting a bit using the pre-determined base correlation value, ρ (and thus accurately detecting tampering or not within the sub-block) can be written as the probability that the linear correlation, r at the decoder, is in the range [**a**, **b**]. That is:

$$p_c = Pr(a \leq r \leq b | s = 1). \tag{3.24}$$

Consider a white Gaussian noise addition of $n_i \sim \mathcal{N}(0, \sigma_n^2)$, which is added to ρ. The new distribution becomes $n_i \sim \mathcal{N}(\mu_r, \sigma_n^2)$. The noise modelled at the receiving end include the HSI as well as other types of noise. Hence the detection statistics, r , at the receiver is of the form:

$$r \sim \mathcal{N}(\mu_r = \rho, \sigma_r^2), \tag{3.25}$$

where:

$$\mu_r = |\rho| s \tag{3.26}$$

and

$$\sigma_r^2 = \frac{\sigma_n^2 + \sigma_x^2}{N \sigma_w^2}. \tag{3.27}$$

It has also been shown by Malvar *et al* in [102] for the generalised ISS that r has a Gaussian distribution of the form $r \sim \mathcal{N}(\mu_r, \sigma_r^2)$. The full expression for Gaussian

[2] Note that we have used r and ρ interchangeably in most parts of this thesis. We have used r to represent the distribution of ρ. Whereas ρ is the absolute mean value of chosen BCV, r represents the possible values it can take with some standard deviations from ρ

probability density function (pdf) in terms of the detection statistics r is given as:

$$f(r) = \frac{1}{\sigma_r\sqrt{2\pi}} e^{-\frac{1}{2}((r-\mu_r)/\sigma_r)^2}. \tag{3.28}$$

But it should be noted that the idea of BER is different for tamper detection mechanism. In fact it is opposite to that of robust watermarking systems. Thus we have a different definition of BER for our purpose:

Definition 3.1. *BER is the ratio of the number of bits **NOT** accurately detected within the bounds of allowable correlation value, $\pm\rho \pm \epsilon$ to the total number of bits embedded in the entire image.*

Hence, we first compute the **probability of accuracy**, p_c and find the complement in order to determine error probability, p_e.

The nature of (3.24) enables us to apply the probability expression for the p.d.f of Gaussian distribution [3] to solve this problem and thus estimate the probability of tampering as well as the trade-offs between fragility and robustness of the C_4S algorithm. The probability that r is within the domain **[a, b]** is therefore, given as:

$$p_c = Pr(a < r < b) = \frac{1}{\sigma_r\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}(\frac{r-\mu_r}{\sigma_r})^2} dr. \tag{3.29}$$

Unfortunately, (3.29), which is the cumulative distribution function (CDF) of Gaussian distribution, cannot be solved analytically for any arbitrary values **[a, b]** . However, when it is in standard normal form, we can use the direct statistical relationships among *Q-function*, *Complementary error function – erfc(.)* and *(Standard) Gaussian functions*, to drive further mathematical analysis. Hence, (3.29) is equivalent to the expression:

$$p_c = \frac{1}{\sigma_r\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{1}{2}(\frac{r-\mu_r}{\sigma_r})^2} dr - \frac{1}{\sigma_r\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}(\frac{r-\mu_r}{\sigma_r})^2} dr. \tag{3.30}$$

[3] K. Gan "Gaussian Probability Distribution" [online] https://www.physics.ohio-state.edu/ gan/teaching/spring04/Chapter3.pdf retrieved 23rd November, 2018.

To convert to standard normal($\sigma_r = 1$), we perform a variable transformation of the form:

$$z = \frac{r - \mu_r}{\sigma_r}. \tag{3.31}$$

Hence, we can re-write (3.30) as:

$$p_c = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-\frac{1}{2}z^2} dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}z^2} dz. \tag{3.32}$$

Given that:

$$Pr(r \leq a) = 1 - Q(\frac{a - \mu_r}{\sigma_r}) = F_r(a), \tag{3.33}$$

where:

$$Q(a) = \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{1}{2}z^2} dz = \frac{1}{2} erfc(\frac{a}{\sqrt{2}}). \tag{3.34}$$

and $F_r(a)$ is the CDF of the normally distributed r evaluated at a.

With the identities now defined in (3.30), (3.33) and (3.34), one can easily evaluate equation (3.29) in terms of the popular complementary error function as:

$$p_c = \frac{1}{2} \left[erfc\left(\frac{a - \mu_r}{\sigma_r \sqrt{2}}\right) - erfc\left(\frac{b - \mu_r}{\sigma_r \sqrt{2}}\right) \right]. \tag{3.35}$$

The actual error probability p_e is the complement of (3.35), which is:

$$p_e = 1 - \frac{1}{2} \left[erfc\left(\frac{a - \mu_r}{\sigma_r \sqrt{2}}\right) - erfc\left(\frac{b - \mu_r}{\sigma_r \sqrt{2}}\right) \right]. \tag{3.36}$$

However, in analogy to Malvar [102], $\sigma_r^2 = (\sigma_n^2 + \sigma_x^2)/(N\sigma_w^2)$ and $\mu_r = \rho \mathbf{s}$. Therefore:

$$p_e = 1 - \frac{1}{2} \left[erfc\left(\frac{a - \rho \mathbf{s}}{\sqrt{2\frac{(\sigma_n^2 + \sigma_x^2)}{N\sigma_w^2}}}\right) - erfc(\frac{b - \rho \mathbf{s}}{\sqrt{2\frac{(\sigma_n^2 + \sigma_x^2)}{N\sigma_w^2}}}) \right]. \tag{3.37}$$

Equation 3.36 shows that the $SNR \leftarrow (\frac{r - \mu_r}{\sigma_r})$ of the detection statistics is now being controlled by the detection tolerance values a and b. Also, for the purpose of symmetry and scalability (as will be seen in Chapter 4) we have assumed the conditions stated in (3.38). For embedding a 0-bit in a sub-block, $s = -1$, while embedding a 1-bit makes

$s = 1$.

$$b = 2r - a \tag{3.38a}$$

$$|b - r| = |r - a| = \epsilon \tag{3.38b}$$

$$\epsilon \geq 0 \tag{3.38c}$$

Bearing the above conditions in mind, remembering that $r = \rho$ when performing empirical computations and also referring back to Inequality 3.23, we can arrive at the error probability definition:

$$p_{e0} = p_{e1} = 1 - \frac{1}{2}\left[erfc\left(\frac{a-\rho}{\sigma_r\sqrt{2}}\right) - erfc\left(\frac{\rho-a}{\sigma_r\sqrt{2}}\right)\right], \tag{3.39}$$

where $p_{e0} = Pr(-b \leq r \leq -a|s = -1)$ and $p_{e1} = Pr(a \leq r \leq b|s = 1)$. Equation (3.39) indicates that the error probability is the same whether a 0-bit or 1-bit is embedded into a sub-block.

As an attacker is expected to introduce tampering that cause overall constant deviation, σ_r from ρ, we plot the Log10 of the error probability as a function of allowable lower bound, a, from the chosen constant correlation, ρ. This is shown in Figure 3.7 for σ_r =0.05,0.5,1.0 and 5.0.

Provided that there is at least one form of attack introduced after the insertion of watermark such that $\sigma_r > 0$, ***stringent tamper detection*** by our algorithm stipulates that if no error tolerance is allowed, then $\epsilon = 0$ and then $a = r = \rho$. The error probability becomes 1 and this is proven by (3.39) and shown by Figure 3.7 as Log10(1) is equal to zero:

$$p_{e01} = 1 - \frac{1}{2}\left[erfc(0) - erfc(0)\right] = 1 - 0 = 1.$$

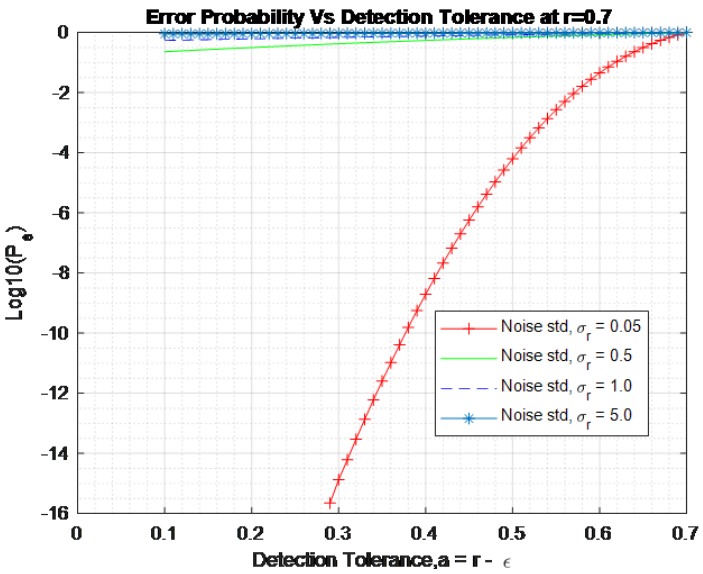


Figure 3.7: Error probability as a function of detection tolerance, a: *with r and attack power (σ) fixed, the error probability decreases towards as the error tolerance ϵ* or detection tolerance, *a* increases.

Equation (3.39) also shows clearly that if there is no tamper detection of any kind, then the error probability would be zero. As indicated by Inequality (3.23), $a \leq r$; so, if $\sigma_r = 0$, (3.39) evaluates to:

$$p_{e01} = 1 - \frac{1}{2}\left[erfc(-\infty) - erfc(\infty)\right] = 1 - 1/2[2 - 0] = 0.$$

This theoretical result is indicated as well by Figure 3.7. The plot for $\sigma_r = 0$, cannot be displayed on the graph because $log_{10}(0) = \infty$.

Using the plot for standard normal distribution in Figure 3.8, one can learn how to fix the lower bound on the detection statistics and thus choose ϵ.

From Figure 3.8, it can be inferred that:

$$Pr(\mu_r - \sigma_r < r < \mu_r + \sigma_r) = Pr(a < r < b) = \frac{1}{\sigma_r\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}\left(\frac{r-\mu_r}{\sigma_r}\right)^2} dr = 0.682 \quad (3.40)$$

[4]source: https://kanbanize.com/blog/normal-gaussian-distribution-over-cycle-time/

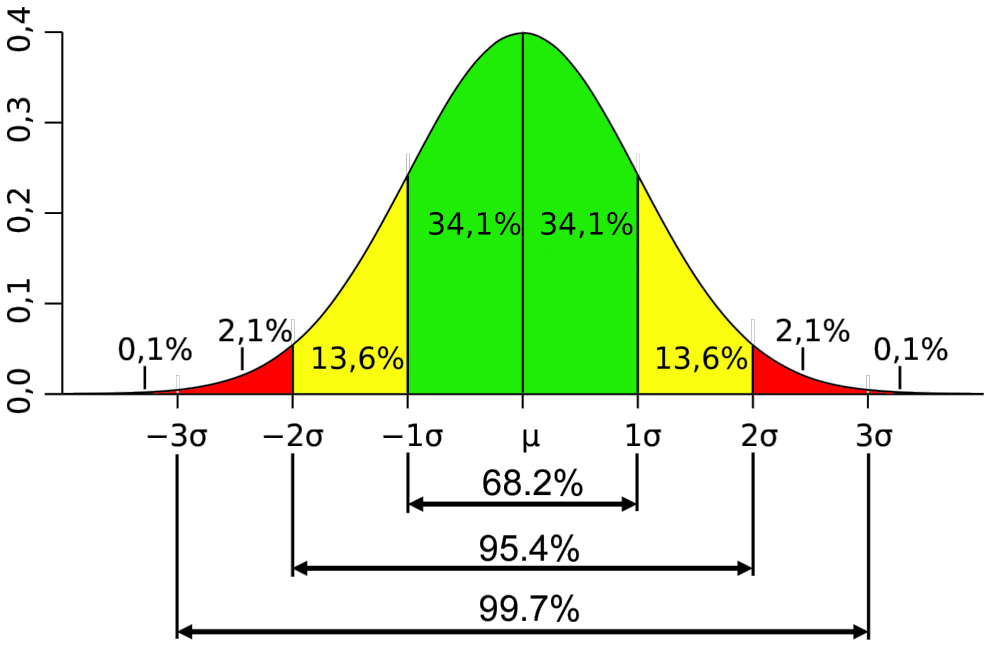


Figure 3.8: Gaussian Cummulative Probability Distribution[4]: *Given that the computed correlation at the receiver has normal distribution, we can set* $\epsilon = n\sigma$ *where* $n = 1, 2, 3$ *in order to control the robustness of the* C_4S *technique.*

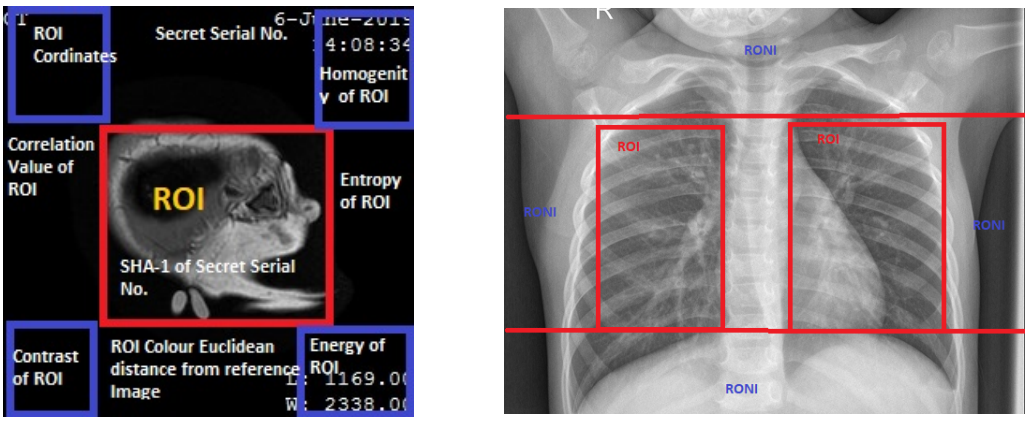
From general statistical point of view of a normal distribution, 68.2% of all the values fall within *one standard deviation* of the mean, 95.4% are within *two standard deviations* from the mean while 99.7% can be located within *three standard deviations* from the mean. This principle was used to inform the constraints assumed in (3.38). One could decide to increase or decrease the detection probability based on the choice of the upper and lower bounds, $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively.

In the next section, we report the results of the empirical experiments to evaluate the C_4S method for tamper and watermark detection.

3.6 Experimental Evaluation

We divided an image into regions of interest (ROI) and regions of non-interest (RONI). We are interested in detecting tampering in ROI. Thus the semi-fragile C_4S is used in ROI while the robust version or any other robust watermarking algorithm is used in the RONI. The ROI image segmentation for different datasets, modalities, and body parts are shown in Figure 3.9.

After the addition of security watermark in ROI, the image features defined in Sections 3.6.1 to 3.6.5 is computed from the ROI and embedded in the RONI as shown in Figure 3.9a. They are computed from the ROI GLCM defined in (3.41).



(a) ADNI MRI (b) Chest X-ray

Figure 3.9: ROI and RONI Selection: *C_4S was used for ROI while robust DWT Watermarking was used in RONI. The result presented focuses on the ROI as the DWT method applied in RONI is not new*

GLCM provides information about the positions of pixel pairs having graylevel values, (i, j) at a distance, d measured in one of the directions, $\theta = 0^o, 45^o, 90^o, 135^o$ about the reference pixel. This is depicted in Figure 3.10 with $d = 1$.

$$GLCM = G_{d,\theta}[i, j] = C_{i,j} \tag{3.41}$$

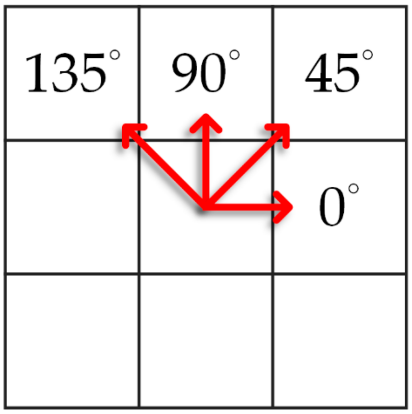


Figure 3.10: GLCM Computation: *GLCM provides information about the positions of pixel pairs having graylevel values,* (i, j) *at a distance,* d *measured in one of the directions,* $\theta = 0^o, 45^o, 90^o, 135^o$ *about the reference pixel.*

3.6.1 ROI Contrast, C_t:

Contrast is a measure of the local variations present in an image.

$$C_t = \sum_i \sum_j (i-j)^2 G_{d,\theta}[i,j] \tag{3.42}$$

3.6.2 ROI Correlation, C_r:

This is a measure of image linearity. A linear surface always has high correlation.

$$C_r = \frac{\sum_i \sum_j \left[G_{d,\theta}[i,j] - \mu_i \mu_j\right]}{\sigma_i \sigma_j} \tag{3.43}$$

where:

$$\mu_i = \sum_i i G_{d,\theta}[i,j] \tag{3.44}$$

$$\sigma_i^2 = \sum_i i^2 G_{d,\theta}[i,j] - \mu_i^2 \tag{3.45}$$

3.6.3 ROI energy, C_e:

Energy is computed from the Angular Second Moment (ASM)

$$C_e = \sqrt{ASM} \tag{3.46}$$

where:

$$ASM = \sum_i \sum_j G_{d,\theta}[i,j]^2 \tag{3.47}$$

3.6.4 ROI Homogeneity, C_h:

Homogeneity refers to how continuous a pixel value spans along a given direction without meeting corners or edges. A homogeneous surface has all the pixels having the same gray value.

$$C_h = \sum_i \sum_j \frac{G_{d,\theta}[i,j]}{1+|i-j|}. \tag{3.48}$$

3.6.5 ROI Entropy, H_X:

This was not computed from the GLCM matrix but from the Claude Shannon's [140] definition of entropy. Given that p(.) is the probability density function (pdf) of the pixels in the image, X, the entropy of X is:

$$H_X = -\sum_i p(i) \log p(i). \tag{3.49}$$

This function is already implemented in MATLAB as *entropy ()* function.

3.6.6 Dataset Description

The choice of the dataset was such that both *In vivo* and *Ex vivo* medical images are represented as well as images of low and high pixel depths. *Ex vivo* imaging is used to

diagnose, monitor and treat infections outside the human body while *In vivo* is for internal body organs. We used sample images of 8-bit to 16-bit JPEG and DICOM images. They came from the specific datasets described below. The ethics approval for the use of these images is included in Appendix A.

- Alzheimer's Disease NeuroImaging (ADNI [5]) Dataset. This is from the Laboratory of NeuroImaging (LONI) hosted by the University of Southern California. It is a neuroimaging dataset provided in .nifti. However, we converted it to DICOM and jpeg and worked on the slices making up a volume rather than raw data.

- OSIRIX MRI [6]: The datasets with the following Alias names were used: BRAINIX(Brain tumor), KNEE (Knee MR), KNIX (Standard knee MRI), MRIX (Standard Thoracic & Lumbar MRI) and Infarctus (Cardiac stress MRI). The importance of this dataset is that it is a 12-bit to 16-bit DICOM standard image. Other datasets were 8-bit grayscale or coloured medical images in jpeg format.

- ISIC [36, 151] Melanoma: The dataset was released by International Skin Imaging Collaboration (ISIC) [7] in 2018 for a challenge targeted towards automatic detection of melanoma skin cancer using dermoscopic images.

- **Chest Xray**[8]: For *In vivo* image validation, the Pneumonia Chest-Xray dataset for the Kaggle Competition was used. It is made of over 3,000 (large) sized X-ray images of patients that are either normal or have been diagnosed with Pneumonia diseases. The same dataset was used in [8] for deep learning image classification of bacteria and viral pneumonia. Some images are so large that they produced ROI of size 1285 x 1241. Each ROI is a little less than half the size of the entire

[5]*Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

[6]https://www.osirix-viewer.com/resources/dicom-image-library/

[7]https://challenge2018.isic-archive.com/

[8]https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia/downloads/chest-xray-pneumonia.zip/2

medical image. This variation in size has led to the normalisation of results obtained for capacity in this research. This dataset will be re-used in Section 5.4 to evaluate the effect of hidden data in medical image autodiagnosis using statistics and Machine learning methods.

It should be noted that all datasets have been properly de-identified before being released by the contributors. These datasets will also be used in Chapter 4 for experimental evaluation.

3.6.7 Experimental Procedure

The algorithms described in Algorithms 1 and 2 were implemented in MATLAB 2017 using the ISIC and the Pneumonia datasets. The robustness of the algorithms were tested by applying attacks both at the local level (8x8 blocks) and the global level (on the entire ROI and the whole image). The first set of experiments were performed by running the algorithms on the images without applying any form of attack before watermark extraction. This enabled us to establish the baseline performance of the algorithm on each of the datasets in terms of BER, embedding capacity and distortion.

After establishing the baseline performance of the algorithms, different types of attacks were applied at various intensities at both the local and global levels. Detailed experiments were performed of different noise powers applied to the images. This helped to study the false positive and false negative performance of the algorithms. The results obtained were averaged over the images that belong to the same dataset. Because of the wide variation in the texture of each dataset, the result for each dataset was reported differently. The MATLAB files that relate to the experiments in this chapter (and Chapter 4) can be found at https://github.com/KingPeter2014/MediHide.

The parameters and results for each attack are shown in Tables 3.3, 3.4 and 3.5. Further description of how specific experiments were carried out is included in the respective subsection of the results presented below in order provide better context for each feature tested on the algorithm. The goal of the experiments is to show to what

extent the algorithms detects other attacks but are not seen as attacks on the image. This means benchmarked on the 40dB distortion threshold and BER of zero.

3.7 Experimental Results

This section presents the empirical results from the evaluation of the C_4S algorithm for tamper detection. First, we present the Steganographic qualities that relate to image fidelity (distortion level). After that, the attack detection capability is presented. The results in this chapter were obtained from the ROI areas as they are the major regions that require integrity protection and verification.

3.7.1 Steganographic Performance

The Steganographic parameters obtained after information hiding in the ROI for different datasets are shown in Table 3.3. The objective is to detect tampering while achieving both low distortion and reasonable information hiding capacity.

As seen in Table 3.3, all the datasets achieved more than 50dB of PSNR, which is well above the 40dB benchmark. This result was due to the adaptive embedding method illustrated in Figure 3.5. The distortion level is minimised by embedding in the CBZ instead of the BZ, whenever distortion in the BZ is considered to be higher. This control embedding method is important because we are dealing with the ROI of the medical images. Furthermore, the SSIM column results show that there is at least 99.63% structural (plus luminance and contrast) similarity between the original and watermarked image. These high figures of merits (PSNR and SSIM) show high image fidelity after adding the tamper detection watermark in the ROI.

To understand the achieved capacity and the tamper detection capability, if the ROI of a melanoma photograph is 512x512, then there are 4096 sub-blocks of size 8x8. Hence, $B = 4096$. Therefore, the steganographic capacity is $0.721 * 4096 = 2953$ bits.

For the OSIRIX MRI DICOM dataset with 16-bit pixel depth, the average PSNR is

Table 3.3: Average ROI Steganographic Parameters. **B** in the *Capacity* column is the bandwidth of the image, which is the number of 8x8 blocks. The fraction **B** that is watermarked with actual EMR data constitutes the capacity of the image.

DataSet	Avg. PSNR (dB)	Avg. SSIM	BER	Capacity (%B)
ADNI	53.42	0.9963	0.00	48.62
X-ray(Normal)	67.05	0.9999	0.00	87.54
X-ray(Pneumonia)	65.54	0.9998	1.05×10^{-5}	89.45
ISIC	52.03	0.9959	0.00	72.10
OSIRIX[MRI 16-bit]	101.59	1.0000	0.00	52.35

100.89dB with an SSIM of 1.0. The distribution of the individual image PSNR is shown in Figure 3.11.

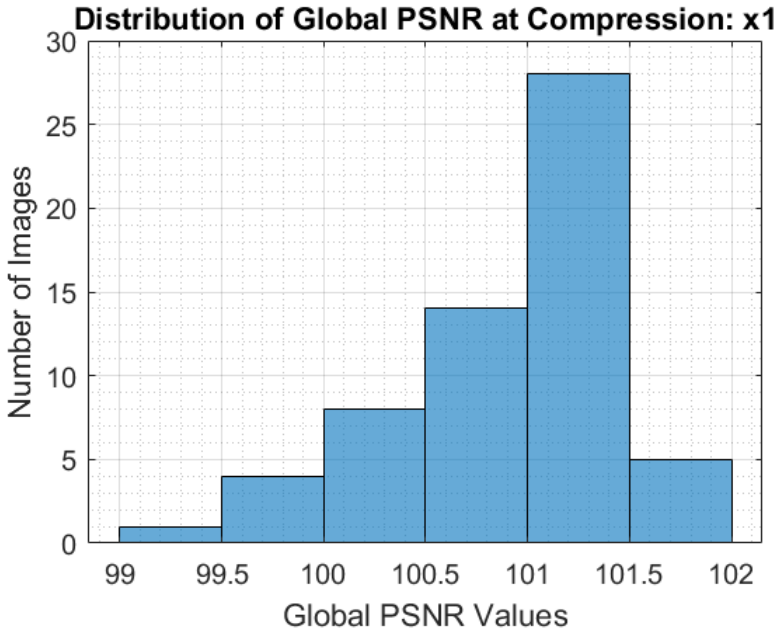


Figure 3.11: PSNR for 16-bit OSIRIX MRI Dataset: *All test images had PSNR above 99.00dB with an average of 100.89dB*

The result in Figure 3.11 shows that none of the MRI images had a PSNR less than 99.00dB. Also, the average value of 100.89dB is higher than the most recent work [131] in the area of medical image steganography at the time of writing this thesis. This result is obtained at comparable capacity (bits per sample).

3.7.2 Tamper Detection Capability

Various signal processing attacks, which may result in image distortion and possible loss of diagnostic information, were executed at the block-level and on the whole image. The results of the detection of the possible tampering by these attacks are presented here.

Figure 3.12 shows how the extracted correlations, $\langle Y, W \rangle$ clearly centres at -0.7 and 0.7 ($= \pm\rho$) for extracted 0 and 1 bits respectively. A significant deviation(more than ϵ) from these two values resulted from intolerable quantisation errors, heavy post-watermarking processing, and intentional attacks. This result shows that the C_4S algorithm correctly detects legitimately inserted watermarks and undue tampering on the image pixels.

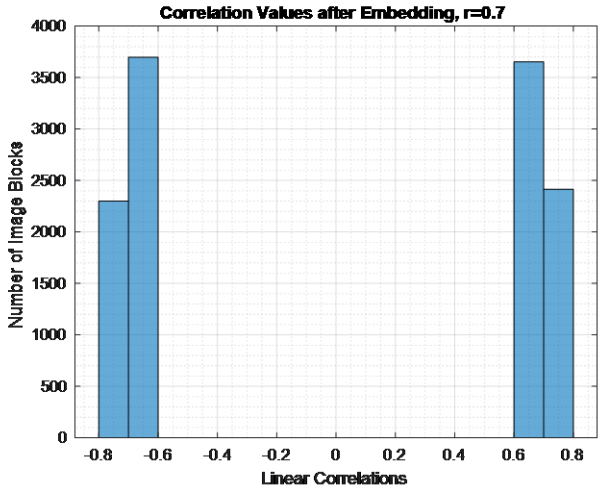


Figure 3.12: Effectiveness of Constant Correlation SS Method: **The extracted correlation values were within the expected value of** ± 0.7

To validate the above claim, we applied malicious and image processing attacks on both the ROI and RONI of the medical image. The result is tabulated in Table 3.4. This experiment was performed on the ISIC Melanoma dataset.

When there is no attack, there is no tampering detected in the ROI. However, with an attack, there are different levels of impact on the ROI of the medical image. The ROI

Table 3.4: ISIC ROI Tamper Detection and RONI robustness to attacks. RTD = ROI Tampering Detected (Figure 3.13), S & P = Salt and Pepper, HE = Histogram Equalisation

Attack	RONI BER	ROI BER	PSNR (dB)	RTD
No attack	0.00	0.00	52.03	0
S & P, 0.005	0.00	0.27	26.79	148
S & P, 0.05	0.13	0.76	17.04	600
Gaussian, 0.05	0.23	0.82	14.85	710
HE	0.00	0.19	12.64	184
Speckle 0.05	0.35	0.43	15.92	269
Poisson	0.00	0.25	24.67	143
JPEG, 50%	0.00	0.42	29.94	37
JPEG, 75%	0.00	0.31	34.26	20
JPEG, 90%	0.00	0.23	34.15	17
Contrast adj.	0.00	0.32	18.91	296

tampering detected (RTD) column in Table 3.4 shows this by having various values for several blocks detected. Figure 3.13 shows the sub-blocks that were detected as tampered when a Gaussian noise attack is applied to a sample image.

In order to account for why some blocks were not marked as tampered, and also to show the robust aspect of the C_4S method, we performed a further in-depth experiment. We decided to perform the Gaussian noise attack on each 8x8 block of an image and with increasing noise variance. Table 3.5 is a summary of the result obtained from a **test image** [9], from the Pneumonia data set. This test image is an ROI of size 1568 x 930, which gave us a total of 22, 785 sub-blocks of size 8x8. The steganographer, as well as the attacker, is constrained by the 40dB quality benchmark. An attack is considered ***truly effective*** if the PSNR of a sub-block is above 40dB, but the watermark could no longer be retrieved correctly in the sub-block; otherwise, it is ineffective. An effective attack leads to false-negative tamper and watermark detection, and that makes the attacker win in the sub-block. Hence, the false negative, FN, is defined (following

[9] https://github.com/KingPeter2014/MediHide/blob/master/IM-0115-0001.jpeg : This was chosen at random for detailed analysis

Figure 3.13: Attacked ROI Detected: *The sub-blocks with severe attacks has been localised and marked with white boxes*

Table 3.5) as:

$$FN = Undetected - Ineffective \tag{3.50}$$

For the robust aspect of the C_4S, we use the concept of ***parameter tuning***. Specifically, we modify the error tolerance, ϵ and repeat watermark decoding on the sub-blocks where effective tampering and wrong watermark bit was decoded. We modified the semi-fragile watermark extraction equation of (3.19) into a more robust extraction equation given by (3.51):

$$\bar{s} = \begin{cases} 1, & \text{if } r = \langle Z, W \rangle = 0.001 \leq \rho \leq 3\rho \\ 0, & \text{if } r = \langle Z, W \rangle = 3\rho \leq \rho \leq -0.001 \end{cases} \tag{3.51}$$

With (3.51), the BER reported in the upper half of Table 3.5 went down to zero. For the lower half of the table there was over 60% reduction but it did not go down to zero. Further increase in the value of error tolerance did not improve BER or FN further. This suggests that the remaining errors from the attack is due to bit flipping caused by change of correlation values from positive to negative and vice versa.

To detect the bit-flipping type of attack, the bits embedded in the ROI should be derived from the cryptographic hash of a known secret word. After the extraction of the embedded bits, a comparison was made between the cryptographic hash of the known

Table 3.5: Pneumonia Normal ROI Gaussian noise attack at different noise variance: *RTD = fraction of block ROI Tampering Detected, Undetected = fraction of ROI block not flagged as tampered, and Ineffective = truly ineffective attacks because watermark was retrieved and the the sub-block PSNR is greater than 40dB*

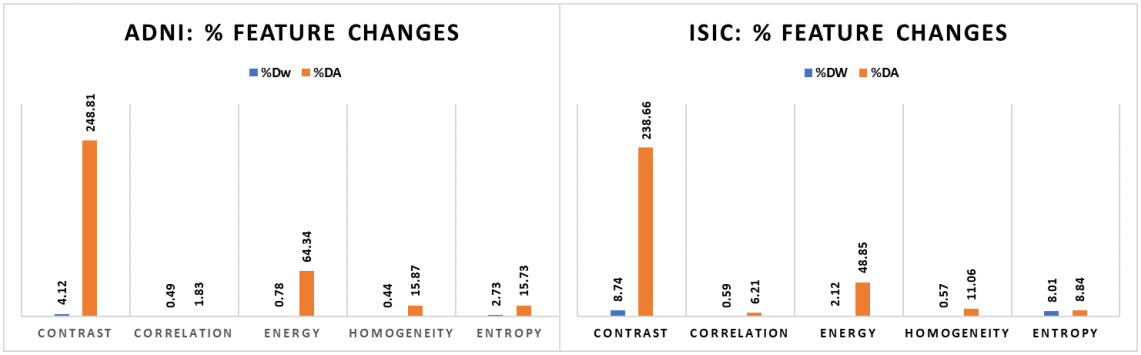
Noise Power, σ	PSNR	RTD	Undetected	BER	Ineffective	FN
0.00000	48.13	0.00	Inf	0	Inf	-
0.00001	47.27	0.08	0.92	0.04	0.92	0.00
0.00010	40.03	0.20	0.80	0.13	0.74	0.05
0.00100	30.02	0.44	0.56	0.31	0.37	0.19
0.01000	20.10	0.76	0.24	0.44	0.12	0.11
0.10000	11.28	0.91	0.09	0.48	0.05	0.05

secret word and extracted bits. This approach validates the result of this algorithm if such an attack is launched against our system.

Effect of Attacks on Textural Features

For a better visual understanding, Figure 3.14a is a plot of the percentage changes in mean and standard deviation (S.D) of the textural features ADNI dataset after watermark addition and after an attack with **contrast adjustment**. Before the attack, Energy is the least affected, while entropy is the most affected. However, after the attack, all features are significantly affected with Contrast changed by approximately 250% from its original value before watermarking.

Similarly, Figure 3.14b has been plotted for ISIC dataset. Similar behaviour to ADNI can be observed. Homogeneity was less affected than Energy before the attack. Entropy was the most affected and even higher than that of ADNI. Nevertheless, after the attack, all features were significantly affected, especially Energy and Contrast.



(a) ADNI ROI (b) ISIC ROI

Figure 3.14: The Effect of (Constrast adjustment) Attack on Textural Features: *% ROI mean feature changes for watermarked and attacked images. $\%D_w$ is change due to watermark while $\%D_A$ is change due to attack*

Table 3.6: Comparison with Existing Techniques. CT= Computerised tomography, OCT = Optical Coherence Tomography, MRI = Magnetic Resonance Imaging

Ref.	Purpose	Dataset	PSNR(dB)	SSIM
[49]	Tampering	CT, MRI, PET	49.41	0.9776
[107]	Privacy	Xray, CT, MRI	45.00	1.0000
[117]	Robust S&P	MRI	41.73	0.9127
[4]	Security	OCT, Fundus	51.95	0.9981
[66]	Autodiagnosis	OCT	65.00	1.0000
C_4S[8-bit]	**Privacy/Tampering**	**X-ray, MRI**	**67.05**	**1**
C_4S[16-bit]	**Privacy/Tampering**	**MRI**	**100.89**	**1**

3.8 Key Findings

The following can be deduced from the design, analysis, theoretical and empirical results presented in this chapter:

1. by pre-determining the correlation at which a watermark bit is embedded in each sub-block, the tampering in that particular sub-block can be determined once there is a significant shift(greater than a defined value, ϵ) from the pre-determined correlation, ρ, at the receiver (See Figure 3.12). This is because such change can only occur if the sub-block data was altered, provided that the same spreading

sequence used at the source for embedding was the same as the one used for extraction at the receiver.

2. To achieve zero BER for watermark detection, more precision on the detection statistic, r, which is now centred on ρ, is required, unlike traditional SS. This feature enabled us to reduce both false positive and false negative to zero provided there is no attack (See Table 3.4).

3. The new algorithm is now more scalable and easily adaptable. By simply tuning the parameter ϵ, one can move from fragile to semi-fragile to robust watermarking while using the same embedding and extraction functions.

4. the new C_4S algorithm can detect the most relevant (95% with an average of 5% False Negative for effective attacks as shown by Table 3.5) attacks on medical images. It was found that the most severe attacks on medical images based on image distortion models are Gaussian noise, Salt & Pepper attack, Contrast adjustment, and Speckle noise in that order (See Table 3.4).

5. With controlled embedding techniques, a high PSNR value of 67.05dB was achieved for 8-bit medical images and 100.89dB for 16-bit medical images while detecting distortion-based medical image attacks (See Tables 3.3 and 3.6). These results are important for ensuring less distortion and hence, less effect on diagnostic features, as will be seen in Chapter 5.

Images with high pixel depth (number of bits per pixel higher than eight) provide more Least Significant Bits (LSBs) that can be manipulated without degrading the image considerably. This is one of the features of DICOM formats for transmitting DICOM-compliant images in Teleradiology.

3.9 Discussion

We established in Section 1.2 that one of the security challenges in teleradiology is verifying the integrity of the medical image and the authenticity of the included patient data after transmission or storage. It was also agreed that spread spectrum methodology had been known mainly for its robustness and resilience in providing confidentiality and availability - two of the three major security primitives - but not in the third primitive, which is integrity. This gap in spread spectrum steganography formed the first research questions in this thesis:

How can SS Steganography offer a solution that combines Medical Image Integrity verification and zero-error watermark detection for text embedding while preserving its robust characteristics for authentication?

The first finding shows that the spread spectrum technique can now be used as an all-in-one security solution that can provide confidentiality through secret watermark embedding of patient records using spreading sequences, availability through jam resistance and now **integrity** through this new constant correlation scheme. The advantages of having a single method that provides all the security requirements are less implementation complexity, ease of maintenance and less fragile interfaces between disparate technologies with the teleradiology system. This simple but fit-for-purpose approach is required for resource-constrained scenarios such as rural settlements and developing countries.

The second research finding above indicates that we can achieve zero BER with the C_4S method. The two major errors that emanate from the noisy nature of an image during watermark detection are False Positive and False-negative. With these two errors eliminated for all sub-blocks in the image, we can determine which block has a watermark or not. The implication of this is that barring any form of attack, text-based EMR can be embedded and accurately extracted.

Theoretical (Section 3.5.3.2) and empirical analysis of the detection scheme prove that the scheme can be dynamically configured as either a fragile, semi-fragile or ro-

bust watermarking scheme. For more robust extraction (required for stronger attack vectors), one simply needs higher tamper tolerance, which is achieved by increasing ϵ (See Figure 3.7). This means that one can make two passes at the decoder: first for tamper detection and second for robust watermark extraction, after increasing, ϵ. This achievement is significant as both the integrity of the image and authenticity of the patient's record and the embedded security sequence can be verified in simple but related computations. The overall outcome is accurate and secure teleradiology and diagnostic service.

However, it is acknowledged that our method does not detect modification at the pixel level and that up 5% False negative may be experienced for some attacks.

The above 5% tamper detection error and other special attacks such as blockwise copy-replace and shifting of correlation from negative to positive and vice versa are not detected as distortion attacks within a sub-block, and they can happen in traditional Spread spectrum IH as well. To solve these problems, We have combined the patient record, hospital logo and their hashes as watermarks. Whenever any of the three attacks above and perturbations occur during storage or transmission, the recomputed hash of the retrieved information and extracted hash will no longer be the same. We then use a distance measure of the difference in position between the original combined watermark w and the tampered combined watermark, $\overline{w}$ to localise the tampered block. This approach implies that advanced geometric and algorithm-specific (white-box) attacks can be detected and localised, thereby improving the safety of medical image scans via linear correlation.

In addition to the above solution, we have used the **feature changes within the ROI** to detect special attacks. The feature values were computed and stored in RONI. If blocks are moved to distances significantly far from there original location and to a region with a different texture, the overall texture parameters of the ROI will begin to change. The use of these features has enabled us to establish the fact that both malicious and non-malicious attacks originating from signal processing can also be detected by the active embedding of these features, though in the RONI of the image.

We compared the results of our algorithm with similar active tamper detection schemes (See Table 3.6) in the area of medical images watermarking, such as [49, 107, 117, 4] and [66]. Our algorithm has better quality in terms of imperceptibility. Our improvement for 8-bit medical images ranged from 3.15% (for [66]) to 60.68% (for [117]). In terms of tamper detection, the signal processing attacks that conform to distortion changes and relevant to medical images were detected and localised (See Figure 3.13).

The overall implication of our algorithm is flexibility and completeness in design for SS Steganography and digital watermarking. As Nyeem [127] noted, the degree of tamper tolerance is application-dependent. Also, Ludewig *et al* [94] posited that the final classification of the usability of an image requires a grading system. Hence, mapping distortion parameters to a doctor's usability decision is a research gap that requires further consideration.

In summary, integrity was achieved in this research for spread spectrum steganography. The use of cryptographic hash and feature changes helped to detect and localise non-distortion but relevant attacks. Watermarks can still be detected even if there is tampering by increasing detection tolerance, ϵ. Hence, the first research question in Section 1.5 is deemed answered in the affirmative.

3.10 Summary

We can see both mathematically and empirically that ***increase in detection probability (and reduction in error probability) increases robustness but decreases tamper detection probability***. This flexibility is the purpose of designing the C_4S algorithm - to perform both semi-fragile and robust Steganography. This was achieved by varying the tolerance parameter, ϵ.

Existing spread-spectrum decoders can still work with our encoder. The legacy spread-spectrum decoders are just the robust version of C_4S - that is, they detect watermark along the entire negative($-\infty$ to 0) or positive space (0 to ∞) of the number line instead of at a specific spectrum ($-x_1$ to $-x_2$ or x_1 to x_2) on the number line.

Our integrity verification scheme's technical approach is to design fragility into the robust nature of spread spectrum technology by dynamically determining a narrow range of correlation value for retrieving a zero or one bit. At the receiver, the blocks whose correlation is outside this narrow range are deemed tampered. This is the constant correlation method.

Spread Spectrum information hiding has great potential for providing higher security and resilience against active attacks; it suffers a low embedding capacity. For example, we can only embed one bit per sub-block. This limits its application to security systems where few amounts of data are required to be hidden, such as Cox *et al* [40] signature application. Medical Image security and Electronic Medical Record (EMR) security seems not to be such an application because more data from patients and radiology reports may need to be embedded.

Chapter 4 will extend the developed theory in this chapter to seek for a new upper bound in Steganographic capacity of SS watermarking. We will discuss how the low distortion required for the ROI tamper detection impacts capacity improvement and the required trade-offs. The evaluation of what our active tamper detection method may mean for autodiagnosis will be explored in Chapter 5.

Chapter 4

Steganographic Capacity Improvement for SS Steganography

This chapter describes the contributions made towards estimating and improving the amount of data that can be embedded and retrieved from medical image scans while minimising image distortion. In Section 4.2, we review the current state of Spread Spectrum (SS) steganographic capacity and the parameters that can be used to measure capacity. After that, we consider the rate-distortion theory (Section 4.3) and the different methods (in Section 4.4) for ascertaining the Steganographic capacity of an information hiding (IH) channel. We then design a new algorithm that improves the SS Steganographic capacity while retaining the tamper detection capability and the C_4S concepts from the previous chapter. These new design concepts and associated algorithms are presented in Section 4.5. An analysis of the entire design follows immediately in Section 4.6. Experiments were designed in Section 4.7 to evaluate the effectiveness of the new method, and the results obtained are presented in Section 4.8. These results lead to the key findings and discussions presented in Sections 4.9 and 4.10, respectively. In the discussion, the second research question was answered as well as the implications of the findings in practice. This chapter concludes with a summary presented in Section 4.11.

