TU Clausthal
Clausthal University of Technology

Bachelor Thesis

by

Emre Cem Elevis

Compression of 3D (Voxel) Textures for Medical Imaging

1st Reviewer: **Prof. Dr. Thorsten Grosch**

2nd Reviewer: **Pro. Dr. Rüdiger Ehlers**

31st March 2025

Institut für Informatik
Clausthal University of Technology

Declaration of Authorship

I have read and understood the guidelines of the Technical University of Clausthal and those published in the ISSE bulletin, including those regarding the use of literature and other sources. I confirm that I have prepared this thesis independently by myself. Any information taken from other sources and being reproduced in this thesis is clearly referenced.

In terms of the general examination regulations, this work has not yet been submitted to any other examination division.

I hereby agree that my bachelor thesis may be exhibited in the institute's library and kept for inspection.

Clausthal-Zellerfeld, 31st March 2025

Location, Date

Emre Cem Elevis

Abstract

In this thesis, we compare various compression methods with a particular interest in medical imaging and standards that are compatible with the most widely used medical imaging format Digital Imaging and Communications in Medicine (DICOM).

We evaluate various compression techniques suitable for medical imaging, based on metrics such as compression efficiency, structural similarity and encoding speed. We explore both lossless and lossy standards that are accepted by various medical communities. 2.7 We implement off-the-shelf video encoding algorithms from the Moving Picture Experts Group (MPEG) to be compatible with the DICOM standard. We also test the various encoding methods to provide insights to the particular compression methods most suitable for various imaging technologies and use-cases needed by the medical imaging community. Through this work, we aim to contribute to the field of medical imaging by providing insights into effective compression strategies that facilitate the efficient handling of large medical datasets, ultimately creating a discourse that may help in better patient outcomes.

We explore and propose a novel experimental encoding scheme suitable for 'intertwining' large volumes of data. This method aims to supplement current compression methods when it comes to archival of medical imaging where quite a lot of the information between patients is repetitive by nature and does not require instant access.

Zusammenfassung

In dieser Arbeit vergleichen wir verschiedene Kompressionsverfahren mit besonderem Fokus auf die medizinische Bildgebung und auf Standards, die mit dem am weitesten verbreiteten medizinischen Bildformat (DICOM) kompatibel sind.

Wir bewerten diverse Kompressionstechniken für die medizinische Bildgebung anhand von Metriken wie Kompressionseffizienz, struktureller Ähnlichkeit und Kodierungsgeschwindigkeit. Dabei untersuchen wir sowohl verlustfreie als auch verlustbehaftete Standards, die in verschiedenen medizinischen Fachkreisen anerkannt sind. 2.7 Darüber hinaus implementieren wir gängige Video-Kodierungsalgorithmen aus dem MPEG-Standard, um sie mit DICOM kompatibel zu machen. Wir testen zudem verschiedene Kodierungsverfahren, um Erkenntnisse darüber zu gewinnen, welche Kompressionsmethoden sich am besten für unterschiedliche bildgebende Technologien und Anwendungsszenarien in der medizinischen Bildgebung eignen. Ziel unserer Arbeit ist es, durch die Analyse effektiver Kompressionsstrategien einen Beitrag zum Umgang mit großen medizinischen Datensätzen zu leisten und damit eine Diskussion anzustoßen, die letztlich zu einer besseren Patientenversorgung beitragen kann.

Ferner untersuchen und schlagen wir ein neuartiges experimentelles Kodierschema vor, das sich zum „Verflechten“ großer Datenmengen eignet. Diese Methode soll bestehende Kompressionsverfahren bei der Archivierung medizinischer Bilder ergänzen, da ein Großteil der Informationen zwischen Patient*innen von Natur aus wiederkehrend ist und nicht jederzeit sofort verfügbar sein muss.

Acknowledgements

Prof. Dr. Grosch has been an ideal professor and thesis supervisor. Alongside Mr. Felix Brüll they have provided me with robust support and freedom to finish my thesis. I would like to express my gratitude for the pydicom project, FFMPEG group and the OPENJPEG project. Without their commitment to free software I would not have been able to finish my work. Most importantly, I am grateful for my family and friends unconditional, and loving support.

Contents

List of Tables

Abbreviations

CR	Encoding Speed
DCT	Discrete Cosine Transform
DICOM	Digital Imaging and Communications in Medicine
ES	Encoding Speed
MPEG	Moving Picture Experts Group
MSE	Mean Squared Error
MSSIM	mean Structural Similarity Index
PSNR	Peak Signal-to-Noise Ratio
QL	Quality Loss
RLE	Run Length Encoding
SSIM	Structural Similarity Index

1 Introduction

In this thesis, various methods for compressing three-dimensional (3D) voxel textures in medical imaging are studied, with particular attention to formats and standards commonly accepted within the medical community. The chosen focus emerges from the fact that medical imaging data ranging from Computed Tomography (CT) scans to detailed pathology slides often exhibits large file sizes, making efficient compression a critical requirement for both archival and clinical workflows. This thesis is based on the work of Emre Cem Elevis, titled *Compression of 3D (Voxel) Textures for Medical Imaging*, which aims to evaluate popular approaches to compression and propose a novel archival method for large-scale medical data.

A core element of this thesis is the thorough comparison of different compression standards, including both classical still-image (e.g., PNG, JPEG-LS, JPEG 2000) and video-based methods (e.g., H.264/AVC, H.265/HEVC, FFV1). These codecs have been explored due to their ability to maintain acceptable image fidelity under high compression ratios. By considering their application within the DICOM standard, the thesis not only measures the raw compression efficiency but also contemplates standardization and compatibility issues pertinent to clinical use. Similar to the widely accepted lossy compression thresholds in diagnostic radiology, as summarized by multiple radiological societies, this research evaluates both lossless and near-lossless techniques, examining whether such techniques remain viable for diagnostic tasks.

The overarching objective is twofold: (1) to investigate compression efficacy, speed, and resultant image quality across several algorithms, and (2) to propose an innovative "intertwining" approach. In this new archival strategy, medically similar scans are grouped, reordered, and collectively encoded to exploit temporal and spatial redundancies. Much like a short-scale feasibility study, the outcomes are intended to shed light on whether medical institutions can reduce storage costs, or gain faster data transfers, without sacrificing clinically relevant details.

Due to the emphasis on real-world feasibility, the analysis of each compression

scheme is carried out on actual 3D voxel data, with size and bit-depth properties representing typical medical scenarios. In addition to quantitative results such as compression ratios, encoding times, PSNR, and SSIM visual assessments of residual artifacts are conducted to ensure that subtle anatomical features are preserved. Through these experiments, the benefits of leveraging advanced video codecs for 3D data and the potential improvements from the novel archival method are pinpointed.

This thesis is organized as follows: chapter 2 discusses the necessary theoretical background, including classic still-image encoding methods, the DICOM standard, and the emergence of 3D (video-based) compression approaches. In chapter 3, the methods, protocols, and performance metrics used in this research are presented. The novel archival strategy is introduced in chapter 4, followed by the empirical results in chapter 5. The implications and limitations of these findings are examined in **??**, while the final conclusion and outlook for further research are given in chapter 6. The purpose of this thesis is to determine optimal compression schemes for large-scale medical imaging, laying groundwork for future standardization and more efficient, clinically safe archiving of medical data.

2 Related Works

2.1 Shannon's Information Theory and Entropy Coding

2.1.1 Shannon's Information Theory

Claude E. Shannon introduced the concept of *information entropy*[20] as a measure of the average amount of information contained in a random variable X. Formally, the Shannon entropy $H(X)$ can be defined as:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\ \log_2\Big(p(x_i)\Big),$$

where $p(x_i)$ is the probability of the i-th symbol. The higher the entropy, the more information is contained in the signal, implying that it is harder to compress.

2.1.2 Entropy Coding

Entropy coding encompasses a family of coding methods designed to represent more probable symbols with fewer bits and less probable symbols with more bits. By matching code lengths to symbol frequencies, these methods can reduce the overall size.

These techniques are often the final step in many lossless compression pipelines. For instance, once data redundancies are minimized by run-length encoding, difference prediction, or dictionary methods, the result is then fed into an entropy coder for additional compression.

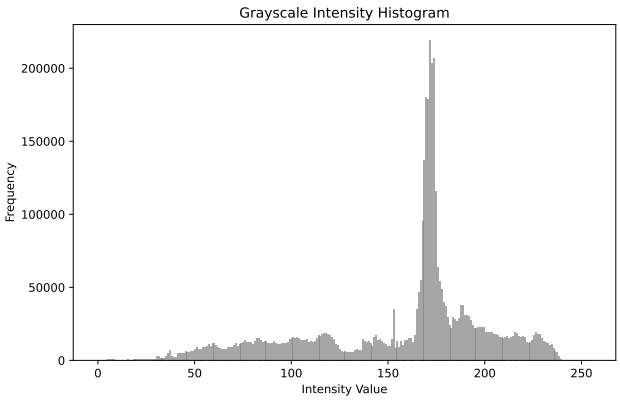

Figure 2.1: *Entropy vs. Probability Distribution of a grayscale teapot image*

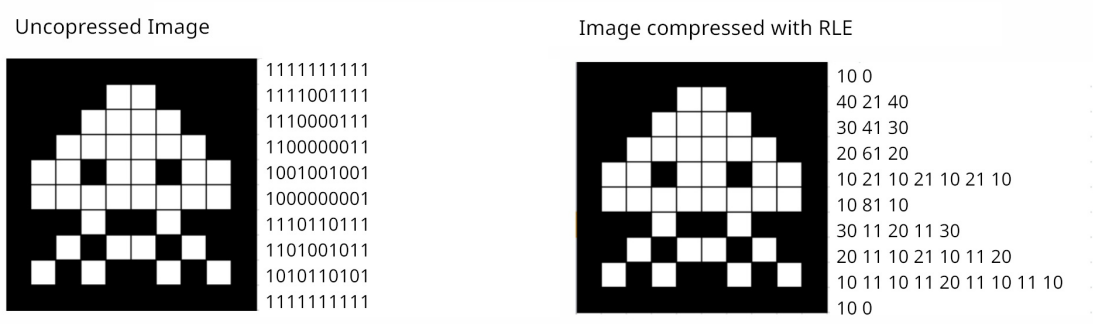

Figure 2.2: (RLE) applied to bitmaps.

2.1.3 Run Length Encoding (RLE)

RLE is a straightforward *lossless* data compression technique that encodes consecutive repeated symbols (or runs) by storing the symbol's value and the number of occurrences. In the context of 3D voxel textures, RLE is especially useful when large regions of homogeneous data exist, which is common in medical imaging datasets.

For example, a sequence of voxels with values `11110000` would be encoded as `4x1, 4x0` using RLE, thereby reducing storage requirements when uniform sequences are prevalent. However, for data with high complexity or frequent changes, RLE often becomes less efficient.

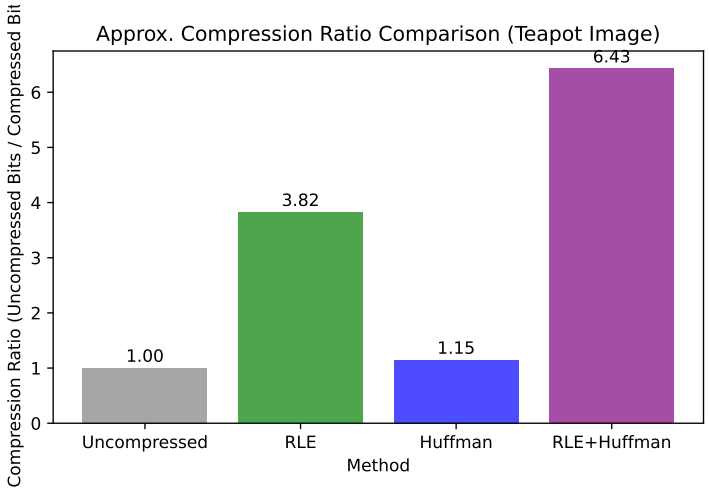

Figure 2.4: Compression comparison of RLE, Huffman and, both combined

2.1.4 Huffman Encoding

Huffman encoding[10] is another *lossless* data compression method that systematically assigns variable-length codes to symbols based on their frequencies. Symbols occurring more frequently receive shorter bit patterns, while those occurring less frequently are assigned longer ones. 2.5 The result is a **prefix condition** code , simplifying unambiguous decoding.

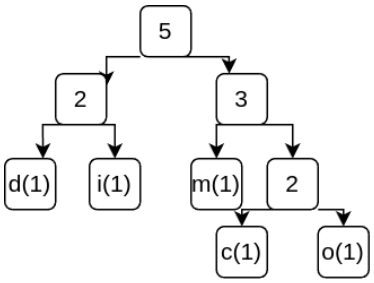

Figure 2.3: An example of a Huffman tree for encoding the string "dicom."

