The generation of MR-based synthetic CT using deep learning for brain radiotherapy

"The only way of discovering the limits of the possible is to venture a little way past them into the impossible" - Arthur C. Clark

Martina Mangion

Supervised by Ms. Chantelle Said

Co-supervised by Dr. Dylan Seychell

Department of Medical Physics

Faculty of Health Sciences

University of Malta

September, 2024

A dissertation submitted in partial fulfilment of the requirements for the degree of M.Sc. in Medical Physics.

L-Università ta' Malta

FACULTY/INSTITUTE/CENTRE/SCHOOL Faculty of Health Sciences, University of Malta

DECLARATIONS BY UNDERGRADUATE STUDENTS

(a) Authenticity of Long Essay/Dissertation

I hereby declare that I am the legitimate author of this Long Essay/Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

(b) Research Code of Practice and Ethics Review Procedures

I declare that I have abided by the University's Research Ethics Review Procedures. Research Ethics & Data Protection form code FHS-2024-00093.

02.2023

WWW.UM.EDU.MT

First edition, November 4, 2024

To all my academic lecturers throughout the years,

whose guidance and knowledge have shaped my journey.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Ms. Chantelle Said, Co-Supervisor Dr. Dylan Seychell and the head of the department Prof. Carmel J. Caruana for their feedback, guidance, and unwavering support throughout my dissertation research.

I would also like to express my heartfelt appreciation to my family, whose love and encouragement have been my rock during moments of doubt and uncertainty. Their support and belief in me have been a driving force in my pursuit of academic excellence.

Lastly, I would like to express my heartfelt gratitude to Isaac for his unwavering support, encouragement and understanding throughout this journey. His patience and constant belief in me have been invaluable.

I am grateful for the opportunity to undertake this research, and I wish to express my gratitude to all those who have contributed to this work.

Abstract

This dissertation evaluates a conditional conditional Generative Adversarial Network (cGAN) algorithm for generating Synthetic Computed Tomography (sCT) images from Magnetic Resonance Imaging (MRI) data, aimed at enhancing Radiotherapy (RT) planning for brain cancer patients by testing a Maltese cohort of 16 patients. This research addresses the need for streamlined alternatives to dual-simulation processes involving both Computed Tomography (CT) and MRI, with the potential to reduce patient fatigue and improve treatment precision.

The study's primary objectives are to assess the algorithm's accuracy in generating structurally and dosimetrically accurate sCT images, examine its effectiveness in replicating dose distributions, and mitigate registration errors that arise from dual-modality simulation. This focus on a minority population, often underrepresented in existing literature, highlights the unique application of the algorithm in a Maltese context.

An experimental methodology was employed, utilising quantitative evaluation metrics to test the algorithm on a local cohort. Image quality was assessed using Mean Absolute Error (MAE), Peak Signal-To-Noise Ratio (PSNR), and Structural Similarity Index Metric (SSIM), while dose accuracy was measured through Dose-Volume Histogram (DVH) and Gamma Pass Rate (GPR) metrics.

Results indicate that the algorithm performs well in generating structurally accurate sCT images, particularly in replicating anatomical features. However, discrepancies in dose metrics were observed, underscoring the need for refinement in handling complex anatomical structures, such as the optic nerves and orbits. This research contributes to the advancement of MRI-based sCT generation, with implications for reducing registration errors and enhancing patient-specific RT planning. Future studies should focus on improving dose distribution accuracy, expanding the minority dataset and optimising the algorithm for broader clinical application.

Keywords: *sCT, MRI, Radiotherapy, Brain Cancer, Dose Metrics, Maltese Cohort*

Contents

List of Figures

List of Tables

List of Abbreviations

1

Introduction to the Study

1.1 | Introduction

This chapter presents the problem statement, background and context, objectives, scope, research methodology summary, ethical considerations and the relevance of the study.

1.2 | Problem Statement

Navigating the complexities of treatment planning, (RT) patients are subjected to a dual-simulation process involving both MRI and CT scans. The accuracy of treatment planning and dose delivery depend on CT-derived electron density data and MRI's superior delineation of soft tissue structures. The integration of both MRI and CT scans into treatment planning has traditionally been the norm to benefit from their respective strengths.