Go To Chapter: 1, 2, 3, 5, 6, 7

4.1 Introduction

In Section 2.3.7, we stated that employing either spreading codes or error-correcting codes (ECC) reduces the number of information bits that can be embedded in the process of spread spectrum (SS) data hiding. This challenge of low capacity has made SS data hiding schemes unsuitable for data-intensive applications such as remote autodiagnosis in teleradiology. However, due to the positive properties (robustness and noise interference resistance) of SS techniques in attack-prone open communication networks, we cannot ignore its use.

Let us use a scenario to illustrate a typical emerging data-intensive scenario in teleradiology autodiagnosis involving no datastore apart from the image itself. Figure 4.1 is a high-level block diagram of a use-case scenario for multiple users in a medical care setting. Four data sources are considered: Patient's Electronic Medical Record (EMR), Radiologist's report, Doctor's report (including prescriptions), and Source authentication data. These sources data are to be multiplexed and embedded into a medical image scan, using some generated secret key and a defined embedding function.

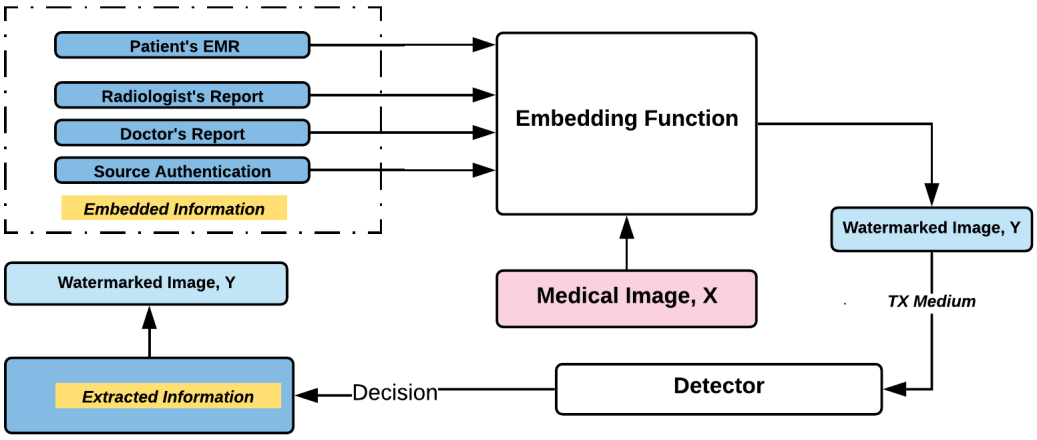


Figure 4.1: Multiplexed Information Hiding in Teleradiology: *Different medical personnels can encode different information using the specific keys assigned to them. This is a data-intensive scenario.*

The output of the embedder (which contains the embedding function) is called *watermarked* or *stego image*. The watermarked image is then transmitted over an open or

insecure network to a given destination. For an SS watermark[1] embedder, the embedding function could be any of classical SS, Improved Additive SS (IASS), Improved Multiplicative SS (IMSS), the Correlation-aware SS (CASS) or the suitable adaptation of any of them. These SS methods affect robustness, security, imperceptibility, and Steganographic capacity in various ways.

At the destination, a detector is needed for the retrieval of each of the data sources from the watermarked image. This retrieval process is said to be blind if neither the original message nor the original medical image is required to extract the watermark. A decision hypothesis to determine the extraction of a zero, a one, or no watermark is required for each of the data sources.

The setting described in the above scenario requires a large amount of data to be embedded at low distortion. Hence, to incorporate the attack-resistant nature of SS technique into network-based applications that require high-capacity data hiding, it is appropriate to find ways to improve the data hiding capacity of SS hiding methods.

Whether we are designing for single or multiple users, there are two major constraints to improving Steganographic capacity. These are ***distortion*** and ***bit error rate (BER)***. For this research, the 40dB of PSNR was used as the **maximum** allowable distortion in local sub-blocks while Zero BER is the maximum allowable error. The zero BER constraint is important because for watermarks generated from text-based messages such as EMR, any error in the ASCII code of a character is easily noticed.

Another challenge for practical application of steganography is the dynamic nature of image content and the volume of data that is required in each case. Although the cover image size could be the same, the image content (complexity) and the amount of data (measured in bits) that are required to be embedded always vary. For example, the size of medical history may vary from patient to patient, but the size of a medical scan from the same radiological machine may always be the same. This dynamic embedding requirement raises a question: *what quantity of data (in bits) could be embedded into a scan*

[1] Again, we use watermark to refer to the binary bits being encoded into the image. This naming is the same for steganography and digital watermarking in this thesis.

without impairing its diagnostic quality? In other words, How can one estimate and maximise this amount of data subject to the accuracy of medical diagnosis? By answering these questions, which originate from the second research question in Section 1.5, we can understand the true meaning of *Steganographic capacity* of medical images.

In order to answer the above questions, we have explored techniques for increasing capacity, especially for images with higher pixel depth, such as 16-bit DICOM (Digital Image Communication in Medicine) images. We extended the C_4S method developed in Chapter 3 for capacity improvement of SS steganography. Specifically, we combine multi-level signaling and channel coding techniques to develop a new method of increasing the data hiding capacity of medical images, especially those of high pixel depth. Further, we explored the volumetric nature of DICOM images, which produces many regions of non-interest across the slices that make up a DICOM volume. These techniques helped us to achieve high-capacity embedding. This increase in data hiding capacity is important for data-intensive applications in teleradiology. Before designing and evaluating the new capacity improvement technique, we first reviewed existing works in hiding capacity improvement and presented a general capacity estimation method for images.

The contribution of this chapter is a new algorithm that is compatible with the existing SS methods but extends its data capacity through a data compression coding technique. The designed method is also compatible with CDMA techniques so that the overall capacity of SS steganography is improved for autodiagnosis (diagnosis without human intervention) applications in teleradiology, especially in rural settings with limited security infrastructure and expert physicians.

4.2 Current State of SS Steganographic Capacity

Table 2.1 identified low data hiding capacity as the major shortcoming of SS steganographic techniques despite other good features highlighted in the last section. In this section, we review the approaches and efforts of various researchers to improve the

data hiding capacity of SS IH methods.

S. Ghoniemy *et al* [55] explored the adaptive embedding strength method and Bit Plane Complexity Segmentation (BPCS) in order to increase the data hiding capacity by 43% from the traditional SS method. Their method and argument show that the estimation of hiding capacity is dependent on *imperceptibility*, *robustness* and *Bit error rate (BER)*. Their method hid data in the Discrete Cosine Transform (DCT) coefficients of the cover image. However, the authors did not perform any mathematical modelling of the relationship among the parameters that determine the Steganographic capacity of an image.

The work of S.Yang *et al* in [161] utilised the multiple access capability of SS technology to improve hiding capacity. This technique is called Code Division Multiple Access (CDMA). CDMA allows different users to be assigned orthogonal codes to embed data. This orthogonality ensures that there is minimal interaction among the co-existing users' data. The researchers included some theoretical modelling for estimating image hiding capacity. They asserted that *Host Signal Interference (HSI)* is one of the major challenges to increasing steganographic capacity using this method. However, the assertion in [39] shows that this HSI problem could be eliminated for some images. It was not clear, though, what type of images could allow total elimination of HSI. What is clear is that, similar to the findings by Ghoniemy *et al* [55], the *accuracy* of retrieved watermark also contributes to the hiding capacity of a method. Hence, it is important to increase detection accuracy in order to improve hiding accuracy.

In what seemed like an attempt to solve the detection problem, R. Hasanah *et al* in [71] used Fast Fourier Transform (FFT) and CDMA to increase hiding capacity and reduce HSI respectively. The researchers took cognizance of the difference between HSI and Multiple Access Interference (MAI) of a CDMA technique. These problems are different and affect watermark retrieval in SS systems. Whereas orthogonal codes will reduce MAI, the researchers proposed a subspace projection method to eliminate HSI. They agreed with researchers in [55] that *image complexity* exploration could help determine the best place in an image to hide in order to maximise capacity. However,

they did not carry out their experiments on large data sets, and their modelling relates to embedding capacity [39] and not Steganographic capacity, which strongly considers distortion.

The researchers in Meerwald [105] and Vicky [152] were concerned with robust watermarking techniques. Whereas [152] applied SS in the DCT domain, [105] utilised both DCT and Discrete Wavelet Transform (DWT) domains. Their works show that steganographic capacity depends on the amount of watermark that could be correctly retrieved after a heavy attack on the watermarked image. Determining what 'heavy' attack means is not easy, and besides, not all cover images with an unreasonably heavy attack will still be useful. Hence, the view from [152] is rather too strong and not applicable to medical images. It was asserted [55, 152], however, that a kind of transform could be applied to coloured images to simultaneously increase hiding capacity, robustness, and imperceptibility.

In what we believe would drastically reduce the retrieved information capacity, Hassan [71] utilised parallel channels to transmit the same watermark. This method introduced the same watermark redundantly, thereby increasing the chance of watermark recovery after attacks. Such a robust system may require significant trade-off with both capacity and imperceptibility. In this work, we did not focus on extreme robustness provided mainly by digital watermarking but on the moderately robust systems provided by steganography.

Before concluding this review, let us consider the factors that are considered significant for capacity estimation from an information-theoretic perspective. These schools of thought are summarised in Table 4.1. From this table, it can be deduced that most researchers who utilise information theory to estimate steganographic capacity have a common belief that an active adversary is often present in a steganographic channel to introduce distortions of some sort.

From the works mentioned above from both theoretical and empirical perspectives, it can be inferred that any of the following constitutes an improvement in Steganographic capacity, especially for an SS steganographic channel:

Table 4.1: Important and Unimportant factors for determining Steganographic capacity under Information Theory: *Capacity is generally determined by the encoder, the channel and the decoder. BER = Bit Error Rate, SNIR = Signal-to-interference-plus-noise ratio*

Authors	Important Factors	Unimportant Factors
Barni [13]	watermark strength, α image features insertion strategy	decoder structure
Gkizeli[56]	SINR BER image distortion level, ϵ	-
Wang [156]	image distortion level, robustness	perfect security
Harmsen [65]	Channel noise, Active attack detection function	Cover signal, encoder distortion
Filler [52]	Fisher Information rate, root rate	embedding and extraction strategy
Zhang [165]	Error Bit Rate (EBR), Channel code	-
Jiang [80]	PN Sequence, insertion strategy	-

1. Increasing the amount of embedded and extracted watermark bits while keeping distortion acceptable for the application.

2. Decreasing the effect of inherent HSI and image noise to decrease the BER of the extracted bits.

3. Increasing the resistance from active attack and transmission noise to the watermarked image to ensure accurate watermark detection and decoding.

When a steganographic approach is adopted instead of watermarking, (1) and (2) above are the most applicable. They ensure that more bits of information are securely embedded into the cover, correctly extracted at the receiver, and any attack on the cover correctly detected at the receiver. In order words, integrity preservation of medical images through minimal distortion and accurate EMR retrieval is the goal of this thesis. As a consequence, we reiterate that the **level of distortion** and **BER** are the major parameters that determine the steganographic capacity of a medical image steganographic

(MIS) system. Robust extraction (3) is not a major focus in this case.

Also, we have also focused on the accuracy of watermark because this research employed text watermark and not image watermark. For text watermarks (such as EMR), less tolerance for detection error is necessary due to its effect on the intelligibility of the retrieved text. Hence, we target 100% accuracy in watermark detection, provided that the cover image is still eligible to be used for medical diagnosis. In the next section, we will discuss the rate-distortion theory - an aspect of information theory that is concerned with finding a balance between the amount of transmitted information and the ability to recover it at the receiver. Although it cannot be applied directly to steganography, we shall find an equivalent analogy for understanding cover distortion and steganographic capacity of an image.

4.3 The Rate-Distortion Theory

Rate-Distortion (RD) theory in information theory determines the minimum number of bits per symbol (Rate, R) that can be transmitted over a channel without exceeding a distortion (D) on the input signal so that it can be accurately reconstructed at the receiver [37]. This field analyses data compression from the standpoint of information theory. As with many computations based on information theory, the key parameter for determining distortion is the change in entropy. Compression or Source coding reduces the entropy of the encoded information [140].

A variant of source coding theorem [57, 140] (also known as **Shannon's first theorem** or **noiseless coding theorem**) establishes an error-free encoding. This theorem ensures that there is no distortion of the message by the noise in the channel. Similarly, in digital watermarking and steganography, error-free information encoding and extraction is necessary. This correctness requirement is why we seek for a Zero BER for watermark retrieval. A watermark insertion strategy that is resistant to image noise will ensure accurate watermark retrieval. This correct retrieval of watermark increases the Steganographic capacity of an algorithm or image.

Moving to a more focused analogy between R-D theory in compression and Capacity-distortion theory in data hiding, there is a different dimension to the problem of noiseless encoding, which exists in watermarking and steganography: the distortion of the (image) channel caused by the inserted watermark. As the image is considered the communication channel in steganography, the watermark's presence is a form of noise whose amplitude distorts the original content of the image. Therefore, even if the message is retrieved without errors, we are still concerned about the degradation caused by the image as a channel. As the amplitude and quantity of the embedded watermark increase, the distortion on the image channel also increases. As we considered medical image cover in this thesis, all capacity increase are subjected to an evaluation of image distortion based on relevant medical image biomarkers (See Section 5.2). However, our approach is to define capacity for the region of interest (ROI) and then for the region of non-interest (RONI). In the RONI, there is far higher tolerance on distortion, and thus more capacity could be achieved. With this approach, some ***rate-distortion (RD) optimisation*** heuristic will be employed to achieve the best results for a given image, especially in the ROI.

In summary, the RD theory in compression is different from RD theory in Information hiding (IH) in the following ways:

1. In IH, we seek to maximise **R** for a given **D**, unlike in compression. In IH, we need to transmit more message bits at the same distortion to the image (channel), while compression seeks to transmit a lower number of message bits for the same distortion to the message.

2. Source coding (compression of a message) occurs before the embedding of the message into an image, which could be again compressed further if necessary.

3. For practical purposes in IH, both **R** and **D** cannot be determined by entropy alone as the distortion on the channel is being considered, not the distortion of the message.

The above differences notwithstanding, the common ground between compression

and IH is that a distortion function is needed to preserve the integrity and intelligibility of both the message and its carrier (image in our case).

In practice, the measurement of rate-distortion performance of an IH algorithm is usually a function of embedding capacity versus PSNR, SSIM, or MSE. These parameters are interpreted as regards other variables used to control capacity and/or robustness [154]. We apply this method of interpretation to present the capacity-distortion (rate-distortion) performance of the proposed algorithm in this chapter. In the next section, we will present the methods of estimating the complexity and capacity of an image. This capacity estimation is independent of the watermark insertion strategy.

4.4 Image Capacity Estimation

The hiding capacity of an image could be determined either as a function of the image content and complexity or as a result of the hiding strategy or algorithm to be used. In the first case, one can estimate the best region of an image to hide in order to achieve maximum security. In such a case, no data or hiding algorithm is being considered. In a second case where an algorithm has been defined, the capacity estimation method could then be determined empirically subject to some criteria. In the following sections, we shall first explore image capacity estimation methods, without having any hiding strategy in mind, and secondly, with the idea of watermark insertion strategy.

4.4.1 Capacity Estimation Before Embedding

Different images are made up of different contents and textures. These variations in features also determine how well an image or section of an image can conceal any modifications to the image. Hence, once could attempt to estimate how well a region of an image can hide data even before embedding. Various factors are considered in determining image-dependent hiding capacity.

The image complexity, which translates into a communication channel noise, affects the amount of information that can be successfully transmitted at an acceptable

distortion on both the message and the channel itself. Image complexity relates to the extent of details contained in an image. It can be seen as the entropy of the image. Given this capacity estimation via image complexity, we leverage the ROI-based image complexity measures proposed by Yaghmaee and Jamzad [160] in conjunction with our computation of image noise with a sequence, to perform capacity estimation before embedding. The ROI method of determining image complexity has been shown to possess a better correlation property with quality degradation and thus with the capacity of an image. The ROI method involves the computation of five (5) complexity metrics and combining them to estimate the capacity of an image or compare the capacity of different regions in an image or two different images.

Let **X** be an image of size $M \times N$. It is then divided into $\boldsymbol{P}$ sub-blocks, $S_i; i = 1, 2, \ldots, P$ each of size $m \times n$. Let a function for average intensity, ***AvgInt()***, of all the pixels in **X** be given as:

$$AvgInt(X) = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} X(i,j) \tag{4.1}$$

Then the ***ROI Score Parameters*** are defined and computed as follows:

1. **Intensity Metric,** M_I: Intensity refers to the brightness of an image. It is often measured from the grey levels present in an image. The blocks of images that lie within the mid intensity range of an image are known to be very sensitive to the human eye [160].

   $$M_I = |AvgInt(S_i) - AvgInt(X)| \tag{4.2}$$

   Repeat for all $S_i \in X$.

2. **Contrast Metric,** M_C - Contrast is the difference in pixel value between adjacent pixels or objects in an image. The human eye is attracted to sub-blocks with high level contrast with respect to surrounding sub-blocks. That sub-block is considered perceptually more important. Let each of the surrounding blocks for the current sub-block be S_{Surr-i}. Then take the average intensity of each S_{Surr-i} to

obtain $AvgSurr_i$. Then, the Constrast metric for current sub-block is given as:

$$M_C = |AvgInt(S_i) - AvgInt(AvgSurr_i)| \tag{4.3}$$

Repeat for all $S_i \in X$.

3. **Location Metric,** M_L - The human eye is attracted to the centre quarter of an image. For each sub-block, the proportion of it that lies within the centre 25% of the image constitutes the M_L. Hence:

$$M_L(S_i) = \frac{centre(S_i)}{Total(S_i)} \tag{4.4}$$

Where $centre(S_i)$ is the number of pixels of the current sub-block located within the centre quarter of the whole image and $Total(S_i)$ is the total number of pixels in the sub-image. This parameter tends to give undue importance to the blocks lying within the centre of the image. This metric can be omitted in this work, and it is also not useful for comparing the capacity of two images as all images have centre quarter component.

4. **Edginess Metric,** M_E - This metric is a count of the edge pixels in an image. First, the Canny edge detection method is applied to the sub-block, then all the edge pixels in the returned edge image are counted. A strong threshold is often recommended to suppress the effect of background noise. A threshold of 0.5 was applied in this work.

5. **Texture Metric,** M_T - This is the variance of the grey levels in the sub-block.

$$M_T(S_i) = var(S_i) \tag{4.5}$$

After these computations, the values for each metric across all sub-blocks is range-

normalised using the equation below:

$$M_{norm} = \frac{M - M_{min}}{M_{max} - M_{min}} \tag{4.6}$$

The normalised forms of these metrics are then utilised to compute the ***Importance Measure (IM)*** for each sub-block as:

$$IM(S_i) = M_I(S_i)^2 + M_C(S_i)^2 + M_L(S_i)^2 + M_E(S_i)^2 + M_T(S_i)^2 \tag{4.7}$$

To illustrate this computation model with medical images, we divided different sample images into 8x8 sub-blocks and computed the (***IM***) of each sub-block. The results are shown in Figures 4.2 and 4.3. At the sub-block level, the sub-block with the highest ***IM*** value is the perceptually most important sub-block and also the most complex region. These regions will hide most watermark bits with little perceived distortion. However, those regions with high ***IM*** are likely to have more HSI that ought to be canceled during the embedding process.

At the image level, the X-ray scan for diagnosing pneumonia has more blocks with $\boldsymbol{IM} > 0.4$ and is considered to have the highest estimated capacity of 2620 bits among the sample test images. We can use this ***IM thresholding*** estimation method to find the region of an image where we can embed information with higher capacity but with minimal distortion. If an image is divided into say four regions, the ***IM*** score of each region can be used to determine where a higher amount of information can be embedded with minimal distortion. Hence, even though we compared different images in Figure 4.2, the same approach can be applied to different regions of the same image, including its ROI region. This approach will be leveraged in our proposed algorithm to increase capacity where applicable.

In Figure 4.3 (b), we have normalised the absolute capacities estimated from 4.3(a) and plotted the normalised capacities against the average ***IM*** for each image. This plot shows that as the average ***IM*** increases, the estimated capacity using the IM thresholding method also increases. Hence, with spread spectrum-based estimation methods,

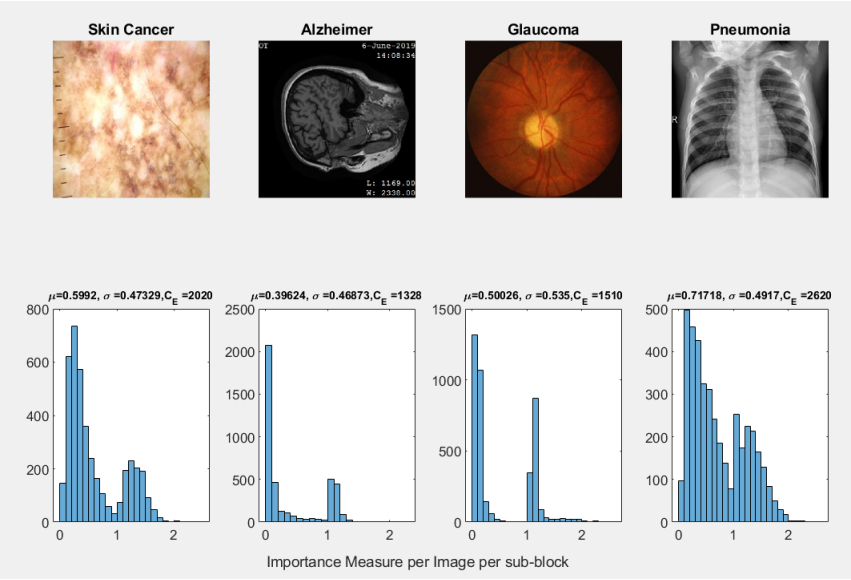


Figure 4.2: Capacity Estimation using Importance Measure (IM): *IM thresholding is used to estimate capacity, C_E. C_E is the number of sub-blocks with IM $>$ 0.4. All images were were normalised to a size of 512 x 512.*

a computation of the average ***IM*** of an image region can help to determine where to embed in order to maximise capacity and minimise distortion. These two goals are desirable in data-intensive medical image applications.

4.4.2 Steganographic Capacity Estimation

Both hiding capacity and steganographic capacity have been used interchangeably to depict the amount of information that could be hidden in a multimedia channel with an acceptably low probability of unauthorised detection and minimal error during authorised detection. Determining this capacity is important for steganography because significant information may need to be communicated while still maintaining the visual fidelity of the cover image.

Apart from the cover distortion level permissible by an application, various parameters have been known to affect Steganographic capacity. Among these are the size of the cover image (Bandwidth, B), watermark insertion strategy, statistical properties of

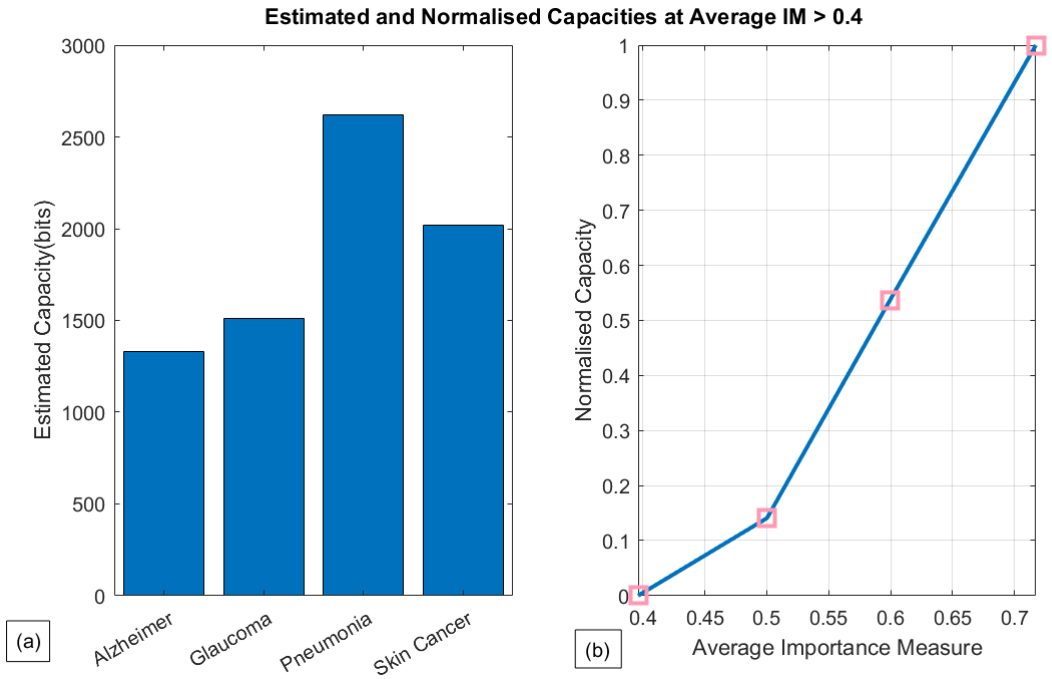


Figure 4.3: Variation of C_E with mean Importance Measure (IM): *The capacity of an image or region of an image is directly proportion to the average of the IM of the sub-blocks in the image or region.*

the cover (content and complexity), watermark embedding strength, channel noise [13, 65, 79], among others. These parameters are incorporated, in certain combinations, into various models and processes to estimate the Steganographic capacity of a given image and/or steganography method. These estimation methods can be generally classified as either theoretical or empirical. Whereas theoretical models are more generalisable than the empirical models, they use parameters that are difficult to measure empirically. An example of a theoretical measure is mutual information. On the other hand, empirical approaches allow for easily measurable properties as well as application-domain parameters for achieving a more accurate estimation. Typical examples of both estimation models are reviewed next.

Information-theoretic Model

Image Steganography is a form of covert communication where the cover image becomes the channel of communication. In considering steganography as communication, no analysis can be provided without reference to the pioneering works of Shannon [140] on Information theory, especially on the bounds on the amount of information that can be transmitted over a noisy channel.

A communication channel is widely modelled as having Additive White Gaussian Noise (AWGN). A steganographic channel follows the pattern of a classical communication channel. However, it has a *steganalyser*, which tries to detect if the information is hidden in the channel. It also has an *attacker* who tries to remove or corrupt this information, if any. These concepts are illustrated in Figure 4.4, which is the same as the tampering and attack model previously described in Section 3.3.2.

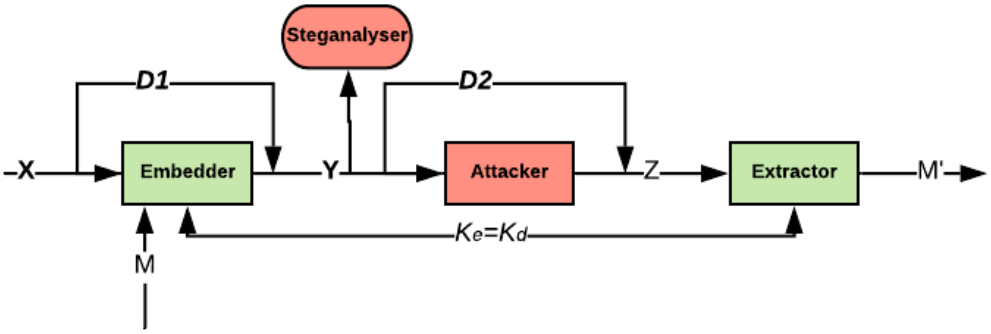


Figure 4.4: Steganographic Channel Modeling: *A distortion D_1 is introduced by the embedding process. This distortion belongs to the steganographic process and hence should be minimised while improving capacity. However, D_2 affects BER and therefore, affects the steganographic capacity from the receiver's perspective.*

For the purpose of theoretical modelling, let **X** (cover image) , **Y** (watermarked image) and **M** (embedded message) be random variables whose *i*th element is denoted by the pdf $f_{xi}(X)$, $f_{yi}(Y)$ and $f_{mi}(M)$ respectively. For an additive Steganographic system such C_4S, the general form of such systems is given as:

$$y_i = x_i + \alpha m_i \tag{4.8}$$

where x, y and m are specific realisations of the random variables **X** , **Y** and **M**. The parameter, α, represents the embedding strength which controls both distortion, D_1 and retrieval accuracy at the decoder. In actual sense, the pdf of the marked data y, is a conditional pdf: $f_y(y|m)$.

In practice, **X** and **M** are either continuous random variables or discrete memoryless channels (DMC) [13]. These represent the general class of cover data in both pixels and transform the domain of images. It should be noted that the DMC model is mostly true when the marked data y and the message m are quantised. For example, DCT coefficients and image pixels are quantised. Also, m is often in the domain, [-1,1]. Furthermore, the memoryless assumption is valid when the cover data x are independent of each other, and if the message bits, m, are independent and identically distributed (i.i.d) random variables. Although image pixels may not be completely i.id (adjacent pixels are sometimes correlated), most transform domain host data are.

With the DMC assumption, the quantisation values will be fixed and both **X** and **M** will be drawn from a finite set of alphabets. Let the input message come from the alphabet, $\bar{M} = m_0, m_1, \ldots, m_{K-1}$ and the output alphabet, $\bar{Y} = y_0, y_1, \ldots, y_{J-1}$. There is a set of transition probabilities that converts input to output through the steganographic operator in question. This transition probabilities are given as: $\mathbf{P} = \rho(\bar{y}_j|\bar{m}_k), j = 0, 1 \ldots J-1, k = 0, 1 \ldots K-1$. As α is a constant in (4.8), it is clear that $\bar{M}$, $\bar{Y}$ and $\rho(y_j|m_k)$ completely describes the DMC steganographic channel, as X can as well be derived from these relations. However, for a blind steganographic system like ours, X, is usually not of interest at the receiver.

The total transition probabilities can be computed by integrating the conditional pdf, $f_y(y|m)$. That is:

$$\rho(\bar{y}_j|\bar{m}_k) = \int_{y_j}^{y_{j+1}} f_y(y|m)dy \tag{4.9}$$

To apply Shannon's theory for the capacity of DMC channels, the mutual infor-

mation, $I(\bar{M};\bar{Y})$, of the watermarked channel is required. This is because ***the channel capacity*** is given by the ***maximisation of this mutual information over all possible sets of inputs*** .

For discrete random variables:

$$I(M;Y) = \sum_{j\in\bar{Y}}\sum_{k\in\bar{M}} \rho(y_j,m_k) log\left(\frac{\rho(y_j,m_k)}{\rho(y_j)\rho(m_k)}\right) \tag{4.10}$$

subject to the constraint:

$$\sum_{k=0}^{K-1}\rho(m_k) = 1, \qquad \rho(m_k) \geq 0 \qquad \forall k. \tag{4.11}$$

For practical use, mutual information is not an accurate measure of distortion because the probability distributions chosen for a given image are probabilistic [60], especially for just an 8x8 pixel block. It often does not correctly represent data samples drawn from that distribution. This inaccuracy has led to the use of a more practical measure called *operational channel capacity*. Operational channel capacity is defined as the highest number of bits that could be transmitted in one channel use without error. This is equal to information channel capacity and can be expressed mathematically as:

$$C_{op} = \log_2 \frac{M}{n} \tag{4.12}$$

Where M is the number of signals sent without error in n transmissions using the same channel. For medical image steganography, each image will be used once as a transmission channel. If there are L message bits to be transmitted, instead of using the image (Channel) L times, the image is divided into L sub-channels, and each channel is used only once. Hence, $n = 1$ while $M = 1, 2, 3, \ldots$. But, if there are multiple uses of the image among different physicians, then one can model n uses of the channel. This second case is relevant for modeling image capacity for traceability and accountability applications. These are not the focus of this thesis.

For complete theoretical evaluation of Steganographic capacity, (4.10) has shown

that the pdf or probability mass function (pmf) of the specific host channel being considered is needed. The common distributions considered in literature are two transform domain covers: *Discrete Cosine Transform (DCT)* and *Discrete Wavelet Transform (DWT)*, and *pixel domain* of natural and medical images. These have been modelled in Section 2.4.1. For some researchers [156, 65], the distribution of the coding sequence is of utmost importance and not that of the cover signal.

Maximum Capacity under Undetectable (MCUU) model

MCUU model is a practical and empirical model for estimating the Steganographic capacity of images. The approach can be applied to different cover images. It involves the extraction of relevant features from the cover image and using these features to estimate Steganographic capacity. This model is based on the assumption that any steganographic system is suitable for an application provided that it provides maximum secrecy for the embedded data. Generally, this secrecy is evaluated by the inability of the human visual system to detect the presence of the embedded data. As humans alone no longer perform certain actions, the view has extended to various features that represent 'vision' in computer systems. In these regards, it is the relationship between these features and how they are altered by an algorithm that defines steganographic capacity.

The MCUU model consists of a Steganography and a Steganalysis System. This approach is illustrated in Figure 4.5 . The steganography system consists of a chosen embedding and extraction function. For steganalysis, designers can select some suitable features and classifiers for the type of algorithm and application concerned. For example, in Jiafa [79], a general steganalysis method which involved three feature extraction methods were used: Markov process-based (MPB), partially ordered Markov model (POMM) and Markov-DCT-based (MDB). These feature extraction methods are specific to JPEG images as the authors analysed the Steganographic capacity of the DCT-based technique. They also used the *libSVM* classifier. Thus, it seems that both

the features extracted and the actual implementation of the classifier may depend on the type of image and steganographic operator involved.

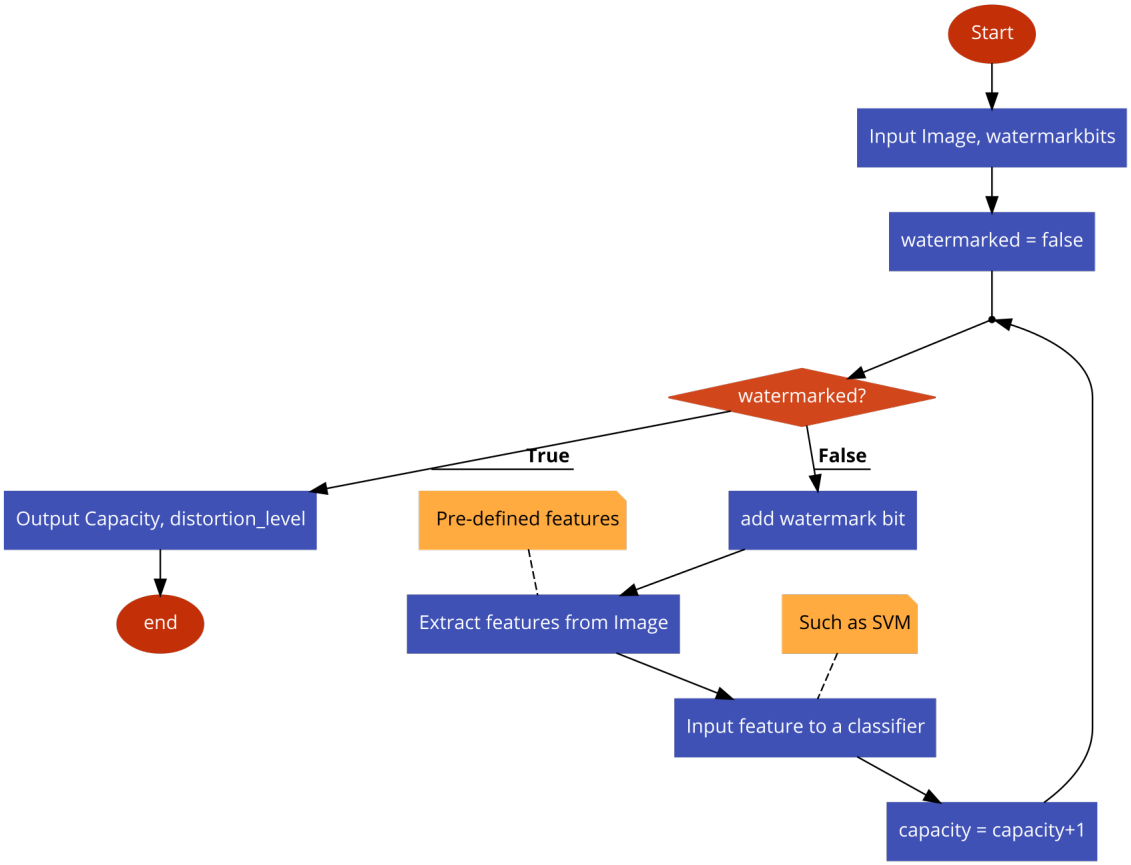


Figure 4.5: MCUU Model: *This is based on the use of an SVM steganalyser to detect watermark and determine if capacity has been exceeded or not.*

Therefore, the steganographic algorithm (steganography operator), nature of the cover image, and the embedding strength are the factors that determine steganographic capacity. Consider a maximum (allowable) change or distortion value Δ_D, at an embedding rate, **cr** per sample and embedding strength α, then one can express total distortion as:

$$\Delta_D = \alpha \sum_{x \in X} cr.P(W|\pi) \tag{4.13}$$

Where $P(W|\pi)$ is PDF of the samples, X, into which the watermark is embedded. It could be a pixel histogram or a chosen PDF for DCT or DWT transform coefficients. Hence, the number of bits that could be embedded until Δ_D is reached becomes the steganographic capacity of the method.

In summary, following from our discussions so far in this chapter and from the factors mentioned in Chapter 2, the capacity of the SS steganography system is subject to the following parameters:

1. **bandwidth, B** of the image - This refers to the number of pixels that make up the image.
2. choice of T_h - The extraction threshold.
3. the HSI of the cover image - the level of interference with respect to the spreading sequence.
4. length of W called Process Gain.
5. the choice of embedding strength, α.
6. complexity of the region of embedding

These parameters determine the number of bits that can be embedded, the distortion level after embedding, and the robustness of message extraction. We will take these parameters and the chosen features of medical images in order to design our algorithms.

4.5 Proposed Method for Capacity Improvement

There are unique features of medical images that have not been explored yet to improve capacity. These are the volumetric nature and the high-pixel depth of medical images. Some medical images are made up of 3-D slices instead of a 2-D image. They are made up of a given number of 2-D slices. The exploration of the different slices provides an opportunity for the embedding of an increased amount of data. This method is voxel-based watermarking, and it is currently under-researched [93]. The major challenges

with this method are lossy compression and geometric distortion [93]. Some 2-D Medical images have a pixel depth of more than 8-bits (12 or 16 bits). This high pixel depth can be utilised to increase the capacity, robustness, and imperceptibility of an SS embedding strategy. The replacement of two or more bits randomly chosen from a 16-bit pixel between the 12^{th} and 16^{th} bit-plane will improve both capacity and robustness. We take cognisance of these unique features and leverage them to achieve higher capacity.

The major concepts behind the improvement in this chapter are multi-level signaling, extending the C_4S technique developed in Chapter 3, and the use of CDMA. We now focus on multi-level signaling adaptation and the extended constant correlation method as we have previously discussed CDMA in Section 2.3.6. The objective of our proposed design is to improve the steganographic capacity while retaining the tamper detection capabilities already established in Chapter 3.

4.5.1 Multilevel Signalling

Multilevel signaling is different from the classic binary signaling system. Instead of encoding a single bit using two symbols, one encodes N-bits using M-symbols, where:

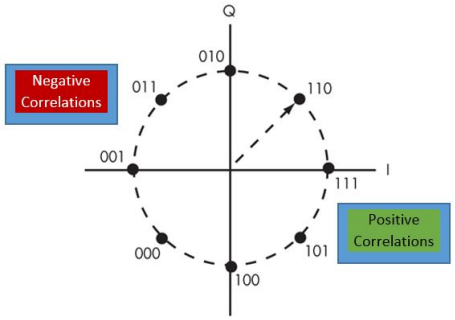
$$M = 2^N \qquad (N > 1) \tag{4.14}$$

For example, two bits are encoded into four symbols (00, 01, 10, and 11). Similarly, three bits is encoded using eight symbols (000,001,010,011,100,101,110,111).

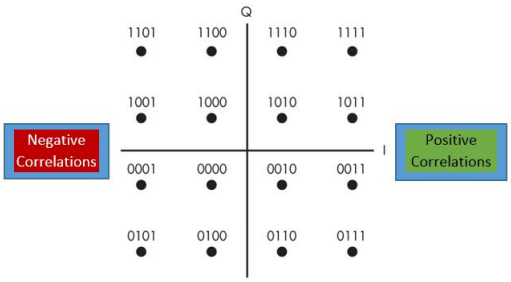
In digital signal processing (DSP), multi-level signaling employs variations in signal amplitude, phase, or both. In Figure 4.6a, an 8PSK (Phase shift keying) uses variation in phase to encode 3 bits per symbol. Eight phases were used to define 3 bits, as shown. The phase position encodes information bits. In Figure 4.6b, both variation in phase and amplitudes were used to encode 4 bits per symbol. Sixteen different phase and amplitude combinations were utilised, one for each 4-bit message.

To achieve similar encoding described above in SS steganography, we have defined correlation values for each possible bit group. The bits on the left will be assigned neg-

ative correlation values, while those on the right will be assigned positive correlation values. By dynamically computing the embedding strength, α, we can ensure that the correct correlation level is computed at the receiving end.



(a) An 8PSK signal with constant amplitude



(b) A 4-bit code: Change is phase and amplitude

Figure 4.6: Constellation diagrams: *Examples of multi-bit representation methods in digital communication. Variation in both phase and amplitude could be used in electrical signals. In our method, phase and amplitude have been converted into linear correlation values betwween the spreading sequence and the image block.*

The ***base correlation value (BCV)*** denoted as ρ, is a secret real-valued decimal ranging from 0.5 to 0.99 for spatial domain embedding and 0.001 to 0.999 for frequency domain of embedding. Higher embedding levels or channels are defined starting from ρ, with channel gap, G, for signal separation between any two levels (See Figure 3.5). This minimum separation is required to ensure that the signal represented by each correlation value is correctly detected at the receiver. For each signal level, an error tolerance, ϵ, is allowed, and it equals $\frac{G}{2}$. These parameters are the same as the ones used in the algorithms in Chapter 3. We will provide further illustrations in Section 4.5.2 that is specific to extended C_4S.

We have observed that similar to information-theoretic channel encoding, multi-level encoding increases the noise in the signal (original image), thereby reducing its quality. The challenge is to find a compromise between capacity and distortion, as is usually the case (see Figure 1.2). The theoretical limit to capacity can be estimated from

the Shannon-Hartley Capacity theorem given by (4.15):

$$C = B.\log_2(1 + \frac{S}{N}), \tag{4.15}$$

where ***C*** is the maximum channel capacity in bits per second, ***B*** is the available bandwidth and ***S/N*** is the received signal-to-noise ratio. Increasing both ***B*** and ***S/N*** increases ***C***. However, ***B*** is often fixed, and either S or N could be manipulated to alter capacity. In our method, S/N is determined by the maximum distortion that would not affect medical diagnosis. In terms of SS steganography, ***C*** represents the embeddable blocks in the image subject to an allowable distortion. Distortion is represented by ***S/N*** (example is PSNR). ***B*** is the total available sample blocks that could be used for embedding. For example, a 512 x512 image that is divided into 8x8 blocks has ***B*** = 4096 samples. However, allowable distortion per sample determines if the sample will be used for embedding or not. The design challenge here is to improve capacity while keeping distortion minimal. Whereas there is no universally accepted benchmark for this distortion in the medical world, we start with the 40dB benchmark established in [49] for medical images.