Huffman encoding is often the final step in compression pipelines (e.g., in PNG's DEFLATE algorithm), where it is combined with other compression methods such as LZ77 or RLE to achieve efficient encoding. 2.4

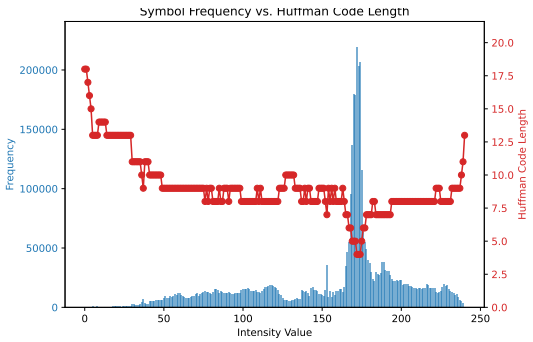

Figure 2.5: Symbol frequencies alongside the length of their Huffman codes

2.2 Digital Imaging and Communications in Medicine (DICOM)

2.2.1 DICOM Group

DICOM (Digital Imaging and Communications in Medicine) is an international standard for medical images and related information. It defines the formats for medical images that can be exchanged with the data and quality necessary for clinical use.

2.2.2 DICOM File Encapsulation

Header: Preamble	
Header: Prefix	
Data Elements	
Group 0010	
Element Tag	
0010: Name 0030: Birthday ...	

Group 0028	
0010: Image Row 0011: Image Columns ...	
Group xxxx	
...	
Pixel Data	
11 00 0D 00 0A 00 08 00 06 00 06 00 08 00 0A 00 00 0D 0F 00 12 00 12 00 10 00 0C 00 08 00 0A 00 09 00 08 00 08 00 08 00 09 300 0A 00 0B 00 08 00 09 00 0A 00 08 00 08 00 0C 00	

Every DICOM file consist of three parts. First, a 128 byte preamble header followed by a 4 byte ASCII prefix of 'D' 'I' 'C' 'M '. Second, the Data Elements which contains ASCII text information both about the patient and the current sample such as patient's name, birthday, ID, sample size, sample location, sample duration etc.

Each data element has a tag field consisting of 2 byte group number and 2 byte element number, a length field and a value field. Important to note is that there are thousands of various DICOM data elements with a well specified meaning. But it is still possible define additional elements called 'private elements' which are indicated with an odd group number while standard DICOM elements are indicated with even numbers Optionally it has a Value Representation (VR) field to specify the listed data type.[6]

Last and third, is the Image Pixel data, this raw data contains formatted information according to the elements of the pixeldata group 0028. Usually image file cannot be decoded without this group's information.

This closely knit combination of sample and additional sample information is

primarily why DICOM standards are so widely used in the integration of digital imaging systems in medicine.

2.2.3 Compression as Utilized in the DICOM Standard

The DICOM standard uses various JPEG protocol identifier data elements to declare the image transfer syntax used. the ones currently in use can be found in the diagram

Keyword	Name
JPEGBaseline8Bit	JPEG Baseline (Process 1): Default Transfer Syntax for Lossy JPEG 8 Bit Image Compression
JPEGExtended12Bit	JPEG Extended (Process 2 & 4): Default Transfer Syntax for Lossy JPEG 12 Bit Image Compression (Process 4 only)
JPEGLosslessSV1	JPEG Lossless, Non-Hierarchical, First-Order Prediction (Process 14 [Selection Value 1]): Default Transfer Syntax for Lossless JPEG Image Compression
JPEGLSLossless	JPEG-LS Lossless Image Compression
JPEGLSNearLossless	JPEG-LS Lossy (Near-Lossless) Image Compression
JPEG2000Lossless	JPEG 2000 Image Compression (Lossless Only)
JPEG2000	JPEG 2000 Image Compression
JPEG2000MCLossless	JPEG 2000 Part 2 Multi-component Image Compression (Lossless Only)
JPEG2000MC	JPEG 2000 Part 2 Multi-component Image Compression
JPIPReferenced	JPIPReferenced
JPIPReferencedDeflate	JPIPReferencedDeflate
HTJ2KLossless	High-Throughput JPEG 2000 Image Compression (Lossless Only)
HTJ2KLosslessRPCL	High-Throughput JPEG 2000 with RPCL Options Image Compression (Lossless Only)
HTJ2K	High-Throughput JPEG 2000 Image Compression
JPIPHTJ2KReferenced	JPIP HTJ2K Referenced
PIPHTJ2KReferenced Deflate	JPIP HTJ2K Referenced Deflate

Figure 2.6: Transfers methods officially recognized by DICOM

2.3 JPEG Standard

As discussed in Section 2.1.2, **Shannon's Information Theory** underpins nearly all modern compression algorithms by highlighting how data redundancy can be reduced through optimal code assignment. The *Joint Photographic Experts Group* (JPEG) leveraged these ideas when developing its original image compression standard in 1992.

JPEG supports both **lossy** and **lossless** compression modes, with the lossy mode being the most common for everyday usecases. In the *lossy* pipeline (often referred to as *baseline JPEG*), each 8×8 block of pixels is transformed into the frequency domain via the **Discrete Cosine Transform (DCT)** . By quantizing the resulting DCT coefficients, many high-frequency

details (which are less perceptible to the human eye) are discarded or reduced, shrinking the data size. Finally, an **entropy coding** step typically Huffman or arithmetic coding, as detailed in Section 2.1.4 encodes the quantized coefficients using fewer bits for more common values. This flow is summarized as:

1. **Block Splitting:** The image is divided into 8×8 blocks.
2. **DCT Transform:** Each block is converted to the frequency domain.
3. **Quantization:** High-frequency components are reduced or set to zero based on a quantization matrix.
4. **Entropy Coding:** The quantized coefficients are compressed using Huffman or arithmetic coding.

Because lossy JPEG exploits characteristics of *human visual perception*, it often achieves high compression ratios with minimal perceptual distortion. Figure 2.7 shows how varying the compression quality impacts the appearance of the *Utah Teapot* image. Figure 2.11 shows the impact on the Structural Similarity Index (SSIM) 3.3 values at varying bitrates while figure 2.10 shows the impact on the Peak Signal-to-Noise Ratio (PSNR).

Figure 2.7: Visual comparison of different JPEG quality settings using the Utah Teapot.[9] Lower quality factors discard more high-frequency data, resulting in visible artifacts.

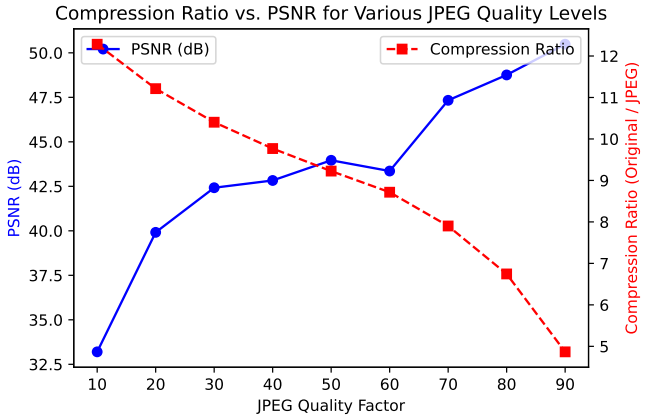

Figure 2.8: *Compression Ratio vs. PSNR:*file size reduction versus the peak signal-to-noise ratio (PSNR) for various JPEG quality factors applied to the Utah Teapot image.

2.3.1 JPEG-LS (Lossless JPEG)

In scenarios requiring **exact** reproduction of pixel data, a *lossless* variant of JPEG referred to as **JPEG-LS** can be used. Rather than discarding high-frequency information, JPEG-LS employs a **predictive** coding scheme combined with entropy coding to achieve lossless compression.

Specifically, it attempts to predict each pixel based on its nearest neighbors, encodes the prediction error, and then uses an entropy coder.2.1.2, JPEG-LS found utility in fields like *medical imaging*, where diagnostic images require precise storage. In practice, JPEG-LS typically outperforms general-purpose compressors like PNG for medical images,5 offering both higher compression and far lower computational cost. For example, hardware implementations of JPEG-LS report compression ratios similar or superior to those of lossless JPEG 2000, and distinctly better than PNG, on both grayscale and color images. This efficiency makes JPEG-LS attractive for medical applications where fast, lossless compression is needed without adding undue processing load. JPEG-LS also supports an optional "near-lossless" mode, allowing a user-defined error tolerance (e.g. a few gray levels) to further improve compression when strictly lossless reconstruction is not required.

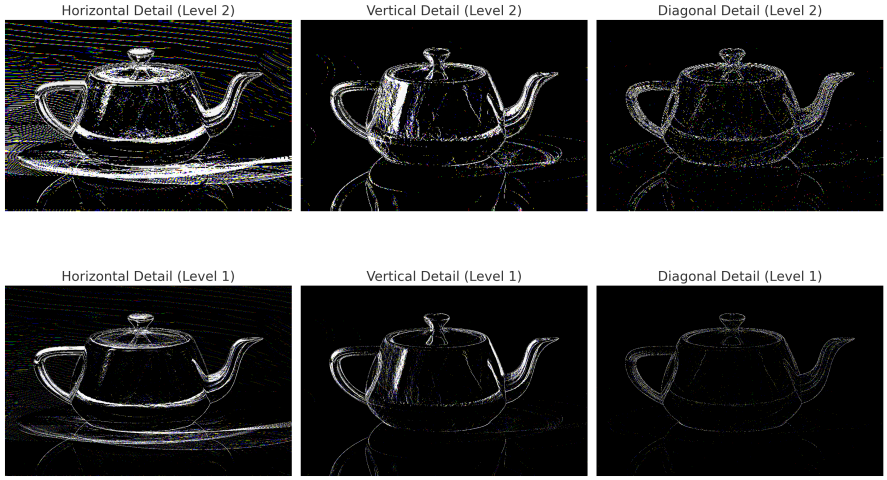

Figure 2.9: An example 2D wavelet transform of the Utah Teapot[9] using an integer wavelet (Le Gall 5/3). Note how the transform sub-bands separate coarse and detail information.

2.3.2 JPEG 2000 and HTJ2K

JPEG 2000 (ISO/IEC 15444) is another important standard, offering state-of-the-art compression performance and rich features for high-bit-depth images. JPEG 2000 uses wavelet transform coding (with a reversible integer wavelet for lossless mode) and advanced bit-plane entropy coding (EBCOT), replacing the 8×8 block DCT-based approach of the original JPEG and Leveraging integer wavelets such as the *Le Gall 5/3* (see Fig. 2.9). In lossless mode, JPEG 2000 provides exact reconstruction while typically achieving better compression ratios than the old lossless DCT based JPEG standard.[2] Moreover, JPEG 2000 supports **any bit depth** and even multi-component (e.g. color) images, and it enables progressive transmission by quality or resolution. Another benefit is **Region of Interest (ROI) Coding:** Specific regions (e.g., a tumor in a medical scan) can be encoded at higher quality than the background, focusing bits where they matter most. Due to these advantages in medicine, JPEG 2000 was incorporated into DICOM and other medical image workflows for both lossless and lossy compression. A known drawback, however, is its computational complexity: encoding/decoding JPEG 2000 is very CPU-intensive.

An important enhancement in this family is *High Throughput JPEG 2000*

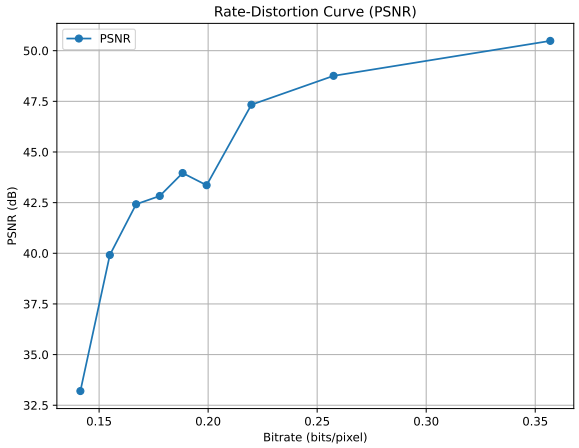

Figure 2.10: JPEG PSNR Comparison at varying quality settings

(HTJ2K), designed to overcome its drawback with encode and decode speeds significantly faster while retaining the compression advantages of wavelet transforms. This improved throughput makes JPEG 2000 more practical for *real-time* or *large-volume* applications frequently encountered in medical imaging.

2.3.3 Implementation in DICOM for Medical Imaging

DICOM) format integrates JPEG, JPEG-LS, and JPEG 2000 (including HTJ2K) as recognized transfer syntaxes for storing and transmitting medical images. (See figure 2.6)

2.4 PNG Standard and Other Lossless Methods

The **Portable Network Graphics (PNG)** format is a widely used lossless image format that employs the *Deflate* compression algorithm, which combines LZ77 dictionary coding with Huffman entropy coding. PNG was originally designed as a patent-free replacement for GIF (Graphics Interchange Format) and supports up to 16-bit grayscale and 48-bit color images, exceeding GIF's 8-bit limitation.