The reliance on CT scans, while justified in RT, introduces several drawbacks. These include exposure to additional radiation and logistical complications in acquiring images from different modalities, possibly across various days. The most critical issue stemming from this dual-dependence is the presence of registration errors. Co-registering and fusing MRI and CT images pose complications in RT planning due to these errors, with anatomical changes between scans further compounding these difficulties.

The limitations of a conventional dual-simulation process and the need for CT scans can be addressed through the generation of an MR-based sCT.

1.3 | Background and Context

Medical imaging plays a fundamental role in the discipline of RT, facilitating precise treatment planning and delivery. The two mentioned imaging modalities, MRI and CT, are imperative tools in RT, especially in the context of treatment planning for brain cancer patients. These technologies enable clinicians to visualise tumours with precision, significantly impacting treatment outcomes.

MRI has been described as "the most important development in medical diagnosis since the discovery of the x-ray" (Westbrook and Talbot, 2018) over a century ago. The MRI modality relies on the interaction of magnetic fields and radio frequency pulses with the body's hydrogen atoms. This interaction produces high-quality, cross-sectional images with exceptional soft-tissue contrast. CT, on the other hand, utilises X-rays to generate detailed anatomical images, offering invaluable insights into bone structures. While CT's electron density data is pivotal for accurate dose calculations in RT, its ability to delineate soft tissue structures in the brain is limited compared to MRI.

The advent of Artificial Intelligence (AI) has ushered in a new era in Medical Physics. In RT, AI's potential to automate and refine treatment planning, by predicting electron density from MRI scans, marks a significant leap forward. This innovation paves the way for more efficient treatment approach, minimising reliance on multiple imaging modalities.

sCT represents a groundbreaking development in medical imaging, where MR images are algorithmically transformed into CT-replicated images. This technique leverages the soft-tissue clarity of MRI and the radiodensity information of CT scans, offering a composite sCT image. The introduction of sCT not only reduces patient exposure to radiation but also streamlines the treatment planning process, while eliminating the challenges of co-registration errors between MRI and CT scans.

The quest for MR-based sCT generation has evolved significantly, driven by the need for more efficient and accurate RT planning methods. Building upon years of research and technological advancements, this study aims to adopt and fine-tune an MR-based sCT generation technique for brain cancer RT using transfer learning.

1.4 | Objectives of the Study

The objectives of the study include:

1. To evaluate the performance of a cGAN algorithm for generating sCT images from MRI as the source modality, in the context of RT treatment planning for brain cancer.

2. To evaluate the quality of sCT generation by conducting tests on a local cohort of Maltese patients diagnosed with brain cancer, a population that represents a minority group.

3. To address registration errors that arise from the dual-simulation processes due to position and anatomical changes of CT and MRI scans, with the generation of synthetic CTs from an MR modality.

4. To assess and quantify the differences between the generated synthetic and original CT images whilst assessing the feasibility of treatment planning from MR-acquired sCT for patients affected by brain tumours.

1.5 | Scope of the Study

This research specifically concentrates on harnessing deep learning techniques to innovate the generation of sCT images from MR scans, with an emphasis on their application in brain RT scenarios.

The study extends its investigation to a local context by employing scans obtained from Maltese patients as the test dataset. This approach not only allows for the validation of the developed models against real-world clinical scenarios but also ensures that the insights and findings are directly applicable to improving patient care within Malta's healthcare system. The geographical focus on Maltese patients aims to provide specific insights into the local healthcare context, addressing and revealing unique regional challenges and opportunities.

The study acknowledges the constraints posed by the available technology and dataset size, yet it seeks to push the boundaries of current medical imaging capabilities. By delineating these boundaries, the study ensures a focused and coherent investigation, while setting a foundation for future research.

1.6 | Ethical Consideration

Obtaining datasets from hospitals, particularly those containing sensitive medical imaging data like MRI and CT scans, requires strict adherence to ethical considerations and protocols. These measures ensure patient privacy, confidentiality, and compliance with relevant regulations.