4.5.2 The Extended Constant Correlation Method

The groups of bits illustrated in Figures 4.6a and 4.6b are computationally represented within an image by constant correlation values starting from ρ, which is the same as the one used in (3.9) and (3.10). This extended method is no longer a binary channel like the one described in Chapter 3. Consequently, not just a zero (0) or a one (1) is transmitted per channel but combinations of 0s and 1s. Similar to multi-level signaling, let each bit group (part of the message to be transmitted) to be sent be encoded by a voltage level, $\rho_o \pm n\epsilon$. Where $n = 1, 2, 3, ...$ and ρ is the BCV. If there are N possible bit groups, then $n = 1, 2, 3, \ldots, \frac{N}{2}$, meaning that at least N quantised levels are required with $\frac{N}{2}$ levels having either positive or negative value. Table 4.2 shows how these define levels of embedding (channels) are combined with different PN sequences assigned to users

or information sources. *Horizontal scaling* involves the use of sequences to represent a group of bits, while *Vertical scaling* involves the use of different signal amplitude levels to encode a group of bits. The combination of these methods gives higher embedding capacity as well as tamper detection capability. A concrete example is provided in Table 4.3. For the horizontal scaling, we use a coding system that maps a subset of bits into a given spreading code. However, this code alone does not define the bits retrieved in a block. What defines the bits to be extracted from a block are the *code and the channel* from which the code was detected.

Table 4.2: Hybrid Methods to increase capacity: *Horizontal plus vertical scaling to achieve higher capacity per block of image.*

	Sequence No(S):				
	S_1	S_2	$\dots$	S_{m-1}	S_m
Levels(L)	n_1		$\dots$		
	n_{-1}		$\dots$		
	n_2		$\dots$		
	n_{-2}		$\dots$		
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	$n_{\frac{n}{2}}$				
	$n_{-\frac{n}{2}}$				

Table 4.3: Concrete Exampe of horizontal and vertical scaling: *Sequences* $= 8(S_1 - S_8)$, *Levels* $= 4(L_1 - L4)$: *A set of 8 sequences can be used to encode 3 bits. In addition to these 3 bits, the signal level (out of the 4 defined levels) at which this sequence is transmitted also encodes 2 extra bits making a total of 5 bits per sequence and per level.*

	$S_1 - S_8$								
$L_1 - L4$		[**000**]	[**001**]	[**010**]	[**011**]	[**100**]	[**101**]	[**110**]	[**111**]
+1	**[00]**	00000	00001	00010	00011	00100	00101	00110	00111
-1	**[01]**	01000	01001	01010	01011	01100	01101	01110	01111
+2	**[10]**	10000	10001	10010	10011	10100	10101	10110	10111
-2	**[11]**	11000	11001	11010	11011	11100	11101	11110	11111

Let us now define the mathematical representation of this embedding strategy. The channels into which groups of bits are embedded are determined by the following equation:

$$\rho_n = \pm\rho \pm n\epsilon \qquad for \qquad n = 0, 2, 4, \ldots, \tag{4.16}$$

where ϵ is the detection tolerance at the receiver, which is expected to be at least equal to the quantisation noise, σ, at the embedding side. This design is illustrated diagrammatically in Figure 4.7.

For example, when 3 binary bits are transmitted per symbol, we need 8 levels to transmit 2^3 bit combinations. That is: 000, 001, 010, 011, 100, 101, 110, 111 . We transit 4 symbols (say 000, 001, 010, 011) as negative correlated symbols at levels $-\rho, -\rho - 2\epsilon, -\rho - 4\epsilon$ and $-\rho - 6\epsilon$ while other symbols are transmitted at $\rho, \rho + 2\epsilon, \rho + 4\epsilon$ and $\rho + 6\epsilon$, respectively.

The vertical arrows in Figure 4.7 show that the amplitude of embedding increases as the number of bits embedded per symbol increases. This trend will cause a variation in local distortion depending on the groups of bits embedded in the sub-block. We can embed 1, 2, or 3 (or more) bits per sample, as shown in Figure 4.7.

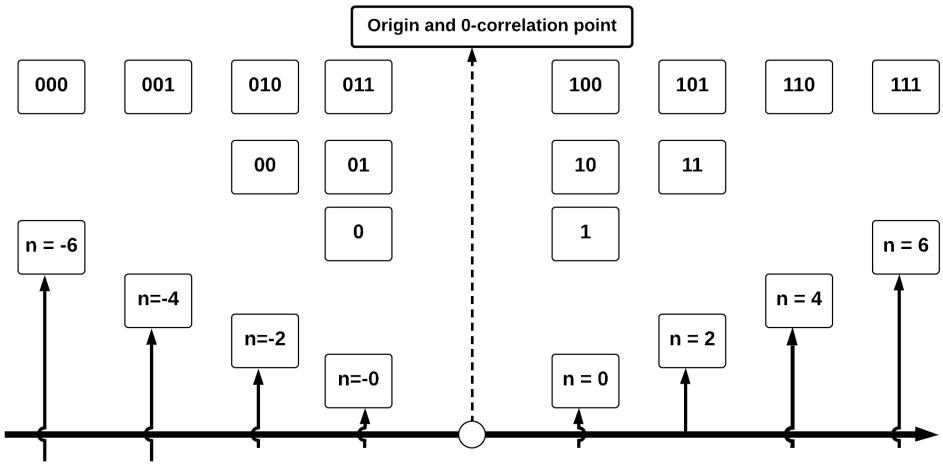


Figure 4.7: Example of extended C_4S Design: *Based on the complexity of the image and maximum permissible distortion for an application, higher number of channel levels could be defined to increase capacity.*

To minimise distortion, we embed bit groups that have a higher probability of having an equal number of ones and zeros closer to the original. For instance, if the binarised EMR or hospital authentication logo is grouped into three bits, we assume that there will be more bit groups containing at least a one or a zero than bit groups containing all ones or all zeros. Thus, bits containing all ones and all zeros are further embedded in the number line as they will introduce more distortions into a sub-block.

During watermark extraction in which tamper detection is a concern, the desired correlation at each sub-block should be ρ_n, depending on the group of bits embedded in the sub-block. However, to tolerate allowable image modification such as JPEG lossless or 100% quality factor compression and few quantisation errors, the actual range of detected correlation value, r, follows the following inequality:

$$(\rho_n + (n-1)\epsilon) \quad \leq \quad r \quad \leq \quad (\rho_n + (n+1))\epsilon \tag{4.17}$$

As in the fundamental C_4S defined in Chapter 3, only ρ is pre-determined and this corresponds to $\rho_n = \rho_0$ in (4.16). Subject to distortion constraints allowed by the application concerned, this M-ary encoding scheme increases capacity by $log_2 M = N$ times.

The process of combining multiple-bit embedding with tamper detection is given in Appendix C.2 as this multi-functionality is not a core requirement of this chapter. However, it worth noting that our design has this capability. In the next section, we will list the algorithms we developed as part of this new method. After that, we will provide mathematical analysis before an empirical evaluation using two datasets.

4.5.3 The Algorithms

In this section, we present the embedding and the extraction algorithms for our steganographic system.

Figure 4.8 shows the flowchart for the watermark generation and embedding algorithm called the compression encoding algorithm (CEA). The inputs to the algorithm includes the original medical image scan **X**, a pseudo-noise sequence **W**, the required

number of bits per sample known as the compression ratio **Cr**, the message bits to be embedded **Msg** (length **L** in bits), and ρ. Error tolerance, ϵ, defaults to 0.5. To insert **Cr** bits per channel, one requires 2^{Cr} levels or $num\_channel$. With tamper detection, the distance between any two channels or embedding levels is $G = 4\epsilon$, otherwise $G = 2\epsilon$. With the help of Figure 4.7, the next is to compute the bit groups that are to be embedded on either side of zero, i.e., 0^-(nzbitvalue) and 0^+ (pzbitvalue). After this, we iterate through the messages and image sub-blocks, computing the dynamic embedding strengths, α, and embedding **Cr** bits in each sub-block until the messages are all embedded. Other parameters and functions used in these algorithms are defined in Table 3.2. The details of the embedding algorithm are shown in Algorithm 3.

The watermark decoding, extraction, and tamper detection algorithm called watermark decompression decoding and tamper detection algorithm (DDTDA). Because of its size, the DDTDA flowchart is in Appendix C.1. However, We present the DDTDA pseudocode in Algorithm 4.

4.6 Analysis of the Extended Method

By adapting Shannon's [140] equation for a classical communication into that of a steganographic channel, the capacity of a channel is given by (4.18).

$$C_w = B.\log\left(1 + \frac{P_{wat}}{P_{host}}\right), \tag{4.18}$$

where C_w is the steganographic channel capacity, P_{wat} is the watermark power, and P_{host} is the power of host data (pixel image or transform coefficients). B is the bandwidth of the image, which according to Nyquist [2] sampling can be given as $B = S_{im}/2$, where S_{im} is the total number of pixels in the image. As B is fixed for an image, a viable means to increase capacity is to increase the number of bits encoded into an image sub-block. We have implemented this technique through some defined correlation levels within

[2] http://www.rctn.org/bruno/npb261/aliasing.pdf

Algorithm 3: Compression Encoding Algorithm (CEA)

Data: X, W, Cr, Msg, ρ,ϵ
Result: Y, PSNR, SSIM, KLD

Divide X into 8x8 sub-blocks, X_i;
$maxDistortion = 40$;
$n = 4$;
$L = length(Msg)$;
$numChannels = 2^{Cr}$;
$channelGap = n * \epsilon$;
$polarisedChannels = numChannels/2$;
$nzbitvalue = polarisedChannels - 1$;
$pzbitvalue = polarisedChannels$;
if *mod(L,cr) != 0* **then**
    Msg = pad(Msg,'0');
end
$msg = substr(Msg, Cr)$;
$subBlockNumber = 1, i = 1$;
while *msg in Msg* **do**
    decMsg = bin2dec(msg);
    if *decMsg less than or equal to nzbitvalue* **then**
        $polarity = -1$;
        $channel = polarisedChannels - decMsg$;
        $ecc = -\rho - (channel - 1) \times channelGap$;
    end
    if *decMsg greater than or equal to pzbitvalue* **then**
        $polarity = 1$;
        $channel = decMsg - polarisedChannels$;
        $ecc = \rho + (channel - 1) \times channelGap$;
    end
    $\alpha = (ecc - \langle X_i, W \rangle)/polarity$;
    compute local distortion, d;
    if *d greater than or equal to maxDistortion* **then**
        $Y_i = X_i + \alpha \times W \times msg$;
        $msg = substr(Msg, Cr)$;
        compute local SSIM and PSNR;
    end
    $subBlockNumber = subBlockNumber + 1$;
    i=i+1;
end
Compute global SSIM, PSNR and KLD ;

Algorithm 4: Decompression Decoding Algorithm (DDA)

Data: Y, Parameters: W,Cr,ρ,ϵ,L
Result: EMR, tamperdetected?,BER

Divide Y into 8x8 sub-blocks, Y_i ;
Get message length, L;
polarisedChannels, $pc = numChannels/2$;
nzbitvalue $= pc - 1$;
pzbitvalue $= pc$;
thisbitgroupvalue $=' 1000000'$;
$i = 1; j = 1$;
while *i less than or equal to L* **do**
    Generate spreading code, W;
    $P = \langle yblock, W \rangle$;
    $binarybits = null$;
    $blockStatus =' NoWatermark'$;
    if P *j*= -$n\rho$ **then**
        while *j less than or equal to nzbitvalue* **do**
            if $P = -j * \rho \pm \epsilon$ **then**
                $thisbitgroupvalue = f(j, P)$;
                $binarybits = binary(thisbitgroupvalue)$;
                $Msg = Msg + binarybits$;
                $blockStatus =' WatermarkFound'$;
            end
            $j = j + 1$;
        end
        if $binarybits == null$ **then**
            reportTampering();
            $blockStatus =' Tampered'$;
        end
    end
    if $P >= n\rho$ **then**
        while *j less than or equal to pzbitvalue - 1* **do**
            if $P = j * \rho \pm \epsilon$ **then**
                $thisbitgroupvalue = f(j, P)$;
                $binarybits = binary(thisbitgroupvalue)$;
                $Msg = Msg + binarybits$;
                $blockStatus =' WatermarkFound'$;
            end
            $j = j + 1$;
        end
        if $binarybits == null$ **then**
            reportTampering();
            $blockStatus =' Tampered'$;
        end
    end
    $i = i + 1$;
    reportBlockStatus(blockStatus) ;
end
Compute Watermark Quality Parameters;
Convert extracted EMR, *Msg* bits to text;
Flag tampered blocks, if any;
Display EMR and tampered version of medical image;

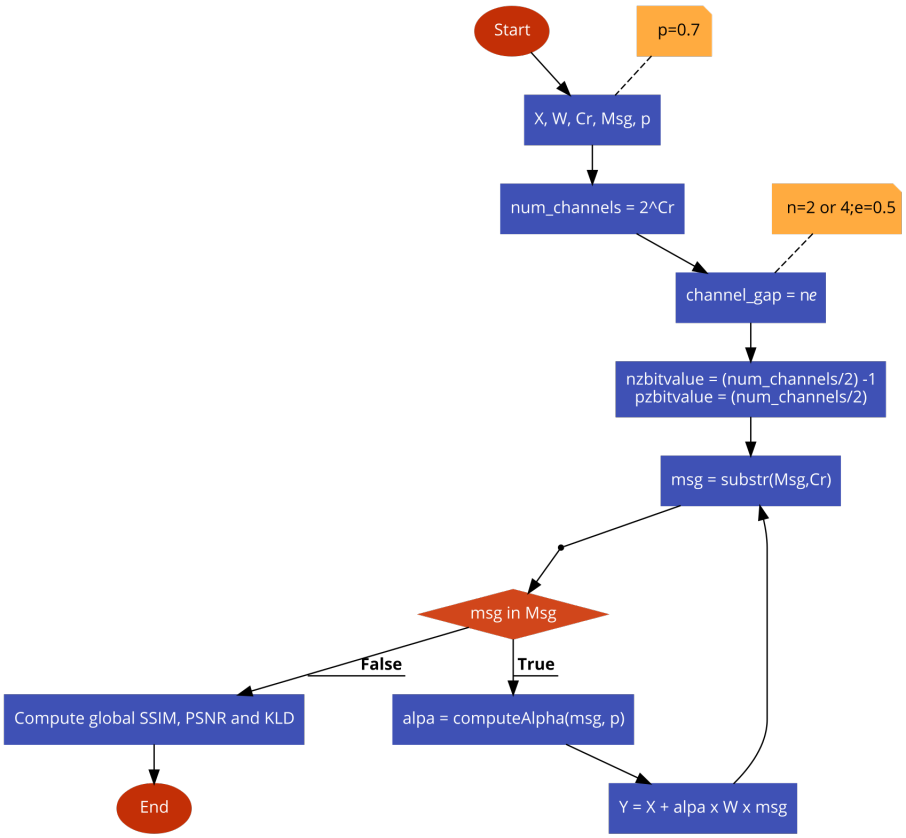


Figure 4.8: Compression Encoding Algorithm(CEA): *'Compression' here refers to the ability to encode more than one bit of information into the spreading sequence embedded in a sub-block*

each image sub-block. We will use this notion of quantised correlation levels or the notion of defined channels within an image to perform our analysis.

Let the distance between any two quantised levels, $G = \lambda\sigma$, where σ is the Root-Mean Square (RMS) value of noise voltage at the receiver. The signal levels for each bit group will then be represented by the series:

$$G_i = \pm\frac{G}{2}, \pm\frac{3G}{2}, \pm\frac{5G}{2}, \ldots, \pm\frac{(M-1)G}{2}, \tag{4.19}$$

where $\boldsymbol{M}$ is the number of signal levels and it is the same as the one used in (4.14). If all symbols are equi-probable in the message, then the average signal power, S, is given

as:

$$S = \frac{2}{M}\left[\left(\frac{3G}{2}\right)^2 + \left(\frac{5G}{2}\right)^2 + \ldots + \left(\frac{(M-1)G}{2}\right)^2\right]. \tag{4.20}$$

By applying the sum of squares of the first $M/2$ odd numbers, Equation 4.20 simplies to:

$$S = \left(\frac{M^2-1}{3}\right)\left(\frac{G}{2}\right)^2 = \frac{M^2-1}{12}(\lambda\sigma)^2. \tag{4.21}$$

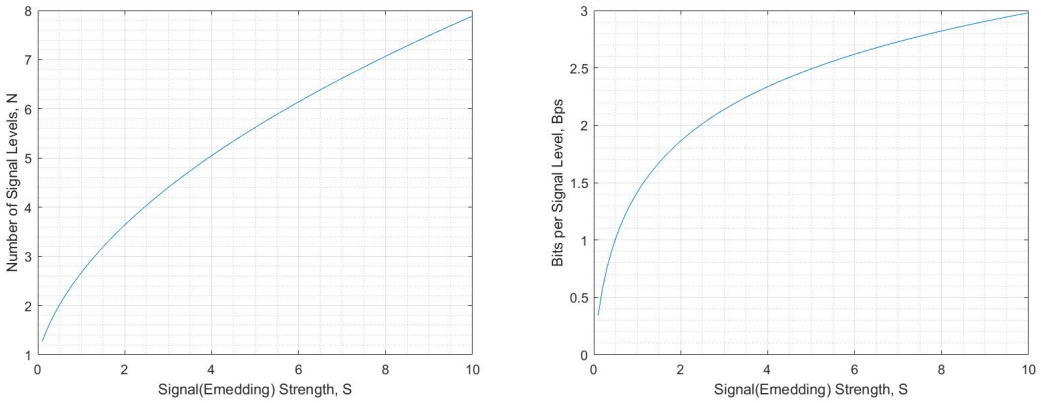
Increasing the number of levels, M, increases signal power, S (equivalent to α in steganography and watermarking). This increased signal power also increases distortion even though this higher signal power increases the probability of correctly detecting the signal (watermark) at the receiver. This trend implies that higher signal power increases robustness but decreases imperceptibility. Hence, capacity improvement is limited by the number of levels or channels accommodated in an image block. Given an average signal power, S for an application, the maximum number of levels, $\boldsymbol{M}$ that can be employed for watermark embedding is:

$$M = \sqrt{\left(1 + \frac{12}{\lambda^2}\frac{S}{\sigma^2}\right)}, \tag{4.22}$$

where $\sigma^2 = N$ (as given in (4.15)) is the noise power in the channel.

The limitation on the number of bits that can be compressed into one image subblock for transmission is determined by (4.22), and it is precisely equal to $\log_2 M$. Hence, according Nyquist rate, this method increases information rate to $2B\log_2 M$.

Now let us consider distortion performance. Given typical distortion level (measured in a quantity of choice) that could be tolerated at a given signal strength S, and given a quantisation noise power, σ (or ϵ in our method) that can be subdued by adequate signal level separation of $G = \lambda\sigma = f(\rho, \epsilon)$, one can compute M and thus $\log_2 M$. According to information theory, λ should be sufficiently large (we chose 4 in the evaluation of our algorithm) to ensure that information is always correctly decoded irrespective of noise.



(a) Signal Strength (S) Vs Number of Levels (b) Signal Strength Vs No. of bits per level (Cr)

Figure 4.9: Signal Strength (S) Vs Channel Capacity : *signal strength as well as steganographic capacity is limited by allowable distortion for an application.*

For illustration purposes, let the maximum value of $\lambda = 4$ and maximum value of quantisation noise, $\sigma = 0.5$, we can then visualise the effect of allowed signal power , S on the number of allowed levels, $\boldsymbol{M}$ as shown in Figure 4.9. S does not grow linearly with the number of bits, Cr, compressed into a sample. At a certain level, an increase in S will not increase steganographic capacity as Cr will become marginally lower as image distortion increases due to a larger S. Hence, signal strength, as well as steganographic capacity, is limited by allowable distortion for an application.

In the case of tamper detection, we only tolerated quantisation noise (± 0.5 for spatial domain). This is because other forms of noise are meant to be detected by the algorithm. Regarding Shannon's theorem, which assumes that information can be transmitted with little or no error if the transmission rate is less than or equal to channel capacity, every other source of error at the decoder is deemed malicious and should be detected.

In the next section on experimental evaluation, we will evaluate our technique empirically and quantify distortion level in terms of image quality assessment (IQA) parameters.

4.7 Experimental Evaluation

In this section, empirical experiments were designed to evaluate the Steganographic capacity improvement algorithms in this chapter. Specifically, the experiments aimed at:

1. Determining the Steganographic capacity of an extended C_4S method.
2. Ascertain the level of distortion of the image.
3. Compare theoretical and empirical Steganographic capacities.

4.7.1 Datasets and Payload

The datasets used for experimental evaluation in this chapter were fully described in Section 3.6.6. Some sample images are shown in Figures 4.11 and 4.12.

We have two types of payload: (i) Randomly generated bits and (ii) Typical payload (See Figure 4.10. We utilised randomly generated bits when testing the full capacity of each image. All the images do not have the same size (bandwidth). Hence, a random payload is generated to fully test the image. The final capacity result is normalised to 1 for each image.

Figure 4.10 reflect typical summary of EMR and radiology records as used by both Qasim *et al* [132] and Al-Haj *et al*[63].

4.7.2 Experimental Procedure

MATLAB R2017b and windows 10, 64 bits 16GB RAM Computer were used to implement and run the algorithms on each of the datasets. Each of the images and image slices (in a DICOM volume) was divided into 8x8 sub-blocks into which one or more (cr $= log_2 M$ =1, 2, ..., 10) bit of ASCII binary patient information was embedded using gold-code spreading sequence.

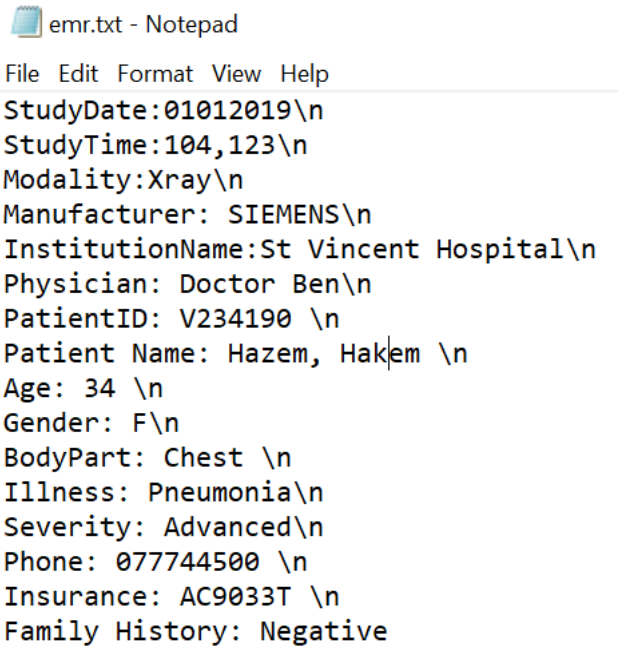


Figure 4.10: Sample Payload: *It includes Electronic Medical Record (EMR), Scan metadata, originating hospital details and initial radiologist interpretation*

The spreading PN code, W, is a gold code generated from $x^6 + x^5 + 1$ and $x^6 + x^5 + x^4 + x + 1$ preferred pairs. Sample Electronic Medical Record (EMR) were converted into binary bits using MATLAB 2017b.

Table 4.4: Watermark Data Sources: *A total of 4,104 bits of information was embedded on the average*

Type	Description	Size (Bits)
Scan Information	Scan metadata such as modality etc.	1448
Patient Note	Special condition about the patient like allergies	800
Radiologist Brief	Initial interpretation by the radiologist	800
Hash Data	ROI image hash or secret hospital hash data for ROI protection	256
Other Biomarkers	Other test data such as blood tests	800

(a) ADNI Sample (b) ISIC Sample

Figure 4.11: Samples of ADNI and ISIC datasets: *The ROI has been defined by the medical experts before the release of the dataset. However, we boxed it into the centre-quarter of the image for easy extraction*

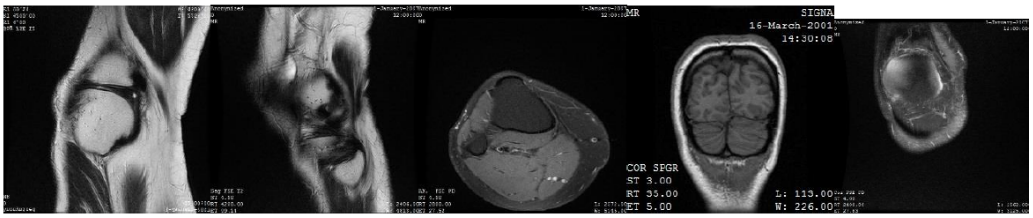


Figure 4.12: OSIRIX Sample MRI images.

4.8 Results

The result of various experiments carried out to investigate the steganographic capacity of the new method in terms of actual hiding capacity as well as distortion is presented in this section. Several experiments were carried out in line with the objectives of this chapter as well as in line with the second research question (See Section 1.5) in this thesis. The exact experimental procedure for each sub-question, as well as the results obtained for such an experiment, is described in each of the sub-sections that follow.

Table 4.5: Steganographic Capacity for C_4S for 16-bit OSIRIX Dataset. NOTE: BER=0, bps = bits per sample

Capacity(bits)	Cr (bps)	PSNR (dB)	SSIM
8,192	2	74.38	0.9999
12,228	3	74.97	0.9999
16, 384	4	74.43	0.9999
20, 480	5	72.26	0.9998
24,576	6	68.84	0.9993

4.8.1 Steganographic Capacity Improvement

In order to measure the Steganographic Capacity that can be reached by our method, the number of bits,Cr, embedded in each 8x8 sub-block, was increased progressively from $Cr = 1, 2, 3, ...$ For each sub-block and value of Cr, the Peak signal-to-noise ratio (PSNR) was measured, among other distortion parameters.

The results about the amount of embedded watermark and the correctness of the extracted watermark are presented here. These are the quantitative measures of Steganographic capacity. The capacity limit, in this case, is measured as the number of bits that can be accommodated in a sub-block with distortion less than 40dB for any medical image sub-block [49]. We applied this benchmark to local sub-blocks instead of the global image used by almost all other researchers. This is to ensure better image quality after watermark embedding.

Table 4.5 and Figure 4.13 show the major Steganographic capacity values for the C_4S method. The results presented here were obtained from the OSIRIX dataset, which is made up of DICOM MRI images of pixel depth between 12 and 16 bits per pixel. The capacity is given in bits and in bits per sample, where a sample is made up of 8x8 pixel blocks of the MRI image.

Figure 4.14 shows the capacity-distortion for the Chest-Xray Pneumonia dataset.

The results show that Steganographic capacity increases with distortion; hence, the decrease in values of PSNR and SSIM as the capacity increases from 2 bps to 6bps. Even

Table 4.6: Steganographic Capacity for C_4S for 8-bit Chest X-ray Pneumonia Dataset. Bps = Bits per sample, S.D = Standard Deviation

Capacity(bits)	Cr (Bps)	PSNR (dB)	SSIM	BER
8,192	2	41.73(S.D=0.1239)	0.9505(SD = 0.0055)	0
16, 384	4	29.0644 (S.D=0.0852)	0.5754(SD = 0.0257)	1.9558e-05

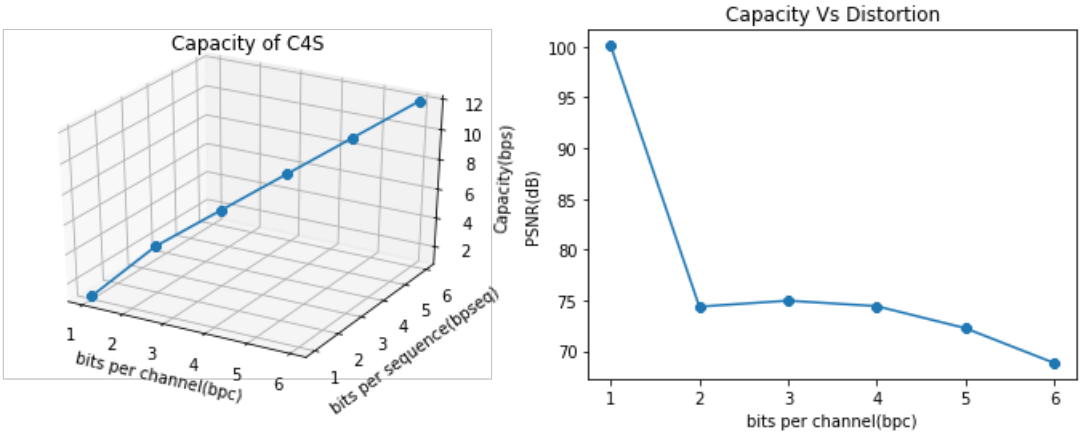


Figure 4.13: Steganographic Capacity for 16-bit DICOM image: *For 16-bit DICOM images decrease in distortion is slow and thus more channels and code sequences can be used to improve steganographic capacity to 12 pbs.*

at 6Bps, the PSNR value is well above the 40dB benchmark given by [49] for medical images.

Furthermore, one can use the entire (approximately) 25% of no-content slices in a DICOM volume for watermarking. If there are K no-content slices and each slice has a Steganographic capacity of Q bits, then a volume of DICOM image can contain $K \times Q$ watermark bits. For instance, in the case of ADNI dataset, there were $K = 22 + 18 = 40$ no-content slices. If the Steganographic capacity of these no-content slices is 12,228 bits (See Table 4.5), then we already have $12228 \times 40 = 489,120$ bits without embedding in the ROI or RONI of the content slices.

In terms of the local sub-blocks whose PSNR went below 40dB, there were only 0.32% for the ISIC dataset, and for ADNI, it is 0.89%.

Figures 4.15 clearly shows that most of the image sub-blocks have a quite high im-

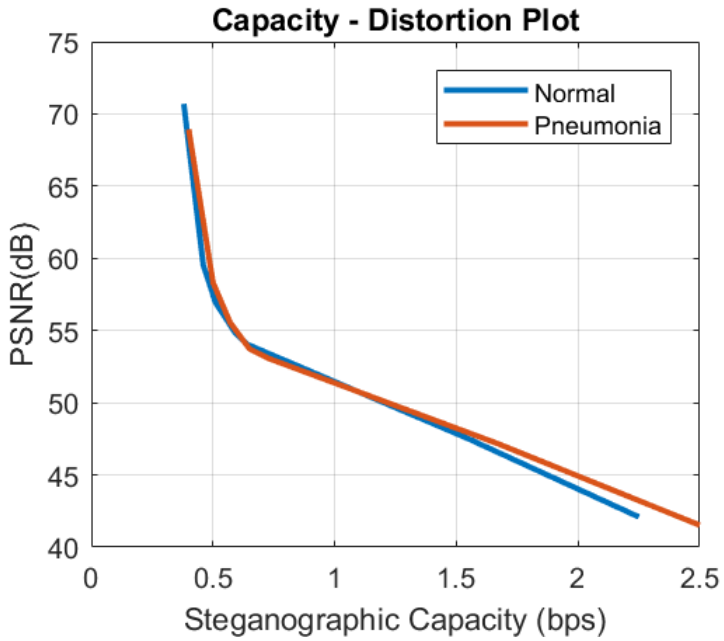


Figure 4.14: Steganographic Capacity for X-ray Pneomonia DataSet: *For both Normal and Pneumoina patients, distortion increases (shown by decrease in PSNR) as the watermark payload increases.*

perceptibility. All PSNR values are above 40 dB, as recommended by [29, 49]. This result shows that this method is very effective for 16-bit DICOM images and does not degrade the image significantly, visually and statistically.

Figure 4.16 extended this concept to multiuser (n = number of users) CDMA embedding as described in Section 2.3.6. It shows the degradation introduced by the CDMA concept in combination with our extended C_4S. The worst-case result for five users using ***3-to-1-bitlength*** compression (that is, 3 bits per user code sequence) gives average PSNR of 69.5 dB for 16-bit DICOM images. This is very high performance as it shows that up to 15 bits (5 users x 3 bits per user) can be embedded in a sub-block without visual image degradation. The major challenge with CDMA, however, is that more distortion is introduced by multiple access interference. As more users embed watermarks either synchronously or asynchronously, the error at the retrieving end increases due to this interference. Also, distortion increases as the total signal amplitude or power from different users sum up. Hence, we have reported the boundary at which retrieval error

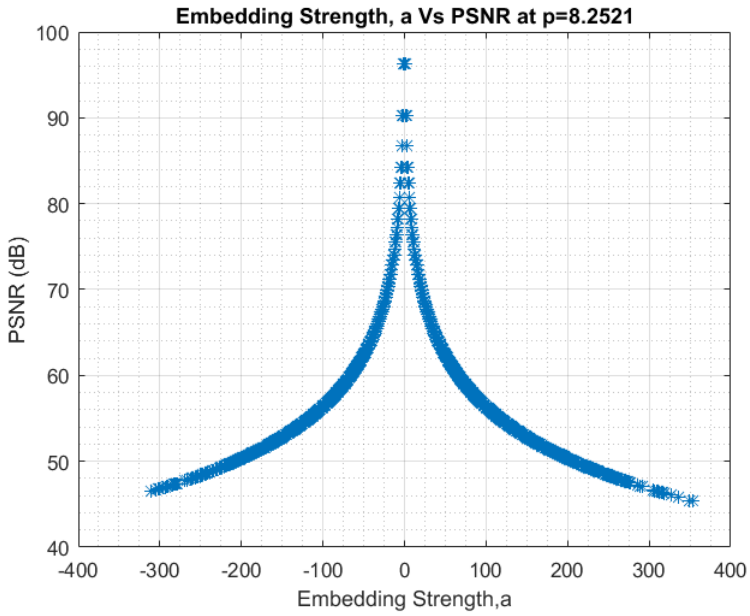


Figure 4.15: Local Statistical Degradation using PSNR:*In each sub-block, α can vary considerably. This introduces micro-level distortion which is directly proportional to α*

became significant.

We represented a character (xter) of text with 8 bits according to the American Standard Code for Information Interchange (ASCII). Hence, dividing the capacity in bits by 8 gives one the capacity of EMR characters that can be embedded. For example, in the X-ray dataset, an image of size 1280 x 960 embedded at $Cr = 2, 3, 4$ gave capacity in bits of 38, 000; 57,600 and 76,800 respectively. This translates to capacity in characters of 4,800; 7,200 and 9,600, respectively. These character capacities are enough for a patient's EMR during remote autodiagnosis.

4.8.2 Impact on Statistical and Visual Imperceptibility

Watermark detection accuracy and higher payload capacity should not significantly compromise imperceptibility of watermark or degrade the image visually and statistically. The degradation and security evaluation criteria are based on measurable quality parameters such as Peak Signal to Noise Ratio (PSNR), Bit Error Rate (BER), Struc-

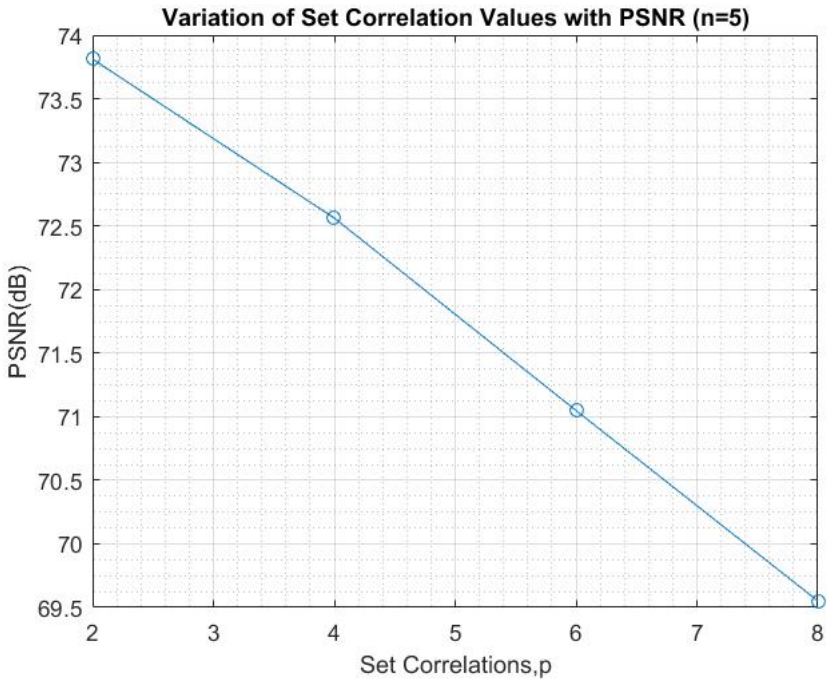


Figure 4.16: Multiple Access (CDMA), n=5, cr =3: *By leveraging CDMA, up to 15bps capacity can be achieved in a sub-block at minimal distortion for a 16-bit MRI DICOM image*

tural Similarity Index Measure (SSIM) and Kullback-Leibler divergence (KLD). KLD is a measure of security between two distributions under noisy conditions [100, 101], and it represents the change in image Probability density function (PDF) between the distributions of the original and watermarked image.

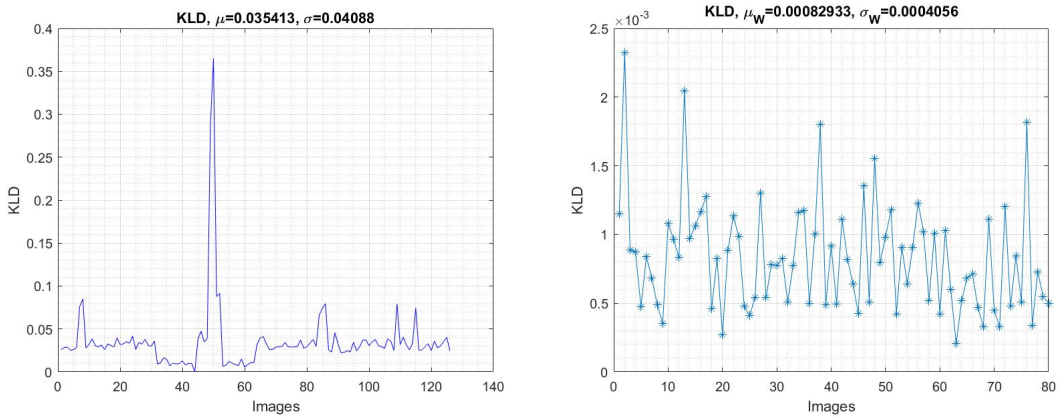
Whereas Table 4.7 presented the mean and Standard deviation (S.D) for KLD, Figures 4.17a, and 4.17b shows the values obtained for randomly sampled images within ADNI and ISIC datasets, respectively.

The lower the KLD, the better. With this in mind, it can be said that the ISIC dataset (*Ex vivo*) scans are more secure than ADNI and BRAINIX (*In vivo*). The change in PDF of the ISIC dataset is smaller than that of ADNI.

Figure 4.18 presents a plot of the distribution of the computed embedding strengths, α, for 153,600 sub-blocks. It shows that 55,300 (47,500 + 7800) ranges from 0 – 5 while 70, 000 (68000 + 2000) ranges from 5 – 10 (absolute values used). This implies that 81.58

Table 4.7: Average ROI Steganographic Parameters, NI = Natural Image: *This is the best-case scenario with 1-bit per pixel as in Chapter 3. Higher PSNR, higher SSIM and lower KLD gives the more imperceptible and more secure result*

DataSet	PSNR (dB)	SSIM	KLD
ADNI (MRI 8-bit)	41.2672	0.9841	0.0354(S. D = 0.0409)
ISIC(NI 8-bit)	45.45	0.9977	8.2933e-04(S.D = 4.0560e-04)
Pneumonia(X-ray)	59.53	0.9986	3.4e-02(S.D = 1.340e-03)
OSIRIX (MRI 16-bit)	100.89	0.9999	5.42e-02(S.D = 1.340e-02)



(a) ADNI KLD (b) ISIC KLD

Figure 4.17: KLD Plots

percent of the sub-blocks might be modified by at most 3-bitplanes (2^3 = 8, nearest to 10) out of the 16-bitplanes of a 16-bit DICOM MRI image.

4.8.3 Comparisons with State-of-the-art

Comparison is made between C_4S method and similar methods within the SS domain and concerning desirable properties in medical image Steganography. For a fair comparison, we consider authors who used SS methods as well as high-pixel depth (between 12 to 16 bits per pixel) DICOM images. These comparisons are presented in Table 4.8. To provide a trade-off between capacity and image quality, Figure 4.19 shows the percentage quality and capacity for some comparable authors.

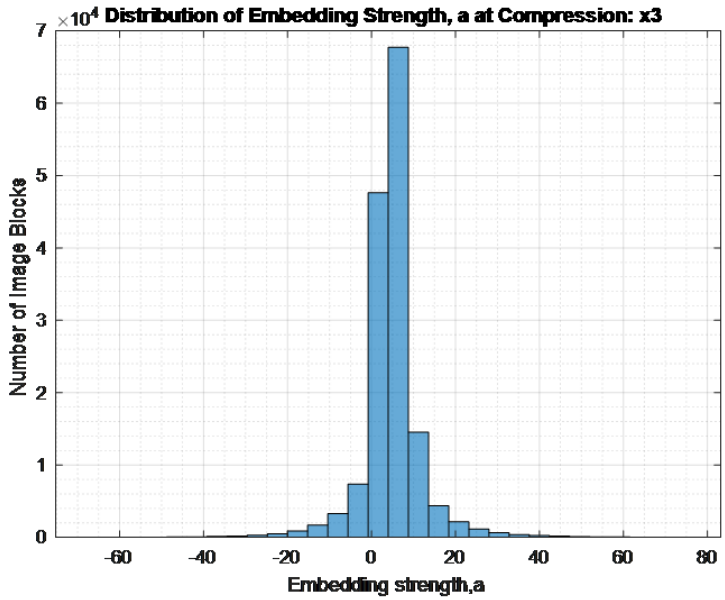


Figure 4.18: Distribution of Embedding Strength

Kumar et al. [88] used DWT spread spectrum watermarking and achieved a maximum capacity of 3,200 bits. This was inferred from the largest binarised logo watermark of size 40×80.

The accuracy of C_4S is better than the results obtained by researchers in [167, 161, 88] and [163]. These works used either traditional SS, Correlation-aware or other SS-related methods. They carried out their research in either spatial or transform (DCT and DWT mainly) domain watermarking. A better algorithm is determined by a combination of **higher Capacity**, **higher PSNR** and **lower BER**.

From a security perspective, C_4S is claimed to be more secure than that of Maity & Maity [101]. The average KLD of our result for MRI is lower than the lowest KLD they obtained for X-ray using their algorithm. This work supports the initial claim that SS steganography would offer higher security by spreading information instead of concentrating them in a single bit.