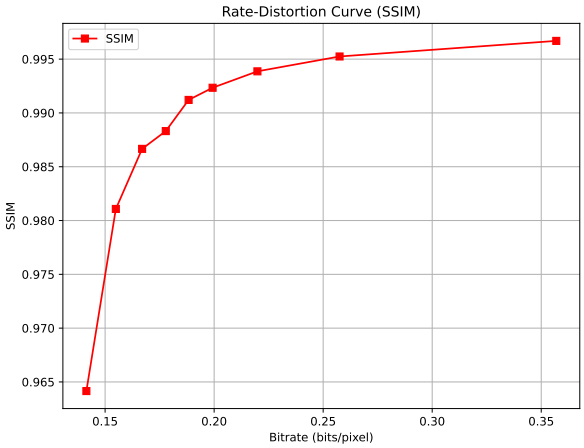

Figure 2.11: JPEG SSIM Comparison at varying quality settings

In practice, however, PNG's compression performance on high bit-depth medical images is often suboptimal. It uses simple predictive filters on image scanlines and then applies the Deflate compressor (as in ZIP/gzip) to the pixel data. This general-purpose approach was not tailored for continuously-toned 16-bit medical images, and studies have found PNG to be less effective than dedicated medical image compressors – especially for images greater than 8 bits per pixel. For example, in a comparison of lossless methods on 16-bit images, PNG achieved significantly lower compression ratios than specialized algorithms. While PNG remains convenient and widely supported, its lack of native support for volumetric (3D) data means that each slice of a medical volume must be stored and compressed independently, forfeiting any cross-slice redundancy benefits.

The **Tagged Image File Format (TIFF)** is another popular format in medical imaging for storing high-resolution, high bit-depth images. TIFF is extremely flexible: it supports grayscale and color depths of 16 bits (and beyond), and it allows multiple images (pages) in a single file (useful for encoding image stacks or slices). For compression, TIFF can employ several lossless schemes. A common option is *Lempel–Ziv–Welch (LZW)* compression, a dictionary-based method historically regarded as the de facto standard compression in TIFF. Another option is *PackBits*, a simple RLE scheme for lossless compression. Modern TIFF extensions even allow Deflate (ZIP) compression as a lossless method. These compression choices

are internal to the TIFF container; unlike video codecs, they compress each image frame independently without spatial or temporal cross-frame prediction. TIFF's ability to handle 16-bit pixels and to losslessly compress images (or leave them uncompressed) without quality loss on re-saving makes it well suited for medical use cases where preserving full fidelity is essential. However, like PNG, standard TIFF compression does not exploit inter-image redundancy in multi-frame data – each slice or frame is compressed in isolation.

2.5 3D Compression (Video-Based Methods)

Inter-frame compression uses motion estimation to exploit temporal redundancy between image frames. In contrast to Intra-Frame compression which exploits the low entropy that is inherent to natural images and encapsulates each frame separately. For our applications in video and volumetric image compression, a series of 2D images(frames) can be compressed more efficiently by leveraging similarities across adjacent frames within the volume, rather than compressing each frame independently. This is the basis of the Moving Picture Experts Group (MPEG) family of video codecs developed by the Moving Picture Experts Group. An **inter-frame** codec will search for matching regions between a current frame and one or more reference frames, and encode the displacement (motion vector) and difference residual instead of the full image content. Figure 2.12 illustrates this process: the encoder finds a block in a reference frame that closely predicts the block in the current frame, and encodes only the motion vector plus the small difference (residual) needed to correct the prediction.(See figure 2.13) Because consecutive medical images (e.g., slices in a CT volume) often exhibit only slight changes, inter-frame prediction can greatly reduce the data to be encoded. The residual errors are typically transformed and entropy-coded, just as in still-image compression, yielding a high compression ratio if the temporal redundancy is high. This video-based approach contrasts with traditional DICOM image compression where each frame is compressed in isolation (e.g., using JPEG on each slice) without exploiting inter-frame correlations. Research has demonstrated substantial gains by compressing medical image sequences as video: one study reported that combining JPEG-LS intra-frame coding with inter-frame motion compensation achieved up to a 77%[11]

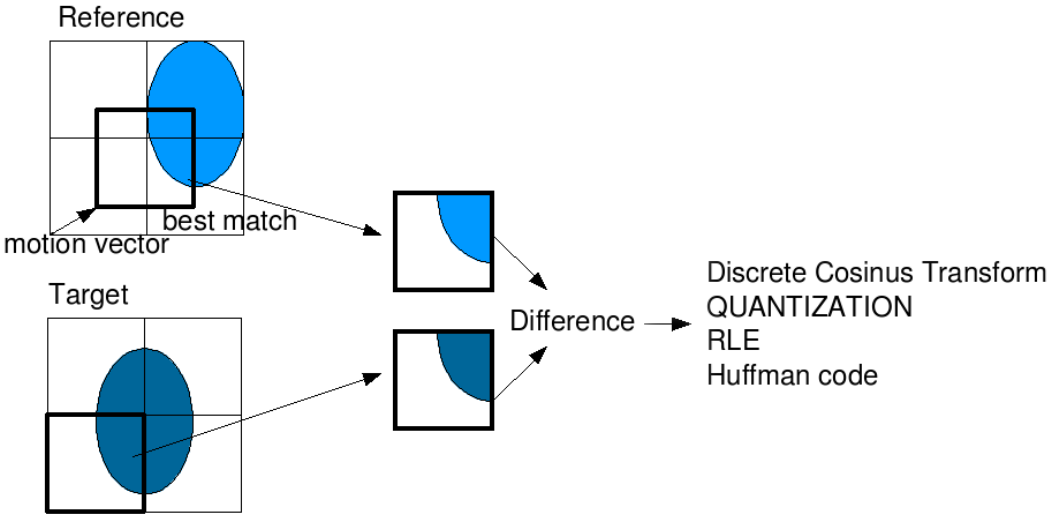

Figure 2.12: Inter-frame prediction process. In this case, there has been an illumination change between the block at the reference frame and the block which is being encoded: this difference will be the prediction error to this block.[15]

reduction in data size for an MRI sequence compared to JPEG-LS alone . Likewise, encoding CT/MR slices in three dimensions (treating the volume as a video) can improve compression by 25%[11] over 2D slice-by-slice methods. These results underscore the potential of 3D and temporal redundancy exploitation in medical image compression.

2.5.1 H.264/AVC and H.265/HEVC

Modern **video codecs** such as **H.264/AVC** and **H.265/HEVC** represent the state of the art in inter-frame compression and have been considered for compressing medical imaging series. H.264/AVC (MPEG-4 Part 10) uses block-based intra-frame coding (an integer DCT transform with entropy coding) and inter-frame motion prediction with multiple reference frames. Its successor H.265/HEVC further refines this by allowing larger and more flexible block structures (see figure2.15), improved motion compensation precision, and more sophisticated entropy coding, achieving roughly 50% better compression than H.264 for the same visual quality in general video applications.

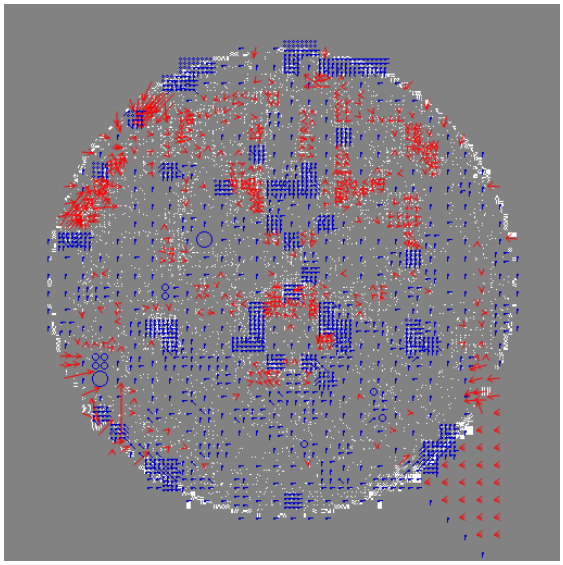

Figure 2.13: Residual information and motion vectors on a CT slice using H264 encoding

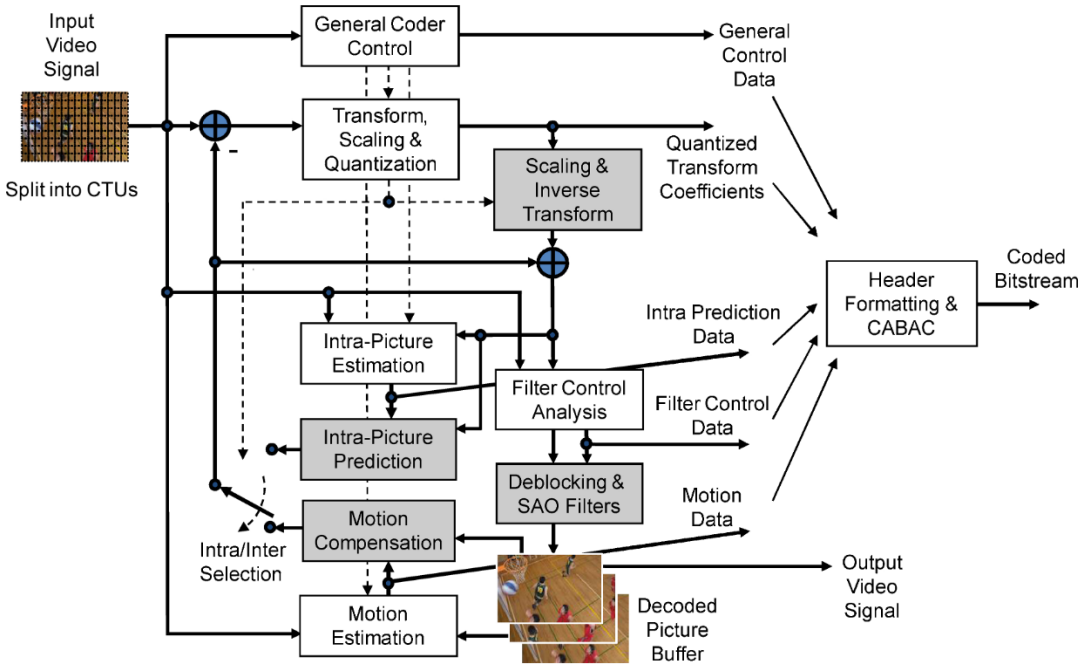

Figure 2.14: Structure of an HEVC encoder[21]

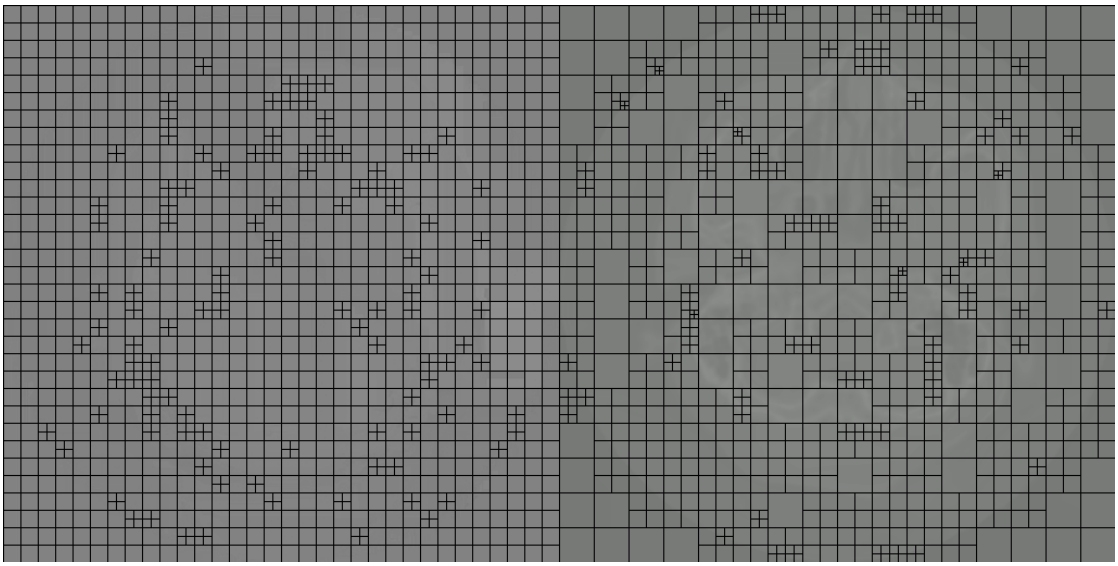

Figure 2.15: Different Grouping strategies employed by H.264 (left) and H.265 (right)

2.5.2 FFV1

FFV1 (FFmpeg Video Codec 1) is a lossless intra-frame video codec. FFV1 is an open-format, lossless codec that employs predictive coding and entropy coding on each frame independently (without inter-frame motion vectors). Originally developed in the FFmpeg project, it has gained popularity in the archival community for its impressive compression density and speed on high-bit-depth content. FFV1 typically achieves compression ratios comparable to or better than JPEG 2000's lossless mode, while requiring less computational overhead. (((Many institutions like the U.S. Library of Congress,[7] have adopted FFV1 for video preservation.)) For medical imaging, FFV1 offers an attractive option when one needs to compress large multi-frame studies (such as angiography runs or surgical videos) without losing any data. It supports high-bit-depth pixel formats (up to 16-bit RGB/YUV and beyond) and is capable of handling grayscale or color frames, which aligns well with the needs of 16-bit grayscale DICOM imagery. Moreover, as a intra-frame codec, FFV1 avoids any temporal artifact propagation and each frame can be decoded independently, a useful property for random access in medical cine loops. But cannot exploit the temporal similarities within the sample to achieve better

Table 2.2: Typical Image Dimensions and Uncompressed File Sizes for Common Medical Imaging Techniques[13]

Modality	Dimensions	Bit Depth	Median Data Size
CT	512, 512, 500	12–16 bits	$\approx 520MB$
MRI	256, 256, 20, 25	12–16 bits	$\approx 130MB$
Pathology	30,000, 30,000	8 bits (color)	$\approx 2.5GB$
PET	128, 128, 40	16 bits	$\approx 32MB$
Radiography	2000, 250)	10–16 bits	$\approx 20MB$
Tomosynthesis	2457, 1890, 50	10–16 bits	$\approx 0.4GB$
Ultrasound	512, 512, -, 50 frames/s	8 bits	$\approx 38MB/s \approx 260MB$

compression ratios.