The dataset used in this study was acquired from deceased patients, and as such, patient consent was not required. This approach fit best with the timeline of the study and adhered to ethical guidelines regarding the use of medical data from deceased individuals. Even though patient consent was not needed, strict ethical protocols were still followed to ensure the responsible use of sensitive medical imaging data, such as MRI and CT scans.

The study also acknowledged the potential risks associated with handling sensitive medical data, even when derived from deceased individuals. Protocols were established to mitigate these risks, including responsible long-term data usage and disposal policies, ensuring that data is handled properly throughout and beyond the duration of the research.

Following ethical approvals, the patient data must be de-identified to safeguard privacy. This process involves an intermediary, operating within the oncology department, who removes or encrypts personal identifiers such as names and contact information to prevent patient identification.

The study acknowledges potential risks to participants and has established protocols to mitigate these risks. Long-term data usage and disposal policies are in place, ensuring data is handled responsibly beyond the duration of this research.

1.7 | Conclusion

Chapter one presented a brief overview of the study. A review of the literature is given in chapter two. The research methodology is covered in Chapter three while the results are covered in Chapter four. The most significant findings of the study are highlighted in Chapter five and finally, the conclusions reached and proposals for additional research are compiled in Chapter six.

2

Literature Review

2.1 | Introduction

Chapter Two presents a critical review of the literature, elucidating the current state of knowledge on the subject and identifying gaps in the existing literature.

This review was initiated using the following electronic research databases: Google Scholar, and HYDI, focusing on works published in the last ten years (2014-2024). This time frame was selected to ensure the review captured the most recent advancements and trends in the field. The search utilized a series of keywords: "Magnetic Resonance Imaging," "Computed Tomography," "Synthetic Computed Tomography Generation," "Deep Learning," and "Brain RT." These keywords were chosen for their direct relevance to the core research areas and objectives of this study, aiming to cover a spectrum of research that intersects at the development and application of innovative imaging and analytical techniques in the treatment of brain cancer.

Selected articles were subject to the following inclusion criteria to be considered for this review: (1) the availability of full-text papers, (2) written in the English language, and (3) studies directly relevant to the scope of this research. Additionally, preference was given to primary research articles over (1) books and case study reports, and (2) papers expressing subjective expert opinions. This selective approach was employed to prioritise empirical evidence and high-quality research findings that could provide robust insights into the development, challenges, and potential of deep learning in the generation of synthetic CT from MRI scans for the purpose of enhancing brain RT planning.

2.2 | Medical Imaging in Brain RT

2.2.1 | The Rationale Behind RT

RT is a century-old medical practice, first recorded in 1899, that uses radiation to treat cancer (Abeloff et al., 2008). At its core, RT leverages the differential radiation sensitivity between cancerous and normal tissues. This therapeutic window enables Medical Phyicist (MP)s to devise optimised treatment plans that maximise tumour control probability while minimising harm to healthy tissue and sensitive organs. Since its inception, RT has witnessed exponential growth, from technological advancement to better understanding of its principles.

Historically, RT was associated with the collateral damage of healthy tissue in its attempt to target cancer cells. However, "technological advancements have allowed for precise and accurate therapy, providing a viable option as a cure for individuals diagnosed with cancer" (Abshire and Lang, 2018).

The introduction of 3D conformal RT marked a significant leap forward, allowing clinicians to shape radiation beams to match the tumour's geometry accurately. This innovation provided greater control over the spatial distribution of radiation, leading to improved targeting of tumours. Subsequent innovations, such as Intensity-Modulated Radiotherapy (IMRT) and Image-Guided Radiotherapy (IGRT), have refined this approach, allowing for more aggressive targeting of tumours while sparing surrounding healthy tissues (Do Huh and Kim, 2020).