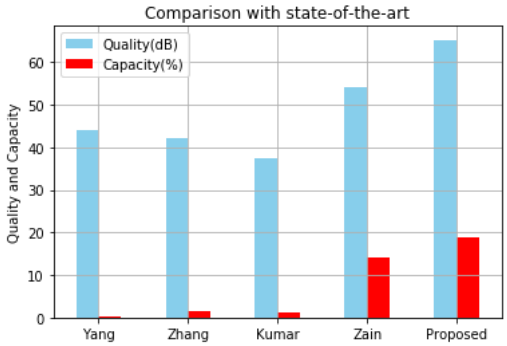


Figure 4.19: Comparison in terms of capacity-distortion trade-off among existing algorithms. Algorithms with higher capacity should also meet the quality requirements for medical diagnosis. Only algorithms with similar spread spectrum approach were compared.

Table 4.8: Comparison with existing SS-based methods for 8-bit X-ray Images: *Capacity is not measured just by the number of bits embedded but also by quality of image after embedding and the BER after extraction, bps = bits per sample, TD=Tamper Detection*

Authors	Capacity[a] (%)	PSNR (dB)	Zero BER?	TD?
Yang *et al*[161]	0.39	44.07	Yes	No
Zhang *et al* [167]	1.56	42.00	No	No
Kumar *et al* [88]	1.22	37.52	No	No
Zain *et al* [164]	14.06	54.15	No	Yes
Proposed Method	18.75	65.19	Yes	Yes

[a]Capacity has been normalised to the percentage of an image instead of the usual unit of bps or bpp. This is because the size of image used for inter and intra-study comparisons differed

4.9 Key Findings

The theoretical and empirical results obtained in this chapter led to the following major findings:

1. Our extended C_4S method can provide Steganographic capacity of as much as 12 bits in a block for 16-bit image. This increases the amount of embedded EMR record with less distortion (See Figure 4.13).

2. In the literature [160, 48], the higher the complexity of an image, the higher the

Table 4.9: Comparison with existing SS-based methods for 16-bit DICOM MRI, bps = bits per sample, TD=Tamper Detection

Authors	Capacity (bps)	PSNR (dB)	SSIM	Zero BER?	TD?
Qasim et al [132]	1	99.94	1	Yes	
Proposed Method	8	100.89	0.9999	Yes	Yes

number of bits that could be embedded without visual degradation. On the contrary, we found that the higher the image complexity, the higher the HSI and thus higher error during retrieval using a spreading sequence. Hence, if one must consider the embedding and the extraction sides' capacity, only sequences and embedding strategies that eliminate HSI at the extraction side can truly claim the increase in capacity with image quality. This is because the embedding strength, α required to cancel noise, will also be low. This finding is in line with the theoretical results in Barni *et al* [13].

3. *In vivo* (ADNI and BRAINIX) and high pixel depth (BRAINIX) medical images have better performance in terms of capacity and distortion than *Ex vivo* coloured and visible light medical image photographs (ISIC). This comparison is made at the grayscale (Luma channel of coloured images) and not in the coloured domain.

4. Finally, about 24.09 % (13.25% from beginning and 10.84% from ending) of medical image slices in a DICOM volume is available for watermark insertion. This observation is for the OSIRIX dataset and thus may vary for different scans. However, extra capacity can be achieved in these regions.

4.10 Discussion

Spreading codes, error correction codes, and the wide-band nature of SS steganography reduce its information hiding capacity when compared with other embedding methods that utilise few pixels or transform coefficients. The purpose of this chapter was to find efficient ways to add more bits of watermark information per sample of the image

instead of embedding just watermark signature or a single bit of information as in the traditional SS methods [38, 40, 104].

In addition to the above constraints for improving steganographic capacity, a more definite constraint for medical image steganography is distortion. As higher embedding strength is required to withstand image noise and attacks in SS, the resultant effect is degradation in image content. For an image-only diagnostic process, where all information is encoded in the image and sent to a third-party, increasing the quantity of information embedded is dependent on the imperceptibility (distortion) requirement of the application.

In order to achieve a higher capacity per sample of an image, we have adopted several methods, including message source coding, multi-channel embedding, and code division, multiple access to achieve this objective.

Undoubtedly, this approach increased the complexity of the steganographic scheme. If 2^N codes are used to achieve N-bits per sample, 2^N searches per sample may likely be performed at the receiver to retrieve a set of correct bit in a sample. To reduce the number of searches and comparisons during decoding, we used the concept of **zoning** of sequences as part of the side information at the encoder and decoder. This concept means that for each transmission, the sender and the receiver already know how to localise the search for a sequence into a given zone. For a code set of (M+1)-sequences, four zones are created with $\frac{M+1}{4}$ sequences in each zone. To increase the number of bits transmitted per sample as well as decrease computation time, we employed a combination of a reduced number of sequences and channel coding. For example, to transmit 5-bits per sample, 2^5 code sequences are needed at the encoder, and the same number of searches or matching is needed at the receiver. In the extended C_4S, we have reduced this drastically using constant correlation channel selection. To transmit at the same rate but reduce computation, use 2^3 code sequences and 2^2 channels. This means that the first 3-bits are assigned to 8 code sequences in one zone, and the remaining two bits are assigned to 4 channels. At the decoder, only eight computations and four comparisons at the maximum, are made. Once a match is made with the pre-defined channel

(that is a pre-defined constant correlation), the 3 bits mapped to that particular code, and the 2 bits mapped to the channel are retrieved. Hence, instead of 32 computations, only 12 are being made using C_4S. As a generalisation:

Theorem 4.1. *In order to reduce* 2^N *computations and code sequences to only* $2^m + 2^{(N-m)}$ *computations and comparisons, where* $m < N$*, use only* 2^m *code sequences and* $2^{(N-m)}$ *correlation (encoding) channels for encoding.*

The reduction in computational time also affects the distortion. The more the correlation channels we have, the more local distortion that is introduced in the image. Thus, increasing capacity through the creation of more correlation channels is limited to 2 bits per channel for 8-bit images and up 6 bits per channel for 16-bit images.

Although our capacity-scaling algorithm introduces some level of computational complexity, this increase in complexity is not a significant challenge in a hospital where batch upload is not a requirement. Thus the full capacity of the algorithm could be used for 16-bit DICOM images to achieve up to 6 bits per spreading code (that is $2^6 = 64$ codes) + 6 bits per channel($2^6 = 64$ channels) giving a maximum capacity of 12 bits per sample (8x8 block sample in our research). This capacity would have required $2^{12} = 4096$ linear correlation computations and comparisons but will now require $2^6 = 64$ computations + $2^6 = 64$ comparisons. Hence, even if the cost of computing linear correlation and cost of comparing two numbers are the same, there will still be a reduction in the cost of about $4096/128 = 32$ times.

In our approach, where the ROI was not entirely avoided, like in [127], the only way to minimise distortion is to control both capacity and robustness. This control was achieved in this research through content-adaptive embedding, where the level of allowable distortion controls the decision to embed and the robustness of embedding. Horizontal (use of more spreading codes instead of using more embedding channels) encoding schemes that do not require increasing embedding strength can be used to extend capacity in the ROI without affecting distortion, provided that the cross-correlation of the sequence with the host data (HSI) is first determined.

As a recommendation, medical images that are used for information hiding should be created at higher pixel depths greater than eight. This ensures better steganographic qualities in terms of capacity, flexibility between robustness and fragility, and imperceptibility.

4.11 Summary

Steganographic capacity is the number of information bits that can be embedded and correctly extracted from a cover multimedia data subject to a distortion constraint on the original cover data. In a relatively recent research [126], computational complexity constraint has been included in this definition. However, this was not considered in this research as our algorithm is not considered a computationally intensive steganographic algorithm.

There are both theoretical and empirical frameworks for estimating the Steganographic capacity of images. Whereas the theoretical framework is largely based on information theory, the empirical approaches are more domain-specific and based on the specific embedding strategy. For practical purposes in medical image steganography, the empirical approach is used as it allows relevant distortion parameters and specific embedding strategies to be considered.

Figure 4.16 has shown that degradation for high pixel depth images is very low for up to five simultaneous users of the channel or data sources, each transmitting 3 bits per symbol. When the concept of CDMA is combined with constant correlation compression encoding, we can significantly extend the steganographic capacity of DICOM images.

In Chapter 5, we will go further to evaluate the effect of embedded watermark on medical images in terms of its use for medical image classification for remote autodiagnosis. The evaluations and synthesis in this chapter and the next chapter will lead to the development of medical image IH frameworks and protocols for telemedicine in Chapter 6.

Chapter 5

Effect of Information Hiding on Image Biomarkers

In this chapter, we evaluate the effect of steganography on computer-aided diagnosis based on statistical changes in medical image features known as biomarkers. Further, we use the result of a Support Vector Machine (SVM) classification of Chest X-ray scans of Normal and Pneumonia patients to ascertain the effect of steganography on classification performance. The aim is to quantify and compare the effect of watermark in ways similar to clinical trials employed in medicine. Section 5.2 provides the definitions and examples of quantifiable image biomarkers while Section 5.3 reviews related works in medical image Information Hiding (IH) evaluation via machine learning. It also reviews recent research that employed textural biomarkers for disease classification. Section 5.4.1 describes the evaluation dataset, while Section 5.4 gives the details of the experimental set up including the statistical profile of the original images and the machine learning evaluation parameters. Section 5.5 presents the results with their implications. Section 5.6 concludes the chapter with a summary of the research findings.

Go To Chapter: 1, 2, 3, 4, 6,7

5.1 Introduction

Automated diagnosis in teleradiology employs image features. These features and disease predictors are not considered while evaluating medical image information hiding (IH) algorithms. Medical image quality assessment during Steganography and water-

marking should include the changes in the image features(biomarkers) that directly affect the diagnostic outcome . Hiding schemes that produce high PSNR values may not always lead to the same diagnostic outcome as the original image [163].

This lack of focus on diagnostic outcomes is probably because the majority of current medical image data hiding algorithms have focused on the reversible and the RONI-only medical image watermarking strategies (see Section 2.4.3). These two approaches provide an easy mechanism to avoid any issues relating to distortion, the complexity of selecting a region of interest (ROI), variation in behaviour of different medical image modalities, and the concerns about the content-specific nature of different images. Whereas we have considered these approaches in this research, our focus in this chapter is to establish empirical facts in the troubled waters of medical image ROI. This study of medical image ROI is justified because it is the most valuable part of a medical image employed for diagnosis and monitoring. It is a known fact [8, 94] that apart from data hiding security techniques, medical images undergo other pre-processing and post-processing operations that also irreversibly affect the ROI. This situation has given few researchers (discussed in Section 5.3.2) the courage to evaluate, in some sense, the effect of security data embedding on diagnostic outcomes, especially in automated scenarios.

The increasing demand for autodiagnosis further highlights the need for a new evaluation technique for medical image steganography. Autodiagnosis involves the use of Artificial Intelligence (AI) and Machine Learning (ML) algorithms for medical diagnosis, treatment, and monitoring. This use of AI/ML has also led to an increase in the digital processing of medical images to increase its computer recognition characteristics [136]. These processes on the medical image scans could occur locally or locally or remotely, affect both the ROI and the RONI (Region of non-interest), and could be performed by either an authorised or unauthorised user. These modifications raise security concerns of which some (integrity and Privacy - the primary concern of this thesis) were addressed in Chapters 3 and 4, where we detected and localised image tampering using spread spectrum techniques. However, in this chapter, we focus on evaluating the im-

pact of the embedded message on disease classification models. This choice is because, in AI/ML systems, little feature changes could lead to a different prediction outcome. Hence, where there no human expert involved, the evaluation methods that relied on the human visual system will no longer be reliable for evaluation and validation. For this reason, evaluation techniques that relate to automated prediction systems such as support vector machines (SVM) are now a necessity. This evaluation is intended to increase the confidence of medical practitioners towards accepting Steganography and watermarking for achieving complementary or alternative layers of security in teleradiology.

Specifically, this chapter quantified the effect of watermarking on the selected textural image biomarkers: Contrast, Correlation, Energy, Homogeneity, and Entropy. Hence, along the line of capacity-distortion (diagnostic distortion) performance analysis, the following questions will be answered: *(i) how does our algorithm affect these selected biomarkers for a medical image? (ii) On what conditions could watermarking/Steganography become adversarial in Machine learning models?*

Both statistical and machine learning approaches were used to answer these questions. The result shows that decisions about IH are strongly related to specific biomarkers being considered. Whereas the embedded watermark bits has little effect on the entropy, the reverse is the case for the contrast. We found out that a balance between robustness and sensitivity is needed. For example, the contrast feature is susceptible to changes and may quickly raise a false alarm, though it generally gives high differentiation between disease classes. On the other hand, the entropy is resilient to little content perturbations and will be a more stable predictor in the noisy or malicious environment. How these predictors' behaviours vary with the parameters of our C_4S algorithms will be studied in this chapter. In the next section, we will define medical image biomarkers and provide some examples too.

5.2 Image Biomarkers

Biomarkers are objectively measured characteristics of biological processes, pathogenic processes, or the pharmacological response to medical treatment [91]. They are often combined in various ways to diagnose or monitor a disease or condition. Biomarkers are generally classified based on how and where they are measured. There are radiological, molecular, histological, and physiological features that could be measured as biomarkers [41]. Radiological characteristics are measured from non-invasive radiological images. These features from medical images are also called *image biomarkers* [91, 41]. In general, medical procedures combine image biomarkers with other biomarkers measured from other parts of the body for diagnosis and monitoring, including during clinical trial of new drugs [41]. Figure 5.1 illustrates this combination.

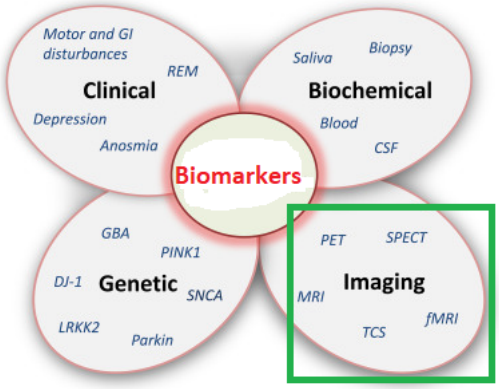


Figure 5.1: Biomarkers: *Biomarkers could be obtained in various ways. In this thesis we are concerned with radiological biomarkers obtained from medical images.*

The importance of image biomarkers in diagnosis can be illustrated with brain image biomarkers. They help in detecting dementia and in assessing the effectiveness of brain tumour therapies [41, 51]. Specifically, Magnetic resonance images (MRI) of the brain are often utilised in this process. Some of the biomarkers measured from brain images include Cerebral Blood Volume (CBV), Microvascular Blood Volume (MBV), and Vessel Caliber Index (VCI). Others are transverse relaxation time (T2) and Apparent Diffusion Coefficient (ADC) [51]. These measurements are often carried out in the

hippocampus of the brain, which is located in the temporal lobe. The primary function of the hippocampus is that of memory and human learning processes. It is known to play a substantial role in the human memory system.

Damage to parts of the limbic system (part of the brain that is important for motivation and emotion) negatively affects the hippocampus and could lead to its increase or decrease in volume. This effect, in turn, could reduce the ability of the brain to store and retrieve consciously stored information [51]. This consequence of a change in volume suggests that a measurement of this change in volume (shrinkage or expansion) of the hippocampus is an example of a brain image biomarker that can be used for diagnosis and monitoring.

Researchers [44, 153] have shown that different kinds of diseases affect the ability of the hippocampus to perform its functions adequately. These impairments manifest in the shrinkage of the standard size of the hippocampus in an individual. The diseases or conditions that affect the size and volume of the hippocampus include Dementia of all types, especially Alzheimer's, Epilepsy, Depression and Stress, Schizophrenia, Bipolar disorders [7, 44, 153], among others. Hence, the quantification of volume, surface area, and thickness of the entire cortical region and its associated regions can help to diagnose and monitor the diseases and conditions mentioned above. The measurement of cortical volume and thickness has been used in medicine to estimate the effects of aging, cognitive abilities, and mental diseases. The quantification is performed after using appropriate methods to segment the brain images into White Matter (WM), Grey matter (GM), and Cerebrospinal Fluid (CSF) [153].

Whereas the size and volume of the hippocampus is not the only indicator for the above diseases, it has been shown by various researchers as reliable biomarkers for these diseases. The work of [153] used the volumes of hippocampus left and right to differentiate among different types of Dementia. Patients with Alzheimer's disease (AD) had up to 25% reduction in hippocampal volume compared with the control individuals. The average hippocampal volume reduction for patients with vascular dementia was recorded as 11per cent, while patients with mixed dementia showed up to 21%

hippocampal volume reduction. It was only a 5% reduction for those with normal pressure hydrocephalus (NPH). According to [44], there is an average of 10% hippocampal volume reduction in people with severe depression than without depression. In [7], the entire cortical volume reduction, surface area, and thickness were used to demonstrate the various stages of Schizophrenia and bipolar disorders. Similarly, it will be useful to evaluate if the inserted watermark bits change these volumes in any way that causes misclassification of the disease stage.

Other image biomarkers surprisingly share boundaries between ordinary images and medical images. These are the **textural features of an image**. For ease of generalisation, we use this set of biomarkers as they are relatively common among medical informatics researchers and core medical experts, especially for machine learning-based disease classification [83, 150]. We will discuss these biomarkers further in Section 5.3.3. Any distortion in these image biomarkers is, therefore, believed to be more relevant to medical practice than the classic image quality assessment (IQA) such as PSNR, SSIM, and KLD (as applied in Chapters 3 and 4). In the next section, we examine current advances in evaluating IH algorithms in terms of diagnostic information loss.

5.3 Data Hiding and Diagnostic Image Quality Evaluation

This review section presents current practice by medical experts for assessing image quality. It goes further to show the limited work in evaluating the effect of IH techniques on image biomarkers used for AI-based diagnosis. Also, it looks at various works that utilised textural biomarkers for classification of diseases in both animals and humans. It concludes with the justification for further evaluation of the effects of watermarking on textural image features in current applications of machine learning algorithms for disease classification.

Table 5.1: Image quality outcomes in Radiology [94]: *The outcomes are subjective in nature even though they have been represented by objective scaling. The objective rating could be arrived at by a doctor's grading or through any of the reliable HVS-based quality parameters used in image processing.*

Scale	Class	Subjective Interpretation
1	**Excellent**	No limitations for Clinical Use
2	**Good**:	Minimal limitations for Clinical Use
3	**Sufficient**	Moderate limitations, no substantial loss of information
4	**Restricted**	relevant limitations, clear loss of information
5	**Poor**	not usable, loss of information, must be repeated.

5.3.1 Current Practice for Medical Image Evaluation

Medical image quality evaluation methods can be classified into: ***physical (Signal processing)***, and ***Observer performance*** methods [94]. The physical or signal processing methods are based on a measure of the difference between the two images. It is popular in image compression and other signal processing such as denoising, contrast adjustment, and image filtering. Image quality parameters such as Peak Signal-to-Noise ratio (PSNR), SSIM, Root Mean Square Error (RMSE), and KLD (see Section 2.6.2) are employed in this type of quality assessment. These parameters can be used to estimate the level of distortion or perceptual degradation introduced into an image. The physical evaluation method is only a preliminary evaluation method in medicine and radiology. Therefore, this nature of the evaluation calls for new methods that are closer to radiology practice. However, when parameters that are highly correlated with the human visual system (HVS) are employed, physical measures are reliable for making subjective diagnostic decisions. This latter case is not often applicable to autodiagnosis, where HVS is not utilised for diagnosis but rather, computer vision. This situation leads to the quest for an evaluation method that could be objectively applied to both HVS and computer vision. The second class - Observer performance methods - utilise objective data to make a subjective decision. Table 5.1 shows the way radiologists classify images from different machines or compare images before and after pre/post-processing.

In current literature ([94, 131]), the Visual Grading Analysis (relative and absolute) is considered the most useful observer performance method. Visual Grading Analysis (VGA) is popularly used in radiology for analysing the reproducibility of anatomical and pathological structures. It is a perceptual method of determining the allowable distortion for a given level of medical image modification[131]. An extended analysis known as Visual Grading Characteristics (VGC) analysis is now considered superior to simple VGA. VGC generates a Receiver Operating Characteristics (ROC) curve based on the subjective assessments from observers, subject to a **specific criterion** [94]. The specific criteria will have to be defined based on the disease condition being diagnosed. VGC studies also use the **ANOVA analysis** (see Section 5.4.2) to compute confidence intervals and *p-values* and serves as a bridge between a completely subjective method and a completely objective method. Hence, IH algorithm designers need to embrace the second method of image quality assessment already in use by medical practitioners. This reason is the motivation for the contribution of this chapter.

In the next section, we will examine existing works that have attempted to evaluate the effect of IH on diagnosis based on specific criteria relevant to a disease condition.

5.3.2 Effect of Watermarks on Diagnosis

Though lots of works have been done on Medical Image watermarking (MIW) and Steganography security techniques [108, 155, 134, 100, 164], the evaluation of the effect of watermarking on diagnosis by the use of image biomarkers seems to be under-researched, especially as regards machine learning models. Existing evaluations by data hiding algorithms designers are limited mainly to the physical methods such as PSNR, Mean Square Error (MSE), Structural Similarity Index Measure (KLD), Kullback-Leibler Divergence (KLD) and other measures of distance used in DSP for images. Predictors and features used in teleradiology and machine-based autodiagnosis are not usually considered; therefore, medical practitioners are not entirely convinced to apply these security techniques. In this section, we highlight a few of the works that exist in

this under-researched area, and in our discussion, we stress why more objective evaluation with diagnostic biomarkers will enhance adoption in mainstream medical data security.

G. Coatrieux *et al* in [35] made a case for why we should not rule out MIW. They described how watermarking could bring evidence in the case of telemedicine litigation through the use of MIW. In this technique, the watermark should not be removed from the medical image, and even many should not be aware of its existence in the first place. They quickly added that it remains a challenge to convince both the legal and medical practitioners that watermarking and Steganography can provide trusted evidence without compromising diagnosis. The difficulty arises from the stringent constraints generally placed on the integrity of data used for medical, military, and legal applications [163]. This situation justifies further research by [163] on digital watermarking techniques that may not violate this stringent constraint. The shortcoming of the research in [163] is that it is subjective, and only a few experts were involved. The criteria used for evaluation do not apply to machine learning algorithms.

J.J. Garcia-Hernandez *et al* in [54] performed an objective evaluation of the impact of watermarking on computer-aided diagnosis in medical imaging. They used about 500 Breast Ultrasound as their dataset; two watermarking algorithms applied on each half of the samples. The evaluation parameters included PSNR, Watson Parameter, and bits-per-pixel. The authors attempted to establish the effect of spread Spectrum DCT (SS-DCT) and High Capacity Data Hiding (HCDH) watermarking on the segmentation and classification accuracy of the lesions in the image. They found out that with an appropriate choice of parameters, both watermarking systems can perform well without any adverse effect on segmentation and classification accuracy. However, SS-DCT could alter the accuracy if high embedding strength is used. Their approach is most related to our work but not in the area of MRI images and pneumonia disease.

In a recent study by [66], some watermarked Fundus eye scans were tested against some models [69, 68, 67, 70] used to classify some Healthy, Macular Edema and Central Serous Chorioretinopathy (CSCR) eyes diseases. The original accuracy of their models

ranged from 95% to 100%. Their results show that there was no difference in classification accuracy for the original and watermarked test set. However, few test data were used (15 to 45). Also, it was not clear if the original model was trained as well with the watermarked training set. Again, this study failed to recognise that if watermarking is adopted for integrity checks, future training sets would contain watermarked data and not just the test data.

Hence, our research intends to fill these gaps by performing extensive evaluation using both statistical and machine learning tools. Specifically, we have explored large data sets for Skin cancer and Pneumonia to study the effect of IH (both digital watermarking and steganography) on medical diagnosis. In the next section, we will discuss the textural image features that we employed as the image biomarkers of interest in this work.

5.3.3 Textural Features as Image Biomarkers

Texture-specific features have been used in several image analysis to study the health of patients. The study by Kim *et al* in [83] used heterogeneity and entropy texture analysis to study the survival outcomes for patients with breast cancer. They established a relationship between the level of entropy and heterogeneity of T2-weighted images and the recurrence-free survival (RFS) rate of patients.

Among other image features, textural image biomarkers were employed by Theek [150] to automatically differentiate among the three types of mouse xenograft tumour models. The used a quantitative approach to achieve this. Up to thirty textural image biomarkers obtained from contrast-enhanced ultrasound scans were used.

Xu [159] extracted 96 features, including 52 features of the gray-level co-occurrence matrix (GLCM) and 44 features of the gray-level run-length matrix (GLRLM) from the ROIs of ultrasound images. Then they used these features to train and classify two liver diseases of hepatocellular carcinoma and liver abscess. Their SVM achieved 88.88% classification accuracy.

In this chapter, we consider the five textural biomarkers that are commonly used in literature [83, 150, 159] for medical image classification for computer-aided diagnosis. They include Energy, Contrast, Correlation, Homogeneity, and Entropy. These are five out of the fourteen Haralick et al. [64] texture feature operators from the Gray-Level Co-Occurrence Matrix (GLCM) of the ROI image [33, 121]. These predictors, which are considered closer to the medical profession than the PSNR and SSIM image quality parameters, could be used for both content and diagnostic information integrity checks.

The mathematical definition of these features has been previously described in Sections 3.6.1 to 3.6.5. In the next section, we will introduce the evaluation method and the experimental procedure.

5.4 Evaluation Method and Experimental Set up

This section is divided into statistical tests set up for performing statistical tests and analysis, and then the set up for SVM model building and evaluation. As we stated at the beginning of this thesis, all models and analyses are based on the existing 40dB benchmark for image distortion [29, 49]. Hence, the number of bits per sample (capacity) was increased from $cr = 1, 2, 3, \ldots$ until visible degradation is noticed, or computed PSNR between original and watermarked image hits 40dB on the average among all the samples in each dataset class. The embedding strength, α, and base correlation values, ρ, were controlled while observing capacity. The dynamic embedding strength, α is the primary parameter that determines both accuracy and distortion. The bounds for this parameter in relation to the algorithms developed in Chapters 3 for ROI and Chapter 4 are established here. Before delving into the main parts of our evaluation method, we will first present the dataset in the next sub-section.

5.4.1 Dataset

The **Chest Xray** dataset has been described in Section 3.6.6. This dataset is suitable for this study because it has the normal images and the pneumonia images as separate

train, test, and validation datasets. Hence, it is suitable for disease profiling, statistical, and machine learning classification for autodiagnosis.

Table 5.2 shows the sample size and classes of the dataset used in this experiment. A large size was sought since a machine learning model will be produced. Our pilot study and statistical power analysis did not require such a large dataset to detect an effect size of 0.01(but 0.5 in entropy because of its higher dynamic range) in a feature. However, the machine learning algorithm does require a large data set. We then leveraged this constraint to perform a high-powered statistical analysis that can detect tiny effect size, as shown in Table 5.4. It is believed that such a robust statistical test would enable the medical community to trust and establish acceptable parameterised limits for medical image security data embedding based on specific image biomarkers.

Table 5.2: Divisions of Data Set: *92.48% for Training and 7.52% for test set*

Dataset	**Training Set**	**Test set**
Chest Xray(Normal)	1241	101
Chest Xray(Pneumonia)	1241	101

There was a two-part evaluation method. One followed the statistical methods of hypothesis testing, while the second is based on observing the changes in performance parameters (accuracy, specificity, precision, and recall) for the SVM classification of Pneumonia using a Chest-X-ray scan.

Each of the dataset class contained 1241 images that were treated (watermarked). Hence, per dataset, 2482 images were used in the evaluation. The original(non-watermarked) normal and Pneumonia sub-classes constitute the **control** group, while their watermarked versions at various levels of embedding strength are the **treated** group. These do not include the 202 images set aside for test set in the case of SVM models. The test set images were not watermarked.

5.4.2 Statistical Tests Setup

The statistical significance test is used to determine the extent to which a claim (hypotheses) can be rejected or accepted. Specifically, the one-way analysis of variance (ANOVA) test was performed to analyse the level of difference between the means of the selected features for original and watermarked images. One-way ANOVA is appropriate in this research because biomarkers are computed differently and considered on their merits before their possible combined effects are considered. The listed textural biomarkers were tested individually. Formally, the null and alternative hypotheses are given as follows.

Definition 5.1. *The null hypothesis,* H_0*:* ***there is no textural difference between original and watermarked image****. That is:* $\mu_o = \mu_w$.

Alternate hypothesis, H_1*:* ***there is a textural difference between original and watermarked image*** *. That is:* $\mu_o \neq \mu_w$.

In this definition, μ_o is the mean of each mentioned feature for original image samples, and μ_w is the mean of any of the mentioned features for watermarked versions of the image samples.

Table 5.3 shows the feature profile of the Chest X-ray image classes (normal and pneumonia) before treatment via information hiding. The computation of these features for each image ROI has been described in Section 3.6. The mean for each of the feature as shown in Table 5.3 was computed using Equation 5.1.

$$\mu_o = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n} \tag{5.1}$$

Where a is any of the considered biomarkers (Homogeneity, Contrast, Entropy, Energy or Correlation) and n is the total number of samples in each test group.

The major factor that affects the difference in mean between the original and watermarked image is the embedding strength,α. However, in our C_4S algorithms in Chapters 3 and 4, the α is determined by the number of bits to be embedded per block,cr as

well as the existing host signal interference (HSI) of the sub-block into which the set of bits will be embedded. This value is, in turn, determined by the correlation value: $\rho \pm n\epsilon$. Nevertheless, α remains the standard distortion parameter for all watermarking and Steganographic algorithm. So even though we vary $\rho \pm n\epsilon$ in this thesis, we compute the average and maximum value of the resultant α. Thus, for fair comparison, either cr or α is considered the ***independent*** variable while the means of the five features(C_t, C_r, C_e, C_h and H) were the ***dependent variables*** being studied.

Table 5.3: Original mean(μ_o) values of the Chest X-ray features before embedding: Δ *is the mean difference while* $\%\Delta$ *is the percentage mean difference between the Pneumonia(P) and Normal(N) feature values.*

Feature	Normal(μ_{oN})	Pneumonia(μ_{oP})	$\Delta = \mu_{oP} - \mu_{oN}$	$\%\Delta$
Contrast	0.0985	0.066	-0.0325	-32.99
Correlation	0.9678	0.9718	0.0040	0.41
Energy	0.1824	0.2439	0.0615	33.78
Homogeneity	0.9511	0.9671	0.016	1.68
Entropy	7.2547	6.9296	-0.3251	-4.48

In Table 5.3, any difference in features is attributed to the fact that the person now has pneumonia instead of being normal. This table suggests that Pneumonia does not have a significant effect on the correlation and homogeneity of the medical image ROI. They seem not to be useful biomarkers to distinguish between healthy (normal) patients and pneumonia patients. However, the variance of the dataset may lead to a different outcome. For entropy, contrast, and energy, the percentage difference in mean between the biomarkers for both experimental groups are higher and all above or around the 5% that is normally used as a significance level (α) in statistical significance testing. However, the variance and bias of the datasets will determine the resultant effect of group treatment and classification accuracy.

Figure 5.2 shows the scatter diagrams for Contrast and Correlation of both classes. Contrast clearly shows less overlap in both classes than correlation. However, it should be noted that our concern is not necessarily the highest accuracy but on how water-

marking would affect any accuracy that is initially obtained without embedding watermarks.

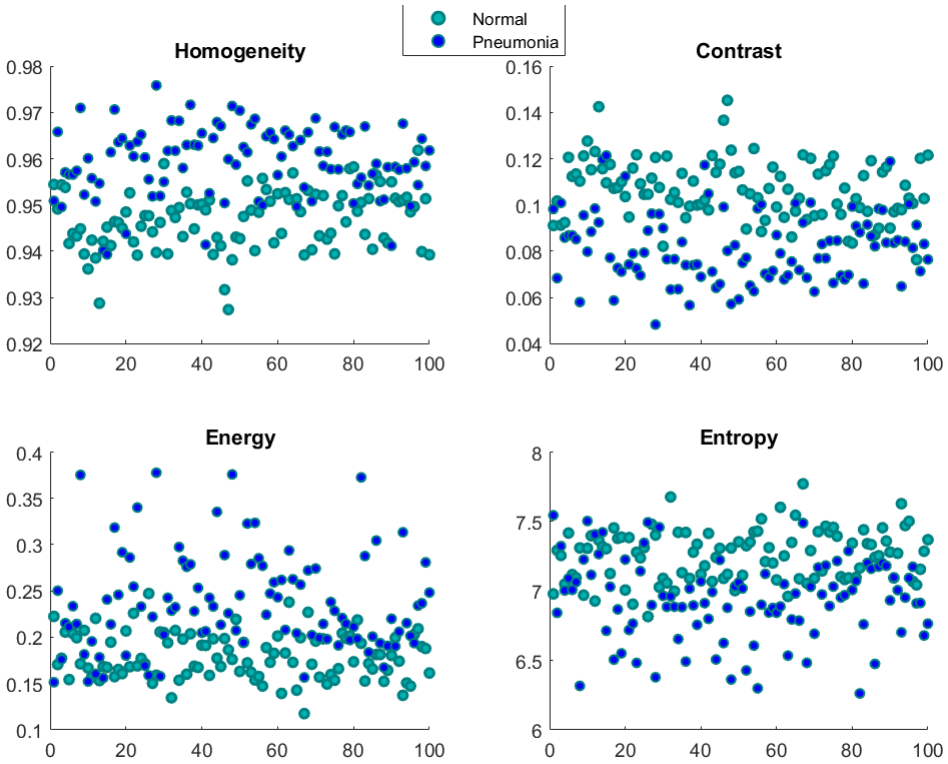
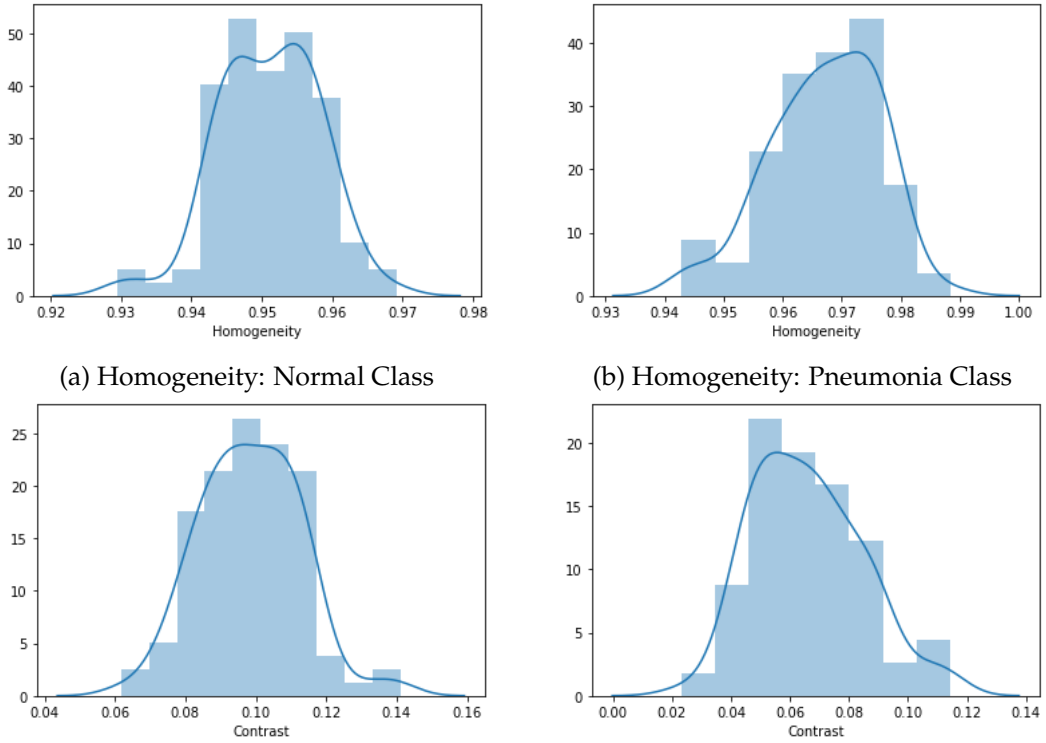


Figure 5.2: Scatter diagram for the selected features (Biomarkers): *There is a considerable separation between the classes. We did not employ any preprocessing to ensure that IH is the only alteration to the X-ray scans. Pre-processing operations could further separate the data points in each of the classes. Only the first 100 data points are shown in this plot.*

To validate the use of the statistical testing parameters in this chapter, Figure 5.3 show the approximately normal distribution of the datasets.

With sample size of 1241 per group (normal or pneumonia for X-ray dataset), significance level(α) of 0.05 used to control false positive (Type I error) in the textural difference between original (control) and watermarked/attacked image (treated), and a statistical power of 0.9, our study shows that we can detect *effect sizes* shown in Table 5.4. The effect size computation is based on the equation:



(a) Homogeneity: Normal Class

(b) Homogeneity: Pneumonia Class

(c) Contrast: Normal Class

(d) Contrast: Pneumonia Class

Figure 5.3: Normal Distribution Estimation: *Contrast and Homogeneity Features*

$$ES_{\mu} = \frac{\mu_{oN} - \mu_{oP}}{\sqrt{\frac{\sigma_{oN}^2 + \sigma_{oP}^2}{2}}}, \tag{5.2}$$

where ***ES*** stands for effect size and σ^2 is the variance of data points in each of the classes. The variance (and standard deviation) of each feature were computed from the 1241 ROI for each class of the dataset and substituted into Equation 5.2.

This observation means that our sample size can detect any effect size in all the features of at least 4.85%. This percentage is enough as our significance level is set to 5%. In simple terms, any change by IH that is less than or greater than the feature of the original image by 5% is considered an insignificant change.

The above choice of parameters has shown why correlation cannot be considered for most of our experiments. In Table 5.3, the mean percentage change for correlation

is 0.47%. In Table 5.4, its minimum detectable effect size is 3.40%. It means that the significance test cannot detect the maximum effect. This behaviour is another pointer that correlation may not be a good biomarker to distinguish between normal and pneumonia patient as pneumonia condition seems not to affect this feature significantly. Our SVM classification performance using correlation would help us conclude on this.

Table 5.4: Minimum detectable effect sizes for our statistical analysis: *The data in this table is for reproducibility and comparison of this and future studies. The $\%\Delta\mu_o$ determines what percent of change that our statistical test can detect. This is increasingly being required in medical statistics.*

Feature	Normal($ES\mu_{oN}$)	$\%\Delta\mu_{oN}$	Pneumonia($ES\mu_{oP}$)	$\%\Delta\mu_{oP}$
Contrast	0.0033	3.16	0.0030	4.85
Correlation	0.0012	1.23	0.0021	3.40
Energy	0.0038	2.23	0.0096	3.43
Homogeneity	0.0017	0.18	0.0019	0.20
Entropy	0.0250	0.34	0.0450	0.67

The power analysis for this study was performed using the Biomath [1] tool from the Department of Pediatrics at the University of Columbia Medical Centre.

5.4.3 SVM Classification Setup

This subsection aims to evaluate the effect of IH on Machine Learning/AI image classification algorithms. As diagnostic processes and medical image processing now use machine learning algorithms based on Support Vector Machines(SVM) and Deep Learning algorithms, it is essential to evaluate how the addition of watermark affects the trained model in its disease prediction and classification using test sets.

SVM is a machine learning algorithm designed to solve a classification problem through a kernel-empowered flexible representation of class boundaries [159]. SVM is popularly used because it can be easily implemented; it has excellent performance on a variety of problems with a little tuning, and it automatically controls complexity to

[1] http://www.biomath.info/power/index.html

reduce overfitting [103].

SVM supports binary(only two classes) classification where each subject is either in the ***positive*** or ***negative*** class. For this study, the positive class is pneumonia, while the negative class is normal. This classification by an SVM is performed through a *kernel function*, Φ, that can map the training examples, x_i into a higher dimensional space. The input data to a classifier are of the form: $Datapoints \implies X_i = [x_1, x_2, x_3, \ldots, x_n]$ with $x_i \in \mathbb{R}^n$. The **labels**, especially for binary classification is such that: $Y_i = y_1, y_2, y_3, \ldots, y_n$ for $y_i \in \{-1, 1\}$. Let ***d*** be the number of feature sets while ***n*** is the number of training points or data points.

For a training example x_i with a corresponding label y_i using ***d*** feature sets, the general form of a classification kernel function is given as [103]:

$$K(x_i, y_i) = \Phi(x_i)^T \Phi(y_i) \tag{5.3}$$

Four types of kernel are in common use:

1. **Linear Kernel**: $K(x_i, y_i) = x_i^T y_i$

2. **Polynomia Kernel**: $K(x_i, y_i) = (x_i, y_i)^d$. Where d is the degree of the polynomial.

3. **Gaussian Kernel or Radial basis function (rbf)**: $K(x_i, y_i) = exp(\frac{||x_i - y_i||^2}{2\sigma^2})$. where σ stands for a window width.

4. **Sigmoid kernel**: $K(x_i, y_i) = tanh(k(x_i y_i) + \varphi)$. Where k and φ are some kernel parameters.

The **Classification rule**: sign(f(x)) where :

$$f(x) = y[w^T x + b] \tag{5.4}$$

Where $w = Weightvector, b = bias$. For perfectly separable cases:

$$f(x) = \begin{cases} \mathbf{w}x + b > 0, & y = 1 \\ \mathbf{w}x + b < 0, & y = -1 \end{cases} \tag{5.5}$$

The classification equation of (5.5) is further visualised in Figure 5.4.

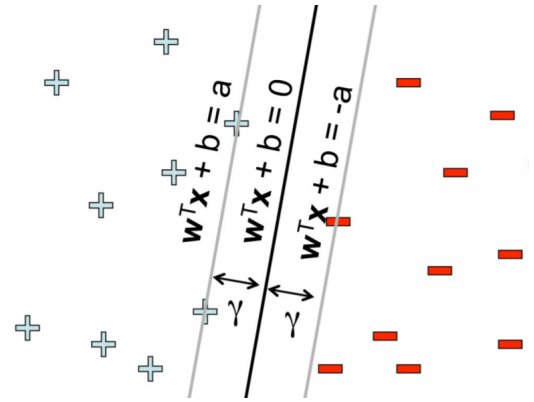


Figure 5.4: SVM Classification hyperplane ($a = 1$): *In a perfectly separable hyperplane for binary classification, the data points lie on opposite sides of a separating plane. γ is the minimum distance from the support vectors for each class. The positive and negative classes are shown. Can distortion introduced by watermarking cause a data point to move from the positive to the negative point or vice-versa?*

An example of Kernel parameter configuration in MATLAB 2019b is :

`('KFold',10,'Cost',[0 2;1 0],'ScoreTransform','sign' );` which specifies to perform 10-fold cross-validation, apply double the penalty to false positives compared to false negatives, and transform the scores using the sign function. Cross-validation prevents over-fitting to training data and allows for better generalisation for future predictions.