2.6 Key Features of Medical Imaging Datasets

2.6.1 Volumetric Data-Sets

Computed Tomography (CT)

CT scans produce high-resolution volumetric data, typically with a fixed 512×512 in-plane resolution. The number of slices varies widely depending on anatomy and protocol. Cardiac CT frequently includes multiple temporal frames to capture motion, generating significantly larger datasets.

Magnetic Resonance Imaging (MRI)

MRI scans commonly yield 3D datasets or 4D cine sequences (especially cardiac MRI). Typical resolutions range between 256×256 to 512×512 pixels. Cardiac cine MRI produces time-resolved datasets to evaluate motion, resulting in moderate-sized data volumes.

Positron Emission Tomography (PET)

PET datasets have comparatively lower spatial resolutions (128×128 to 256×256) but may include temporal sequences (e.g., cardiac gating). Data volume remains

moderate due to lower resolution and fewer slices compared to CT or MRI.

Tomosynthesis (3D Mammography)

Breast tomosynthesis produces limited-angle 3D reconstructions from multiple X-ray projections. Typical studies generate moderate-sized data volumes with approximately 50 thin slices, each having high spatial resolution.

2.6.2 Non Volumetric Data-Sets

By their characteristics, they are not fit well for inter-frame encoding, they are usually outside the scope of this paper.

Digital Pathology (Whole Slide Imaging)

Whole slide imaging generates extremely high-resolution 2D color images, often exceeding billions of pixels per image. High-resolution scanning at 20× magnification easily produces files exceeding gigabytes in size, necessitating compression and specialized storage solutions.

Radiography (Chest X-Ray)

Digital radiography provides high-resolution single-frame images, commonly around 2000×2500 pixels. These datasets are relatively small in size, facilitating rapid processing and storage.

Ultrasound (Cardiac)

Cardiac ultrasound generates real-time sequences at high frame rates (30–60 fps). Typically producing color images with moderate resolution (512×512 pixels), the data rate is high (38 MB/s uncompressed), requiring efficient compression and management strategies.

2.7 Lossy Compression Standards, Globally Accepted in Medical Imaging Community

There are many precedents among physicians of all fields that perfect reconstruction of the probe information is inconsequential to the diagnostic procedure. This is often because the very subtle details the compression encoders remove are finer than the inherent noise within the measurement modality. For this reason, many physicians and medical organizations around the world have accepted certain lossy algorithms where a loss in quality is tolerated. We follow the diagnostic guidelines outlined by the German Rontgen Society and supporting European societies.[12, 18, 22]

2.7.1 Medically Acceptable Compression Ratios

Medical imaging organizations in Germany, the UK, and Canada have all issued clinically tested maximum compression thresholds for diagnostically acceptable image quality. These thresholds vary by modality and anatomy. For example, CT images of the body and chest tolerate higher compression ratios than brain CT or digital mammography. MRI and PET imaging also permit moderate lossy compression, especially for non-critical diagnostic tasks.[16, 17]

The German Rontgen Society recommends a maximum of 8:1 compression for body CT, 5:1 for brain CT, and up to 10:1 for general MRI use.[12] The European Society of Radiology (ESR) echoes these recommendations and emphasizes a principle of Diagnostically Acceptable Irreversible Compression (DAIC), where compression is permitted only if it does not interfere with clinical diagnosis.[18] For mammography and tomosynthesis, either lossless or near-lossless compression is advised to preserve microcalcifications and fine tissue structures.[8, 22]

As summarized in Table 2.3, most modalities allow for significant lossy compression under clinical evaluation. Nevertheless, the final decision lies with the interpreting radiologist, who must verify that diagnostic content is preserved.[16, 18]

Table 2.3: Examples of Diagnostically Acceptable Lossy Compression Ratios by Modality

Imaging Modality (Region)	**Recommended Max. Ratio**
CT (Body/Chest)	10:1–15:1
CT (Brain/Head)	5:1–12:1
MRI (General)	5:1–10:1
PET / Nuclear Medicine	Up to 10:1
Digital Radiography (Chest/Skeletal)	10:1–30:1
Mammography (2D)	15:1–20:1
Tomosynthesis (3D Breast)	Near-lossless only

2.7.2 Bit-Depth Reduction in Clinical Compression

Medical images are frequently acquired in 12- to 16-bit grayscale formats. However, not all bits carry clinically significant information. Downsampling from 16 bits to 10 or 12 bits has been shown to retain sufficient dynamic range and detail for diagnostic use, especially since diagnostic monitors and human perception are limited to approximately 10-bit grayscale resolution.[19, 22]

European and Canadian guidelines support this form of bit-depth quantization when the resulting images remain indistinguishable from originals after contrast adjustment. However, reductions below 10-bit are discouraged for primary diagnosis, as they may obscure low-contrast features in soft tissue or pathology.[16]

2.7.3 Transcoding for Compression: FFmpeg Encoding Best Practices

For long-term storage and transmission, encoding series of CT, MRI, or PET images using video compression techniques (e.g., H.264/AVC or H.265/HEVC) is gaining traction. Since these codecs exploit inter-frame redundancy, they yield higher compression ratios for volumetric scans.[19]

To ensure medically acceptable compression, encoding must retain bit-depth and avoid chroma subsampling that alters grayscale precision. Below are example FFmpeg commands for converting 16-bit grayscale PNG slices into high-bit-depth video sequences:

H.264 High 10 Profile (10-bit):

```
ffmpeg -f image2 -framerate 1 -i slice_%03d.png
-c:v libx264
-profile:v high10 -pix_fmt yuv420p10le -preset veryslow
-crf 20 -g 30 -tune grain -colorspace bt709 -color_trc bt709
-color_primaries bt709 output_avc10.mp4
```

H.265 Main 12 Profile (12-bit):

```
ffmpeg -f image2 -framerate 1 -i slice_%03d.png -c:v libx265 -profile:v
main12 -pix_fmt yuv420p12le -preset slow -crf 22 -g 30 -x265-params "monochrome=1"
-colorspace bt709 -color_trc bt709 -color_primaries bt709 output_hevc12.mkv
```

2.8 Conclusion

Overall, the integration of lossless and lossy still compression methods from PNG, TIFF to JPEG-LS and JPEG2000, and the video-based codecs like H.264/H.265 and FFV1 for sequences, provides a robust toolkit for handling many medical usecases. Each method offers a trade-off between complexity, compression efficiency, and compatibility. There appears to be no widely accepted codec that can fully leverage the spatial and temporal redundancies inherent in medical imaging data while guaranteeing reconstruction standards of diagnostically critical information.

3 Methodology

3.1 Research Questions

RQ1: What are the quantifiable benefits and trade-offs to re-encode old medical information with brand new compression techniques?
RQ2: Are there encapsulation methods that more suitable to certain type of medical imaging technologies? How feasibly can they be integrated with the current DICOM standard?

3.2 Research Protocol

3.2.1 Various schemas used for testing

Method	Bit Depth per Component	Overview
JPEG	8	Lossless
JPEG 2000	1 to 38	Lossless
JPEG LS	1, 4 ,8 ,16	Lossless
JPEG LS	1, 4 ,8 ,16	Near Lossless
PNG	1, 4 ,8 ,16	Lossless (no transparency channel)
H.264	10	Very Slow Preset, 4:2:0 YUV
H.264	10	Very Slow Preset, 4:4:4 YUV
H.264	10	Very Slow Preset, 4:2:0 YUV -crf 20 -g 30 -tune grain
H.265	12	Medium Preset, 4:4:4 YUV
H.265	12	Very Slow Preset, 4:4:4 YUV
H.265	12	Slow Preset, 4:2:0 YUV -crf 22 -g 30
FFV1	1, 4 ,8 ,16	Lossless ,gray16be

3.2.2 Sample Set for Testing

Uncompressed DICOM images found from public sources combined with newly encapsulated DICOM files created from base medical images. All images used, have been gray scaled.

3.2.3 Dependant Variables

Variable	Explanation
CR	a compression ratio denoted to the second decimal
ES	encoding speed measured in ms
QL	quality loss measured in MSSIM and PSNR

3.3 Performance Metrics

3.3.1 MSE

The Mean Squared Error Mean Squared Error (MSE) is a commonly used quantitative metric to measure the difference between an original image I and its reconstructed version I'. The MSE is calculated as the average squared intensity differences between corresponding pixels:

$$MSE(I, I') = \frac{1}{N} \sum_{i=1}^{N} (I_i - I'_i)^2 \tag{3.1}$$

where N is the total number of pixels, and I_i and I'_i represent the intensity values of the i-th pixel in the original and reconstructed images, respectively. Lower MSE values indicate higher similarity and better reconstruction quality.

3.3.2 PSNR

The Peak Signal-to-Noise Ratio PSNR is another widely used metric derived from MSE, typically employed to quantify reconstruction quality. PSNR is defined as:

$$PSNR(I, I') = 10 \cdot \log_{10} \left(\frac{L^2}{MSE(I, I')} \right) \tag{3.2}$$

where L is the dynamic range of pixel intensities (commonly $2^{\text{bits per pixel}} - 1$). PSNR values are expressed in decibels (dB), with higher PSNR indicating better image quality.

3.3.3 SSIM

The reconstruction quality is measured using the Structural Similarity Index Measure SSIM, which evaluates perceptual similarity rather than relying solely on absolute intensity differences. Unlike absolute difference metrics, such as MSE and PSNR, SSIM considers luminance, contrast, and structural information, aiming to align better with human visual perception. SSIM values range between -1 and 1, with values closer to 1 indicating higher similarity, and values closer to -1 indicating higher anti-correlation, while 0 indicating no similarity.

For images I and reconstructed image I', SSIM is defined as:

$$SSIM(I, I') = \frac{(2\mu_I\mu_{I'} + c_1)(2\sigma_{II'} + c_2)}{(\mu_I^2 + \mu_{I'}^2 + c_1)(\sigma_I^2 + \sigma_{I'}^2 + c_2)} \tag{3.3}$$

where:

- μ_I and $\mu_{I'}$ are the averages of I and I'
- σ_I^2 and $\sigma_{I'}^2$ are the variances of I and I'
- $\sigma_{II'}$ is the covariance between I and I'
- $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are constants to stabilize the division
- L is the dynamic range of pixel values (typically $2^{\text{bits per pixel}} - 1$)
- $k_1 = 0.01$ and $k_2 = 0.03$ by default

SSIM is typically computed locally using a sliding window approach, and the mean SSIM (MSSIM) over the entire image is defined as:

$$MSSIM(I, I') = \frac{1}{M} \sum_{j=1}^{M} SSIM(I_j, I'_j) \tag{3.4}$$

where M is the number of local windows in the image, and I_j, I'_j represent image contents within the j-th local window.

3.3.4 Compression Ratio (CR)

The Compression Ratio (CR) quantifies the effectiveness of a compression algorithm by comparing the size of the original data to that of the compressed data. It is defined as:

$$CR = \frac{S_o}{S_c} \tag{3.5}$$

where:

- S_o is the size (in bits or bytes) of the original data,
- S_c is the size (in bits or bytes) of the compressed data.

A higher CR value indicates more efficient compression, as a larger original file is represented by a smaller compressed file.