Technological advancements have significantly enhanced the efficacy of brain radiotherapy. Techniques such as stereotactic radiosurgery (SRS) and stereotactic radiotherapy (SRT) allow for the delivery of high-dose radiation to precise target areas. These methods are particularly beneficial for treating small to medium-sized brain tumours, including those in inoperable locations. Each case is patient-specific and each treatment plan is personalised by keeping all these characteristics in mind.

One advancement that is currently heavily studied is the integration of AI into healthcare. AI, particularly deep learning, has shown significant promise in various medical applications, including image analysis and treatment planning. In RT, AI algorithms can enhance image segmentation, improve the accuracy of sCT generation from MRI and overall optimise treatment plans. This integration has the potential to reduce treatment planning time, increase precision, and ultimately improve patient outcomes. Ongoing studies are focused on validating AI-driven approaches in clinical settings and addressing challenges such as data quality, algorithm transparency, and clinical integration.

With a population of 500,000, Malta records nearly 2,000 new cancer cases annually, with a 5% yearly increase (International Atomic Energy Agency, 2024). Cancer-related deaths now account for over 30% of all mortalities in the Maltese islands (Ministry for Health, Malta, 2023). The healthcare system faces unique challenges in managing cancer care, necessitating innovative strategies to optimise treatment. The Ministry for Health, Malta (2023) emphasises the country's commitment to high-quality cancer treatment, exemplified by recent advancements in RT technologies, including the MR-Linac (TVM News, 2024).

2.2.2 | The Rationale behind the Brain

The brain was chosen as the focal organ for this study due to a number compelling reasons. Firstly, brain tumours present unique challenges in RT that make them an ideal subject for investigating advanced imaging approaches and co-registration techniques. The intricate anatomy and critical functions of the brain necessitate extremely precise targeting.

Secondly, The brain is a relatively stationary organ enclosed within the skull, which provides stability and protection. Unlike other organs in the body, the brain is not prone to significant movement or changes in size between scans (Quanyue Xu, 2023). This consistency ensures minimal variability between multi-modal images, allowing deep learning models to more accurately train and learn the characteristics of brain tumours, further enhancing the reliability of co-registration processes and improves the precision of treatment planning.

Several anatomical regions have been explored for MR-based sCT generation, extending the research beyond brain cancer radiotherapy. The pelvis has been a focal point, particularly in prostate and gynecological cancer treatment as outline by Autret et al. (2023). Additionally, abdominal and liver regions evaluated by Fu et al. (2020), demonstrated promising dosimetric accuracy while assessing two deep learning models. Research into breast cancer treatment has also highlighted the potential of sCT generation for MRI-guided RT, where accurate differentiation between tissues is required (Jeon et al., 2019). These studies underscore the growing interest in applying sCT generation across diverse anatomical regions, though further research is needed to optimise its clinical implementation in each area.

Focusing on the brain in this study contributes valuable insights to the field of neuro-oncology. By addressing the challenges of brain tumour RT, the study aims to enhance the understanding and application of advanced imaging techniques, ultimately improving patient care and outcomes in neuro-oncology.

2.2.3 | A Spectrum of Treatable Cases

Modern-day RT stands capable of treating a wide array of cancers across different stages and complexities. Its applicability ranges from localised, early-stage tumours to advanced cancers, offering both curative and palliative benefits. This spectrum is broadened by RT's effectiveness in treating both benign and malignant tumours, each with unique considerations and objectives in treatment.

Benign tumours, while not cancerous, can be life-threatening if their location or size affects vital functions, this is of particular importance in the brain. RT for benign tumours aims at controlling growth, often in cases where surgical removal poses too high a risk. Malignant tumours, characterised by uncontrolled growth and the potential to invade nearby tissues or metastasize, require a more aggressive approach.

Primary brain tumours originate within the brain and can be either benign or malignant. The treatment approach varies based on the tumour's type, size, location, and aggressiveness. Gliomas are the most common type of primary brain tumours, originating from glial cells. Glioblastoma multiforme is an aggressive form of glioma that often requires a combination of surgery, RT, and chemotherapy (Aldoghachi et al., 2022)). While surgery is often the primary treatment, RT is utilised for inoperable cases or when residual tumour tissue remains post-surgery (Yamanaka et al., 2017). Meningiomas are typically benign tumours arising from the meninges, the protective layers surrounding the brain and spinal cord. "Adjuvant radiotherapy treatment is reserved for newly diagnosed cases of grade II and grade III meningiomas in cases of recurrent disease or when surgery is not radical or feasible" (Caccese et al., 2023).