So the SVM is testing if a subject is positive ($y = 1$) for pneumonia or normal ($y = -1$). This test results in four possible outcomes of a classification or prediction algorithm: True Positive (TP), False Positive (FP), True negative (TN), and False negative (FN). These outcomes form what is called a **Confusion Matrix**, C, such that:

$$C = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$$

TP is a correctly predicted positive class; FP is a false positive prediction where a negative class is predicted as a positive class. FN is the opposite of FP, where a positive class is predicted as a negative class, while TN is when a negative class is predicted as a negative class.

First, a ***baseline model*** is developed for medical image classification using the original dataset without any watermark inserted in the ROI. After this, different number of bits of watermarks are applied to the original image. The corresponding watermarked images are used as training data to create a new model for each of the watermark levels. Each of the models is used to classify a test set, which is not part of the training set. The result of the classification will help to generate the confusion matrices.

With the confusion matrices generated above, other performance parameters for a machine learning algorithm, such as accuracy, specificity, recall(sensitivity) and precision can be defined. The equations below define the performance parameters used in this experiment.

Accuracy is a measure of the ratio of all correct predictions, whether positive or negative, to the entire test set. It is not a good parameter if the number of subjects in each class is not the same.

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{5.6}$$

In this study, accuracy is the ratio of the sum of patients correctly diagnosed(predicted) as normal and those correctly diagnosed as pneumonia to the total number of patients that arrived for diagnosis (subjects used for test set). We have limited our study to balanced training and test sets to avoid the complexity and bias introduced by unbalanced datasets.

Specificity is the True Negative Rate (TNR) as it is the ratio of the number of correct

negative predictions divided by the total number of the negative class (N).

$$specificity = \frac{TN}{TN + FP} \tag{5.7}$$

It is the ratio of the number of people correctly predicted as normal to the total number of normal people. It monitors the supposed assurance that patients are not predicted as having a disease that they do not have [96].

Recall or Sensitivity or True Positive Rate (TPR) is the ratio of correct positive predictions to the total number of the positive class (P).

$$recall = \frac{TP}{TP + FN} \tag{5.8}$$

This is the ratio of the number of patients correctly predicted as having pneumonia to the total number of people who really have pneumonia. Sensitivity and specificity are often used to ascertain the quantitative ability of a diagnostic test [8, 96].

Precision or Positive Prediction Value (PPV) is the ratio of correct positive predictions to the total positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{5.9}$$

In our experiment, this is the ratio of the number of patients who actually have pneumonia to the number of those predicted as having pneumonia.

Area Under the Receiver Operation Characteristic Curve (AUC) is a graphical measure of the performance of a machine learning model. The Receiver Operation Characteristic (ROC) Curve is a design tool in machine learning models. The area under this curve is AUC, and its value ranges from 0 to 1, with 1 being the best and 0, the worst performance.

The magnitude of changes in any of the performance parameters due to the models created from watermarked images with respect to the baseline model will be recorded in the results section.

5.5 Results

The statistical measures of biomarker location (mean), variation, and significance tests concerning the watermark embedding parameters that distort images are first presented. This is followed by the results from the SVM classification models created from the watermarked and non-watermarked training images, using each of the biomarkers as the classification feature.

5.5.1 Statistical Results

In statistics, the measure of location is a typical value that summarises and represents the data points in a test group. Mean, mode and median are the most commonly used measure of location. The **mean** of each of the selected biomarkers for the test groups (normal and pneumonia) was used as the measure of location. The measure of variation or spread helps to tell us how similar or varied the data points in a group are. The most common measures of spread are quartile, range, variance, and standard deviation. This quantity typically shows the scatter of the data points and how far they are from the mean value. The **Standard deviation** was used in this case to represent spread or variation. Hence, in the following results, we present how IH may have changed these statistics for a given biomarker and then, if such changes are significant or not. A quick and straightforward way to study effects is to study the change in location and spread of the dataset's feature values due to a cause (watermark in this case).

5.5.1.1 Statistical Effect on Location and Spread of Biomarkers

Initially, we first embedded watermark in all sub-blocks without trying to control distortion in a specific sub-block, irrespective of its complexity. Each of the dataset classes (normal and pneumonia) were studied separately using this approach.

Table 5.5 presents the means, standard deviation, and percentage difference between the mean features of the original and watermarked *normal* patient dataset. All

the sub-blocks in the ROI have been watermarked at 1 bit per sample(bps).

Table 5.5: XrayNormal: *Average ROI Textural Feature changes due to Steganograpgy. $Avg.V_o$ = Average Original feature value, $Avg.V_W$ = Average Watermarked feature value, $\%\Delta$ = percentage change, S.D = Standard Deviation*

Feature	$Avg.V_o$	$Avg.V_W$	$Avg.\Delta$
Contrast	0.0985(S.D =0.0143)	0.1046 (S.D = 0.0137)	0.0061(6.19%)
Correlation	0.9678 (S.D =0.0098)	0.9659(S.D =0.0100)	-0.0019 (-0.20%)
Energy	0.1824(S.D =0.0251)	0.1804 (S.D =0.0249)	-0.0020(-1.10%)
Homogeneity	0.9511 (S.D =0.0071)	0.9482(S.D =0.0067)	-0.0029(-0.31%)
Entropy	7.2547(S.D =0.1816)	7.2562 (S.D = 0.1819)	0.0015(0.02%)

The result shows the maximum mean difference in contrast and minimum mean difference in entropy. Contrast also had the greatest difference in the standard deviation of about six units compared to others. However, it should be noted that no feature had more than 1.5% change in mean value apart from the contrast feature.

Table 5.6: XrayPneumonia: *Average ROI Textural Feature changes due to Steganograpgy. Only the Contrast feature changed by more than 5%.*

Feature	$Avg.V_o$	$Avg.V_W$	$Avg.\Delta$
Contrast	0.066(S.D =0.0185)	0.0724 (S.D = 0.0165)	(9.69%)
Correlation	0.9718 (S.D =0.0110)	0.9688 (S.D = 0.0117)	(-0.31%)
Energy	0.2439(S.D =0.0552)	0.2408 (S.D =0.0538)	(-1.27%)
Homogeneity	0.9671 (S.D =0.0092)	0.9639(S.D =0.0081)	(-0.33%)
Entropy	6.9296(S.D =0.2880)	6.9313 (S.D = 0.2880)	(0.02%)

Table 5.6 shows similar results for the *Pneumonia* patients group. Like the *Normal* class, there is a higher effect in both location and spread of the contrast feature. Again, apart from contrast, the percentage mean difference is less than 1.5% for all other features.

For both experiments above, the embedding strength, α, was not controlled, and it considerably varied from 0.50 to 16.73.

This variation in embedding strength in each sub-block is visualised in Figure 5.5 as

a function of PSNR, which is a measure of image quality. This figure demonstrates the effects at the micro-level in each sub-block. It shows that embedding strength decreases with image quality.

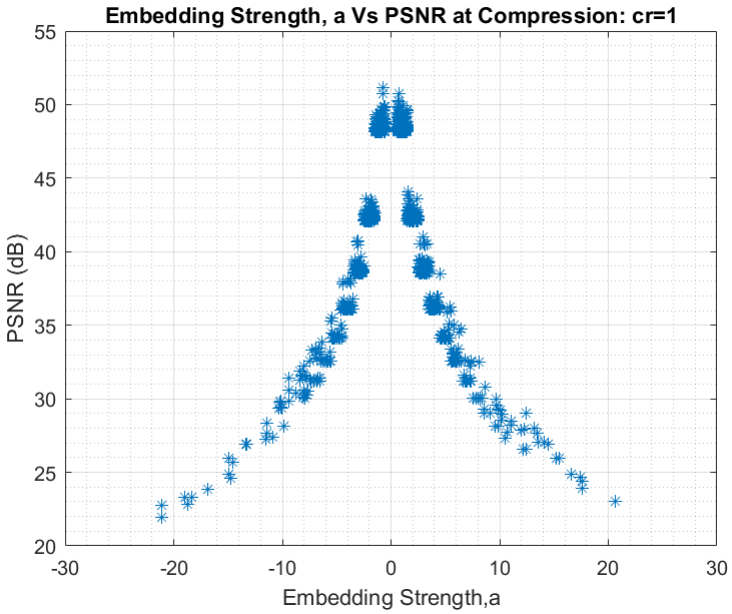


Figure 5.5: Uncontrolled Embedding Strength, α: *This leads to larger distortion in some blocks as shown by low PSNR*

Further statistical significance test results are presented in Table 5.7 to further analyse and understand the distributional effect of watermark embedding, especially for the C_4S algorithm as validated using this statistical framework. The end goal is to characterise IH algorithms based on the embedding strength, α - the primary parameter that determines distortion, robustness, and accuracy of watermark retrieval.

5.5.1.2 Significance Test Results

The ***p-value*** is the major statistic that determines if a hypothesis should be accepted or rejected. The general rule is given thus:

Reject the null hypothesis only if the p-value is less than the alpha level (0.05). However, there is an increasing demand to include other statistics such as Confidence

level, Confidence Interval, and effect size while reporting medical statistics [25]. We have reported most of the other statistics in Section 5.4.2 while exploring the datasets. Hence, we will report only the p-values in the results that follow.

Table 5.7 presents the statistical significance tests for the textural features for normal (healthy) and pneumonia classes. It shows that, for the uncontrolled embedding approach, the null hypothesis cannot be rejected for correlation, energy, and entropy but are rejected for contrast and homogeneity. This assertion is because the **p-values** for correlation, energy, and entropy are above the alpha level of 0.05.

Table 5.7: Normal (Left) and Pneumonia (Right): *ANOVA for one bit per sample at average embedding strength,* $\alpha = 0.89$.

Biomarker	P-value
Contrast	0.0047
Correlation	0.1945
Energy	0.5641
Homogeneity	0.0046
Entropy	0.9589

Biomarker	P-value
Contrast	0.0103
Correlation	0.0625
Energy	0.688
Homogeneity	0.0095
Entropy	0.9669

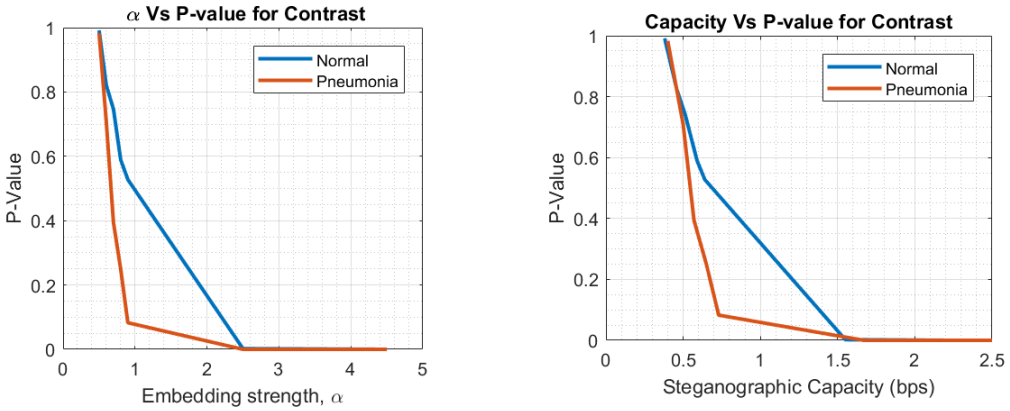
On the other hand, the **p-values** for contrast and homogeneity are below the alpha level, and thus the null hypothesis will be rejected. Hence, at the average embedding strength of 0.888 (S.D = 0.97), some features of the medical image are significantly affected (contrast and homogeneity), while others are not significantly affected (energy and entropy). Correlation is at the borderline as the p-value is close to 0.05, especially for the Pneumonia test group. These results also indicate that the Pneumonia class is affected more by watermarking than the normal class.

Controlled Embedding

The results so far are not based on controlled embedding strength. Those results showed that further optimisation of the algorithm is required for biomarkers that behave like contrast and homogeneity. This requirement led to further controlled (see Section 3.4.2)

for controlled embedding based on complementary bit zone) experiments on the allowable dynamic embedding strength, α. These experiments followed a stepwise increment of maximum α before embedding and not based on average α. The following results were obtained from these controlled experiments.

The variation of contrast feature as a result of embedding strength and steganographic capacity is shown in Figure 5.6. The significance of this change as a result of a maximum α and payload capacity is visualised in Figure 5.6.



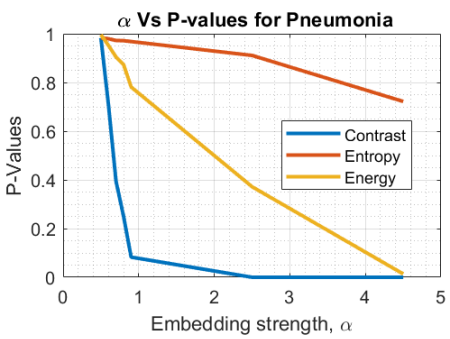
(a) Embedding Strength

(b) Steganographic Capacity(bps)

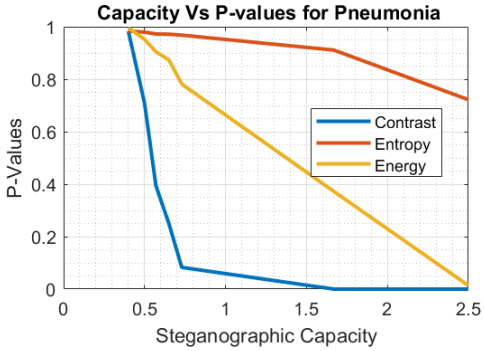
Figure 5.6: Statistical effect (*p-values*) of Watermarking Parameters on the Contrast feature of Normal and Pneumonia dataset : *The Constrast feature of Pneumonia dataset degrades faster than the Normal dataset.*

We concentrate further on results that relate to the pneumonia disease condition. Figure 5.7 shows variation of the *p-values* with embedding strength, α and steganographic capacity for the various features being considered at this stage.

The trend for α correlates with that of Steganographic capacity. Both results in 5.7 show that the sensitivity of various biomarkers to IH differs. Generally, as α increases (Figure 5.7a), the *p-value* falls. For example, using the 0.05 rejection criteria for statistical tests, it is seen that the null hypothesis for contrast will be rejected at α of 1.6; for energy, it is at α of 4.2 while entropy will be rejected at a very large value of α (which can be got by interpolation). Although specific *p-value* level could be selected for different applications, our experiment helps to choose α based on statistical results quickly.



(a) Embedding strength Vs P-values



(b) Capacity(bps) Vs P-values

Figure 5.7: Watermarking parameters Vs *p-values* for Pneumonia: *An increase in α or embedding capacity decreases the probability of accepting the null hypothesis. The rate varies for different image biomarkers*

Current works ([94, 131]) show that such decision in medicine is largely subjective at the moment, and the *p-value* remains only an indicator and should be combined with effect size and confidence interval in order to make final decisions.

Similar results for steganographic capacity, as shown by Figure 5.7b, helps one to easily estimate capacity based on the *p-value* of the statistical test. For example, using our C_4S algorithm, for a 512 x 512 ROI using an encoding scheme of 3 bits per sequence, the maximum capacity for which the null hypothesis will not be rejected for the energy biomarker is $2.4 \times 3 \times 4096 = 29,491$ bits. This capacity will reduce for biomarkers that behave like contrast (that is, $1.6 \times 3 \times 4096 = 19,661$ bits) or increase considerably for biomarkers that behave like entropy.

Hence, the trends presented so far serve as trade-off charts. The *p-value* accepted for different application may differ. This will serve as a black-box test for every steganographic algorithm. For medical image IH security, higher *p-value* may be required. However, the image quality classification presented in [94] does not have a sharp cut-off for medical images but simply tries to rate the quality on a scale and then state to what extent such image could be used for medical applications.

In the next section, we present the results that relate to autodiagnosis using SVM classification.

5.5.2 SVM Results

With the original datasets and the images generated from the controlled (Table 5.8) and uncontrolled (Tables 5.5 and 5.6) embedding experiments reported in Section 5.5.1.2 above, various SVM models were trained. First, a baseline or benchmark model was trained using the original images(Cr=0), watermarked images at one bit per sample (Cr=1) embedded in all sub-blocks, and then watermarked images at two bits per sample (Cr=2). The performance results for controlled embedding into the X-ray Pneumonia dataset are presented in Table 5.8. The models where $Cr = 0$ is the ***baseline model*** because no watermark was added to the images before feature extraction and model training.

Table 5.8: SVM performance with contrast as training feature at various embedding capacity, Cr is the number of bits embedded in a sample by the C_4S method.

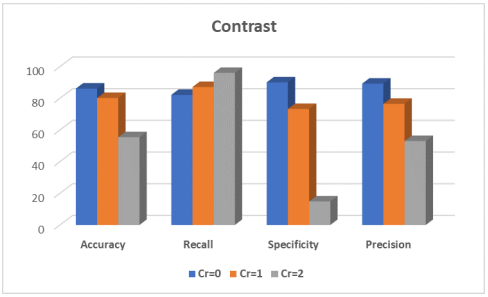
Cr	Accuracy (%)	Specificity(%)	Recall(%)	Precision(%)
0	86.14	82.18	90.1	89.25
1	80.2	87.13	73.27	76.52
2	55.45	96.04	14.85	53.01

During the experiment, we noticed that the baseline performance of correlation is not up to 70%. Even though it is above the 50% for a mere random guess in binary classification, it is considered too low for a medical classification algorithm. For this reason, the correlation biomarker is not considered in the results presented in this chapter.

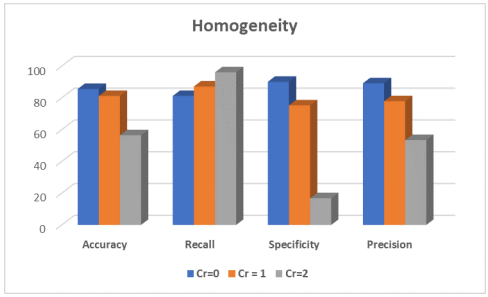
Next, we present the comparison of the models'performances as a function of the watermark payload (The embedding strength, α was not controlled in these experiments). These are shown in Figures 5.8 and 5.9.

Figure 5.8 shows that Contrast and Homogeneity are good classification features for pneumonia, but they are not robust to watermark embedding. There were significant changes in accuracy for lower distortion in the image. This calls for adequate control of both watermarking strengths, α, and payload size.

Figure 5.9 shows that energy and entropy features are least affected by watermark

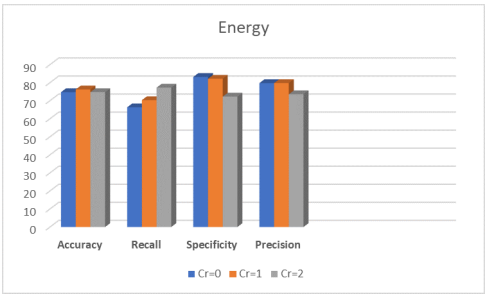


(a) Contrast

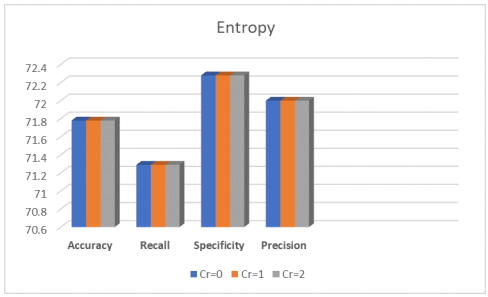


(b) Homogeneity

Figure 5.8: Effects of watermark bits on Contrast and Homogeneity Models. $Cr = 0$ *indicates that no watermark was added to the image while* $Cr = 1$ *and* $Cr = 2$ *means that 1 and 2 bits, respectively, of watermark has been added in each 8x8 block subject to some distortion rates.*



(a) Energy



(b) Entropy

Figure 5.9: Effects of watermark embedding on Energy and Entropy Models: *Energy and Entropy are more robust features than Contrast and Homogeneity. The classification performance is not affected by increase in watermark payload when using entropy as a prediction feature.*

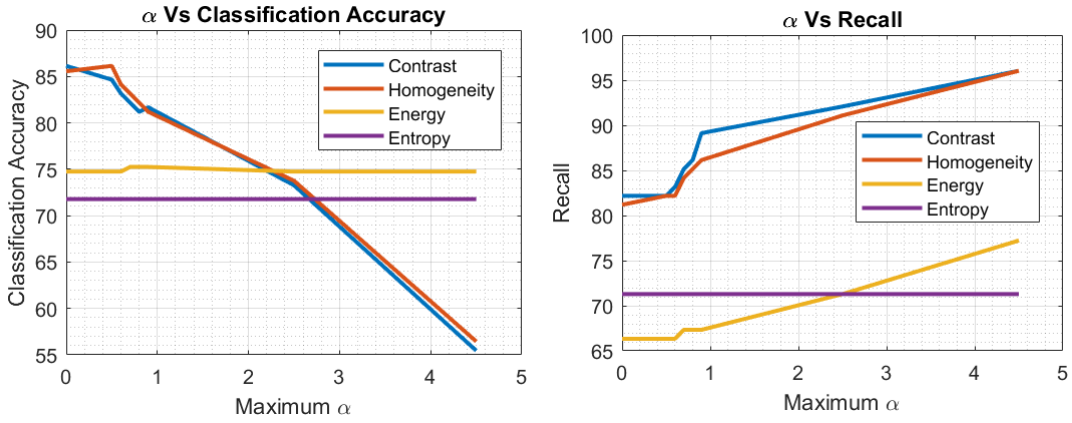
insertion. Entropy is not affected at all. The value of accuracy, recall, specificity, and precision remained the same for all payloads. There was a slight change in these parameters in terms of energy.

The negative result of Figure 5.8 called for further controlled experiments. A large standard deviation or variation in α introduced significant distortion in some sub-blocks, which led to a large change in contrast and homogeneity within different areas of the image. This led to further experiments where the independent variable was the maximum embedding strengthα_{max} and not the payload. This produced more models than previously. The results for models based on contrast feature is shown in Table 5.9.

Table 5.9: SVM performance with contrast as training feature at various watermark embedding strengths,α

α	Accuracy (%)	Recall(%)	Specificity(%)	Precision(%)
0	86.14	82.18	90.1	89.25
0.5	84.65	82.18	87.13	86.46
0.6	83.17	83.17	83.17	83.17
0.7	82.18	85.15	79.21	80.37
0.8	81.19	86.14	76.24	78.38
0.9	81.68	89.13	76.24	78.57
2.5	73.27	92.08	54.46	66.91
4.5	55.45	96.04	14.85	53.01

In order to visualise the above effects of α on accuracy, recall, specificity and precision, Figures 5.10 and 5.11 are presented. These figures simultaneously present the results about contrast, homogeneity, energy and entropy features for the purpose of comparison.



(a) Accuracy

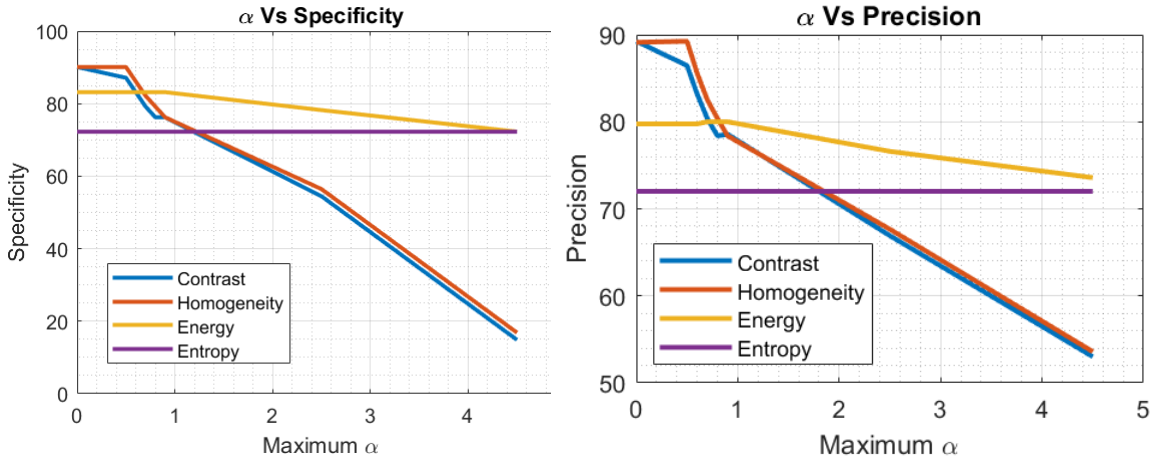
(b) Recall

Figure 5.10: Effects of watermark embedding on Accuracy and Recall: *The maximum alpha determines the maximum distortion that can be introduced in each sub-block.*

Figure 5.10a shows a decline in accuracy for contrast and homogeneity, an increase in accuracy for energy, and constant accuracy for entropy. For homogeneity, it can be seen that the accuracy at the maximum α of 0.5 is the same as the accuracy of the base-

line model. This result shows that some biomarkers are more stable to image changes than others.

The result of 5.10b for the recall is an interesting one. Recall can be interpreted as the accuracy of the positive class. For contrast, homogeneity, and energy, recall increases with maximum embedding strength. This means that more accurate predictions of people having pneumonia are being made. As noted by Kermany *et al.* in [8], false-negative (Type II error) result is far more serious in disease classification. Hence, an increase in recall reduces type II error and ensures that the maximum number of people having pneumonia is correctly referred for further investigations.



(a) Specificity (b) Precision

Figure 5.11: Effects of watermark embedding on Specificity and Precision: *Energy and Entropy are considered robust features while Homogeneity and Contrast are non-robust features for the C_4S watermark embedding scheme*

Specificity is the accuracy of the negative class. Figure 5.11a indicates that as the embedding strength increases, the specificity also decreases. This means that the false alarm (Type I error) rate may also increase. The implication is that normal patients could be referred for further investigations. However, for homogeneity, this does not occur until $\alpha = 0.6$, for energy, it is until $\alpha = 0.9$, while ***for entropy, it never occurred*** for the range of α used in the experiment. The contrast had the most rapid negative response.

Precision had a similar trend with specificity. This is shown by 5.11b. If more type I error occurs, then the precision of a prediction would decrease. This does not stop recall from increasing provided that more accurate positive predictions are also being made.

5.5.3 Comparison with Existing Methods

In this work, we have introduced the ***Statistical and Machine-learning*** evaluation methods for applications that are both automated and will be used in environments where medical experts are not readily available. As the dataset used has already incorporated the subjective knowledge of experts, the initial human input has been taken care-of as currently being used in the existing (VGC) methods.

But then, the superiority of our method lies in the following:

1. We can incorporate the opinions of more experts and historical patient scans in one study. In [164], only three experts were involved, whereas in [94], about five experts were involved. A recent study in [8] managed to get seven expert evaluations. These are opposed to the hundreds of experts and exactly 1,284 patients whose data were used in this study.

2. The AUC score is just a single score in *Observer performance* evaluation. In our method, a more specific score relating to diagnostic effects and their consequences could be used. These include specificity, accuracy, recall, and precision, among others. Greater insights can be drawn with the new evaluation method.

3. Testing several **specific criterion** as mentioned by Ludewig *et al* in [94] will be very tedious for human experts. With our method, automated models could be used to extract the specific features that could be used to test each criterion, and both the individual and relational effects could be evaluated repeatedly.

4. For remote, developing, and rural areas where experts are not accessible, our method becomes an efficient way to transfer and utilise the knowledge of experts

in image evaluation. Existing models can always be retrained as more expert information becomes available without the expert's presence.

5.6 Key Findings

There are four major key findings than can be inferred from the results of this study:

1. Empirically, in the spatial (pixel) domain of embedding, with embedding strength, $\alpha < 2.6$, the local (block-level) distortion in terms of PSNR is greater than 40dB. However, this could still cause significant changes in some features used for training machine learning algorithms. Hence, the 40dB benchmark in [49], is not reliable for feature-specific evaluation in teleradiology.

2. Whereas the C_4S Algorithm has improvements in terms of Steganographic Capacity, accurate watermark detection as well as tamper detection, at high embedding strength, some biomarkers are affected. Hence, the embedding strength should be limited to 2.0 for 8-bit images and about 4.5 for 16-bit DICOM images.

3. Not all medical image biomarkers (features) are equally affected by the same strength or quantity of watermark. Some are not affected at all, while others could be either positively or negatively affected. Different image modalities behave differently, as well.

4. Generally, Watermark can be embedded in the region of interest (ROI) for tamper detection by controlled methods without significantly changing the diagnostic outcome from an SVM classifier. This is because distortion is a major constraint for medical image security using data hiding techniques.

5.7 Discussion

One of the current challenges in adopting Steganography and Digital Watermarking in Teleradiology is a general belief that they could alter medical image scans and there-

fore raise ethical and legal concerns, including diagnostic integrity [127]. We believe that the reason for this challenge is related to the evaluation mechanism employed by algorithm developers to validate their applicability in medicine. As a solution, digital watermarking and Steganography should use biological and medical parameters, known as biomarkers (instead of only the physical image quality assessments such as PSNR, SSIM, KLD), to evaluate the effect of watermarking/Steganography on medical images.

Zain *et al* [163] have used subjective and objective pieces of evidence to show that the wide claim that steganography and digital watermarking affects diagnosis is not true. However, in the recent times of computer vision and machine learning where human visual system are being replaced by computer systems and where adversarial machine learning is also emerging, further investigations by Garcia *et al* [54] suggests otherwise. The evidence produced by the later showed that embedding strength could affect the accuracy of a classification algorithm. In this chapter, we have gone further to quantify how this embedding strength and payload size per sub-block (Communication rate) affects some biomarkers used to classify pneumonia diseases. This because machine learning, which uses images for training models, is the emerging autodiagnosis technique for disease prediction. Hence, it is the best context for evaluating the effect of Steganography in the medical image autodiagnosis system.

According to the findings in this chapter, it cannot be refuted that watermarking or Steganography does not change the micro-elements that make up a medical image. Without such micro-changes, no information will be encoded in the first place. However, it is an unfounded generalisation to assert that any modification of any type changes the diagnostic outcome. We argue that focus should be shifted to specific features of diagnostic decisions instead of the image as a whole.

It can be seen from Figures 5.10 and 5.11 that certain features (biomarkers) are vulnerable (Homogeneity and Contrast) to watermarking while others(Energy and Entropy) are not. SVMs, as well as deep neural networks (DNN), have two features that they can recognise: robust and non-robust. Frequently, the non-robust features are

subtle, but representatives and serve good features to predict a class. However, they are easy to subvert as little pixel changes tend to shift them to another class. On the other hand, robust features continue to deliver correct results even when the pixels are changed by small amounts. This interpretation is in line with a recent report by Edwards in a recent Communications of the ACM [46].

Interpreting why recall improved as the embedding strength increases prompted for more investigations. It was observed that the negative class and the positive class responded differently to watermark embedding. In general, the decision boundary shifted to favour correct classification of positive class (higher recall). Correspondingly, the reduction in accuracy resulted from the misclassification of the negative class(lower specificity). However, as noted by Kermany *et al* in [8] , false-negative (Type II error) result is *far more serious* in disease classification. Hence, an increase in recall reduces type II error and ensures that the maximum number of people having pneumonia is correctly referred for further investigations. Hence, Steganographic methods affect biomarkers based on the type of image feature that is modified by the embedding process.

Further observations show that for the significantly affected biomarkers, there is a decrease in type II error and a corresponding increase in Type I error (false alarm). Whereas all errors are important to be eliminated from a medical system, the implication of high payload is that some *Normal patients* are likely to be referred for second-level examination when they are not supposed to be referred. This result might be considered more acceptable than the opposite where a *Pneumonia* is declared *Normal* by the autodiagnostic system. At an embedding strength of more than 2.5, the accuracy of the watermarked model would reduce beyond 70% benchmark and could be considered adversarial. Therefore, for any new steganographic or watermarking system, it is important to establish the parameter boundaries beyond which the algorithm becomes adversarial for the biomarkers of interest. This practice is more important for the non-robust features or biomarkers.

The findings in this study should be digested together with the fact that current machine learning methods use other forms of image pre-processing and data augmen-

tation before feature extraction. It will be an interesting study to find out how the pre-processing and data argumentation compares with information hiding and whether they can override both the positive (in terms of energy) and negative (in terms of contrast) effects on classification accuracy, recall (sensitivity), specificity and precision.

From the foregoing, we recommend, therefore, that at high embedding strength and high watermark payload, watermark insertion should be restricted to the region of non-interest. Also, for the time being, autodiagnosis should be applicable where its predictions are more accurate than the human expertise available. An Example is in low-resource dwellings [8], such as remote rural areas.

5.8 Summary

Distortion is an important primitive in Steganography and watermarking. It becomes even more important and domain-specific when we deal with Medical image IH. The distortion effect on the specific image biomarker that relates to a particular disease or condition needs to be evaluated. In situations where machine learning and autodiagnosis are involved, further experiments based on machine learning models are required. The mere use of high PSNR or SSIM as an evaluation parameter is not enough.

In this Chapter, we chose a limited number of biomarkers that apply to a wide range of diseases to evaluate the effect of our IH algorithms. For biomarkers that behave as entropy or energy, there is higher flexibility on payload and distortion. However, for biomarkers that behave like contrast, our evaluation results have shown that the results obtained by [164] and [49] in terms of PSNR may not hold except for human visual system model for diagnosis. This is not the case for computer vision and, hence, brings the need for further investigation in the area of Steganography and watermarking design in computer vision and autodiagnosis systems.

Some effects of our data hiding algorithms are positive as they increase the accuracy of predicting the positive class. For the negative effect, further improvement is required in this respect, and only watermarking at a low embedding strength and payload is

recommended. As a last resort, reversible or RONI-only watermarking is applicable for such biomarkers.

It can be deduced from this whole thesis that robust AI in medicine will include a good mix of machine learning techniques with rule-based expert systems. Both systems should also know when to say: *I do not have enough information or confidence to decide this medical case, but this is what I think at x confidence score.* This means that irrespective of the evidence provided in this research, machine learning and AI for digital health is still at the stage of partial automation and cannot replace humans yet.

Nyeem in [127] highlighted the need to develop a framework that can dynamically select the appropriate IH algorithm that fits the properties of an image scan to minimise any effect on medical image biomarkers and, thus, on the diagnostic outcome. This contribution is a basis for evaluating Steganography and digital watermarking using biomarkers for autodiagnosis. The next chapter proposes a wider framework that can be used to evaluate and select any data hiding algorithm for security in teleradiology, including the human-based, machine learning, and other autodiagnostic approaches. This software framework was to enable dynamic selection of IH algorithms and parameters based on the service required and the level of distortion allowable for a medical application.

Chapter 6

Medical Image Information Hiding Software Framework

Various researchers have proposed hundreds of algorithms for use in medical image watermarking and steganography (MIW/S). Also, well defined theoretical and mathematical frameworks exist for modeling watermarking and Steganographic systems. However, detailed software design patterns are not yet in place to provide concrete structures for widespread implementation, adoption, and scaling of MIW/S. In this chapter, we identify and analyse existing conceptual frameworks that can serve as the base for designing a generalised software framework. In Section 6.2, we review the existing conceptual and mathematical frameworks to identify implementation gaps. Section 6.3 focuses on software concepts and the design principles that enable extensible and adaptable system designs. After these, we propose, design and implement a new software framework in 6.4. To test and evaluate this software system, we then propose new criteria for evaluating steganographic algorithms and then apply these criteria to evaluate some specific algorithms, including the ones developed in Chapters 3 and 4. The results of this evaluation are presented in Section 6.5. Key findings are discussed in Section 6.7 before a concluding summary is presented in Section 6.8.

Go To Chapter: 1,2, 3, 4, 5, 7

6.1 Introduction

Information hiding (Watermarking and Steganography) has permeated many other multimedia industries, but not the health industry. This is the case for the health industry because it has numerous strict regulations and procedures that allow only a slow

introduction of new technologies into the field [106]. The available lee-ways require standards and agreed frameworks for accepting a new technique. This stringency is evident in the various stages of clinical trials in medicine. The adoption of Medical Image Information Hiding (IH) techniques by medical experts also requires a unified and algorithm-agnostic framework for developing, testing, evaluation, and adoption of the associated algorithms. In such a context, the definition of architecture is generalised, but the specific implementation could vary. This design approach would provide a standard interface for evaluation and selection of medical image watermarking and steganographic (MIW/S) algorithms irrespective of medical image modality, disease type, and internal algorithm implementation strategy.

In Structural Engineering, a framework is an essential supporting skeleton required to construct a building, car, bridge, and the likes. A system is built on this basic structure. In literature [47, 3, 146], frameworks are often synonymous with system architectures. In theoretical research and applied sciences, frameworks may range from mathematical/theoretical, conceptual, analytical to Software frameworks. In particular, software frameworks often combine other structures to create a practical implementation of ideas. This chapter builds on existing mathematical and conceptual MIW/S frameworks to develop a practical software implementation of MIW/S for autodiagnosis in teleradiology.

The non-existence of standard frameworks and protocols are the significant challenges that have hampered the adoption of MIW/S security techniques in the mainstream e-health system [131]. Different works cited in this thesis and those we have developed in Chapters 3 - 5 show that there are lots of algorithms with potential applications in teleradiology. In addition to these algorithms, there are lots of evaluation methods in existence. However, little or no universal practical frameworks and standards exist in watermarking [131] as it is in cryptography. Although most cryptographic algorithms, such as Data Encryption Standard(DES) and Advanced Encryption Standard (AES), still have limitations, cryptography has more developed standards and frameworks than information hiding. Hence, similar standard frameworks in data hid-

ing security are required to ensure a systematic adoption of the technology in medical practice.

In this chapter, we first review the existing and generalised MIW/S frameworks in literature and build on these to propose a new conceptual framework. From this extended conceptual framework, we design a generalised software framework that is compatible with the current or future algorithm that exposes the same standard interface, such as the one we define in this chapter.

Specifically, we made the following contributions in this chapter. We:

1. reviewed and extended existing conceptual and mathematical frameworks that relate to MIW/S, especially those based on **Operational Determinism** [127].

2. proposed and evaluated new criteria for MIW/S algorithm selection based on service fulfillment, image quality, and diagnostic correctness.

3. designed and developed a software framework that is agnostic to the specific medical image information hiding (IH) algorithm, incorporates the above contributions, and is extensible by design.

The software framework consists of software interfaces, modules, functions, and data structures. The set of inputs to and outputs from these software structures are expected to be standardised across present and future watermarking and steganographic algorithms. The internal implementation of those algorithms may vary, however. The framework was evaluated in the light of Medical image IH requirements previously mentioned in Chapter 2. In the next section, we will review, analyse, and identify new inputs, functions, and outputs that have been omitted from existing watermarking and steganographic models and conceptual frameworks.

6.2 Review of Existing Frameworks in MIW/S

The existing requirements and algorithms for medical image watermarking and steganography(MIW/S) are the beginning of the quest to define a standard framework. The

general functional requirements that are common to these algorithms provide a base for defining the security services that should be provided by these algorithms. The general requirements and existing algorithms were reviewed in Sections 2.4.2 and 2.4.3 respectively . In this section, we focus on generalised functional and operational specifications that will enable any algorithms to be implemented to provide a specific service that fulfills any of those requirements. This requirement-service mapping will enable one to define a standard but scalable software framework. The focus of this review is to identify the ***inputs***, the ***outputs*** and the ***generic operations*** that represent generic functionalities seen in most existing watermarking and steganographic algorithms. This approach is known as **Operational Determinism** [127], and it helps in the unification and standardisation of operational processes.

6.2.1 Parameter Analysis and Operational Determinism

Mathematical models operate on parameters to get results. This phenomenon directly allows us to identify the initial parameters that serve as inputs and the final parameters that serve as the outputs. There are also intermediate inputs and outputs towards the final output. We are not concerned with these intermediaries as they could vary considerably depending on the algorithm. These internal operations and intermediate variables are grouped and represented by a function, f. Hence, we are interested in the general form shown in (6.1):

$$y = f(x), \tag{6.1}$$

where both x and y could be a number, a vector or a matrix and f is a function (a set of operations) that transforms x into y.

A generalisation based on defining input, output, and functions is essential in this work because it completely maps to the same approach used in software abstraction. One is not to be concerned by the specific implementation of these functions but more on the nature and properties of the inputs they take and the outputs they provide. This

approach further allows separation of concern among medical experts and biomedical engineers. They are involved in the design, development, and maintenance of Health Management Systems (HMS), though from different perspectives. For example, a standard output image from any MIW algorithm should be evaluated for any changes in diagnostic information without the medical expert's awareness of the algorithms.

Many partial models (addressing only specific aspects of IH) had existed in literature [112, 13, 12, 116, 90, 2] since last two decades. One can deduce some commonality in input, output, and functions among these authors in their formal mathematical models and their proof-of-concept algorithms. Barni *et. al* [12] combined the concepts of information theory, cryptography, and signal processing to provide a general security framework for robust watermarking. Only two major functions were considered in this model: Embedding and Decoding functions. The major input/output includes original cover, watermark, watermarked cover, embedding key, and detection key. The attack model was restricted to robust watermarking systems only. Reversible watermarking, integrity checks, and Steganography were not included in this model.

The modeling of Steganography as a game was put forward by Moulin *et. al* in [116]. They used game theory to estimate hiding capacity in an optimal attack context. All forms of cover data, including image, audio, and video, were considered. The concept of **side information** was also presented. They used composite data to represent the watermarked output. *Encoder* and *decoder* were the major functions involved. Also in the steganographic space, Mittelholzer in [112] included *key* as input data and *stego-channel* as a function related to attacks. The attack functions and distortion functions were not given due consideration in both frameworks.

The work of Nyeem [127] can be seen as the most recent unifying work with almost comprehensive identification of input, output, and functions from mathematical models or frameworks in previous research works. However, the author focused on listing functions and parameters that are most relevant to robust watermarking and not Steganography. This limited approach has left functionalities, such as integrity checks, to be ignored.

After due consideration of various partial definitions, it can be deduced that a **key-based** IH scheme is made up of 8-tuples $(\mathbb{I}, \mathbb{M}, \mathbb{W}, \mathbb{K}, G, E, D, EX)$ where:

- $\mathbb{I}$ - the input space of all host data. For an original medical image scan X, $X \in \mathbb{I}$.
- $\mathbb{M}$ - the input space of all messages. For any plain text or image data , m, to be embedded into X, $m \in \mathbb{M}$.
- $\mathbb{W}$ - the output space of generated watermarks, $w \in \mathbb{W}$. The generating function,G, uses K_g to transform m into w.
- $\mathbb{K}$ - the key space with the following special subsets: watermark generation keys, K_g; embedding keys, K_e; decoding keys, K_d and extraction keys, K_{ex} (i.e., K_g,K_e,K_d,$K_{ex} \subset \mathbb{K}$).
- G - Watermark generation function. $w = G(K_g, m)$
- E - watermark embedding function. $\bar{X} = E(K_e, w, X)$.
- D - watermark decoding function. $\bar{w} = D(K_d, \bar{X})$.
- EX - watermark extraction function. $\bar{m} = EX(K_{ex}, \bar{w})$. $\bar{m}$ is supposed to be the same as the embedded message m. However, sometimes intentional and unintentional attacks may result to a modified message being extracted.