3.3.5 Analysis Plan

A pipeline that will encode, decode and analyze the various methods with the dataset has been created to log the dependent variables. Another pipeline to visualize the dissimilarities by generating heatmaps based on individual SSIM and MSE calculations have been created to look deeper in the regions effected by the compression. As well as the tools to group and analyze similar images so further insights to the data-encoding pairs can be found. Storage gains possible with the standards, and more importantly the similarity metrics with the uncompressed have been explored and their feasibility have been outlined with the possible implementation steps to accommodate the medical use-cases.

3.3.6 Testing Environment and Hardware

Python 3.10.4 and the following libraries combined with the FFMPEG codecs have been used to conduct this paper.
numpy, opencv-python, scikit-image, imagecodecs, pandas, pillow-jpls, OPENJPEG, $\text{jpeg}_{ls}, glymur$,
$The computer used for all the testing sis an AMD Ryzen Threadripper 2990WX32 - Core Processor 3.00GHz with 128GB RAM running x64 Windows 10.$

4 Proposed Archival Method

We propose relatively a novel method for archiving similar (i.e., images taken of the same body part using the same camera calibration) 3D medical scans together by "intertwining" similar slices together, utilizing modern video encoding techniques such as H.264/AVC and HEVC 2.5.1. These codecs are engineered to leverage similarities between consecutive frames, and intertwining the frames across different but similar samples increases the overall interframe redundancy, thus improving CR. This is most suitable as an archival method for modalities where variance between samples are low such as skeleton and head CT's. Although our approach might not achieve the absolute compression ratios of specialized domain-specific methods, it benefits from the robustness, reproducibility, and continued support provided by off-the-shelf encoding systems.

4.0.1 Exploiting Temporal Redundancy

In both H.264 and HEVC, once the spatial representation is obtained via the DCT and DST, the codecs exploit temporal redundancy using interframe prediction. To quantify the similarity between successive frames, we compare their transform coefficient matrices. Let $T_1(u, v)$ and $T_2(u, v)$ denote the transform coefficients (obtained via either DCT or DST) for two successive frames. A similarity metric can be formulated as:

$$S(T_1, T_2) = \frac{\sum_{u,v} |T_1(u, v) - T_2(u, v)|^2}{\sqrt{\sum_{u,v} |T_1(u, v)|^2 \sum_{u,v} |T_2(u, v)|^2}} \tag{4.1}$$

We consider two frames sufficiently similar for interframe encoding if:

$$S(T_1, T_2) \leq \epsilon \tag{4.2}$$

where ϵ is a predefined similarity threshold. When this condition is met, the encoder reorders the frames that are most similar next to each other, achieving significant compression by capitalizing on temporal redundancies. It is important

to note that this metric is performance intensive, And should be calculated for a few key frames with the same positioning across different samples. An initial grouping based on additional factors such as modality, sex, age, weight, position of the slice is also necessary.

4.1 Compression Scheme

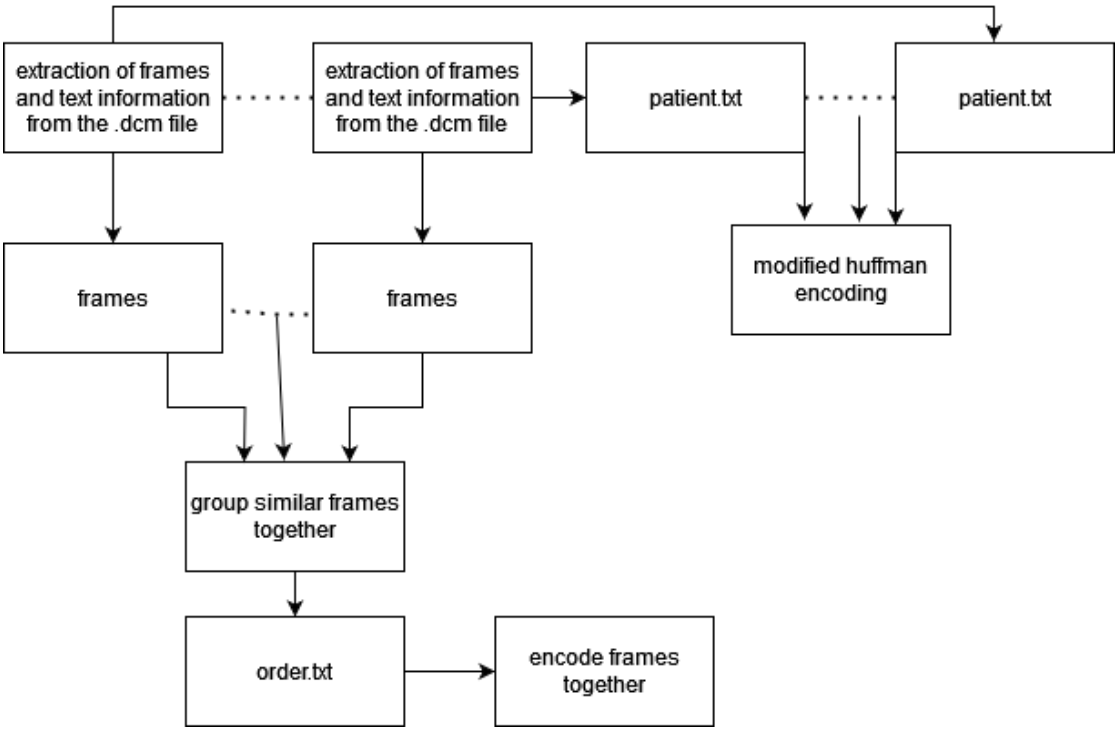

Figure 4.1: Compression Schema

Our compression approach consists of three main steps:

- **Extraction** Frames and the relavent patient information is extracted.
- **Grouping:** After a grouping based on keywords and contextual information, similar enough (see 4.1) scans are found and their frames are ”intertwined” by finding an efficient ordering.
- **Encoding:** The frames are encoded together with the chosen parameters.

Further improvements such as encoding the the patient information and the scan together similar to[1] is possible,

4.2 Implementation and File Organization

Our archive structure is organized as:

```
Archive
    order.txt     # Order information
    patients.txt  # Patient information
    encoded-video # .avi .mkv or .mp4 depending on the format used
```

4.3 Quality Assessment

We use theSSIM and PSNR to measure the quality of our compressed images.

5 Results

5.1 Volumetric Information

5.1.1 Human CT scans

21 different head CT scans as well as 4 breast tomographies have been encoded with the various methods shown in 3

Breast Tomography

4 different Breast Tomography scans have been tested

Intertwined Data

21 different Head CT scans have been encoded into a single file to provide the results.

5.2 Non-Volumetric Information

5.2.1 Human X-Ray Scans

15 different X-Ray scans have been tested.

5.2.2 Pathology Scans

2 Slices one on the liver and the other on the breast have been tested.

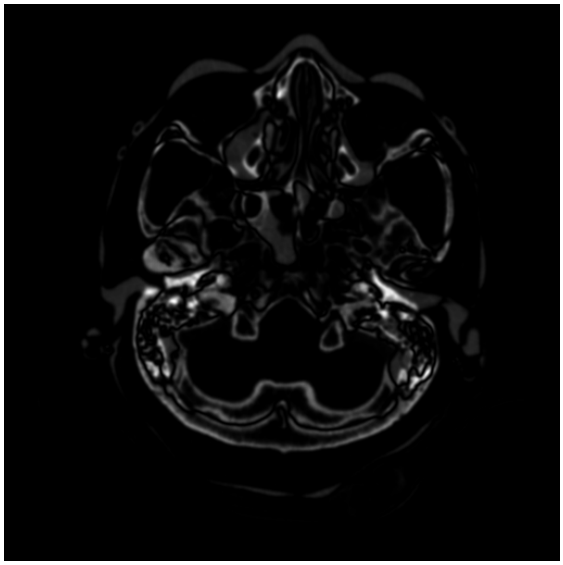

Figure 5.1: An overlayed heatmap to display the structural changes between 5 frames.