Each case is patient-specific, necessitating a personalised approach to treatment planning. This takes into account not just the physical characteristics of the tumour, but also the patient's overall health, potential risks and genetic markers that may influence the tumour's response to treatment.

2.2.4 | The Role of Medical Devices in RT

Medical devices are integral to the precision and efficacy of RT. The incorporation of advanced medical technologies has markedly improved clinicians' capacity to accurately image and delineate tumours. This section examines the various medical devices employed in RT and their contributions to this clinical area.

2.2.4.1 | Linear Accelerator (LINAC)

"Radiation therapy machines currently used in clinical practice are the result of evolution over a long period of time. Various areas of science, such as medical physics, mechanical engineering, and computer engineering, have contributed to the continual development of the technology"(Do Huh and Kim, 2020). RT machines, known as LINACs, are the foundation of contemporary RT. These devices produce high-energy X-rays or electrons that are precisely directed at tumor sites. A LINAC can deliver various types of radiation, including IMRT and Stereotactic Body Radiation Therapy (SBRT), which facilitate highly targeted treatments that minimise harm to surrounding healthy tissues (Akino et al., 2019).

2.2.4.2 | Imaging Devices

Advanced imaging devices are essential for accurate tumour localisation and treatment planning. Technologies such as CT and MRI provide detailed images of tumours and surrounding anatomy. These imaging modalities are integrated into the RT workflow to guide the precise delivery of radiation doses.

CT is recognised as the standard imaging modality for RT planning, as it offers a three-dimensional visualisation of the tumour and electron-density data, which is essential for dose calculations in radiotherapy planning (Beaton et al., 2019). Conventional CT imaging data underpins the approach of administering a relatively uniform radiation dose to the well-defined target volume of the tumour (AAPM Task Group 111, 2021).

MRI has become increasingly important in the field of RT due to its superior contrast resolution compared to conventional CT. This improved contrast resolution allows for more precise segmentation of the target area, leading to better treatment outcomes

(Chandarana et al., 2018).

Current clinical practice in RT incorporates both CT and MRI to leverage the individual strengths of each imaging modality. The combination of CT's quantitative capabilities and MRI's qualitative imaging allows for a comprehensive approach to radiotherapy planning.

2.2.4.3 | IGRT

One of the key advantages of IGRT is its ability to adapt to changes in the patient's anatomy or tumour size during the treatment course. It allows for the real-time visualisation of tumours during treatment, ensuring that radiation is accurately targeted.

Malta recently introduced the MR-LINAC, with acceptance testing, commissioning and validation completed in July 2024, marking a significant milestone in the country's dedication to RT care while also highlighting the need for continuous advancements in this clinical area. An MR-LINAC combines MRI with a LINAC to offer real-time imaging during the treatment session. The MR-LINAC is considered an IGRT technology due to its ability to provide continuous imaging and adapt treatment based on real-time changes in tumour position and anatomy (Elekta, 2024).

2.3 | Challenges in MRI and CT Image Analysis

The integration of MRI and CT imaging in RT provides a comprehensive view of the tumour and surrounding tissues, enhancing the precision of treatment plans. However, several challenges need to be addressed to realise these benefits fully. sCT generation from MRI data aims to provide consistent and accurate imaging data, eliminating registration errors and logistical issues, by using image synthesis (Huijben et al., 2024). This study aims to address the following challenges in MRI and CT image analysis.

2.3.1 | Anatomical Variations Over Time

One of the primary challenges in using both MRI and CT scans is the anatomical changes that can occur between different scanning sessions. Since scans are often taken on different days, variations in patient anatomy can arise due to changes in positioning, natural physiological processes, or the progression of the disease. The brain, however, is a relatively stationary organ, making it less susceptible to such variations. Nevertheless, even slight differences in head positioning or changes in brain morphology over time in between scans can introduce challenges in aligning images from different modalities.