Concerning Kerckhoff's principle[1] of security, watermarking and steganographic algorithms cannot be kept secret but the key should. Hence, in Table 6.1, the seed used to generate the security keys is the only secret that needs to be kept. Although the message might be encrypted or kept secret, the security in key-based systems does not rely on data hiding only but also on the key's secrecy. However, as it is in cryptography, the complexity of the hiding algorithm can contribute to security. Still, in as much as this algorithm is made public, it could be figured out by someone. At least a key in the sending side and another key in the receiving party must be made primary and remain secret and independent of the message or the cover image.

[1] https://blog.cloudflare.com/a-note-about-kerckhoffs-principle/

Despite the above useful model by [127], some generalisable aspects are still lacking. We have summarised these missing aspects as follows:

1. Availability of side information (SI) to define the protocols for accurate embedding and extraction. The right format for SI is critical and requires standardisation across algorithms. This format specification is related to the definitions of protocol header in the internet protocols.

2. Reversible Watermarking function(WR). WR functions need to be specified as they have become an essential aspect of medical image watermarking due to some ethical requirements.

3. Distortion function (DF). For non-reversible watermarking algorithms, the DF is required for capacity control and robustness control as these features are proportional to distortion, which affects diagnostic information.

4. Attack function (AF). Image transmission and storage have inherent and external attacks that deteriorate information. This attack function may differ based on the use-case. A unified AF is required to enable various attacks to be simulated via the same interface. Even if new attacks emerge in the future (as they always do), one should test these new attacks on existing algorithms via the same interface.

5. An integrity verification function in the form of an Integrity checker(IC). This service has become an essential utility provided by watermarking and Steganography [63]. There are various ways this could be achieved, but a generic input is either the original, the watermarked image, or both. For blind verification methods, the original image is not required.

6. Image Authentication (IA). Another essential utility related to integrity is the source image and watermark authentication function

Hence, whereas other functions may exist as mentioned earlier, the goal of IH is to satisfy the requirements of a specific application while providing services in the form

of confidentiality, integrity, authentication, availability of information and traceability of actions. Inputs, outputs, and functions needed to provide these desired outcomes should be included in a MIW/S framework. Some of these are relevant to both watermarking and Steganography and could be generalised to medical and non-medical images.

Table 6.1 is a summary of the sets of functions with inputs and outputs that represents all watermarking and steganographic algorithms found in literature, including the ones we have newly identified. We refined and extended the parameters and operations (functions) identified by [127] but only for watermarking systems. We argue that neither watermarking nor Steganography is an end in itself but a means of providing some services and utilities. The functions that directly relate to this range of utilities should be part of a software framework.

6.2.2 Conceptual Frameworks

Conceptual frameworks are concerned with the skeletal architecture through which disparate sub-systems could be integrated into a complete system. Medical image IH algorithms need to integrate and co-exist with teleradiology and health information systems. We review the relevant conceptual frameworks below.

The conceptual framework proposed and presented by Qasim [131] and the one presented by Singh [143] are very relevant in defining the place of watermarking and Steganography in medical image transmission and storage architecture. According to [131], there are three phases in a teleradiology application where MIW may be required: *Acquisition*, *Archiving* and *Viewing* phases. However, according to [31], *Transmission* network is a vital component of digital health including teleradiology.

In order to design a relevant framework for MIW in Teleradiology, it is essential to review, at least briefly, the Picture Archiving and Communication System(PACS) being used in mainstream teleradiology. Digital medical image archiving and communication system standards used by PACS is based on the Digital Image Communication in

Table 6.1: Summary of generalised watermarking/steganographic model: functions, inputs and outputs. SD= Standard Deviation, X = host image

Function (**f**)	Inputs($\boldsymbol{x}$)	Outputs ($\boldsymbol{y}$)
Keys, K(.)	X message, m seed, s	generation key, K_g embedding key, K_e decoding key, K_d extraction key, K_{ex}, reversal key, K_r
Watermark Generation, G(.)	X message, m generation key, K_g	watermark, w
Watermark Embedding, E(.)	host image, X watermark, w embedding key, K_e	watermarked image, $\bar{X}$
Attack Function, AF(.)	watermarked image, $\bar{X}$ mean of attack, μ S.D of attack, σ Class of attack, *MODE*	attacked version of $\bar{X}$, $\bar{Y}$
Watermark Decoding, D(.)	decoding key, K_d watermarked image, $\bar{X}$	estimated watermark, $\bar{w}$
Watermark Extraction, EX(.)	estimated watermark, $\bar{w}$ extraction key, K_{ex}	estimated message, $\bar{m}$
Watermark Removal, WR(.)	watermarked image, $\bar{X}$ reversal key, K_r	host image, X
Distortion Function, DF(.)	watermarked image, $\bar{X}$ host image, X	Quality Parameter, Q
Integrity checker,IC(.)	watermarked image, $\bar{X}$ host image, X Side Information, SI	Yes/No Location in Image Parameter, T
Image Authentication,IA(.)	watermarked image, $\bar{X}$ host image, X Side Information, SI	Yes/No Explanation Parameter, T

Medicine (DICOM) [75] standard. This standard and PACS marked the beginning of telemedicine and teleradiology [31].

PACS's major components consist of an image and data acquisition gateway, a PACS

server and archive, and several display workstations (WSs) integrated by digital networks[75]. One of the recent open-source implementation [23] of PACS is shown in Figure 6.1. Image data are transmitted between acquisition devices and servers at Mega Packets per Second (MPPS).

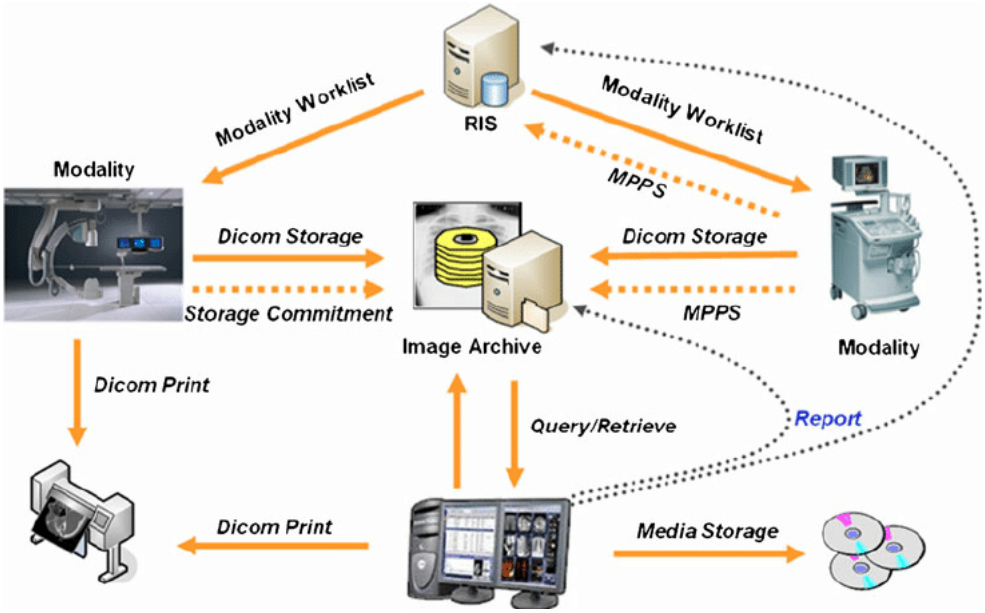


Figure 6.1: PACS Workflow[23].

The major components of a PACS framework include:

1. **Image acquisition modules** - Different digital modules are used to acquire images of different modalities. These range from film digitisers, Computed Tomography (CT) Scanners, Ultrasound scanners, Nuclear medicine modules, Computed Radiography to Magnetic Resonance Imaging (MRI) scanners, among others. The embedding function, E(.), could be called here to perform integrity watermark insertion, which will later be verified by a call to the IC(.) function. For evaluation, the AF(.) function would have been called using relevant attack 'MODEs'.

2. **Data Management System** - This is generally a server computer that controls the PACS network, acquisition devices, and image storage devices. It manages image archiving. This is a point where integrity and authentication data could be added

before archiving. A call to an appropriate security algorithm could happen here.

3. **Transmission network** - There are local and remote PACS systems. Moreover, in this decade, the peer-to-peer PACS system has been proposed [23]. Data for images, control signals, and text data such as patient records are transmitted via the networks.

4. **Image Display Stations** - Physicians interact with the PACS system using the display stations, which are made up of computers with local storage and some user interfaces. The image editing and actual diagnosis occur at the display stations.

5. **Printing Stations** - At some point, hard copies of scans may need to be printed. An example is the printing of an X-ray film. In recent times, however, scan data are commonly distributed or stored in electronic copy.

6. **Interfaces to other Systems** - Imaging is simply a part of the entire Health management system. Hence, it will need to interface with other patient care management systems such as Radiology Information System (RIS) and Hospital Information system (HIS). It is also envisaged that a new system such as Medical Image IH and other utility services can also interface to the PACS system through similar interfaces. This gives relevance to the proposed framework we envisage.

The PACS system is a software-hardware integration system. Some of the existing ones are legacy systems, while new ones are being developed. Any utility system such as the MIW/S framework will have to consider the existing architecture and identify where to hook into the PACS system. This assumption enables us to survey some relevant existing MIW/S frameworks that considered PACS in their designs.

A high-level framework that is suitable for determining where to call a MIW algorithm for security utility was proposed by Borra *et al* in [18]. They mentioned three points of security need in a standard teleradiology system:

1. data storage within a single hospital or its associated intranets,

2. online or offline transfer of data from one hospital to another,

3. remote hospital or third-party diagnosis place.

Although the authors later focused on their specific algorithm, this work suggested various points where various hooks could link an existing teleradiology system with corresponding MIW algorithms that can provide security and other utilities needed by the teleradiology system.

The work of Qasim *et al* in [132] and [131] clearly differentiated between a MIW algorithm and a MIW framework, respectively. In [132], they designed a reversible watermarking algorithm that they used as a case study in [131] to illustrate the place of their previous algorithm in an entire PACS. The framework proposed by [131] is shown in Figure 6.2.

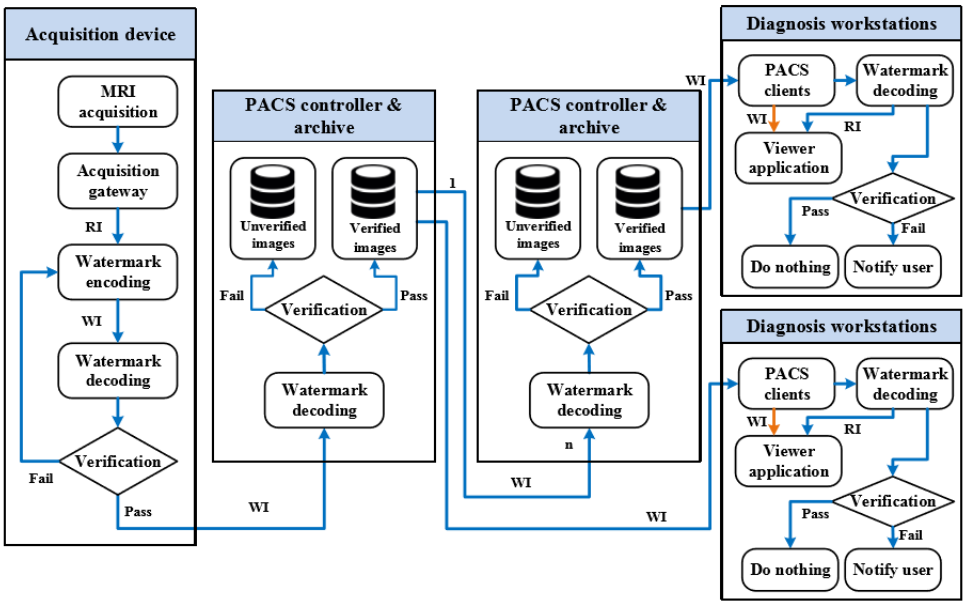


Figure 6.2: Integration of MIW Algorithm into PACS[131].

Undoubtedly, the work is an attempt to design a practical framework that could be considered for integration into the aforementioned PACS components. However, the framework is limited to their previous algorithm, considers only MRI modality, and assumes that the MIW algorithm is an integral part of PACS. It does not give room for

other algorithms that are suitable for other security features or relevant to other image modalities.

The very conflicting requirements in teleradiology, especially the ones related to ethics and legality, have made a generalised framework even more challenging. This has also made auto-diagnosis increasingly mere experimental research without core practical implementation and also produced more conflicting frameworks if considered only at the algorithmic level. A comparison of the algorithmic frameworks from [142] with the one from [66] shows two opposite ideologies that claim the same end result in terms of imperceptibility. Both watermarked Fundus images but using zero watermarking and ROI/RONI watermarking, respectively. In order words, Singh [142] did not add any watermark while [66] added watermark in both ROI and RONI. However, [66] (just as [164] who also used ROI/RONI) was able to perform correct autodiagnosis with the watermarked images. This leads to the need for a framework that considers the end-goal of correct diagnosis as the significant criteria for evaluation when the supporting utilities from the MIW algorithm are achieved. How this would affect existing policy is unclear, but there is clear evidence that MIW in themselves does not affect diagnosis in general.

In their assertions, both Singh [142] and Hassan *et al* [66] had a framework that agreed that MIW algorithms are needed after image acquisition, transmission and storage. There was also a common understanding that imperceptibility is of paramount importance. Furthermore, the ability to conduct autodiagnosis correctly, whether an image is watermarked or not, should be a major concern provided security requirements are also achieved using MIW [164, 34, 142, 66]. This assertion was reflected in a quick evaluation and implementation framework by Hassan *et al* in [66]. In their work, they introduced a machine-learning model previously developed to re-classify a watermarked medical image scan before transmission. The idea was to ensure that the classification system would do the same work as the physician if the physician were unavailable. We have previously pointed out the shortcoming of their work in Section 5.3. However, their approach is very relevant in an automation evaluation for MIW and

is similar to the work we have done in Chapter 5.

The specific strengths and shortcomings of the reviewed algorithms and frameworks are shown in Table 6.2. However, it is found that most of the existing frameworks have the following common shortcomings:

1. They are largely linked to the specific algorithm developed by the authors and thus are not easily generalisable.

2. Few authors consider the existing PACS system already in use by medical practitioners.

3. They are exploratory, and no software design details are given for future implementation and refinement.

4. algorithm-agnostic but disease-specific evaluation methods are often not considered.

After conceptual designs and representations in the problem domain, software Frameworks and designs are necessary to aid the implementation of the solutions. Frameworks are independent of specific algorithms that would be implemented in the future or already in existence. The work of Babel *et al* in [11] defined generalisable protocols and implementable security frameworks for medical image compression, security, and transmission. Their designs considered and defined protocols that could be implemented via any MIW/S or cryptographic algorithm. Protocols that relate to secure bitstream communication, including source-channel coding, were also considered. The design is suitable for emerging demands in medical image coding in both lossy and lossless formats, together with integrity and speed for low-power devices in eHealth. They used their Local Adaptive Resolution (LAR) codec as a case study to demonstrate medical image coding that could enable content integrity and compression. However, no software interfaces were provided to accommodate alternative solutions or to integrate with the existing PACS system. In the next section, we will explore the components of a software framework.

Table 6.2: Shortcomings of Existing Frameworks

Author	Unique Features:☑	Shortcomings: [✗]
Zain *et al* [164]	Supports doctor's subjective evaluation	**No Generalised Software Framework**, Autodiagnosis not considered, Did not consider existing PACS, Algorithm-Specific Conceptual framework
Abunyeem [127]	Generalised theoretical models	**No Generalised Software Framework**, Autodiagnosis not considered, Did not consider existing PACS, Algorithm-Specific Conceptual framework
Singh & Dutta [142]	Zero-watermarking Concept and lossless	**No Generalised Software Framework**, Auto-diagnosis not considered, Imprecise Conceptual Framework for PACS Integration
Al-Haj *et al* [63]	Considered Confidentiality, Integrity and Authenticity	**No Generalised Software Framework**,
Qasim *et al* [131]	Algorithm-Agnostic Conceptual Framework	**No Generalised Software Framework**, Auto-diagnosis not considered, No Separation of MIW and PACS
Hassan *et al* [66]	Supports Auto-diagnosis evaluation	**No Generalised Software Framework**,Auto-diagnosis not considered, Did not consider existing PACS, Algorithm-Specific Conceptual framework
Borra *et al*[18]	Special MIW algorithm for colured images	**No Generalised Software Framework**, Auto-diagnosis not considered, Shallow consideration of existing PACS, Algorithm-Specific Conceptual framework

6.3 Components of a Software Framework

In software engineering, a framework is the abstraction of guidelines and protocols that provide generic functionality but upon which selective changes could be made to create application-specific solutions that are fit for purpose.

Frameworks enable us to create a non-modifiable base but extensible interfaces. This particular feature makes it possible for specific predefined protocols and standards to be implemented already in this framework. Such well-defined standards will not be overridden but could be extended to provide better capability and functions.

6.3.1 Frozen Spots

Most of the details to be defined in this chapter are mainly the frozen components and communication links in our framework. They generally determine what functions could be called, and they are rarely called by other modules. This feature is utilised to implement not-really flexible medical ethics, standards, and specific disease diagnostic requirements. Also, stringent access control mechanisms could be implemented as frozen spots and cannot be modified by future developers.

The Hollywood Principle[2] or **inversion of control** is a crucial feature of software frameworks, especially in their frozen layer. This principle ensures that external components do not call the framework, but the framework calls them when needed. This can be used to ensure that only the core medical applications call the MIW and MIS functions. MIW algorithms are utilities in this case and thus are called when needed. The parent medical-related functions would form the frozen part of the framework with the validation APIs to grade the IH algorithm based on Table 5.1. Different design patterns [47], such as singleton, factory pattern, and service location patterns, can be used to implement frozen components of a framework. This is important to ensure that the generalised business logic is re-used instead of the detailed implementation that the hot spots would provide.

[2] http://latemar.science.unitn.it/segue_userFiles/2016WebArchitectures/IOC-DI15.pptx.pdf

6.3.2 Hot Spots

The flexibility of a framework depends on predefined spots in the framework where adaptation is possible. This is opposed to the frozen spots with lots of rigidity in design, access, and modification. We re-use specific hot spots through the rigid and generalised frozen spots in the framework. The classes that implement a specific strategy, known as specialisation, for a generalised interface or superclass relates to hot spots. Hence, hot spots are practically implemented in object-oriented software frameworks through the use of inheritance and interfaces. In the **white-box**, hot-spot of a framework, abstract methods without meaningful default implementations can be defined in the superclass in the frozen spot part of the framework. Then the specific applications implementing the method will override the abstract method in the inherited class. The abstract methods are known as the hot-spots through which the framework could be implemented and concretely used to solve a real-world problem. Interfaces are pure abstract classes, as all its methods do not have any meaningful implementation.

Black-box hot spots use composition instead of inheritance. Hence, the useful methods in the superclass are used as-as though new methods can be written to add specific functionality.

Hot-spots can be used to implement algorithm-specific and application-specific requirements. For instance, new algorithms, new medical or engineering quality measures, and new disease biomarkers could be implemented at the hot spots.

As there are many new MIW algorithms, the MIW component of our framework is white-boxed and will use only interfaces. Also, signal processing quality measures such as PSNR, SSIM, KLD, and other parameters that are widely accepted in literature are implemented and black-boxed. Furthermore, medical practice has well-defined ethics, policies, and codes. These aspects of the framework will also be black-boxed, and developers are not allowed to extend at will. These concepts of good framework design are incorporated based on the domain expert's recommendation. Hence, those relatively fixed aspects of engineering and medical quality assurance are only composed of the new medical IH algorithm. The definition of the black-boxes in the framework

requires a joint forum between IH algorithm developers and medical technicians and experts.

In summary, new frameworks like ours are mostly white-boxed. This is because what is fixed, especially in the medical domain, is not programmatically available, and more consultations are needed. As the framework matures, it will evolve into a mostly black-boxed framework with regular major reviews when significant policy changes occur.

6.3.3 Relevant Software Design Principles

For a dynamic system like software, a good design is very crucial. Provisions for new features, algorithms, and integrations should be adequately made. Different design approaches are recommended for different types of software. As ours is almost entirely automated and does not focus on user interaction, we focus on the principles that enable system-to-system interaction rather than user-to-system interaction.

Firstly, the use of software design patterns [47] is widespread in the field of software engineering. Software design patterns are well-known solutions to recurring problems [3]. They help us to avoid reinventing the wheel. However, design patterns are not high-level but low-level principles on implementing codes and instantiating and calling objects. Though there are several design patterns, seven of these have been identified as most important: Singleton, Factory Method, Strategy, Observer, Builder, Adapter, and State [4].

The high-level Software design principles are summarised in Figure 6.3. We explain the relationship between these guidelines and business goals.

There should be **separation of concern** at both the business level and implementation level. For example, the concern of medical experts is different from that of computer scientists. Medical diagnosis and related processes are the major concerns of the medical team. However, they would require services such as secure transmission,

[3] https://www.educative.io/courses/software-design-patterns-best-practices
[4] https://medium.com/educative/the-7-most-important-software-design-patterns-d60e546afb0e

image integrity, patient data privacy, and access authorisation. Scientists will provide these. The module handling each functionality should be separated from each other by design, with loose coupling when accessing each other's service. This improves reusability, maintainability, and quality assurance through testing. **Information Hiding** can be used to ensure that each module access the minimum amount of information from other modules. This allows each module to keep changing its internal implementation without affecting other modules connected to it.

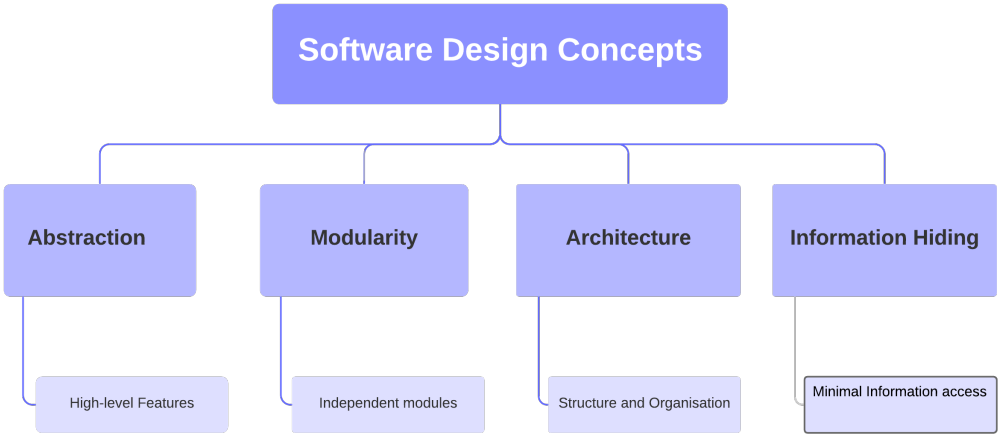


Figure 6.3: Major high-level Software Design Principles

Another notable design principle is that of **interface abstraction** from the actual algorithm. The abstraction layer will decide which components are needed to fulfill the task based on the user's inputs. For example, different types of IH algorithms provide different utilities for medical experts. Patient data privacy, source data authentication, image integrity, and copyright protection are some of these services. All these should be subject to low distortion. As one algorithm cannot optimally provide these services, different algorithms have been developed to provide just one or two of these services maximally. Based on the service requested by the medical expert, the right algorithm would be called. However, the medical expert is expected to make his/her request the same way regardless of the algorithm implemented. To achieve this separation of concerns, one needs to apply the principle of algorithm abstraction using a software

interface.

The design of good Software architecture is essential in a scalable and high-performance software system. Architecture is the structure and organisation of software modules. It defines the vertical and horizontal relationships among modules as well as the number of inputs and outputs going in (fan-in) or going out (fan-out) of a particular module. Architecture can affect the functional and non-functional requirements of a system. For example, an access control module should be placed at the top of the software architecture, and all entry-point modules would fan into this module. Software architecture should be designed to ensure that the primary function of a software system is easily achieved without compromising other critical non-functional features such as speed, security, and maintainability.

Simplicity is a virtue in software design. Hence, a **modular design approach** is paramount. Each Software component should have a well-defined function. Other components can then interact with these separate entities to access the functionality through an access method. Therefore, we relied on **component** and **sequence** diagrams to illustrate the component modules and their interactions within the framework. The data structures and messages within a component will be illustrated using **class** diagrams. These standard diagrams are created according to the Universal Modelling Language (UML). The UML is the standard language for visualizing, specifying, constructing, and documenting the artifacts of a software system [133].

6.4 The Proposed Frameworks

We present a conceptual framework and, from it, design a software framework ***to unify the design, implementation, evaluation, and integration*** of medical image algorithms into teleradiology. Firstly, the conceptual framework is a generalised framework that incorporates the idea in Qasim *et al* in [131] but has a separation of concern. This separation of concern enables independent operation of existing teleradiological systems and IH frameworks but also allows seamless integration of both systems through a

loose-coupled *middle layer*. Secondly, a software framework that enables existing and future MIW algorithms to be easily evaluated and integrated into teleradiology through a standardised evaluation method agreed by both steganographic security engineers and radiologists is also presented.

6.4.1 The Conceptual Framework

This generalised framework is based on the fact that watermarking and Steganography provide certain security services to a teleradiology system. We consider these points at which services from an IH framework is provided as being decoupled from the teleradiology system (PACS) itself. Hence, the PACS system only calls the IH service when required. Figure 6.4 presents the layered framework.

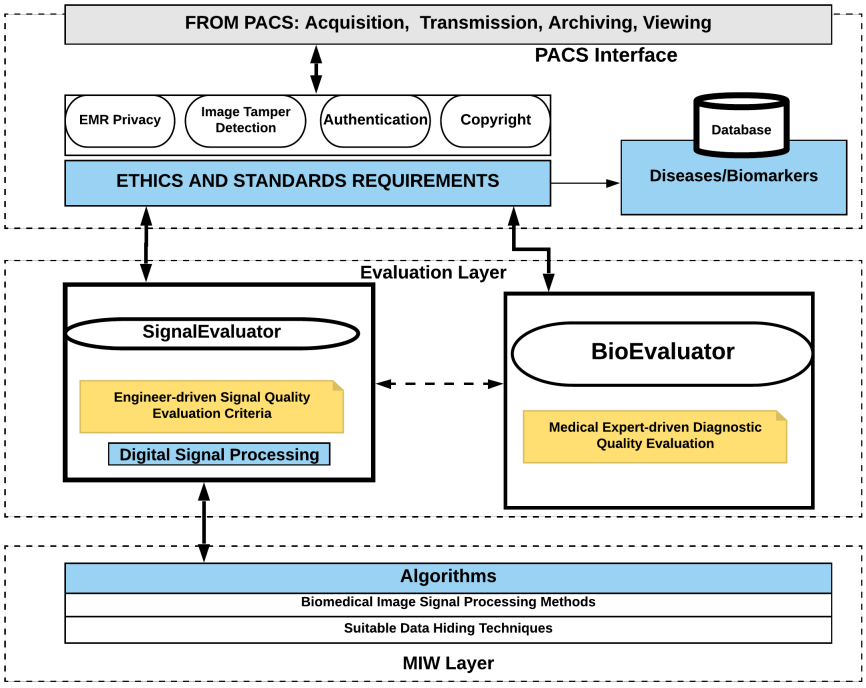


Figure 6.4: Conceptual MIW Framework: A logical view. *The conceptual design lays emphasis on the need to properly evaluate (by the Evaluation Layer) watermarked medical image(from MIW Layer) against the diagnostic information fidelity, subject to ethics and standards defined by the medical experts and sent with the security service request via the PACSInterface.*

There is a need to evaluate and dynamically select an algorithm that is the best fit for service and application, as stressed by [127]. Hence, the PACS (or any external system) makes requests to the IH framework by specifying the **required *services* subject to a *set of constraints*** on the input image in terms of prevailing ethics, standards and specific disease modality. Other required details could be included in the request.

The description of the various layers of the proposed framework is given below.

1. **PACS Interface** - Our framework is an independent system but integrates with other systems through Application Programming Interface (API) and webhooks. The external systems will have endpoints in the framework that will enable them to request specific security service (Privacy, tamper detection, authentication, and copyright protection) of their choice. On the other hand, the webhooks will be used to notify the PACS or other external interfaces when new algorithms are added to our framework. The external PACS can initiate an evaluation process for the new algorithm by performing a mock-up service request to our MIW framework. All service requests originate from the *PACSInterface*.

2. **Evaluation Layer** - This layer quality service delivery to the client (PACSInterface). It ensures the highest level of compliance by taking requests from PACS or other external systems, validating the inputs, and performing some pre-tests to ensure that the appropriate algorithm is selected from the *MIW* layer. However, the two-way communication between its components: *SignalEvaluator* and the *BioEvaluator* are not for the joint execution of this task, but the re-use of common operations. The *SignalEvaluator* component exclusively provides the actual security service requested while the *BioEvaluator* Component independently evaluates the diagnostic integrity of the returned and watermarked medical image subject to ethics, standards, and diagnostic requirements. The decision concerning whether to use the service rendered or not is independent of the *SignalEvaluator* component. This design decision is necessary because medical experts may have to use subject reasoning in some cases. This means that the system can be

set either to **auto** or **manual** mode of operation. In the **auto** mode, no human intervention is required, unlike in the manual mode. The request for bio-evaluation is not from the *SignalEvaluator* but from the *PACSInterface*. This enables the medical experts to modify the evaluation framework continuously as it suits medical practice.

3. **MIW Layer** - This layer contains the actual algorithms and signals processing functions that provide the required security services. This includes past, present, and future algorithms that provide related security services. We will not discuss this layer in many details as they formed the bulk part of this thesis (Chapters 1 to 4). However, a specific implementation of two more algorithms will be utilised in the experimental algorithms case study presented in Section 6.5.

Further details about the components in each layer and their relationships are provided in the core software design presented next.

6.4.2 The Software Framework

The software framework is designed directly from the conceptual framework presented in Figure 6.4. In this section, the Unified Modeling Language (UML) tools are used to present both the high-level and low-level specifications and designs for the framework. Specifically, we utilised the **Component** diagrams, **Class** diagrams and **sequence** diagrams. The system's larger components are described using the component diagrams, which also shows the classes in each component. The relationship among classes is shown using class diagrams, while the interactions among the major classes in order to solve a problem are shown using sequence diagrams.

6.4.2.1 Software Components

The four major components (packages) and associated modules are grouped in a layered architecture in Figure 6.5. These include: ***PACSInterface***, ***SignalEvaluator***, ***BioE-***

valuator and ***MIWAlgorithms***.

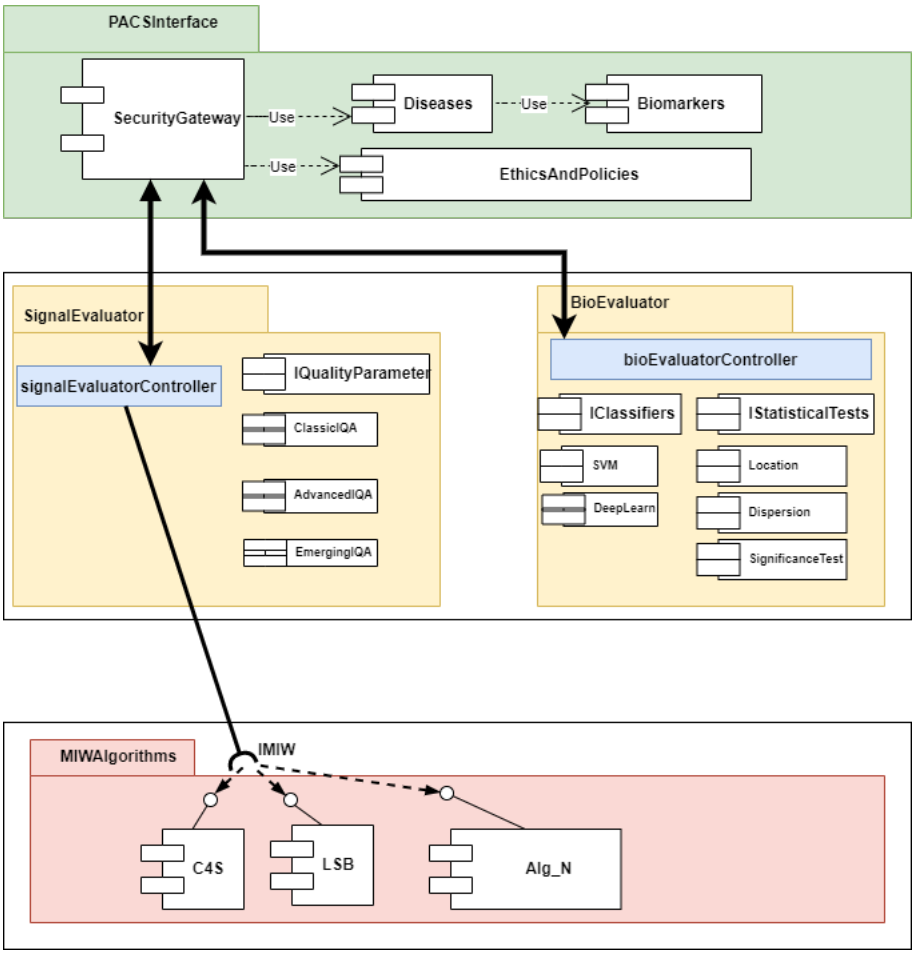


Figure 6.5: Software Components in the Software Framework: *To enable unification and dynamic selection of algorithms, the signalEvaluator requires a software interface (IMIW) to access each concrete class that implements this interface.*

Further description of these components is given below.

1. ***PACSInterface*** - For existing and future PACS systems, an API interface and endpoints crafted to send and receive data from our framework is needed. The format and details of each endpoint depending on what service or combination of services are required from our MIW framework. As a start, we have defined four major service requests from a PACS system: *Privacy* of patient's record, *Integrity* of the archived or transmitted scan, *Authenticity* of the information sources and

Copyright protection of the image and its content. The input data come from this layer. These include the actual scan, type of disease, constraints for watermarking and the electronic data to be watermarked. An important part of this component is a reference to *medical ethics, standards, and policies*. This provides the constraints to be considered for algorithm selection.

2. ***SignalEvaluator*** - As mentioned by Nyeem in [127], there is a need for a framework that dynamically selects an optimised algorithm to deliver watermarking and steganographic security services. This component's primary function is to combine diagnostic requirements of the specific disease, its modality constraint, and the digital signal processing techniques to pre-evaluate the existing algorithms and then select the best algorithm. Random blocks from the image (sampling method) are chosen for this algorithm selection process.

3. ***BioEvaluator*** - This component is dedicated to medical practice. As shown in Ludewig *et al* [94], the medical technicians already have parameters for evaluating the quality of scans produced from different machines and in different settings. They also know more about scan modalities as well as the specific disease biomarkers that can be computed from a diagnostic scan for disease using any of the image modalities. Furthermore, the modern era of AI, machine learning, and autodiagnosis are already taking place in medicine. Hence, this module is dedicated to the evaluation of the watermarked medical image scan using medically-defined parameters. The quality parameters used here are different and independent of those defined in ***SignalEvaluator***.

4. ***MIWAlgorithms*** - This is the greatest white-box hotspot of our framework. It defines a standard interface that any MIW algorithm designer must implement. It is a repository for new algorithms that could provide a better security service for the radiologist. The actual methods in the software interface(IMIW) have been defined in Table 6.1.

6.4.2.2 Software Classes and their Interactions

Each of the components defined in Figure 6.5 is made of classes that implement some core functionalities of the software framework. As a use case, the radiologist at the local hospital selects a patient's X-ray scan and the corresponding medical record. This information will now be submitted to the *PACSInterface* through the part of the framework shown in Figure 6.6.

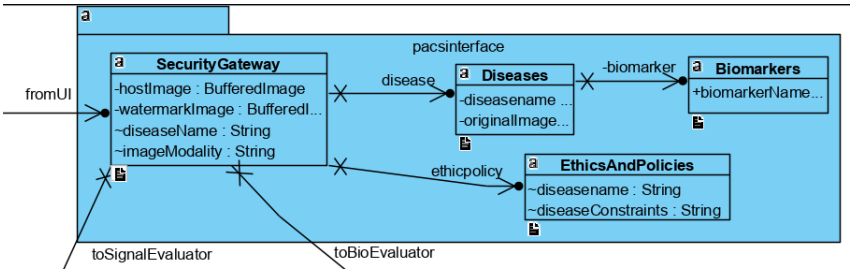


Figure 6.6: External PACS Interface for data hiding security service request

There are four major modules in the ***PACSInterface***. The *SecurityGateway* is a controller class that coordinates data input, the formulation of specific data related to the disease in question (pediatric pneumonia in our case), and then the handing off and reception of the medical data for the provision of the required security. It receives medical image scan and patient data from the User Interface (UI), such as connected medical scan equipment and Hospital Information System (HIS). It then uses the *Diseases* class to store disease name, severity, image, and non-image biomarkers, associated with the disease. The *Biomarker* class is a knowledge base of different diseases and their corresponding biomarkers for its prediction and classification. After gathering this information, the *SecurityGateway* will then query the *EthicsAndPolicies* class to obtain general and disease-specific constraints that could be applicable to image modification and patient data privacy. These policies are usually codified in a JSON key/value pair. The gathered information is then passed to the ***SignalEvaluator*** module, which selects the best algorithm required to provide the integrity and data privacy security services subject to the defined constraints included with the request.

The ***SignalEvaluator*** module is shown in Figure 6.7 with its component classes. As

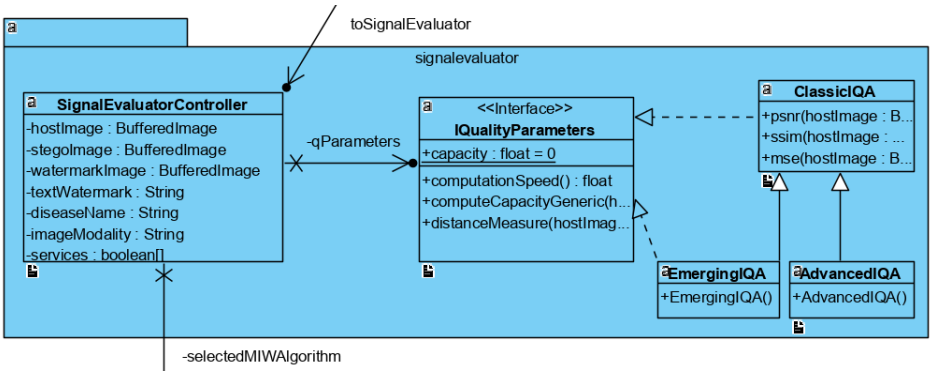


Figure 6.7: The SignalEvaluator Component: *classes for quality assessment implemented through an interface to ensure that new IQA can be added in the future. It also contains a method for MIW algorithm selection*

usual, there is a controller class - *SignalEvaluatorController*. It coordinates the internal and external functions of this component. The major function of this module is to accept inputs and requests from the ***PACSInterface*** and uses both classic, advanced and emerging image quality assessment (IQA) parameters to determine the best algorithm to fulfil the security service requests (integrity and privacy in our case). Hence, for each of the candidate algorithms existing in ***MIWAlgorithm*** module, the quality parameters are computed, compared and then the best algorithm is selected to fulfill the request.

The ***MIWAlgorithm*** module contains an interface called *IMIW*, which every new data hiding algorithm must implement. The methods in this interface originated from the analysis and synthesis presented in Table 6.1. The translation into class methods is shown in Figure 6.8.

In this chapter, only our *C4S*, *ZainEtal* [163] and LSB replacement [10] algorithms were implemented for case study analysis. The concrete implementation of this interface by these few algorithms is shown in Figure 6.9.

The dynamic selection of the best algorithm to fulfill a request was achieved in this research using the **Strategy Pattern** in Software Engineering. In the Strategy pattern, the behaviour of a particular class can be changed at run time. The ***signalEvaluatorController*** class in the middle layer of the framework takes the request from the teleradiologist through the *PACSInterface* layer, randomly selects sub-blocks of image and

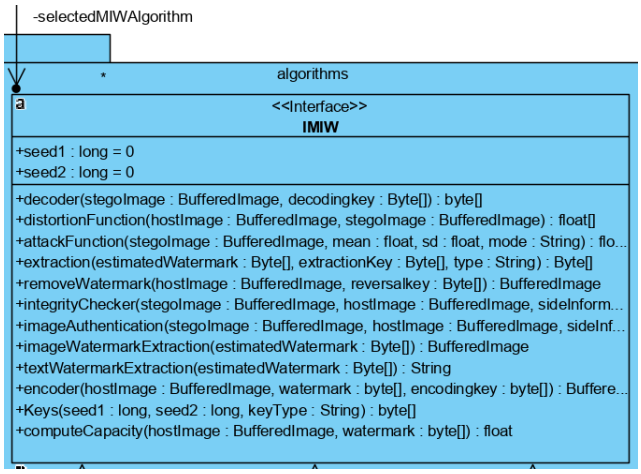


Figure 6.8: The IMIW Interface: *the major software interface that unifies all data hiding algorithm implementation. It is a programatic representation of Table 6.1*

then uses them to run across the different IH algorithms in the MIW/S layer in order to determine the best algorithm (strategy) for achieving the best result for the radiologist. This dynamic algorithm selection can only happen if the individual watermarking algorithms are implemented via the *IMIW* interface.

During the pre-selection by *SignalEvaluatorController*, an array of the IMIW interface is created, and each element of the interface is instantiated with a concrete algorithm. Hence, the different methods for each of the algorithms to be appropriately called.

After the watermark encoding to provide privacy and integrity, the watermarked version of the image is returned to the ***PACSInterface*** module. On receiving the watermarked image, the ***BioEvaluator*** module is contacted for a second evaluation to ensure that the medical experts are satisfied with the image quality. The classes in the ***BioEvaluator*** are shown in Figure 6.10.

Again, the ***BioEvaluator*** has a controller class, *BioEvaluatorController*, which handles communication between the internal classes and requests from outside the module. This ensures encapsulation and cohesive class design. It is made up of two major evaluation software3 interfaces: *IStatisticalTests* and *IClassifiers*. It has been shown that in radiology, statistical tests of various kinds are used to determine the quality of different images from different machines and parameter settings [25, 94]. The existing sta-

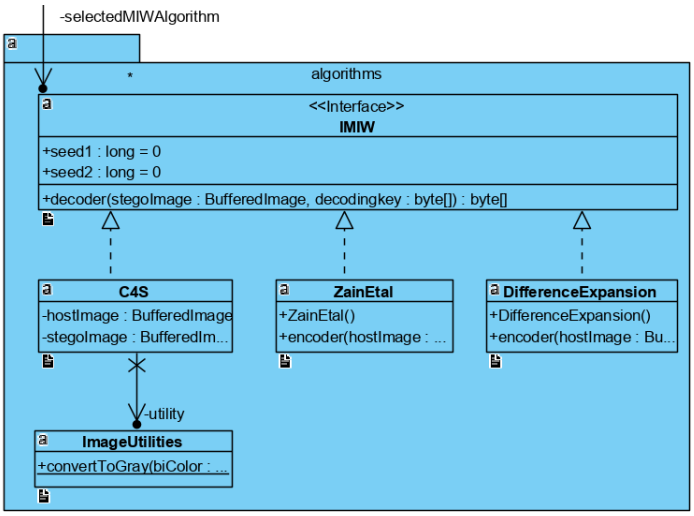


Figure 6.9: The Concrete Implementation IMIW interface: *The C4S, ZainEtal, LSB and dummy DifferenceExpansion algorithms were implemented in this research*

tistical tests include measure of location (implemented in *LocationTest* class), measure of dispersion or spread (implemented in *DispersionTest* class), and different statistical significance tests (implemented in *SignificanceTest* class).