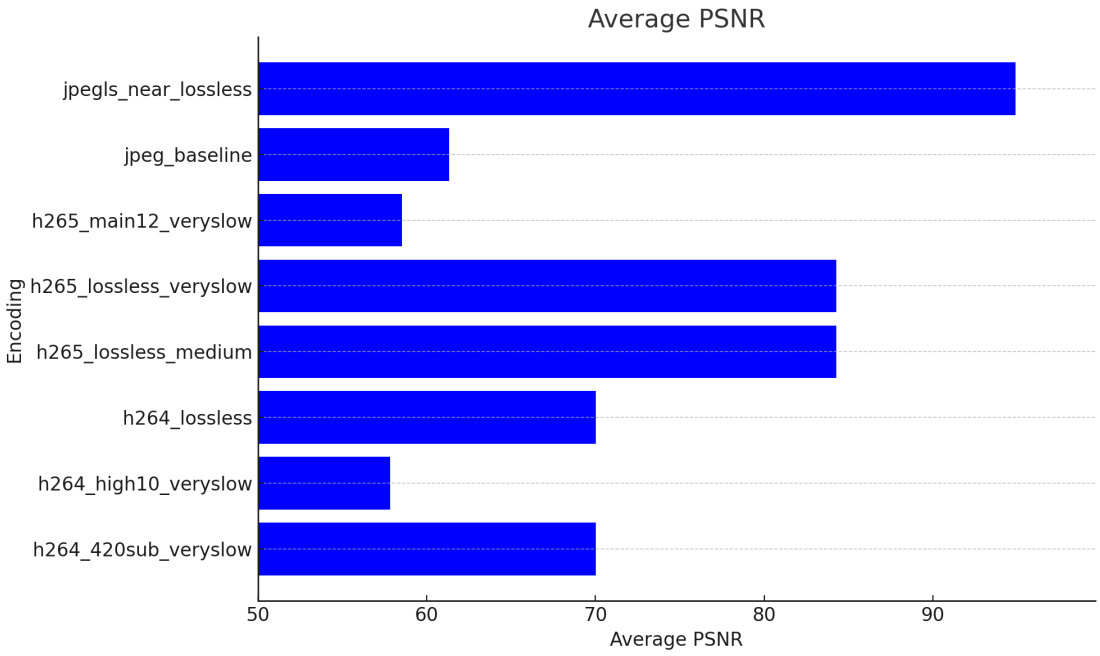

Figure 5.2: Averaged PSNR Values for Head CT

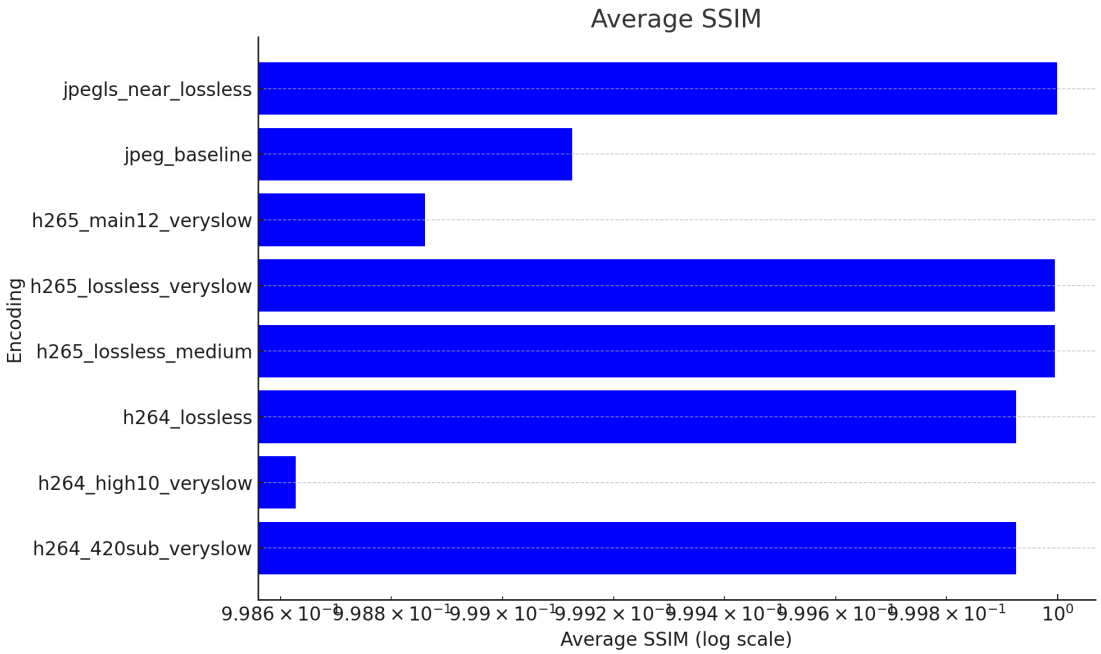

Figure 5.3: Averaged SSIM Values for Head CT

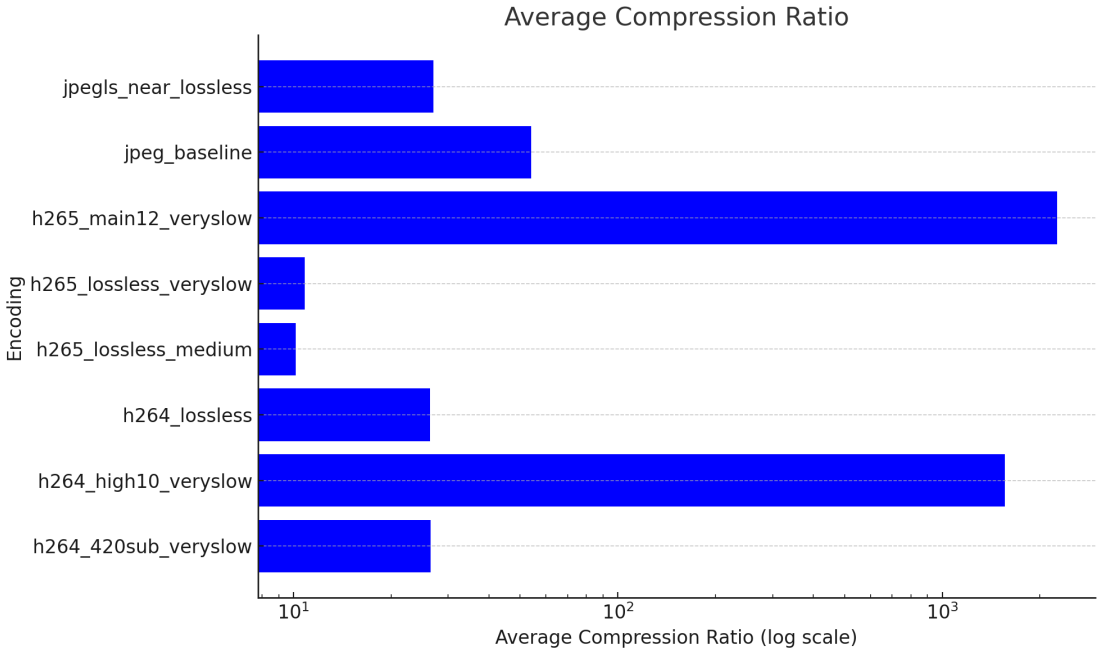

Figure 5.4: Averaged Compression Ratio for Head CT

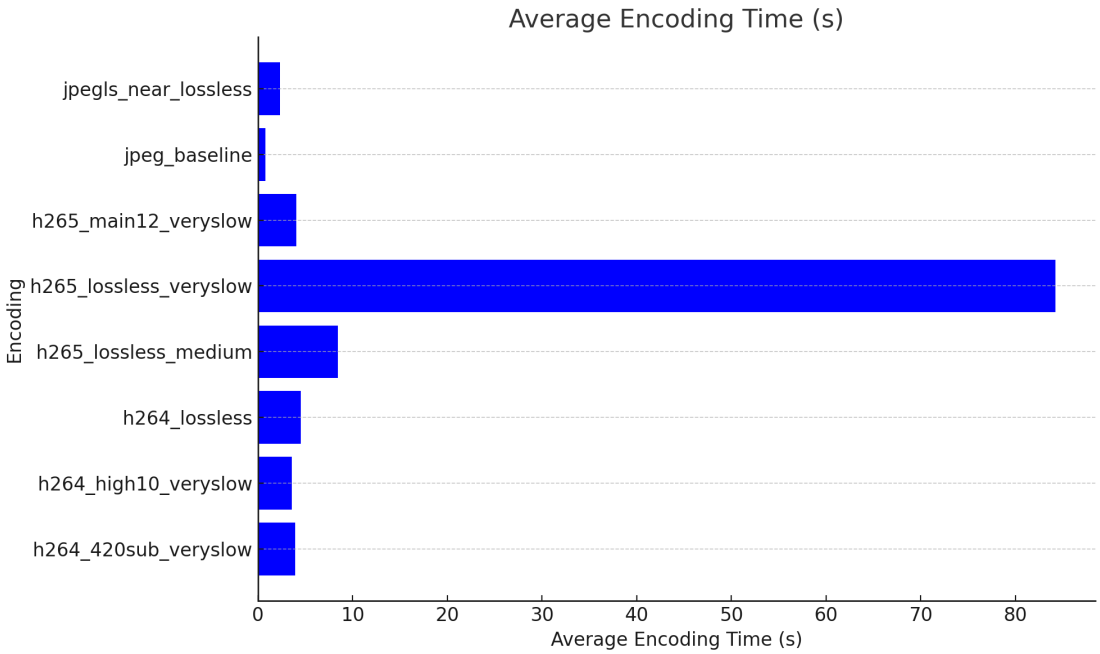

Figure 5.5: Averaged Encoding Time for Head CT

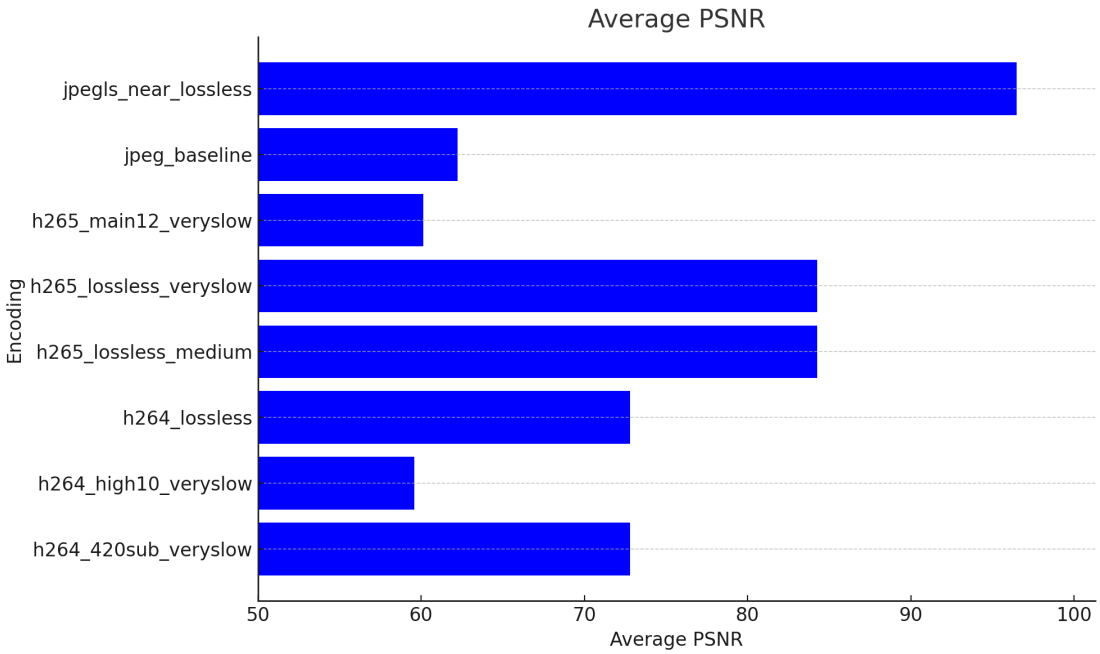

Figure 5.6: Averaged PSNR values for Breast Tomography scans

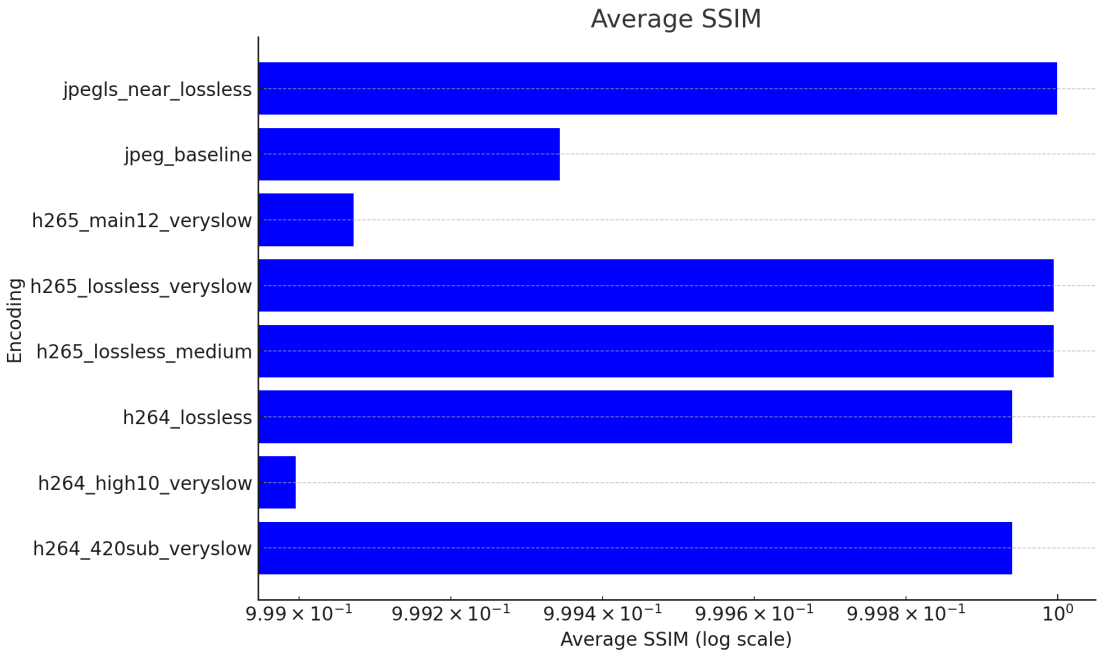

Figure 5.7: Averaged SSIM values for Breast Tomography scans

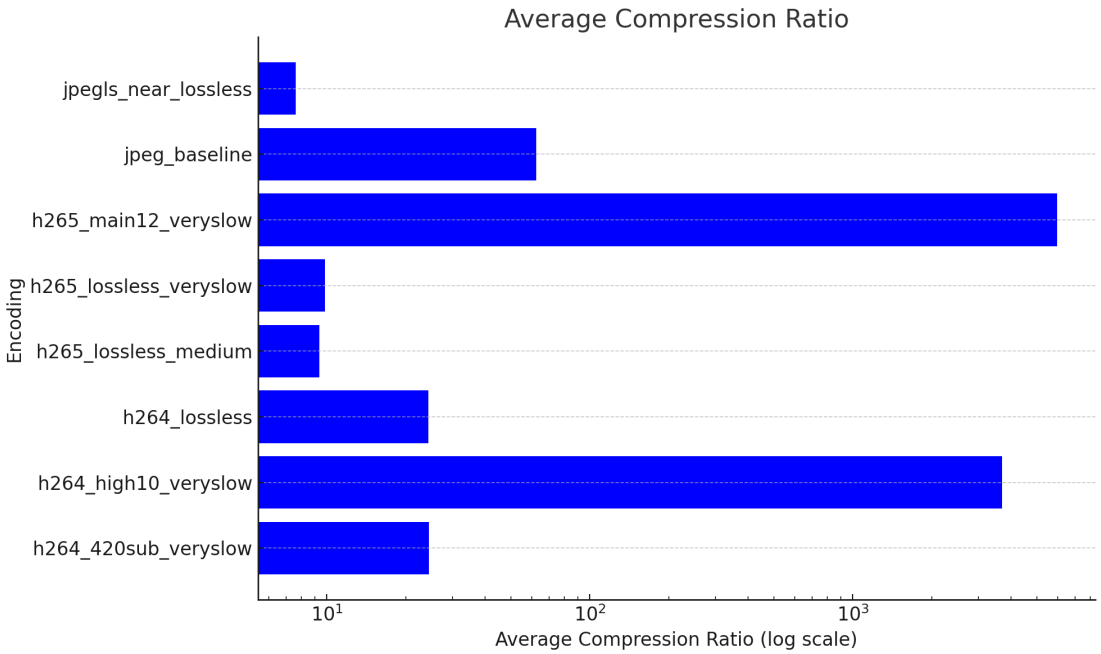

Figure 5.8: Averaged Compression Ratios for Breast Tomography scans

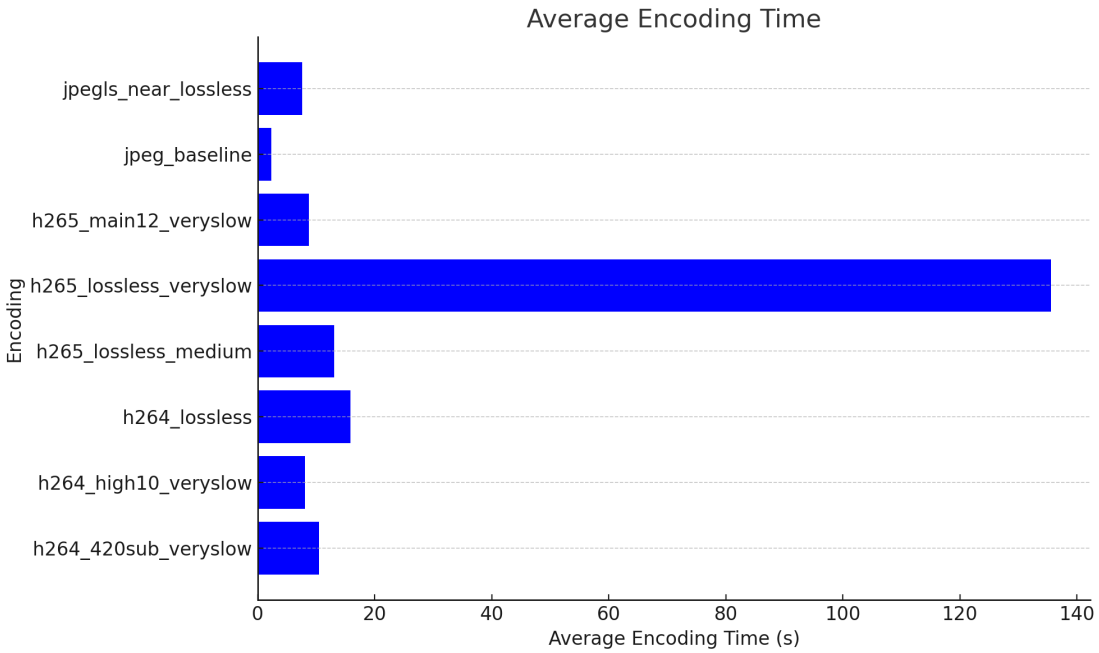

Figure 5.9: Averaged Encoding Time for Breast Tomography scans

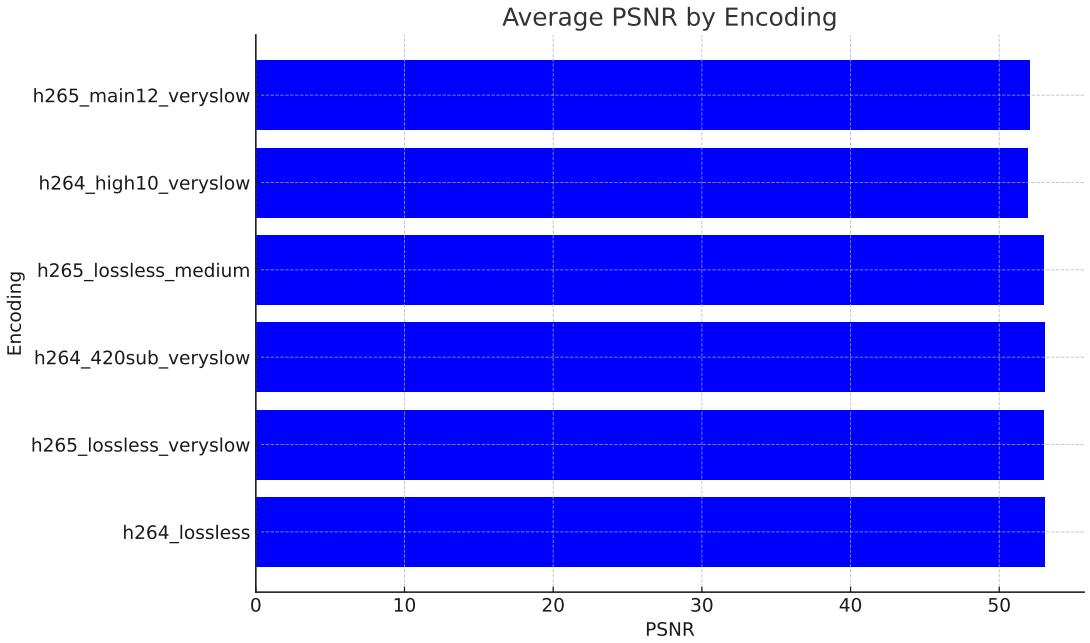

Figure 5.10: PSNR values of the Proposed Intertwining Method

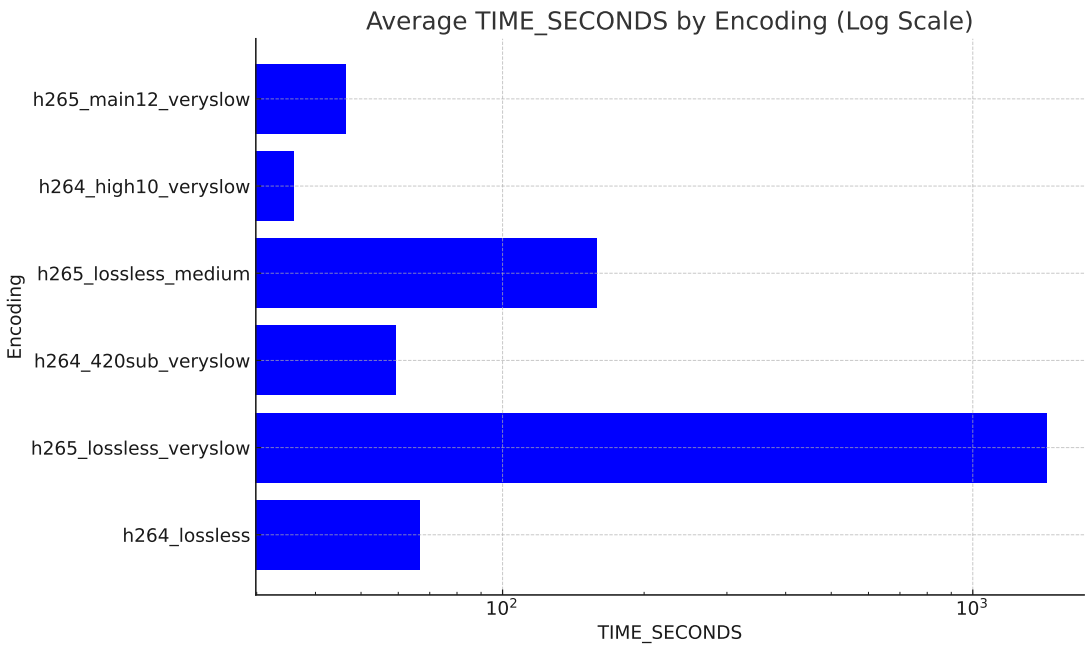

Figure 5.11: Encoding Time of the Proposed Intertwining Method

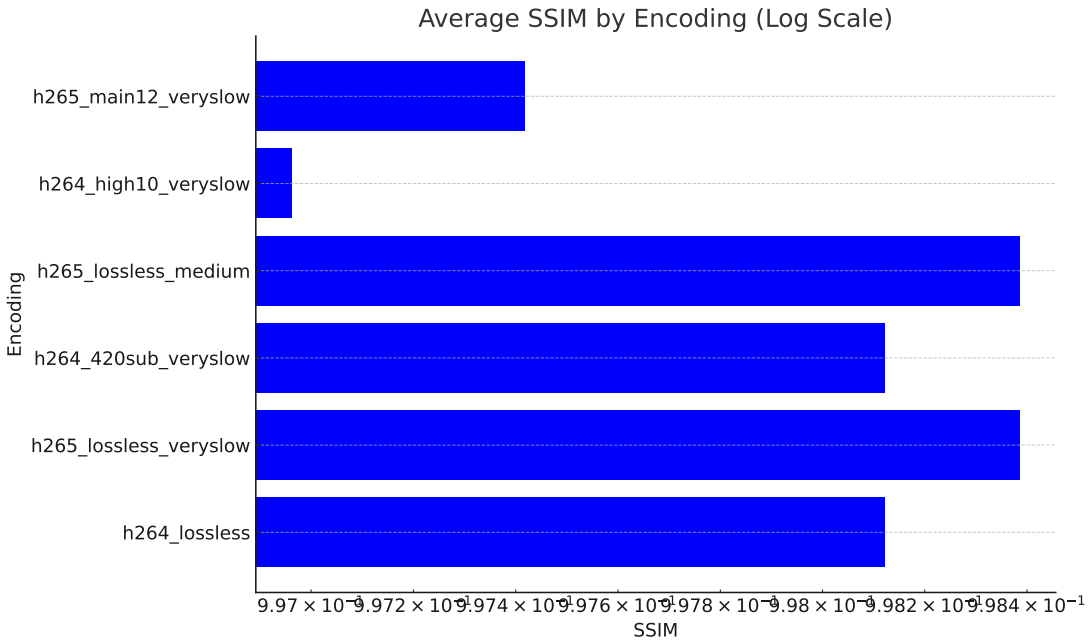

Figure 5.12: SSIM values of the Proposed Intertwining Method

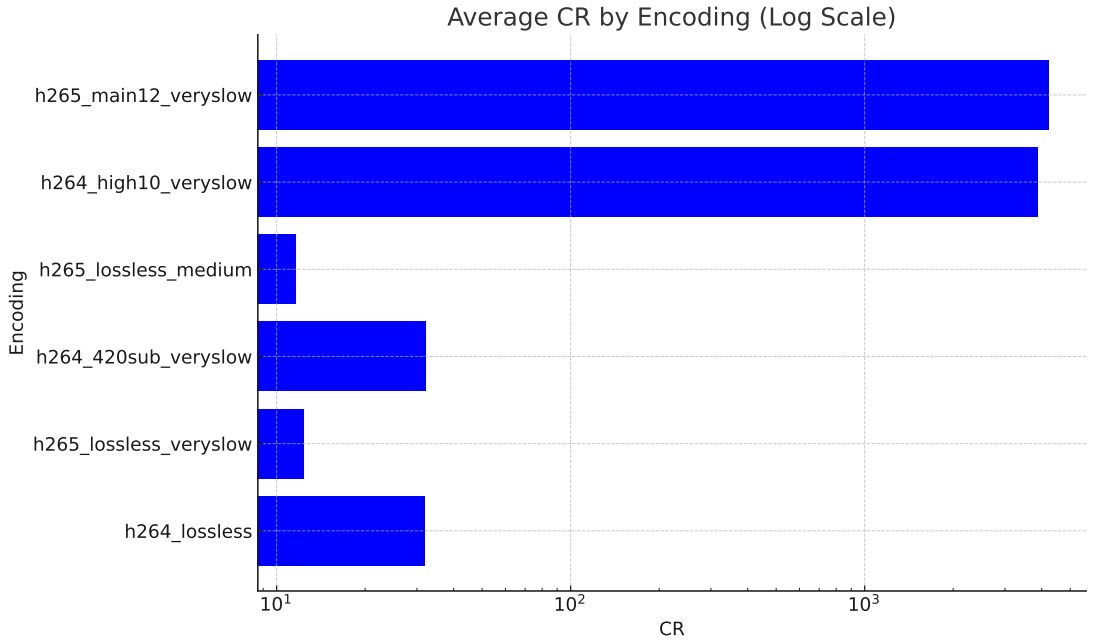

Figure 5.13: Compression Ratio of the Proposed Intertwining Method

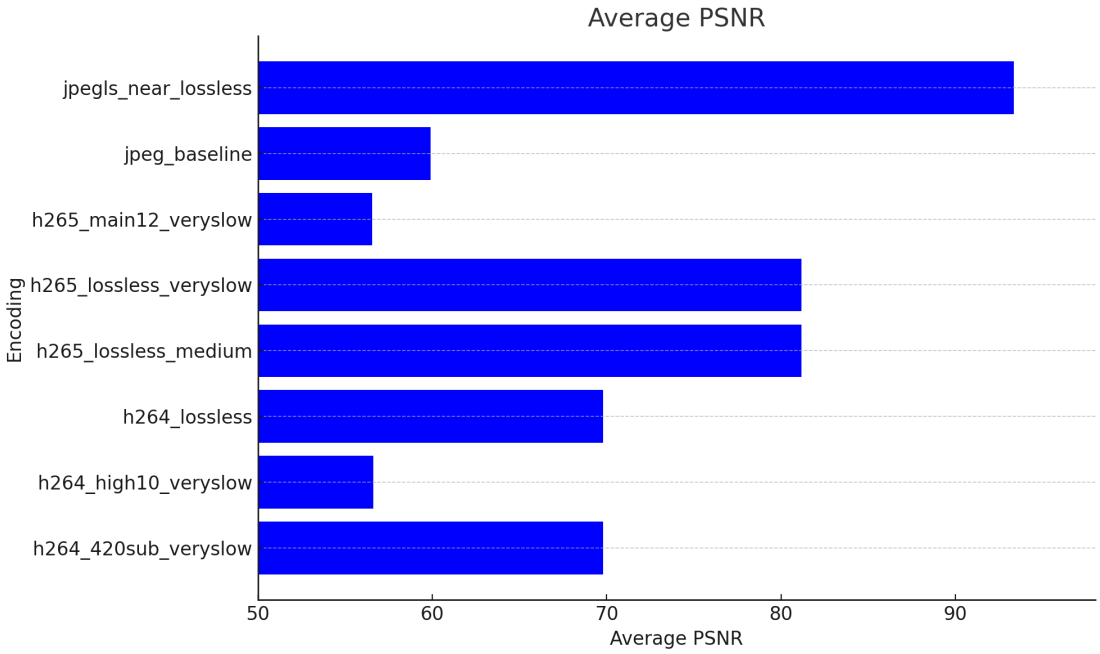

Figure 5.14: Averaged PSNR values for Human X-Ray scans

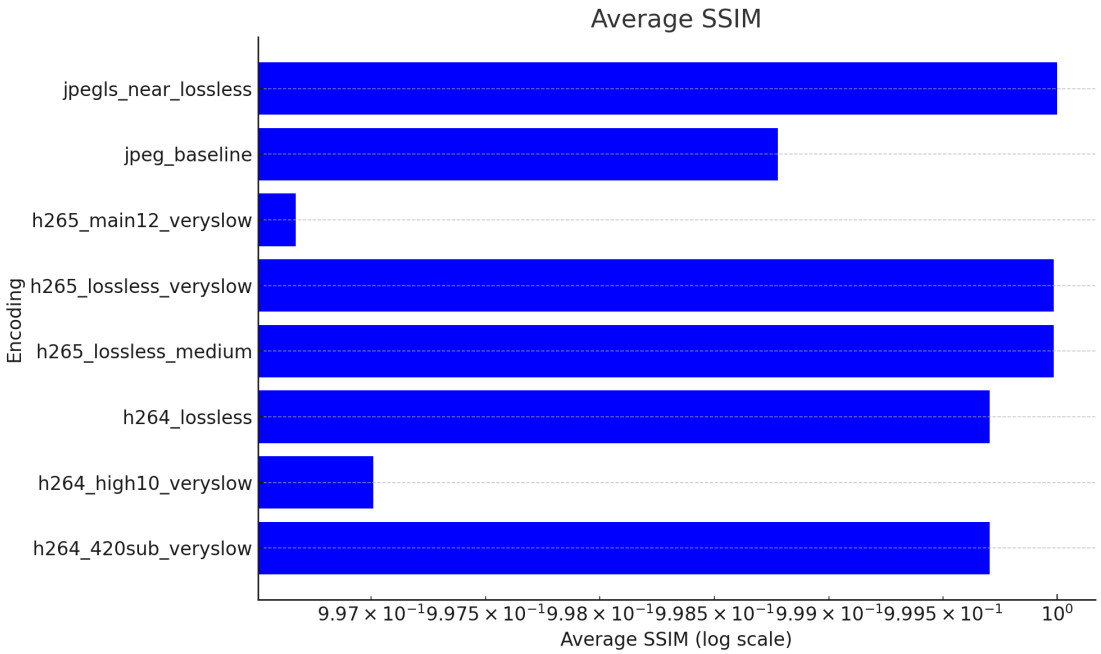

Figure 5.15: Averaged SSIM values for Human X-Ray scans

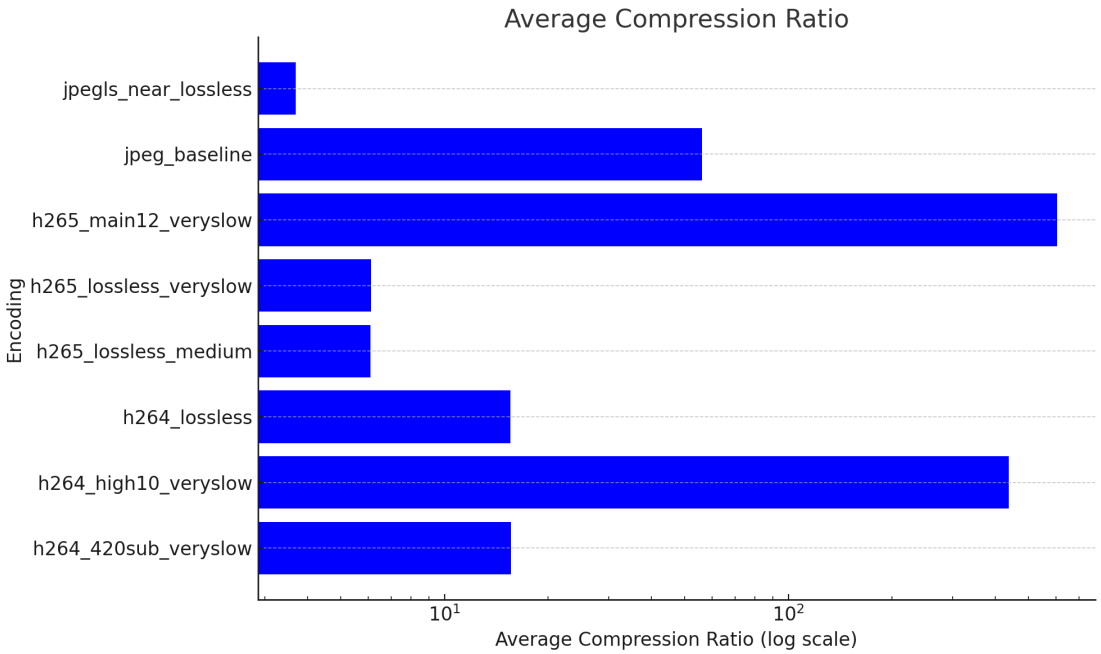

Figure 5.16: Averaged Compression Ratios for Human X-Ray scans

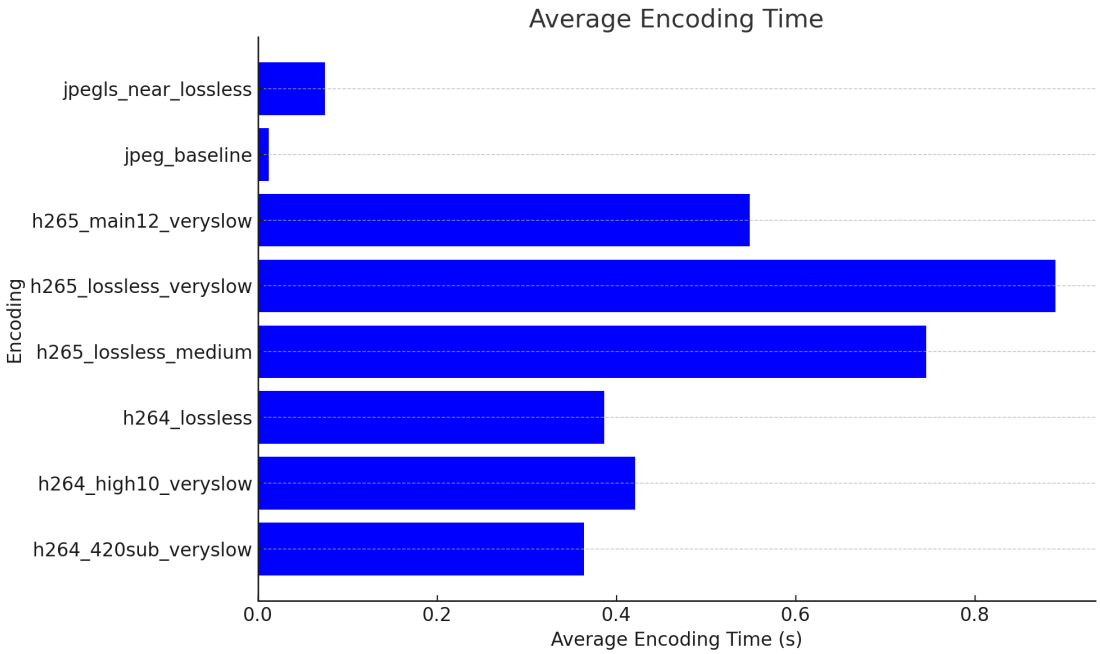

Figure 5.17: Averaged Encoding Time for Human X-Ray scans

5.3 Discussion

The results in Chapter 5 highlight the benefits of using advanced compression methods for volumetric CT scans. In Figures 5.2–5.5 show that modern encoders (e.g., H. 264/H. 265) can achieve a significantly higher compression ratio without seriously compromising image fidelity, as indicated by the favorable PSNR and SSIM values.

For the *intertwined* dataset (Figures 5.10–5.13), we combined 21 different CT scans into one continuous stream. This approach further capitalizes on slice-to-slice similarities across multiple patients. Figure 5.10 confirms that the overall PSNR stays consistently high, while the compression ratio (Figure 5.13) is higher than simpler, single-volume encodings. However, the encoding time (Figure 5.11) can be longer, so this method might be best suited for archival or offline workloads.