2.3.2 | Co-Registration

Co-registration is the process of aligning multi-modal images to offer a more detailed view of both the anatomical and functional features of the tumour and surrounding tissues (Velesaca et al., 2024). In RT, this pre-processing step reduces the impact of patient movement and variations in positioning across scans. It plays an important role in merging data from CT and MRI, enabling the creation of a more precise and comprehensive treatment plan.

Software algorithms are used to perform co-registration, aligning images based on common anatomical landmarks or fiducial markers (Velesaca et al., 2024). Despite its benefits, co-registration presents several challenges. This pre-processing step is not without error; inaccuracies can arise from patient movement, differences in patient positioning, and anatomical changes between scans (Alam et al., 2016). Moreover, additional pre-processing steps are often necessary to achieve a good registration, but these steps can introduce their own set of errors.

2.3.3 | Logistical Issues

The need for patients to undergo multiple imaging sessions, often carried out on different days, can complicate the process. Coordinating these sessions to minimise time between scans while ensuring the patient's comfort and safety adds to the complexity. Moreover, differences in equipment calibration and imaging protocols between MRI and CT machines can lead to inconsistencies in image quality and resolution.

2.3.4 | Time and Resource Constraints

The integration of MRI and CT imaging is resource-intensive, both in terms of time and technology (Alam et al., 2016). Preparing for and performing multiple imaging sessions require significant time from both patients and healthcare providers. Additionally, the need for specialised software and hardware to handle and process these images adds to the treatment planning pipeline. Developing a more effective workflow that optimises the use of time and resources without compromising the quality of the imaging data is essential for improving patient care.

2.3.5 | The Need for Extensive Datasets

To overcome the challenges of developing and validating advanced imaging algorithms, extensive datasets are required. These datasets must include a wide variety of patient anatomies, imaging modalities, and clinical scenarios to train and validate advanced algorithms effectively. Large and diverse datasets ensure that the developed solutions are robust, generalisable, and capable of handling the variability seen in clinical practice (Jeon et al., 2023).

However, the collection and utilisation of extensive datasets present several issues. The General Data Protection Regulation (GDPR) imposes strict regulations on the handling of personal data, this includes medical images. Ensuring compliance with these regulations can be challenging, as it requires anonymisation processes and secure data storage solutions to protect patient privacy (Diaz et al., 2021). The complexity of these processes is further compounded by the need to obtain informed consent from patients, which can be time-consuming and difficult to manage (Dankar et al., 2019).

"A major challenge is building a refined dataset with large comprehensive clinical and imaging data"(Jeon et al., 2023). For this reason, accessibility to large and diverse image datasets is often limited. Many medical institutions have substantial datasets, but sharing these data outside their organisation can be difficult due to privacy concerns, lack of standardisation, and logistical barriers (Khalid et al., 2023). This limited accessibility hinders the development and validation of advanced imaging algorithms, as researchers may struggle to obtain necessary data to train and test their models adequately (Thakur and Thakur, 2024).

To address these challenges, several initiatives have emerged to provide open-source datasets for the research community. Notable examples include the Brain Tumor Segmentation (BraTS) challenge, which focuses on the segmentation of brain tumours using MRI scans, and the SynthRAD challenge, which provides synthetic radiology images for algorithm development.

The ethical and logistical challenges associated with dataset collection and sharing continue to require careful consideration and innovative solutions. As the field progresses, it is imperative to explore new methods for balancing data accessibility with patient privacy, potentially through advancements in secure data sharing technologies and more streamlined consent processes. Researchers and clinicians must work together to develop frameworks that facilitate the secure and ethical sharing of medical data (Khalifa and Albadawy, 2024)

2.4 | The application of Deep Learning in Medical Physics

Deep learning, a branch of machine learning, has impacted various domains in healthcare, particularly, medical physics (Shen et al., 2020). Unlike traditional machine learning, which relies on handcrafted features, deep learning models "learn multilevel representations directly from raw input data" (Cui et al., 2020), making them highly effective for complex data-driven tasks. By leveraging large datasets and advanced algorithms, these models have brought significant advancements in medical imaging, diagnostic procedures, and treatment planning.