For the sake of a more transparent presentation and simplicity, the sequence diagram, which shows the interaction between only the significant classes in the proposed framework, is shown in Figure 6.11. The essential messages or operations between the main objects and modules in each layer were considered.

The source files for the executable architecture code for this framework can be found at: https://github.com/KingPeter2014/MedStegFramework/tree/master/src.

6.5 Experimental Results

To illustrate the usage of this framework, selected algorithms were implemented to generate data that illustrate how the unified ***MediSteg*** framework works.

We have implemented our C_4S algorithm (Chapters 3 and 4), that of Zain *et al* [164] and LSB replacement steganographic algorithm [10] as the case study algorithms. There are three major challenges that this case study experiment focused on:

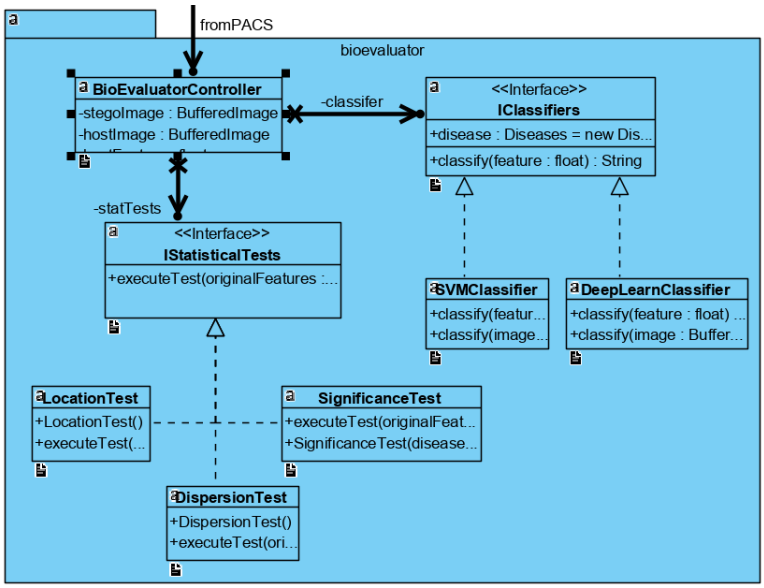


Figure 6.10: The BioEvaluator module: *Utilises medical image biomarkers to perform medical statistical and machine learning evaluations on watermarked medical image. This validates the applicability of data hiding security technique for autodiagnosis*

1. Preservation of the privacy of the patient record.

2. Preservation of the integrity of the medical image scan

3. Preservation of the diagnostic information to ensure accurate classification at the remote AI system.

The Graphical User Interface (GUI) used for testing the executable architecture of the proposed framework is shown in Figure 6.12. The input data from this GUI is passed to the *SecurityGateway* class of the *PACSInterface* component.

The summary of the result is shown in Table 6.3. The scores for each of the criteria were computed as follows:

1. **Integrity** - Integrity is computed as a percentage of the number of sub-blocks with correct detection of attacks. An algorithm that offers integrity verification service is expected to implement the *IntegrityChecker, IC(.)* function as described in Table 6.1. If this function is not implemented by the algorithm, then the *Integrity* score

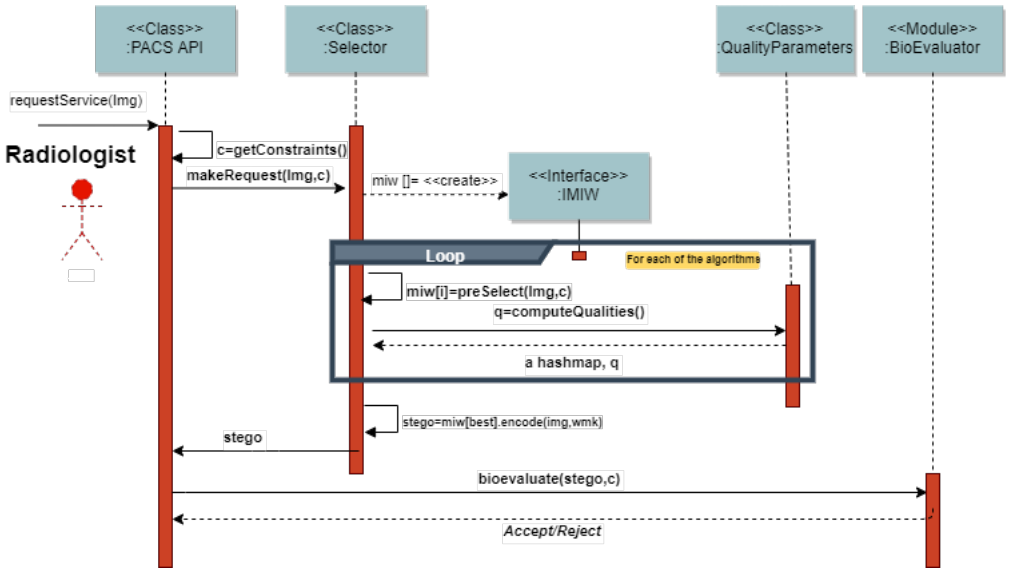


Figure 6.11: MIW Object Interaction

is for such algorithm is 0%. Hence, if a total of A sub-blocks were attacked in ROI and B blocks detected by the IC() function of the algorithm, then the *Integrity* score is computed as:

$$Integrity = \begin{cases} 0, & \text{if } IC = Null \\ \frac{B}{A} \times 100, & Otherwise \end{cases} \tag{6.2}$$

In this research, we have used *Contrast Adjustment* and *Guassian noise* of various strengths as the attacks that we want to detect. This is because they are common attacks that change pixel intensity. Other attacks demonstrated in Section 3.3.2 can as well be tested for and then an average taken.

2. **Privacy** - We have combined capacity and bit error rate (BER) to compute the privacy criteria. However, we have defined capacity differently for this purpose. Let **M** be the total number of bits of information to be hidden and let **C** be the amount of information that can be embedded into the image at allowable distortion,D_1,

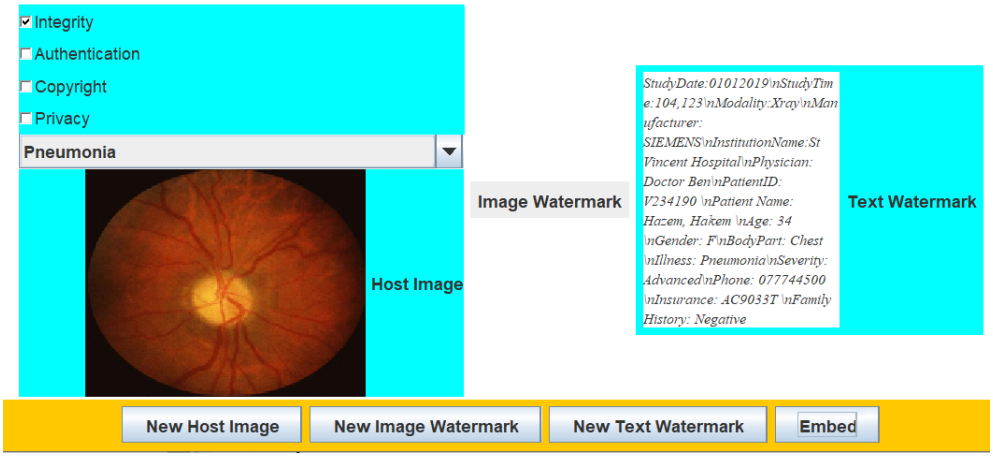


Figure 6.12: User Interface for Testing Proposed Framework

then we define capacity thus:

$$Capacity = \begin{cases} 100, & \text{if } C >= M \\ \frac{C}{M} \times 100, & Otherwise \end{cases} \tag{6.3}$$

After computing capacity, then the privacy parameter is computed as:

$$Privacy = \frac{BER + Capacity}{2} \tag{6.4}$$

The BER must be expressed in percent and not as a fraction. This definition captures the ability to securely hide the EMR data as well as correctly decode the EMR at the receiving end with or without attacks.

3. **Distortion** - The distortion parameter utilised here is the PSNR. Huynh-Thu and Ghanbari [76] has proven that PSNR remains an effective quality measure between two images with the same original content, even though it might not be a good parameter to measure the quality of entirely different images. Hence, it is a valid measure as we are using different algorithms on the same image and then comparing their quality. However, even though it is computed using the equation

described in Section 2.6.2, we have modified how it is used as a distortion parameter in comparison to other algorithms. Let there be N selectable MIW algorithms in our framework. Let the PSNR-based distortion as a result of using i^{th} algorithm be D_i, then we compute the *Distortion* criteria thus:

$$Distortion = \begin{cases} 0, & \text{if } D_i < 40dB \\ 100, & \text{if } D_i = \infty \\ \frac{D_i}{\sum_{i=1}^{N} D_i} \times 100, & Otherwise \end{cases} \tag{6.5}$$

Note that in our computation, higher *Distortion* percentage does not mean higher distortion but lower distortion instead. This reverse computation is necessary to ensure that the algorithm that comes out with the highest average total score in Table 6.3 is actually the best algorithm to provide the service.

4. **Location Change** - For every quantifiable medical image biomarker used to classify or characterise a disease, a value is computed by the radiographic machine or a software system. The percentage deviation from the original value for an individual as a result of watermark addition is computed here. Let **X** be the original biomarker value before watermarking and **Y**, the recomputed value after watermarking, then compute the biomarker's location change for a medical image scan as:

$$LocationChange = 100 - \frac{|Y - X|}{Y} \times 100 \tag{6.6}$$

Hence, algorithms with a larger deviation from the original biomarker value would be penalised with a lower score.

5. **Dispersion Score** - According to Chaleunvong [25] of the World Health Organisation(WHO), medicine is a quantitative science but not as exact as physics or chemistry. Also, he mentioned that diagnosis and treatment are based on proba-

bility. Hence, for every quantifiable biomarker used to classify disease, there is a range for such statistics for a *normal*, a *benign*, and a *malignant* case. For example, a normal systolic B.P. ranges from 100-140mm of Hg. We now define the *Dispersion-Score* as a binary-valued score of either 0 or 100. Let the original image biomarker before watermarking be of the class **X** because it lies within the range X_1 and X_2. Let the recomputed value of the biomarker after watermarking be **Y**. Then:

$$DispersionScore = \begin{cases} 0, & \text{if } Y < X_1 | Y > X_2 \\ 100, & Otherwise \end{cases} \tag{6.7}$$

The rationale behind this scoring is that the security service provided by data hiding techniques should remain valid provided that it maintains diagnosis within its original class. Conversely, if there is a shift in class due to watermark embedding, the IH algorithm should be severely penalised. This scoring, unlike *LocationChange*, is not based on the magnitude of the change but on its tendency to alter disease classification by either humans or autodiagnostic systems.

After profiling the samples chosen from our dataset, the following range was computed for the homogeneity feature: Normal($X_1 = 0.920$ to $X_2 = 0.955$) and Pneumonia($X_1 = 0.960$ to $X_2 = 0.990$). Hence, if the homogeneity of an image was within one of these ranges before watermarking, the value must remain within that particular range to get a Dispersion Score of 100. Otherwise, it will be zero.

6. **SVM Accuracy** - The Accuracy score for an SVM classification was defined and utilised in Equation 5.6. Firstly, a model is built from the watermarked images generated by each of the algorithms. Then, their classification and prediction accuracy are used as one of the criteria. However, we recommend that the average accuracy, recall, and precision in place of this score to handle unbalanced data.

7. **Attack Response** - Irrespective of the service being provided, the retrieval of em-

bedded watermark after attacks remain important. The ability to retrieve information after a specified attack is what we call *AttackResponse*. This is defined mathematically as:

$$AttackResponse = 100 - BER, \tag{6.8}$$

where *BER* is expressed as a percent and not a fraction.

8. **Average Score** - It is obvious at this stage that we have generally scaled individual scores from 0 to 100. It is also very likely that new parameters may be introduced in the future or some, which may appear as redundant, removed. For consistency and accuracy, the *AverageScore* is used for final algorithm selection so that it will not be affected by the absence of a parameter in the final computation. Hence:

$$AverageScore = \frac{SumOfIndividualCriteriaScores}{NumberOfCriteriaConsidered} \tag{6.9}$$

We are aware that higher weighting may need to be assigned to given criteria when necessary, but this was not considered for this thesis.

We evaluated this framework using integrity and privacy as the services being delivered by the three candidate steganography algorithms whose performances are shown in Table 6.3. Hence, the goal is to use our framework to select the best algorithm to fulfill the mentioned security services. The attack that the medical expert is interested in detecting is *Contrast adjustment*. The biomarker of interest to the radiologist is *Homogeneity* because it gave the best accuracy during classification in Chapter 5.

From Table 6.3, we can see that ZainEtal [163] scored poorly in most of the criteria. The very low score was because it failed the two critical tests set by both the signal processing engineers and medical experts. First, it scored zero for distortion because the 37.1dB is below the minimum set for medical images [49, 29]. Secondly, after watermarking, the homogeneity of the medical image fell outside the range for 'normal' patient, which it originally was.

On the other hand, C_4S had the highest overall score and passed the critical tests

Table 6.3: Algorithm Selection Criteria via the Unified MediSteg Framework. Based on predefined scoring techniques, the best algorithm will be selected to deliver some security services for data security in teleradiology. For ZainEtal, the 37.01dB for distortion was converted to 0dB before average is computed

	Algorithms		
Criteria	C_4S	ZainEtal [163]	LSB [10]
Integrity	97.77	0.00	0.00
Privacy	100.00	50.00	100.00
Distortion	50.96	37.01	51.43
Location Change	99.92	94.03	100.00
Dispersion Score	100.00	0.00	100.00
SVM Accuracy	86.13	57.92	86.63
Attack Response	68.00	0.00	0.00
Average Score	**86.11**	**28.85**	**62.58**

apart from providing the required security services. LSB [10] has an excellent performance in many aspects, but the algorithm was designed to be fragile and provide privacy for an almost already secure and noise-free environment. It was not designed for integrity checks. The LSB replacement method has an inadequate response to attacks. With contrast adjustment, the length of watermark stored in the first 32 pixels of the image changed from **'0330'** to **'[B'**. Because the later character cannot be converted to a number, it is no longer possible to retrieve the watermark nor localise other attacks. Thus, although it has a higher PSNR than the C_4S algorithm, it does not robustly provide integrity checks.

For the C_4S algorithm, we have separated watermark retrieval from integrity checks. We do not need the length of the embedded watermark to check tamper detection. Tamper detection is based on the agreed value ρ, which is either robustly embedded in the RONI or encoded in the encoder and decoder systems. Hence, the C_4S is selected to provide this service.

It should be noted that C_4S did not have the highest score in all criteria. There are three major criteria that an algorithm needs to satisfy to have a chance of being selected:

1. Successfully provides the requested security service(s).

2. fulfill at least the 40dB minimum for *distortion*, set by DSP engineers.

3. Score 100% on *DispersionScore* according to the range set by the radiologist or medical experts.

Having fulfilled the above, then higher scores in these and other criteria, which the radiologists provided as part of the **EthicsAndPolicy Constraints**, will aid in the emergence of an algorithm. All constraints will have to be translated to a measurable quantity and normalised to 100%, as we have done in Table 6.3.

6.6 Key findings

Below are the major findings from the results and synthesis in this chapter.

1. The existing specifications for information hiding (IH) inputs, outputs, and functions do not state the specific security services (integrity, privacy, authentication, or copyright protection) they provide. We have refined these algorithmic parameters and included new specifications that encompass both Steganography and digital watermarking. (See Table 6.1).

2. To remove the barriers to the adoption of data hiding techniques for providing medical image security in practice, a unified model and framework are required for algorithm design, implementation, evaluation, and evolution. Therefore, we identified the necessary software abstraction techniques - interface, inheritance, and polymorphism - to ensure that medical experts can focus on security service requests and diagnostic quality evaluations. Simultaneously, algorithm designers can easily add new algorithms that can satisfy these medical security services

without being concerned about evaluation and integration tools for the new algorithms. This is guaranteed, provided that each new algorithm exposes common inputs, outputs, and functions as defined in our framework. This approach unifies the wide variation in the implementation details and makes existing and future algorithms to be evaluated and utilised maximally. (Section 6.4.2)

3. We designed our algorithm selection parameters and software framework in fulfillment of the need for algorithm selection framework, which was a research gap mentioned in [127], and in fulfillment of the diagnostic concerns often raised by physicians. This approach increases the confidence of both providers and consumers of autodiagnostic healthcare systems in teleradiology (Section 6.5).

6.7 Discussion

We believe that a framework provides the guidelines, benchmarks, and protocols upon which design, implementation, and evaluation of new technologies can thrive. A framework also includes actual tools with which inventions and innovations can be implemented and tested before deployment. Unfortunately, technologies are often developed before end-users begin to think about frameworks for their adoption and standardisation. Lack of standards could lead to chaotic and non-compatible implementations.

We cannot overemphasize the importance of standard protocols and tools for the widespread adoption of a technique or technology. As an analogy, the adoption of the internet as a global communication network is due to the standardisation of the protocols and frameworks such as the Internet Protocol(IP), Transport Control Protocol (TCP) and IP security. Without these standard specifications and uniform implementation across hardware and software devices, it would have been difficult to establish a global communication network using devices and software from different vendors. Furthermore, Castro [24] posits that one of the major challenges that slow down the adoption of new technologies is the lack of frameworks and necessary tools. To make any useful technology to be widely adopted for practical use, a framework and a set of

standards are required to exist. This concern, which is currently applicable to medical image steganography, led to this chapter's contributions.

In our first research finding, the existing specification of watermarking/steganographic inputs, outputs, and functions was updated to state the specific security service they provide clearly. We argue that providing functions for embedding and extraction is not enough to specify a data hiding algorithm. Such limited provision blurs the fact that end-users (teleradiologists) are more likely to think in terms of the services they require than in how they are provided. This led to an updated specification in Table 6.1. Unlike the specification by Abunyeem [127] (Chapter 3), our updated specification included specific functions that provide specific security services. Also, because data hiding security techniques are not limited to digital watermarking as in Nyeem [127], security functions provided through Steganography were included. The implication of this is that even though computer scientists will still be concerned about the subtle differences between watermarking and Steganography, these details will be abstracted away from the teleradiologists. Hence, teleradiologists will only be concerned about a single framework that provides all their security services. With these details abstracted away from the end-users, they will feel comfortable to use *these fewer services* and not *those numerous algorithms*.

In response to the fourth research question in Section 1.5, we went further to identify the software abstractions that could be used to unify the practical implementation of data hiding algorithms in teleradiology. We recognised that for wider adoption, both a software framework and a robust evaluation mechanism are indispensable. We argue that the use of a common *Software interface* for all underlying watermarking and steganographic algorithms is a strong unifying design. With this design approach, every designer must expose the same algorithmic parameters(input, output, and functions) to the outside world but could implement other details required to make the performance of his algorithms better differently. For example, a teleradiologist simply calls an *integrityChecker()* function but will not care if it is from the C_4S, *ZainEtal* or *LSB* algorithm. For consistency, it is recommended that even if an algorithm designer is

not going to implement an *integrityChecker()* function, they must include the function in their software module and simply return a *null* value from the function. With this standard implementation, all algorithms can be evaluated against the same standard parameters and criteria, which we have defined in Section 6.5.

Apart from enabling a unified framework for faster adoption, software abstractions (interfaces), and hierarchical design through inheritance are employed in design to ensure maintainability, scalability, and reusability. Maintainability will increase adoption as the overhead cost will be reduced. Scalability will ensure that new challenges arising in the future can be easily incorporated into the framework, while reusability ensures that the efforts and time spent creating a similar system is minimised [97]. Hence, once the medical community has access to the top-level interface, they will not need to know if new algorithms are being added or removed from the framework. They can continue to access new and improved security services through the same interface.

As a direct response to the research gap raised by Nyeem in [127] (a need to develop a framework that can dynamically select the best IH algorithm that fits the properties of a medical image), we went further to design the two evaluation layers in our **MediSteg** framework: ***signaEvaluator*** and ***bioEvaluator***. Whereas the ***signaEvaluator*** is made up of functions already existing in image processing and applications, we provided the ***signaEvaluator*** as a new and necessary module that should consist of quality assessments that strongly relate to the parameters used for disease diagnosis and treatment. For example, the value of biomarkers is used by medical experts to diagnose a disease. Hence, one of the reliable evaluations that we have proposed is *DispersionScore*. This is a score that measures if the value of a biomarker of interest has gone outside the accepted range for the disease condition after watermarking. This type of measure has not been considered before now, but it is very significant in diagnosis [25]. The importance of the new measures that conform with the medical diagnostic procedure is to give more control to the medical experts in choosing when to accept or reject the security services obtained, subject to the potential effect of added security watermark on the medical image biomarkers, and hence, diagnostic outcome. Therefore, the results in Table 6.3, it is

clear how the best algorithm can be selected for a particular purpose using our framework. This contribution is believed to have filled this research gap raised by Nyeem [127].

In comparison, a closer and more recent attempt to solve the above gap for algorithm selection is found in [131]. The frameworks and solutions are primarily defined at the conceptual level. We have moved closer to practice in this work by improving and then converting the existing IH models and conceptual frameworks into an implementable software framework.

To further address the barrier to the adoption of data hiding security techniques in teleradiology, we have also included health ethics and policies in the security service requests to ensure that legal and ethical concerns are considered while selecting algorithms while evaluating the watermarked image. This is often not taken into consideration by existing systems. Whereas one believes that data hiding can provide confidentiality, integrity, and availability security services as required by different laws such as HIPPA [73], it should be noted that any significant modification to the image could cause diagnostic failure. Hence, we have taken this constraint into account through a ***distortionlimit*** parameter that can be retrieved for each image modality and disease using the **EthicsAndPolicies** class.

Finally, the intended impact of this chapter's contribution is to move closer to practice because many MIW algorithms exist, but not much is implemented in hospitals and teleradiology. We have shown that the *inputs, outputs, functions, services, and service access interface* are the necessary features that need to be unified across all data hiding algorithms. With this unified framework and relevant evaluation criteria, *a lower barrier to usage* will be provided. The provision is through the guidelines, baselines, tools, and protocols provided by and deduced from the framework. With the barrier to entry and experimentation removed, adoption and evolution will be guaranteed.

6.8 Summary

This chapter focused tremendously on bringing the entire contributions in this thesis and that of the works of others into practical use. Specifically, it updated the systematic models initiated by Nyeem [127] by including more specific inputs, outputs, and functions, and then defining ways that they can be implemented in a practical software framework. In [127], only theoretical definitions known as *mathematical formalism and operational determinism* were given. Our framework has taken this systematic approach of data hiding development closer to practice by defining a framework for its implementation and evaluation.

We have evolved new evaluation techniques from existing quality assessment parameters for both computer scientists and medical experts. There are new ways of interpreting the statistical and biomarker data generated from signal processing and medical image steganography. This new approach is intended to take cognisance of and reward parameters that ensure that all the conflicting goals in medical image security via Steganography are achieved optimally.

The new evaluation technique and the unified algorithm selection framework are believed to increase the adoption of Steganography and other data hiding security techniques in teleradiology. This is because they utilise the evaluation methods that are very close to those used in clinical trials for new drugs and because they give the medical community a choice of algorithms via a single framework. All these will reduce the barrier to trying out new algorithms until the best is found.

With this framework in place, and through its future refinement, it is hoped that a standard will evolve towards secure autodiagnosis in teleradiology with a low implementation cost.

Chapter 7
Conclusion and Future Works

This chapter contains summary of contributions, thesis conclusion, recommendations for practice and future works.

Go To Chapter: 1, 2, 3, 4, 5, 6

7.1 Summary of the Main Contributions

This thesis presented a research that aimed at improving tamper detection and data hiding capacity of spread spectrum (SS) medical image steganography. Then, it provided a robust evaluation framework to determine the applicability of any information hiding (IH) security scheme for autodiagnosis in teleradiology. SS steganography has been identified as one of the robust security schemes that can overcome active multimedia security attacks in open networks and noisy environments. However, these robust features have limited its use to only tamper-resistance and low data capacity applications, thereby reducing its applicability to diverse conditions such as where tamper detection and higher volume of data needs to be transmitted. In addition to this challenge, the emerging security and privacy needs in teleradiology over open networks and poor security infrastructure makes teleradiology a good candidate for evaluating and validating any new SS IH algorithm. This challenge of improving spread spectrum steganography and using it to solve emerging security challenges in teleradiology motivated this research. Hence, we investigated and designed a new spread spectrum based tamper detection and data capacity improved algorithm, and a generalised evaluation

framework for any data hiding security technique.

In the course of solving the challenges posed by this thesis, the following main contributions were made:

1. The design, analysis and evaluation of the Spread Spectrum Constant Correlation Compression Coding Scheme (C_4S) for medical Image Integrity verification and zero Bit Error Rate (BER) watermark detection. From the perspective of SS steganography, this dynamic algorithm, which could serve as a semi-fragile or robust watermarking algorithm, will now provide both tamper-resistance and tamper detection capabilities depending on the configuration of the error tolerance parameter,ϵ. For teleradiology, this is necessary for ensuring correct diagnosis over a network that is vulnerable to malicious data modification attacks. Tamper resistance feature would ensure embedded EMR is retrieved despite attacks, while tamper detection determines whether a scan is still reliable as diagnostic data.

2. The design and evaluation of a new capacity improvement algorithm by extending the C_4S method and implementing amplitude modulation techniques. Both theoretical spread spectrum methods, as well as heuristics specific to medical images, were combined in order to increase the amount of data that can be added and successfully extracted from medical images without distorting diagnostic information. Without heuristics, one can achieve up to 12 bits per sample for 16-bit DICOM images instead of 1 bit per sample or simple identification code signature in classic ([38, 40]) SS watermarking techniques. This result is mainly for the region of non-interest of medical images.

3. A statistical as well as machine learning evaluation of the effect of watermarking on diagnostic biomarkers. The significance is to gain direct insight into any adversarial threat that medical image watermarking may pose for remote autodiagnosis using machine learning models. This knowledge will help to create better medical image hiding algorithms. This study established that the design and eval-

uation of watermarking for the human visual system are different for that which will create images for use in machine vision and computer-aided autodiagnosis.

4. Analysis of existing models and then the design and evaluation of a new unified software framework for selecting suitable medical image data hiding techniques to provide a given security service. The significance of this contribution is to bridge the gap between theory and practice. This software framework is a practical implementation of existing theories, models and conceptual frameworks. With this, it is easy to establish what works and what needs to be redesigned (Chapter 6). In a more specific sense, this contribution advanced the formal and systematic watermarking design methods in Nyeem [127] into a formalized software design and implementation framework. This advancement implies a gradual progress towards the adoption of data hiding in medicine through the availability of software tools that evolved from a systematic design and not from ad hoc algorithms.

This thesis, in line with the above specific original contributions, advances knowledge and enhances practice in the field of spread spectrum medical image steganography, in particular, and multimedia security applications, in general.

7.2 Conclusions

Based on the empirical analysis of different medical image modalities and the systematic analysis and evaluation of current practice in implementing and adopting medical image data hiding schemes, the following major conclusions were drawn:

1. ***Image content (complexity), correlation properties of spreading code and allowable distortion for an application are the important factors to be considered when improving both tamper detection and Steganographic capacity of spread spectrum methods***. The results indicate that controlled embedding designs should be sought, especially in the region of interest (ROI) of medical images.

2. ***By incorporating relevant image biomarkers used for medical diagnosis during algorithm evaluation, one can robustly evaluate an algorithm for use in teleradiology security***. The results revealed that whereas some of these image biomarkers (features) are robust to steganographic algorithms, others are fragile. Quantifying the effects of steganography on these biomarkers through a unified and standard framework is the key to future adoption and the use of steganographic algorithms in providing security for telemedical data utilized during remote diagnosis. Before this research, these image features measured by radiologists for disease diagnosis are not being considered by steganographers while evaluating data hiding algorithms.

The evidence and approach provided by this research are relevant to addressing the concerns of both the patients and the medical experts involved in telediagnosis. The patients do like to get quick service without compromising their privacy. The medical experts are committed to providing these services correctly as well. Hence, the means of providing quick, secure service should not compromise accuracy. In this thesis, the security services being referred to, in practice, are the integrity of the transmitted medical image scan and the privacy of the associated medical record (EMR) embedded into it. With this practice, an external medical expert (or AI system) who is not part of an internal Hospital Information System (HIS), can be handed over the minimal information required to diagnose a patient's case. This approach is what we have termed a **one-image medical record**. It will contain the medical image scan, into which patient ID, scan report and other test results are embedded. The remote expert or AI system would have a lightweight application that extracts this information provided it has the requisite secret key for data extraction and integrity verification. The significance of this thesis is in designing such algorithms and ensuring that it is fit-for-purpose.

Unlike many other researchers in the last decade who completely avoided the region of interest of medical images or focused only on the reversible data hiding schemes, we have researched the ROI embedding more thoroughly. Our approach is justified by the fact that authorized post-processing and unauthorized ROI attacks are still per-

formed on the ROI despite stipulated legal and ethical concerns in medicine. We strongly advocate that the only practice necessary for adopting medical image IH security techniques is to subject each candidate algorithm to a benchmarked unified and rigorous scrutiny framework. If an algorithm could pass similar statistical and diagnostic tests as other re-construction and pre/post-processing algorithms in radiology, we do not see any reason why it cannot be approved by a medical ethics committee to be used to provide security services in teleradiology.

Hence, we recommend that the frameworks developed in this thesis should be used as a baseline tool for uniformly evaluating and adopting information hiding security algorithms for use for autodiagnosis in teleradiology.

7.3 Future Research

There are various open questions and design challenges that have been made more evident by this research.

1. **Designing a reversible data hiding algorithm based on the** C_4S: This would enable one to perform both active and passive detection of medical image tampering while increasing hiding capacity as well as preserving the diagnostic quality of the medical image. To achieve this, one needs to use the HSI knowledge concerning the spreading sequence to come up with a single embedding strength α, that could be used across all sub-blocks. Parameter tuning of error tolerance, ϵ, could still help to reduce false negative and false positives of watermark detection.

2. **Design of Machine Learning-based Steganographic algorithms**: An investigation is necessary towards (i) designing future steganographic algorithms that hide information in the model itself without compromising the classification accuracy of future predictions, (ii) designing machine learning models that are not sensitive to selected features that are so fragile such that any little modification will lead to the adversarial effect. In the first case, although more works have been

done in the use of deep learning for steganalysis, preliminary works [1] show that deep learning can as well be used for steganography. In the second case, a closer study into feature engineering for deep learning algorithms is required.

3. **Extension of the *MediSteg* framework**: To include more algorithms and then to integrate it with existing proprietary and open-source PACS such as Orthanc [2], Dicoogle, OHIF, EasyPACS, ProtonPACS, CentricityPACS, among others. Integrating this framework will provide a secure service for users who cannot pay for PACS services or maintain PACS infrastructure.

4. **As a recommendation**: The Working committees on digital health standards and ethics and the DICOM standards committee should incorporate the use of data hiding security schemes in teleradiology. As standard bodies and institutions begin to consider standards and ethics relating to e-Health, it is important to consider and incorporate steganography and digital watermarking security techniques. This quest now has increased importance as symposiums and conferences on Digital health as a service are being held in this 2020 [3].

[1] https://arxiv.org/abs/1812.05725
[2] https://medevel.com/10-open-source-pacs-dicom/
[3] https://conferences.computer.org/services/2020/symposia/dhaass.html

Appendices

Appendix A
CSIRO Ethics Approval

This ethics approval was granted by the low-risk panel of the CSIRO Health and Medical Human Research Ethics Committee (CHMHREC). This research was considered low risk because no medical image data was directly collected from humans. The medical images are publicly available and de-identified scans. The ethics approval is attached on the next page.

SCIENCE IMPACT & POLICY
www.csiro.au

Ecosciences Precinct, Dutton Park QLD 4102
GPO BOX 2583, Brisbane QLD 4001, Australia
T (07) 3833 5615 • ABN 41 687 119 230

Peter Eze
CSIRO Data 61
Door 34, Goods Shed North, Village St
Docklands VIC 3008

15th August 2019

Dear Peter,

Re: CSIRO Health and Medical Human Research Ethics Committee (CHMHREC) – Proposal 2019_061_LR
"A Robust and Reliable Tele-medicaldata Security and Authentication System using Spread Spectrum Steganography"

Thank you for the above submission which was considered by the low risk review panel of the CSIRO Health and Medical Human Research Ethics Committee (CHMHREC). I am pleased to grant approval for the project to proceed.

Please note that this approval expires 9th July 2020, the completion date nominated by you. The CHMHREC must be informed of any significant alterations to the protocol, changes to the project team or to the completion date. All serious adverse events must also be reported to the CHMHREC coordinator as soon as possible.

At the completion of the project it is a requirement that a Final Report be completed by you and submitted to CHMHREC. Where a project exceeds 12 months duration, a report must be submitted annually on the anniversary of this approval. A copy of the report form can be obtained from the MyCSIRO Human Research Ethics webpage or by contacting the CHMHREC coordinator.

I wish you success with your project and thank you for your application.

Yours sincerely,

For **Assoc Professor Brian Stoffell**
Chair, CSIRO Health and Medical Human Research Ethics Committee

NHMRC Registered Committee Number and Name:
EC00187 CSIRO Health and Medical Human Research Ethics Committee

Appendix B

Security Concepts in Watermarking and Steganography

The algorithms and methods developed so far in Sections 3 and 4.1 is based on the modification of the basic SS watermarking techniques. Hence, the idea is to have an extra layer of security for telemedicine, especially in Teleradiology. However, the form of security tackled in this thesis does not strongly relate to passive attacks on the medical image data. The form of security that focuses on key extraction is not the focus of this thesis. We have added this appendix to briefly discuss this form of security which is often related to cryptographic security rather than steganographic security.

Both old [26] and new [15] results in Spread Spectrum Steganographic Security have shown that the general assumption of security for Spread Spectrum in wireless communication are not the same when the same technology is applied to information hiding security systems. This is because of the issues of Blind Source Separation (BSS) techniques and the problems of equivalent keys in digital watermarking. Also, for more hiding capacity using either more bits per user or multiplexing many users' data into a single block, transparency and thus security is put at higher risk. These have made it a necessity to properly evaluate the Steganographic security of any watermarking system that is based on SS techniques.

The definition of Steganographic Security was given in [166] as a measure of detectability, robustness, and difficulty of extraction. In this work, detectability was called 'Security' while the difficulty of extraction was called 'Privacy'. The robustness was

more related to hiding capacity. They found consistency between Security and Privacy but a reverse variation between these terms and hiding capacity. In most Steganographic literature, the security of a stego system refers to detectability of the presence of a watermark, and not necessarily its extraction [21, 115]. Though some more formal definitions will be given for Steganographic Security, we will be considering it as both detection and an extraction problem. In this work, the difficulty, measured by some parameters, associated with distinguishing between a watermarked and a non-watermarked image and the ability to extract either the embedding key or the message itself is called Steganographic Security.

Definitions of Steganographic Security

The prisoners' problem [141] provides a scenario for understanding the concept of security during invisible communication. If Alice and Bob are criminals put in two different cells and they need to plan escape, they can only communicate through the prison warden, Wendy, in an invisible manner. If Wendy scrutinises the innocuous message and suspects any secret content, she supresses all communications even without knowing what the secret message is. A secure Steganographic system allows Bob and Alice to use an innocuous cover to hide communication about their escape plane without any suspicion by Wendy.

The above example defines detectability as the major primitive for defining Steganographic security. In modern cyberspace and due to greater availability of computing power, attackers are not only interested in detection but also in the retrieval of the information hidden and the secret key used for hiding. Hence, there are various definitions in literature that needs to be considered in some formal sense.

Definition B.1 (Unconditionally Secure[81])**.** *Let the Cover $\boldsymbol{X}$, Message $\boldsymbol{M}$ and Secret Key, $\boldsymbol{W}$ be random variables. The output of the embedding process, $\boldsymbol{Y}$, is also a random variable. Then the Steganographic system is secure if the knowledge of $\boldsymbol{X}$ and $\boldsymbol{Y}$ does not reveal any information*

about ***M*** *or* ***W*** . That is, the Mutual Information, **I(.)** is,

$$\mathbf{I\ (M;\ X|Y)\ =0} \tag{B.1}$$

Definition B.2 (Relative Entropy Measure[21]). *A Steganographic system is* $\epsilon -$ *Secure if* ***RE (X||Y)***$\leq\ \epsilon$. *If* $\epsilon\ =\ 0$ *then the Steganographic system is perfectly secure. Where* ***RE (.)*** *denotes the relative entropy between the distributions of* ***X*** *and* ***Y***.

Another suitable discriminator could be used instead of relative entropy.

Definition B.3 (Active Warden[50]). *Given a permitted* ***critical distortion***, ***dc***, *for both hider, P1 and attacker,P2, a Steganographic System is secure if the strategy, S1 chosen by the hider always maximises the channel capacity (payoff) irrespective of the attacker strategy, S2.*

The amount of distortion introduced by the attacker does not reduce the capacity of the Steganographic channel after introducing the permissible distortion. The goal of *P1* is to maximise the payoff while the goal of *P2* is to minimise the payoff; all are subject to the distortion constraint.

Another emerging definition of Steganographic security is the one that relates to the ability to retrieve the actual hidden information and not just detecting or changing the hidden data. This is considered more of a cryptanalytic problem than a Steganalytic problem [50]. However, this aspect is important in Teleradiology as it relates to the privacy of sensitive patient record embedded in the medical image. In view of this, we shall consider a fourth and fifth definitions of Steganographic Security.

Definition B.4 (Active Steganalyst[26]). *For a given spatial Stego data* ***Y***, *a Steganographic System is secure if neither the secret key* ***W***, *Message* ***M*** *nor the original cover* ***X*** *can be nearly accurately estimated by some means by an active steganalyst who has some reasonable computational resources and time.*

In this definition, Kerckhof's principle is completely assumed and some well-defined mathematical assumptions are assumed to be true for both the known and unknown Steganographic primitives involves.

Definition B.5 (Effective Key Length[15]). *A given Steganographic system is secure if the probability* ***P*** *that an adversary can use a brute-force method to find an effective key (exact or equivalent) of length , l , that gives him/her access to the Steganographic channel is less than a small value,* ϵ.

This effective key length is given as:

$$l = -log_2 P(bits) \tag{B.2}$$

In this research we are interested to less extent in Definitions B.1 and B.2 as they have been described to be overly theoretic and do not lead to the current methods used by practical steganalysts. Definitions B.3 will not be considered as the watermarking system is for tamper detection and information hiding but not tamper-resistant watermarking system. Definitions B.4 and B.5 are the main focus of this research as they relate mostly to the privacy of patient data as transmitted with medical image scans in Teleradiology.

Appendix C
Detailed Algorithms and Illustrations

C.1 Watermark Decoding and Tamper Detection Algorithm

Figure C.1 is a complete flowchart of the decoding and tamper detection component of the C_4S algorithm.

C.2 Extended C_4S with Tamper Detection

Here we want to preserve tamper detection and accurate watermark detection at higher capacity and possibly the same distortion rate. In this approach, we increase the gap between any two embedding channels in order to create a ***deep well***, that an adversary could fall into. Such deep well exists between any two embedding levels. To do this we modify Equation 4.16 into C.1:

$$\rho_n = \pm\rho \pm n\epsilon \qquad for \qquad n = 0, 4, 8, \ldots \tag{C.1}$$

The implication, however, is that the capacity already achieved in Equation 4.16 will be halved or distortion is doubled in order to maintain the same capacity. This widened gap between embedding levels is shown in Figure C.2

The deep well requires a 'bridge' of at least 2ϵ to cross it. Any correlation that lies within this bridge length is detected as a tampering where an attacker is trying to move some bit groups from one side of the well to the other side. A successful movement,

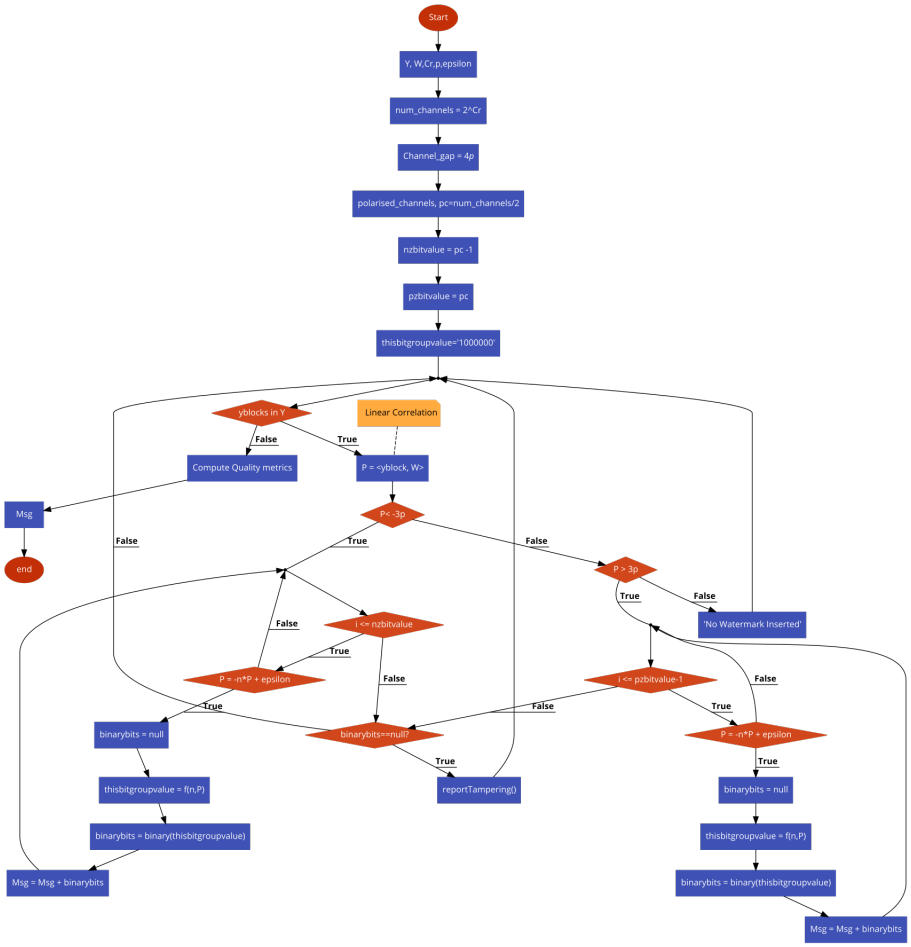


Figure C.1: Watermark Decoding and Tamper Detection Algorithm (DDTDA)

however, could be achieved but this is supposed to cause both higher distortion on the watermarked image, as well as bit errors in the EMR or cryptographic hash used to generate the embedded watermark.