In contrast, non-volumetric images such as single-frame X-rays or pathology slides 5.16 show little benefit from "video-like" strategies. Their compression largely relies on standard 2D methods like JPEG 2000 or JPEG-LS, which still maintain strong diagnostic fidelity in a smaller file size. Overall, these findings emphasize the importance of matching the compression strategy to the imaging modality and clinical use, striking the right balance between file size savings and diagnostic integrity.

6 Conclusion

This thesis explored the landscape of 3D voxel texture compression in medical imaging, with a focus on formats compatible with the DICOM standard. Through a comparative analysis of still-image codecs 2.3, video-based methods 2.5.1, and a novel intertwined archival strategy 4, we demonstrated the potential of modern encoders to significantly reduce storage requirements while preserving diagnostic quality.

Empirical results 5 showed that inter-frame redundancy can be effectively exploited in volumetric datasets like CT scans, especially when similar frames from multiple patients are grouped and encoded collectively. Our proposed method yielded favorable compression ratios 5.5 and high structural similarity scores 5.3, though with increased encoding times suggesting its best fit in archival scenarios rather than real-time applications.

For non-volumetric data , traditional 2D approaches like JPEG 2000 and JPEG-LS remain sufficient and more efficient.5.16 The results reinforce the importance of aligning compression strategy with imaging modality and clinical use case.

Future work should evaluate clinical usability through observer studies and work toward standardizing acceptable levels of compression for regulatory approval.

References

[1] Trupti N. Baraskar and Vijay R. Mankar. The dicom image compression and patient data integration using run length and huffman encoder. In Sudhakar Radhakrishnan and Muhammad Sarfraz, editors, *Coding Theory*, chapter 12. IntechOpen, Rijeka, 2019.

[2] David Clunie. Lossless compression of grayscale medical images - effectiveness of traditional and state of the art approaches. volume 3980, 02 2000.

[3] F. F. Cunha, A. M. Fernandes, and N. M. Rosa. Lossy image compression in a preclinical multimodal imaging study. *Journal of Digital Imaging*, 36(5):2331–2344, 2023.

[4] DICOM. Ps3.10 2025a - media storage and file format for media interchange, 2025.