Cui et al. (2020) identifies medical physicists as the ideal professionals to address pressing challenges in modern radiation oncology and cutting-edge tools powered by deep learning. These models are particularly suited to address challenges related to medical imaging, where the precise analysis of data is of utmost importance. As a result, the integration of deep learning into medical physics is driving innovations that not only improve the quality and speed of medical procedures but also enhance their overall accuracy.

2.4.1 | Computer Vision

Before the conceptualisation of machine learning, computer vision had already made significant strides in medical physics. Traditional computer vision techniques, which relied on manually designed algorithms and feature extraction, played an important role in the early stages of medical image processing, diagnostics and treatment planning (Ker et al., 2017).

One common technique was edge detection, where methods like the Canny and Sobel filters were used to identify boundaries within medical images. These early approaches were particularly useful in segmenting anatomical structures, such as organs or tumours, from surrounding tissues (Mahdi et al., 2022). Though limited by the need for manual tuning and the complexity of medical images, these techniques laid the groundwork for more sophisticated automated systems (Barragán-Montero et al., 2021).

Despite the success of these early techniques, they were often limited by their dependency on predefined rules and hand-crafted feature extraction techniques (Ker et al., 2017). The complexity and variability of medical images meant that such rule-based systems struggled to generalise across different patients or imaging conditions. Additionally, these methods required significant human expertise to fine-tune parameters and were sensitive to noise and artifacts (Shen et al., 2020).

The limitations of classical computer vision methods spurred the shift towards integrating machine learning, where algorithms could learn features automatically from data. Early machine learning models, such as decision trees and support vector machines, to be applied alongside classical computer vision techniques to improve performance (Zapata-Cortes et al., 2024). This transition marked the beginning of a new era of AI-driven medical imaging, leading to more robust and generalisable solutions in the field.

2.4.2 | The Data Required for ML to work

The performance and reliability of ML algorithms depend heavily on the quality, quantity and diversity of the data used for training and validation. High-quality data ensures that the model learns accurate and relevant features, while larger datasets help improve generalisation and reduce overfitting (Goodfellow, 2016).

Developing image-to-image translation models typically requires a training dataset containing multiple pairs of registered images, such as MR and CT images for each patient. This voxel-level alignment ensures that the models learn corresponding anatomical features across the two imaging modalities.

As discussed in section 2.3 the process of obtaining paired data can be challenging due to logistical, ethical and technical constraints. Given these limitations, there has been a growing interest in developing models that can be trained with unpaired data. In this approach, a model is trained with separate sets of MR and CT images from different patients (Zhu et al., 2017). While unpaired data methods allow for easier data acquisition and larger training datasets, they present additional challenges. Without direct voxel-to-voxel correspondence between images, the model must infer relationships between MR and CT images, which increases the complexity of the task. As a result, performance may suffer when compared to models trained on paired data due to the domain gap between the two modalities (Jin et al., 2019).

In response to the scarcity of high-quality paired data, researchers are increasingly exploring techniques such as synthetic data generation, domain adaptation, and semi-supervised learning. These methods aim to reduce the reliance on large paired datasets by either generating synthetic pairs that approximate real-world data or by leveraging unlabelled data in conjunction with a smaller labeled set. As these techniques evolve, integrating them with traditional machine learning approaches holds promise for improving machine learning outcomes, even in data-scarce environments.

Moreover, in image-to-image translation tasks, the choice between grayscale and RGB images impacts complexity and performance of the algorithm. Grayscale images, used in CT and MRI scans, consist of a single channel, making them computationally simpler and suitable for tasks focused on structural details. RGB images introduce more information through three colour channels (Van Den Oord et al., 2016). These are mainly beneficial in tasks involving colour and texture.