We then present the algorithms that were implemented for the extended C4S for capacity improvement and tamper detection based on spread spectrum steganography.

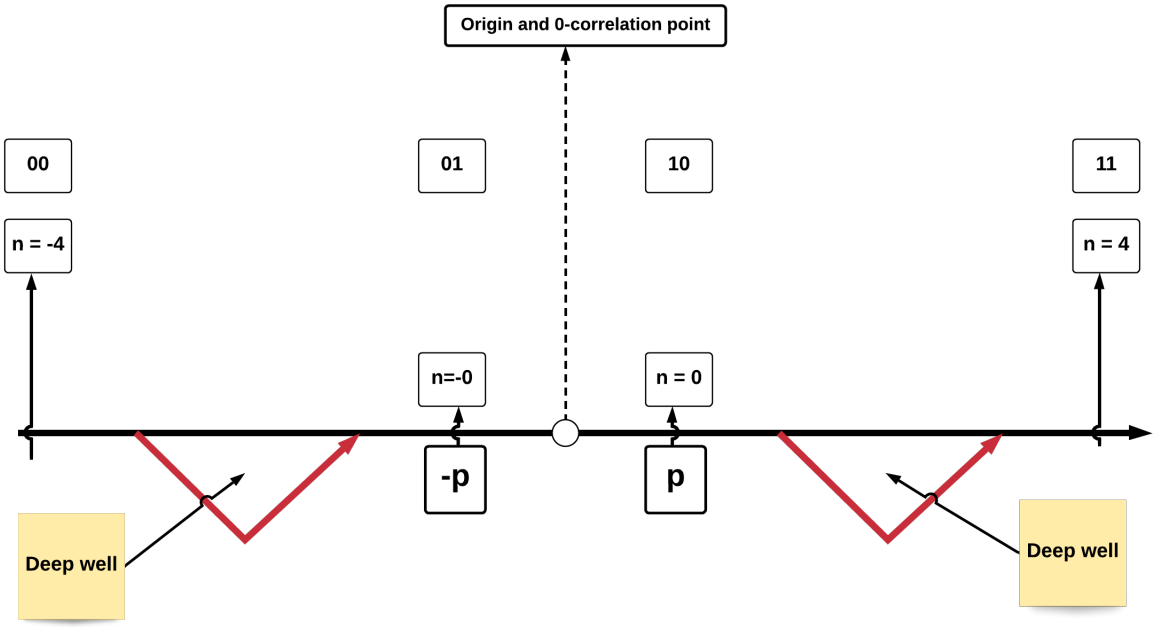


Figure C.2: Extended C_4S with Tamper Detection

Appendix D
Further Mathematical Identities

D.1 K-Space and Radon Transforms

These transforms that are particularly useful in medical image acquisition, reconstruction and processing will be briefly discussed here.

K-space is the frequency domain of an MRI image. It is the raw data obtained after the interaction between the magnetic field of strength, β_0 and the ***gyromagnetic ratio*** , γ of the spinning hydrogen nucleus contain in the human body being imaged. The resultant frequency value is given by Larmor's equation as:

$$f_0 = \beta * \gamma \tag{D.1}$$

For hydrogen nucleus, $\gamma = 43 MHz/Tesla$ [45]. A typical field strength used in MRI imaging is $\beta_0 = 1.5 Tesla$. As shown in figure D.1, the collection of spatial frequency data from different parts of the body forms the k-space data, whose Fourier transform gives the MRI image in pixel domain. This transformation is NOT a one-to-one mapping from K-space pixel K(x,y) to MRI spatial image pixel, M(x,y). Each point in the K-space contains spatial frequency and phase information about every pixel in the final MRI scan.

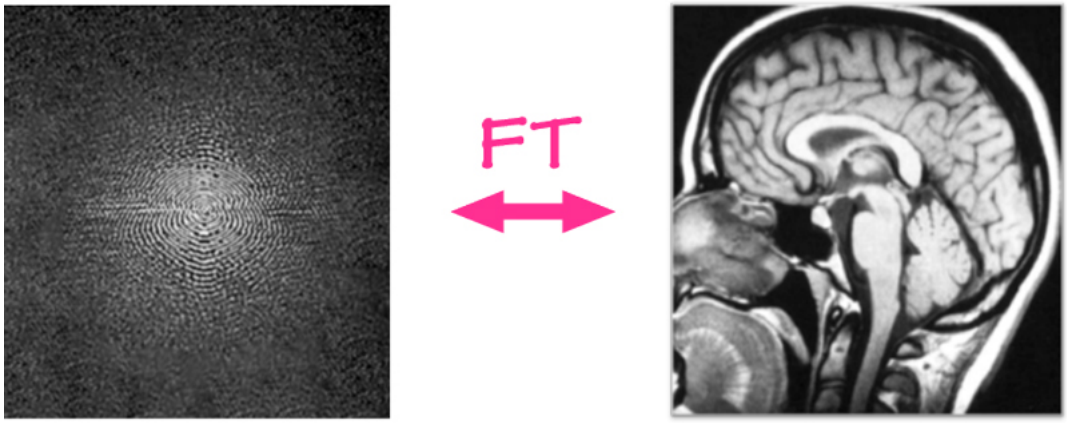


Figure D.1: Fourier Transform of K-space data gives MRI (source: http://mriquestions.com/what-is-k-space.html)

Appendix E

MediSteg Class Diagram

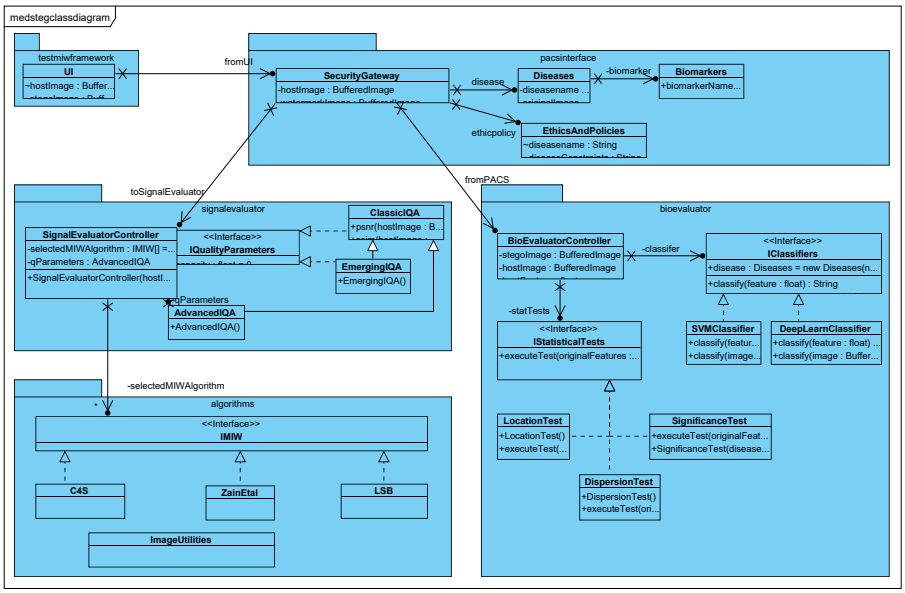


Figure E.1: Complete MediSteg Class Diagram

References

[1] E.H Abd, H.J Abdulwahed, and H.A Mohammed. "The Use of Discrete Cosine Transformation (DCT) In Information Hiding Process." In: *Journal of Kerbala University*. 10.1 (2012).

[2] A. Adelsbach, S. Katzenbeisser, and A.-R. Sadeghi. "A computational model for watermark robustness". In: *Proceedings of the Information Hiding, Springer*. 2007, pp. 145–160.

[3] D. Adom, E.K Hussein, and J.A Agyem. "Theoretical and Conceptual Framework: Manadatory Ingredients of a quality research". In: 7.1 (2018), pp. 438–441. ISSN: 2277-8179.

[4] R. Ahmed, B. Hassan, and B. Li. "Robust hybrid watermarking for security of medical medical images in computer-aided diagnosis based telemedicine applications." In: *IEEE International Symposium on signal processing, Information Technology (ISSPIT)*. 2018.

[5] C. Saini et al. "Digital Image Forgery Detection using Correlation Coefficients". In: 129.14 (Nov. 2015), pp. 17–23.

[6] Holub et al. "Universal distortion function for steganography in an arbitrary domain". In: *EURASIP Journal on Information Security* 1.1 (2014), pp. 1–13. URL: http://jis.eurasipjournals.com/content/2014/1/1.

[7] L.M Rimol et al. “Cortical Volume, Surface Area and Thickness in Schizopre-nia and Bipolar Disorders”. In: *BIOL Psychiatry, Society of Biological Psychiatry* 71 (2012), pp. 552–560. URL: http://dx.doi.org/10.5402/2013/174524.

[8] D. S Kermany et al. “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning”. In: *Cell Press, Elsevier Inc.* 172 (2018), pp. 1122–1131. DOI: 10.1016/j.cell.2018.02.010.

[9] C. Altera. “Gold Code Generator Reference Design”. In: *Altera Corporation* (2003), pp. 1–16. URL: http://www.cs.cmu.edu/~sensing-sensors/readings/Gold_Code_Generator-an295.pdf.

[10] Image Analyst. “Encoding text into image gray levels”. In: *MATLAB Central* (2020), pp. 1–4. URL: https://www.mathworks.com/matlabcentral/fileexchange/54155-encoding-text-into-image-gray-levels.

[11] M. Babel et al. “Preserving data integrity of encoded medical images: the LAR compression framework”. In: 7 (2012), pp. 1–35. URL: https://hal.archives-ouvertes.fr/hal-00658130/document.

[12] M. Barni, F. Bartolini, and T. Furon. “A general framework for robust water-marking security”. In: *Signal Processing* 83.10 (2003), pp. 2069–2084.

[13] M. Barni et al. “Capacity of Full Frame DCT Image Watermarks”. In: *IEEE Trans-actions on Image Processing* 9.8 (2000), pp. 1450–1455.

[14] M.S Barts. “The only thing that is black and white in MS is your MRI Image”. In: *BartsMS Blog* (2015). URL: http://multiple-sclerosis-research.blogspot.com/2015/01/the-only-thing-that-is-black-and-white.html.

[15] P. Bas and T. Furon. “A New Measure of Watermarking Security: The Effective Key Length”. In: *IEEE Transactions in Information Forensicsand Security* 8.8 (2015), pp. 348–1317. ISSN: 1306-1021.

[16] W. Bender et al. “Techniques for Data Hiding”. In: *IBM System Journals* 35.3 (1996).

[17] M. Bhaskaranand and J.D Gibson. "DISTRIBUTIONS OF 3D DCT COEFFICIENTS FOR VIDEO". In: *ICASSP 2009*. 2009, pp. 793–796.

[18] S. Borra et al. "Secure transmission and integrity verification of color radiological images using fast discrete curvelet transform and compressive sensing". In: 12 (2019), pp. 35–48. DOI: 10.1007/j.smhl.2018.02.001.

[19] M. Brady. *Basics of MRI (accessed 27th September, 2018)*. 2004. URL: http://www.robots.ox.ac.uk/~jmb/lectures/medimanallecture1.pdf.

[20] T. Brando, M.P Queluz, and A. Rodrigues. "On the Use of Error Correction Codes in Spread Spectrum Based Image Watermarking". In: *H.Y Shum, M. Liao and S.F Chang (Eds.): PCM 2001, LNCS 2195*. 2001, pp. 630–637.

[21] C. Cachin. "An information-theoretic model for steganography". In: *Aucsmith, D. (ed.) IH 1998. LNCS, Springer, Heidelberg*. Vol. 1525. 1998, pp. 306–318.

[22] F. Caoa and X.Q. Zhoub H.K. Huanga. "Medical image security in a HIPAA mandated PACS environment". In: *Computerized Medical Imaging and Graphics* 27 (2003), pp. 185–196.

[23] C. Carlos et al. "Dicoogle - An open source peer-to-peer PACS". In: *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology* 24 (Oct. 2010), pp. 848–56. DOI: 10.1007/s10278-010-9347-9.

[24] P. Castro et al. "The Rise of Serverless Computing". In: 62.12 (2019), pp. 44–54. URL: 10.1145/3368454.

[25] K. Chaleunvong. "Medical Statistics:Training Course in Reproductive Health Research". In: 62 (2009), pp. 1–44. URL: https://www.gfmer.ch/Activites_internationales_Fr/Laos/PDF/Medical_statistics_Chaleunvong_Laos_2009.pdf.

[26] R. Chandremouli. "A Mathematical Approach to Steganalysis". In: *Proceedings of SPIE Security and Watermarking of Multimedia ContentsIV, California*. 2002.

[27] S. Chaplot, L.M Patnaik, and N.R Jagannathan. "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network". In: *Biomedical Signal Processing and Control* 1 (2006), pp. 86–92. URL: doi:10.1016/j.bspc.2006.05.002.

[28] B. Charif and AI. Awad. "Towards smooth organisational adoption of cloud computing a customer-provider security adaptation". In: *Computer Fraud and Security* 2016 (Apr. 2016), pp. 7–15. DOI: http://dx.doi.org/10.1016/S1361-3723(16)30016-1.

[29] K. Chen and T.V Ramabadran. "Near-lossless compression of medical images through entropy coded DPCM." In: *IEEE Transactions in Medical Imaging* 13.3 (1994), pp. 538–548.

[30] Y.H Chen and J.C Chen. "A new multiple audio watermarking algorithm applying DS-CDMA". In: *International conference on Machine Learning and Cybernetics*. 2009, pp. 2205–2210.

[31] R. H Choplin, J.M Boehme, and C.D Maynard. "Picture Archiving and Communication Systems: An Overview". In: (1992), pp. 127–129.

[32] M. M. H. Chowdhury and A. Khatun. "Image Compression Using Discrete Wavelet Transform". In: *International Journal of Computer Science, ISSN (Online): 1694-0814* 9.4 (2012), pp. 327–330.

[33] D.A Clausi and M.D Jernigan. "A Fast Method to Determine Co-Occurence Texture Features". In: *IEEE Transactions on Geoscience and Remote Sensing* 36.1 (1998), pp. 298–300.

[34] G. Coatrieux et al. "A Review of Image Watermarking Applications in HealthCare." In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2006, pp. 275–282. URL: DOI:10.1109/IEMBS.2006.259305.

[35] G. Coatrieux et al. "Watermarking: a new way to bring evidence in case of telemedicine litigation." In: *European Federation for Medical Informatics.* (2011), pp. 611–615.

[36] Noel C. F. Codella et al. “Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI)”. In: *International Skin Imaging Collaboration (ISIC)*. arXiv:1710.05006. 2017.

[37] T.M Cover and J.A Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc, 1991, pp. 336–373. ISBN: 0-471-20061-1. URL: https://www.sns.ias.edu/pitp2/2012files/Cover_and_Thomas_Chptr13.pdf.

[38] I. Cox, M. Miller, and A. McKellips. “Watermarking as communications with side information”. In: *Proceedings of IEEE*. 1999, pp. 1127–1141. DOI: 10.1109/5.771068.

[39] I. Cox et al. *Digital Watermarking and Steganography*. Elsevier Science,ProQuest Ebook Central, 2007, pp. 12–14.

[40] I.J Cox et al. “Secure Spread Spectrum Watermarking for Images, audio and video”. In: *Proceedings of the IEEE International Conference on Image Processing, Lausanne, Switzerland*. 1996, pp. 243–246.

[41] J. Deng and Y. Wang. “Quantitative magnetic resonance imaging biomarkers in oncological clinical trials: Current techniques and standardization challenges”. In: *Chronic Diseases and Translational Medicine* 3 (2017), pp. 8–20. URL: http://dx.doi.org/10.1016/j.cdtm.2017.02.002.

[42] V. Dhore and P.M Arfat. “Secure Spread Spectrum Data Embedding and Extraction.” In: *International Journal of Science and Research* 4.1 (2015), pp. 743–747.

[43] H.S. Al-Dmour. *Enhancing Information Hiding and Segmentation for Medical Images using Novel Steganography and Clustering Fusion Techniques: PhD Thesis*. 2018.

[44] D. Dresden. “What is the Hippocampus?” In: *Medical News Today, 7th December 2017* (2011). URL: https://www.medicalnewstoday.com/articles/313295.php.

[45] N. Dwork and J. Pauly. *Introduction to MRI Signal Equation*. Stanford University, 2018. URL: http://www.nicholasdwork.com/teaching/1706ee102a/lectures/lectureAppMRI.pdf.

[46] C. Edwards. "Malevolent Machine Learning". In: *Communications of the ACM* 62.12 (2019), pp. 13–15. URL: 10.1145/3365573.

[47] N.M Edwin. "Software Grameworks, Architectural and Design Patterns". In: (2014), pp. 670–678. DOI: 10.4236/jsea.2014.78061.

[48] R. English. "Comparison of High Capacity Steganography Techniques". In: *IEEE International Conference of Soft Computing and Pattern recognition*. 2010, pp. 448–453.

[49] R. Eswaraiah and E.S Reddy. "Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest". In: *IET image Process* 9.8 (2015), pp. 615–625. DOI: 10.1049/iet-ipr.2014.0986.

[50] J.M Ettinger. "Steganalysis and Game Equilibria". In: *Second International Workshops on Information Hiding*. Vol. 1525. 1998, pp. 319–328.

[51] C.T Farrar et al. "Sensitivity of MRI Tumor Biomarkers to VEGFR Inhibitor Therapy in an Orthotopic Mouse Glioma Model". In: *PLoS ONE* 6.3 (2011). ISSN: e17228. URL: doi:10.1371/journal.pone.0017228.

[52] T. Filler and J. Fridrich. "Fisher Information Determines Capacity of -Secure Steganography". In: *Katzenbeisser S., Sadeghi AR. (eds) Information Hiding. IH 2009. Lecture Notes in Computer Science,Springer, Berlin, Heidelberg vol 5806*. Vol. 5806. 2009, pp. 32–47.

[53] S.M Finlayson et al. "Adversarial Attacks Against Medical Deep Learning Systems". In: *Pre-Print, Arxiv:1804:05296v2* (2018), pp. 1–15.

[54] J. J. Garcia-Hernandez, W. Gomez-Flores, and J. Rubio-Loyola. "Analysis of the impact of digital watermarking on computer-aided diagnosis in medical imaging". In: *Computers in Biology and Medicine* 68 (2016), pp. 37–48.

[55] S. Ghoniemy, O.H Karam, and O. Ibrahim. "Robust and Large Hiding Capacity Steganography using Spread Spectrum and Discrete Cosine Transform." In: *International Journal of Image Processing and Visual Communication*. 1.4 (2013). ISSN: 2319-1724.

[56] M. Gkizeli, D.A Pados, and M.J Medley. "SINR, Bit Error Rate and Shannon Capacity Optimised Spread-Spectrum Steganography". In: *Proceedings of International Conference on Image Processing (ICIP*. 2004, pp. 1561–1564.

[57] M. Goemans. *Shannon's noiseless coding theorem*. June 2015. URL: http://math.mit.edu/~goemans/18310S15/noiseless-coding.pdf.

[58] R. C Gonzalez, R. E Woods, and S. L Eddins. "Digital image processing using MATLAB". In: 624 (2004), pp. 1–12.

[59] E.S Gopi. *Digital Signal Processing for Medical Imaging Using Matlab*. Springer Science+Business Media New York, 2013. ISBN: 978-1-4614-3140-4.

[60] B. Goyal et al. "Noise Issues Prevailing in Various Types of Medical Images". In: *Biomedical and Pharmacology Journal* 11.3 (2018), pp. 1228–1237. DOI: 10.13005/bpj/1484.

[61] M Grzenda, J. Furtak AI Awad, and J. Legierski. "Advances in Network Systems: Architectures, Security, and Applications". In: *Advances in Intelligent Systems and Computing* (2017).

[62] X. Guo and T.G Zhuang. "A lossless watermarking scheme for enhancing security of Medical data in PACS". In: *SPIE Medical Imaging, 2003*. 2003, pp. 350–359.

[63] A. Al-Haj, A. Mohammad, and A. Amer. "Crypto-Watermarking of Transmitted Medical Images". In: *Journal of Digital Imaging* 30 (2017), pp. 26–38. DOI: https://doi.org/10.1007/s10278-016-9901-1.

[64] R.M Haralick, K. Shanmugan, and I. Dinstein. "Textural Features for Image Classification". In: *IEEE Transactions On Systems, Man and Cybernetics* 3.6 (1973), pp. 610–621.

[65] J.J Harmsen and W.A Pearlman. "Capacity of Steganographic Channels". In: *IEEE Transactions on Information Theory* 55.4 (2009), pp. 1775–1792.

[66] B. Hassan, R. Ahmed, and O. Hassan. "An Imperceptible Medical Image Watermarking Framework for Automated Diagnosis of Retinal Pathologies in an eHealth Arrangement". In: 7 (2019), pp. 69758–69775. DOI: 10.1109/ACCESS.2019.2919381.

[67] B. Hassan, R. Ahmed, and B. Li. "Computer aided diagnosis of idiopathic central serous chorioretinopathy". In: *Proceedings of 2nd IEEE Advanced Management, Communication, Automation Control (IMCEC)*. May 2018, pp. 824–828.

[68] B. Hassan and G. Raja. "Fully automated assessment of macular edema using optical coherence tomography (OCT) images". In: *Proceedings of International Conference of Intelligent System Engineering*. Jan. 2016, pp. 5–9.

[69] B. Hassan et al. "Structure tensor based automated detection of macular edema and central seous retinopathy using optical coherence tomography images". In: 33.4 (2016), pp. 455–465.

[70] B. Hassan et al. "Automated retinal edema detection from fundus and optical coherence tomography scans". In: *Proceedings of 5th ICCAR , Beijing, China*. 2019, pp. 1–6.

[71] R. Hassan and M.S Arifianto. "A High Payload Reversible Watermarking Scheme based on OFDMA-CDMA". In: *Proceedings of 10th International Conference on Telecommunications and Applications, Denpasar, Indonesia, 6-7th October*. 2016.

[72] X. He, T. Zhu, and G. Yang. "A geometrical attack resistant image watermarking algorithm based on histogram modification". In: *Multidimensional System signal Processing* 26 (2015), pp. 291–306. DOI: 10.1007/s11045-013-0257-0.

[73] US Department of Health and Human Services. *The HIPPA Privacy Rule (accessed 3rd March 2017)*. URL: https://www.hhs.gov/hipaa/for-professionals/privacy.

[74] H. Huang. *Contribution to the integrity control of medical images, Ph.D Thesis*. 2011. URL: https://portail.telecom-bretagne.eu/publi/public/fic_download.jsp?id=7622.

[75] H.K Huang. *PACS and Imaging Informatics: Basic Principles and Applications*. (book). Wiley, New York., 2009.

[76] Q. Huynh-Thu and M. Ghanbari. "Scope of validity of PSNR in image/video quality assessment". In: *ELECTRONICS LETTERS* 44.13 (2008), pp. 26–38. DOI: https://doi.org/10.1049/el:20080522.

[77] National Instruments. *Understanding Spread Spectrum for Communications(accessed 2nd March, 2017)*. URL: http://www.ni.com/white-paper/4450/en/.

[78] V. Jabade and S. Gengaje. "Modelling of Geometric attacks for Digital Image Watermarking". In: *International Journal of Innovations in Engineering Research and Technology* 3.3 (2016), pp. 1–8. ISSN: 2394-3696.

[79] M. Jiafa et al. "A method to Estimate the Steganographic Capacity in DCT Domain Based on MCUU Model". In: *Wuhan University Journal of Natural Sciences* 21.4 (2016), pp. 283–290.

[80] N. Jiang and J. Wang. "The Theoretical Limits of Watermark Spread Spectrum Sequence". In: *The Scientific World Journal, Hindawi Publishing Corporation* 2014.4 (2014), p. 6.

[81] S. Katzenbeissser and F.A.P Petitcolas. "Defining Security in Steganographic Systems". In: *Proceedings of SPIE International Society for Optical Engineering*. 2002.

[82] Y. Ke et al. "Generative Steganography with Kerckhoffs Principle based on Generative Adversarial Networks". In: *Arxiv* (2018), pp. 1–5. URL: https://arxiv.org/ftp/arxiv/papers/1711/1711.04916.pdf.

[83] J. Kim et al. "Breast Cancer Heterogeneity: MR Imaging Texture Analysis and Survival Outcomes". In: *Free Access Breast Imaging* 282.3 (2016), pp. 665–675. DOI: https://doi.org/10.1148/radiol.2016160261. URL: https://pubs.rsna.org/doi/full/10.1148/radiol.2016160261.

[84] T.H. Kim, A. Stoica, and R.S. Chang (Eds.) "Medical Imaging: A Review". In: *Springer-Verlag Berlin Heidelberg* (2010), pp. 504–516.

[85] J. King and A.I Awad. "A distributed security mechanism for resource constrained IoT devices". In: *Informatica (Slovenia)* 40.1 (2016), pp. 133–143.

[86] D. Kirovski and H. S. Malvar. "Spread-spectrum watermarking of audio signals". In: *IEEE Transactions on Signal Processing* 51.4 (Apr. 2003), pp. 1020–1033. ISSN: 1941-0476. DOI: 10.1109/TSP.2003.809384.

[87] M.O Kolawole. *Satellite Communication Engineering, 2017)*. CRC Press, Australia. URL: https://www.routledge.com/Satellite-Communication-Engineering/Kolawole/p/book/9781138075351.

[88] B. Kumar et al. "Secure Spread-Spectrum Watermarking for Telemedicine Applications". In: *Journal of Information Security.* 2 (2011), pp. 91–98. URL: DOI:%2010.3233/978-1-60750-806-9-611.

[89] E.Y Lam and J.W Goodman. "A Mathematical Analysis of the DCT Coefficient Distributions for Images". In: 9.10 (Oct. 2000), pp. 1661–1666.

[90] Q. Li, N. Memon, and H. T. Sencar. "Security issues in watermarking applications - a deeper look". In: *Proceedings of the ACM International Workshop on Multimedia Contents Protection and Security (MCPS)* (2007), pp. 23–28.

[91] S. Liang and G. Shen. "Biomarkers of Glioma". In: *Molecular Targets of CNS Tumors, Dr. Miklos Garami (Ed.)* (2011). URL: http://www.intechopen.com/books/molecular-targets-of-cns-tumors/biomarkers-of-glioma.

[92] C. Lin. *Watermarking and Digital Signature Techniques for Multimedia Authentication and Copyright Protection: PhD Thesis*. 2000.

[93] G. Louizis, A. Tefas, and I. Pitas. *Copyright Protection of 3D Images Using Watermarks of Specific Spatial Structure.*

[94] E. Ludewig, A. Richter, and M. Frame. "Diagnostic imaging evaluating image quality using visual grading characteristic (VGC) analysis". In: *Vertinary Research Communications* 34.5 (2010), pp. 473–479. ISSN: 1573-7446.

[95] H T Lutz. "Basics of Ultrasound". In: *Handbook of experimental pharmacology* (2008), pp. 1–19.

[96] D. G Altman andJ. M. Bland. "Diagnostic tests: Sensitivity and specificity". In: 308.6943 (1994), p. 1552. DOI: 10.1136/bmj.308.6943.1552.

[97] A.A Macedo et al. "A Health Surveillance Software Framework to deliver information on preventive healthcare strategies". In: *Journal of Biomedical Informatics* 62 (2016), pp. 159–170. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2016.06.002. URL: http://www.sciencedirect.com/science/article/pii/S1532046416300454.

[98] H. Maitre. *Statistical Properties of Images*. 2004.

[99] A. Maity et al. "A Comparative Study on Approaches to Speckle Noise Reduction in Images". In: *2015 International Conference on Computational Intelligence and Networks*. 2015, pp. 148–155. DOI: 10.1109/CINE.2015.36.

[100] H.K Maity and S.P Maity. "Intelligent Modified Difference Expansion for Reversible Watermarking." In: *International Journal of Multimedia and Its Applications (IJMA)* 4.4 (2012), pp. 83–85.

[101] H.K Maity and S.P Maity. "Joint Robust and Reversible Watermarking for Medical Images." In: *2nd International Conference on Communication, Computing and Security, Procdia Technology*. 2012, pp. 275–282.

[102] H.S. Malvar and D.A.F. Florencio. "Improved Spread Spectrum: A New Modulation Technique for Robust Watermarking". In: *IEEE Transaction on Signal Processing* 51.4 (2003), pp. 898–905.

[103] A. Mariette and K. Rahul. "Support Vector Machines for Classification". In: Jan. 2015, pp. 39–66. ISBN: 978-1-4302-5989-3. DOI: 10.1007/978-1-4302-5990-9_3.

[104] L.M Marvel, C.G Boncelet, and C.T Retter. "Spread Spectrum Image Steganography". In: *IEEE Transactions on Image Processing* 8.8 (1999), pp. 1075–1083.

[105] P. Meerwald and A. Uhl. "Scalability evaluation of blind spread-spectrum image watermarking." In: *H.J Kim and S. Katzenbeisser, and A.T.S Ho (Eds.): IWDW 2008, LNCS 5450*. 2009, pp. 61–75.

[106] D. Meetoo, R. Rylance, and H.A Abuhaimid. "Health care in a technological world". In: *British Journal of Nursing* 27.20 (2018), pp. 1172–1177. DOI: https://doi.org/10.1155/2010/851920.

[107] A. Mehto and N. Mehra. "Adaptive lossless medical image watermarking algorithm based on DCT DWT". In: (2016), pp. 88–94.

[108] N.A Memon and S.A.M Gilani. "Watermarking of Chest CT scan Medical Images Authentication." In: *Journal of Computer Mathematics.* 88.2 (2010), pp. 265–280.

[109] P.C.M Mintas. "Code Selection with Application to Spread Spectrum Systems Based on Correlation Properties". In: *International Journal of Advanced Research in Computer and Communication Engineering* 4.6 (2015), pp. 348–351. ISSN: 2278-1021.

[110] Y. Mirsky et al. "CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning". In: *Proceedings of 28th USENIX Security Symposium (USENIX Security 2019)*. June 2019, pp. 1–19. URL: https://arxiv.org/pdf/1901.03597.pdf.

[111] M. Mishra and M.C Adhikary. "Digital Image Tamper Detection Techniques - A Comprehensive Study". In: *International Journal of Computer Science and Business Informatics* 2.1 (2013), pp. 1–12.

[112] T. Mittelholzer. "An information-theoretic approach to steganography and watermarking". In: *Proceedings of the Information Hiding, A. Ptzmann, Ed.* 2000, pp. 1–16.

[113] G.D Moody, M. Siponen, and S. Pahnila. "Toward a unified model of information security policy compliance". In: 42.1 (Aug. 2018), pp. 2829–2839.

[114] M. Moorthi and R. Amutha. "A Near Lossless Compression Method for Medical Images". In: *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012)*. 2012, pp. 39–44.

[115] I.S. Moskowitz, L. Chang, and R.E Newman. *Capacity is the wrong paradigm (accessed 4th October 2018)*. URL: http://chacs.nrl.navy.mil/publications/~/2002moskowitzcapacity.pdf.

[116] P. Moulin and J. A. O'Sullivan. "Information-theoretic analysis of information hiding". In: *IEEE Transactions on Information Theory* 49.3 (2003), pp. 563–593.

[117] S. M Mousavi et al. "A robust medical image watermarking against salt and pepper noise for brain MRI images". In: 76.7 (2017), pp. 10313–10342.

[118] S.M Mousavi, A. Naghsh, and S.A.R Abu-Bakar. "Watermarking Techniques used in Medical Images: A Survey". In: *Society for Imaging Informatics in Medicine.* 27.6 (2014), pp. 714–729. DOI: doi:10.1007/s10278-014-9700-5.

[119] J. Muthuswamy. *Biomedical Signal Analysis:Standard Handbook Of Biomedical Engineering and Design*. McGraw-Hill, 2004. URL: www.digitalengineeringlibrary.com.

[120] N.Manikandaprabu, S.Pavithra, and V.N.Thilagamani. "Data Hiding in Color Images." In: *International Journal of Novel Research in Engineering and Pharmaceutical Sciences.* 1.5 (2014), pp. 1–7.

[121] L. Nanni et al. "Different Approaches for Extracting Information from the Co-Occurrence Matrix". In: *PLOSOne* 8.12 (2013), pp. 1–9. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3873395/pdf/pone.0083554.pdf.

[122] G.T. Narayanan. *A STUDY OF PROBABILITY DISTRIBUTIONS OF DCT COEFFICIENTS IN JPEG COMPRESSION, Masters Thesis*. 2010. URL: http://archives.njit.edu/vol01/etd/2010s/2010/njit-etd2010-064/njit-etd2010-064.pdf.

[123] M. Naseem et al. “Spread Spectrum based Invertible Watermarking for Medical Images using RNS and Chaos.” In: *International Arab Journal of Information Technology*. 13.2 (2016), pp. 223–231.

[124] Greenbone Networks. *Confidential Patients data freely accessible on the Internet: Information Security Report*. 2019. URL: https://www.greenbone.net/wp-content/uploads/CyberResilienceReport_EN.pdf.

[125] T.T Nguyen and H.D Tuan. “AA Modified Spatial Spread Spectrum Method for Digital Image Watermarking”. In: *IEEE 2nd International Conference Communication and Electronics, ICCE Hoi an Vitenam*. 2008, pp. 282–287.

[126] H. Nyeem, W. Boles, and C. Boyd. “A Review of Medical Image Watermarking Requirements for Teleradiology.” In: *Journal of Digital Imaging*. 26.2 (2013), pp. 326–343.

[127] H. M.A Nyeem. *A digital Watermarking Framework with application to Medical Image Security*. Ph.D Thesis, School of Electrical and Computer Science, Queensland University of Technology, Australia, 2014.

[128] A.M.J. Paans. “Positron emission tomography”. In: *Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, The Netherlands* (), pp. 363–382. URL: https://cds.cern.ch/record/1005065/files/p363.pdf.

[129] W. Pan et al. “An Additive and Lossless Watermarking Method Based on Invariant Image Approximation and Haar Wavelet Transform.” In: *32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31-Sept. 4*. 2010, pp. 4740–4743.

[130] R.L Pickholtz, D.L Schilling, and L.B Milstein. “Theory of Spread Spectrum Communication - A Tutorial”. In: *IEEE Transactions on Communication* 30.5 (1982), pp. 855–884.

[131] A.F Qasim, R. Aspin, and F. Meziane. “Integration of Digital Watermarking Technique into Medical Image Systems”. In: *10th IEEE International Conference*

Dependable Systems, Services and Technologies, Leeds, United Kingdom. June 2019, pp. 202–207.

[132] A.F Qasim et al. "ROI-based reversible watermarking scheme for ensuring the integrity and authenticity of DICOM MR images". In: (2019), pp. 16433–16463. DOI: 10.1007/s11042-018-7029-7.

[133] Terry Quatrani. *Introduction to UML 2.0 (accessed 21st October, 2019)*. 2005. URL: https://www.omg.org/news/meetings/workshops/MDA-SOA-WS_Manual/00-T4_Matthews.pdf.

[134] O.M Al-Quershi and B.E Khoo. "Authentication and Data Hiding using a hybrid ROI-based watermarking schemes for DICOM images." In: *Journal of Digital Imaging*. 24.1 (2011), pp. 114–125.

[135] RadiologyInfo. "Positron Emission Tomography - Computed Tomography (PET/CT)". In: *2018 Radiological Society of North America* (2018), pp. 1–7. ISSN: 20191-4397. URL: https://www.radiologyinfo.org/en/pdf/pet.pdf.

[136] A. Rocek and V. Zatloukal. "Comparative Study of Negative Aspects Elimination of Medical Image Watermarking Methods". In: *International Journal of Information and Electronics Engineering* 5.1 (2015), pp. 6–9.

[137] S. Rohini and V. Bairagi. "Lossless Medical Image Security". In: *International Journal of Applied Engineering Research* 1.3 (2006), pp. 536–541. ISSN: 0976-4259.

[138] A. Salman and M. Najim. *Wavelet Transform and DSP Applications (accessed 29th September, 2018)*. 2005. URL: https://www.slideshare.net/farohalolya/wavelet-transform-and-dsp-applications.

[139] J.L Semmlow. *Biosignal and Biomedical Image Processing MATLA B-Based Applications*. Marcel Dekker, Inc., Cimarron Road, Monticello, New York 12701, U.S.A., 2004.

[140] C.E Shannon. "A mathematical theory of communication". In: *Bell Syst. Tech. J. Reprinted in C.E. Shannon and W. Weaver The Mathematical Theory of Communication; University Illinois Press: Champaign, IL, USA, 1949* 27 (1948), pp. 379–423.

[141] G. Simons. "The Prisoners Problems and Subliminal Channel". In: *Advances in Cryptography, CRYPTO 83* (1984), pp. 51–67.

[142] A. Singh and M.K Dutta. "A robust zero-watermarking scheme for tele-ophthamological applications". In: (2017), pp. 1–14. DOI: 10.1016/j.jksuci.2017.12.008.

[143] S. Singh, V.S Rathore, and R. singh. "Hybrid NSCT domain multiple watermarking for medical images". In: *Multimedia tools Applications* 76 (2017), pp. 3557–3575. DOI: 10.1007/s11042-016-3885-1.

[144] A. Siper, R. Farley, and C. Lombardo. "The Rise of Steganography". In: *Proceedings of Student/Faculty Research Day, CSIS, Pace University*. 2005, pp. 1–7. URL: http://csis.pace.edu/~ctappert/srd2005/d1.pdf.

[145] S.W Smith. *The Scientist and Engineer's Guide to Digital Signal Processing (2nd Ed.)* California Technical Publishing, 1999.

[146] J. Son. *Integrated Provisioning of Compute and Network Resources in Software-Defined Cloud Data Centers: Ph.D Thesis*. 2018.

[147] W. Stallings. *Cryptography and Network Security: Principles and Practice.* 6th ed. Pearson Education Limited Essex, England, 2014.

[148] A. Sulleyman. *NHS Cyber Attack: Why Stolen medical information is so much more valuable than financial data (accessed 8th October, 2018)*. URL: https://www.independent.co.uk/life-style/gadgets-and-tech/news/nhs-cyber-attack-medical-data-records-stolen-why-so-valuable-to-sell-financial-a7733171.html.

[149] M. S. Taha et al. "Combination of Steganography and Cryptography: A short Survey". In: *IOP Conference Series: Materials Science and Engineering* 518 (2019), p. 052003. DOI: 10.1088/1757-899x/518/5/052003. URL: https://doi.org/10.1088%2F1757-899x%2F518%2F5%2F052003.

[150] B. Theek et al. “Radiomic analysis of contrast-enhanced ultrasound data”. In: *Scientific Reports* 8.11359 (2018), pp. 1–9. DOI: DOI:10.1038/s41598-018-29653-7. URL: https://www.nature.com/articles/s41598-018-29653-7.pdf.

[151] P. Tschandl, C. Rosendahl, and H. Kittler. “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions”. In: *Science Data* 5 (2018).

[152] P. Vicky and P. Udaya. “Novel Authentication Scheme with Pseudorandom Sequences in the Frequency Domain of Images”. In: *Proceedings of Eighth International Workshop on Signal Design and its Applications in Communications (IWSDA)*. 2017.

[153] A. Vijayakumar and A. Vijayakumar. “Comparison of Hippocampal Volume in Dimentia Subtypes”. In: *ISRN Radiology, Hindawi Publishing Corporation* (2013). URL: http://dx.doi.org/10.5402/2013/174524.

[154] M.A Waheed. *Interpolation based adaptive reversible data hiding for digital image applications: A Masters of Science Thesis*. Department of Electrical, Electronics, Communication, Military Institute of Science, and Technology, Dhaka, 2018.

[155] A. Wakatani. “Digital Watermarking for ROI Medical images by using compressed signature image.” In: *Proceedings of the 35th Annual Hawai International Conference*. 2002, pp. 2043–2048.

[156] Y. Wang and P. Moulin. “Perfectly Secure Steganography: Capacity, Error Exponents, and Code Constructions”. In: *IEEE Transactions on Information Theory* 54.6 (June 2008), pp. 2706–2722. ISSN: 1557-9654. DOI: 10.1109/TIT.2008.921684.

[157] R. B. Wolfgang and E. J. Delp. “A Watermark for Digital Images”. In: *IEEE International Conference on Image Processing, Laussane, Switzerland*. 1996.

[158] T.A Wong. “Spreading Sequences”. In: *Wireless Information Networking Group, University of Florida* (2017). URL: http://wireless.ece.ufl.edu/twong/Notes/CDMA/ch3.pdf.

[159] Sendren Sheng-Dong Xu et al. "Classification of Liver Diseases Based on Ultrasound Image Texture Features". In: *Applied Sciences* 9.2 (2019). ISSN: 2076-3417. DOI: 10.3390/app9020342. URL: https://www.mdpi.com/2076-3417/9/2/342.

[160] F. Yaghmaee and M. Jamzad. "Estimating Watermarking Capacity in Grayscale Images based on Image complexity". In: *EURASIP Journal on Advances in Signal Processing* 2010.13 (2010), pp. 1–9. DOI: https://doi.org/10.1155/2010/851920.

[161] S. Yang, Z. Song, and J.H Park. "A High-Capacity CDMA Watermarking Scheme Based on Orthogonal Pseudorandom Sequence Subspace Projection". In: *5th FTRA International Conference on Multimedia and Ubiquitos Engineering, IEEE Computer Society*. 2011.

[162] H. Yin. *Spread Spectrum Techniques In Lecture Notes in Radio Communication*. Feb. 2005. URL: http://www.comlab.hut.fi/opetus/333/2004_2005_slides/Spread_spectrum.pdf.

[163] J. M. Zain, A. R. M. Fauzi, and A. A. Aziz. "Clinical Evaluation of Watermarked Medical Images". In: *International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY*. 2006, pp. 5459–5462.

[164] J.M Zain and M. Clarke. "Reversible Region of Non-interest (RONI) watermarking for Authentication of DICOM images." In: *International Journal of Computer Science and Network Security.* 7.9 (2007), pp. 19–28.

[165] F. Zhang and X. Zhang. "EBR Analysis of Digital Image Watermarking". In: *S. Patnaik and X. Li (eds.), Proceedings of International Conference on Computer Science and Information Technology, Advances in Intelligent Systems and Computing*. 2014, pp. 11–18.

[166] W. Zhang and S. Li. "Security Measurements of Steganographic Systems". In: *Jakobsson M., Yung M., Zhou J. (eds) Applied Cryptography and Network Security. ACNS 2004. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg* 3089 (2004).

[167] X. Zhang, Z.J Wang, and X. Wang. "Correlation-and-bit-aware additive spread spectrum data hiding for Laplacian distributed host image signal". In: *Signal Processing: Image Communication* 29 (2014), pp. 1171–1180. ISSN: 0976-4259.

[168] J. Zhong. *Watermark Embedding and Detection*. Ph.D Thesis, Department of Computer Science and Engineering, Shanghai Jiaotong University, 2007.