[5] DICOM. Ps3.14 2025a - grayscale standard display function, 2025.

[6] DICOM. Ps3.5 2025a - data structures and encoding, 2025.

[7] FADGI. Digital file formats for videotape reformatting, 2011. `https://www.digitizationguidelines.gov/guidelines/video_reformatting_compare.html` [Accessed: (24/03/2025)].

[8] U.S. Food and Drug Administration. Mammography quality standards act (mqsa), 2023.

[9] Gnash. Utah tea pot, 2008. `https://commons.wikimedia.org/wiki/File:Utah_teapot_simple_2.png` [Accessed: (24/03/2025)].

[10] David A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, 1952.

[11] F. Liu, M. Hernandez-Cabronero, V. Sanchez, M. W. Marcellin, and A. Bilgin. The current role of image compression standards in medical imaging. *Information*, 8(4):131, Dec 2017. Epub 2017 Oct 19. PMID: 34671488; PMCID: PMC8525863.

[12] R. Loose, M. Wucherer, G. Weisser, and A. Uhl. Compression of digital images in radiology: results of a consensus conference. *Fortschritte auf dem Gebiet der Rontgenstrahlen und der bildgebenden Verfahren*, 181(1):32–37, 2009.

[13] Inc medDream. The most common dicom features in a medical study, 2025. `https://www.dicomlibrary.com/dicom/study-structure/` [Accessed: (24/03/2025)].

[14] S. G. Miaou, F. S. Ke, and S. C. Chen. A lossless compression method for medical image sequences using jpeg-ls and interframe coding, 2009.

[15] new.limit. Example of interframe prediction process, 2008. `https://commons.wikimedia.org/wiki/File:Interframe_prediction.png`.

[16] Canadian Association of Radiologists. Guidelines for irreversible compression in diagnostic imaging. 2020.

[17] Royal College of Radiologists. Pacs and ris: Guidance on compression, 2011. `{https://www.rcr.ac.uk/publication/guidance-compression-medical-images}`.

[18] European Society of Radiology. Usability of irreversible image compression in radiological imaging. *Insights into Imaging*, 2(2):103–115, 2011.

[19] S. S. Parikh, Z. Yaniv, and B. Lowekamp. High bit-depth medical image compression with hevc. *IEEE Journal of Biomedical and Health Informatics*, 22(2):552–560, 2018.

[20] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[21] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1649–1668, 2012.

[22] I. A. Urbaniak. Using compressed jpeg and jpeg2000 medical images in deep learning: A review. *Applied Sciences*, 14(22):10524, 2024.

[23] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, 1977.