2.4.3 | Implementation of Deep learning in the medical field

Deep learning has become a transformative tool in healthcare, offering the ability to automatically learn complex patterns from large datasets. At the core of deep learning are neural networks—computational models that mimic the structure of the human brain, with neurons, or nodes, arranged in interconnected layers. These networks learn to recognise patterns in data by adjusting the strength of the connections between neurons, improving their ability to make predictions or classifications over time.

As Lee et al. (2017) notes, the increasing availability of healthcare data has led to a growing interest in applying deep learning to handle these vast datasets efficiently. This technology has found widespread application in medical imaging, diagnostics and treatment planning. By leveraging vast amounts of data, deep learning models can assist in automating tasks of detection, segmentation and classification in medical processes (Jin et al., 2021).

Among the deep learning architectures driving these advancements is the CNN. This architecture is described as "the most successful type of models for image analysis to date" (Litjens et al., 2017). CNNs are particularly well-suited for analysing medical images, as they excel at recognising spatial hierarchies in data through the use of convolutional layers. These layers apply filters to the input images, extracting features such as edges, textures and shapes. The strength of CNNs lies in their ability to automatically learn these features from raw data, eliminating the need for manual feature extraction, which was common in traditional machine learning approaches (LeCun et al., 1998).

Krizhevsky et al. (2012) introduced the famous AlexNet architecture, which changed the trajectory of CNNs for image classification tasks. It demonstrated how deep learning could achieve promising results in image analysis tasks. In subsequent years, further progress has been made using related but deeper architectures. As the network goes deeper, it learns more abstract and complex features (Deng et al., 2014).

As observed in figure 2.1, the architecture of the CNN model typically consists of several convolutional layers followed by pooling layers, which reduce the dimensionality of the data while retaining important features. The final layers are usually fully connected, allowing the network to make predictions based on the learned features. In the context of medical imaging, CNNs have been highly effective in tasks such as tumor detection, organ segmentation, and even the generation of synthetic images (Shen et al., 2020).

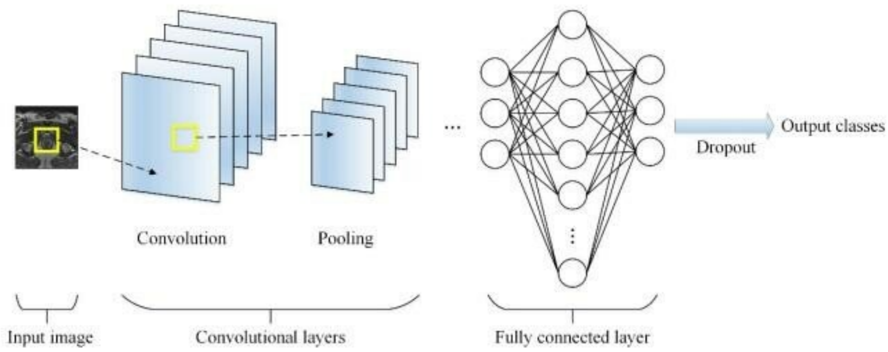


Figure 2.1: A typical CNN framework for image analysis, as illustrated by Dutta et al. (2020)

Another powerful deep learning model making significant strides in healthcare is the GAN. Initially introduced by Goodfellow et al. (2014), GANs are composed of two neural networks, a generator and a discriminator, that work in opposition to each other. The generator attempts to create realistic synthetic data, such as medical images, while the discriminator tries to distinguish between real and generated data. Over time, both networks improve, resulting in the generator producing increasingly realistic images. This ability to generate has applications beyond imaging, contributing to great data augmentation and training model capabilities.

A noteworthy extension of GANs is the cGAN, which were "introduced as a novel way to train a generative model" (Mirza, 2014). This model introduces a conditional input, which could be a class label or another data modality, such as an MRI scan used to generate an sCT image. cGANs have shown promise in improving image synthesis tasks by incorporating contextual information, allowing for more precise generation of synthetic images based on specific input data. This model can learn the relationship between paired datasets, such as MRI-to-CT mappings (Isola et al., 2017).

The success of cGANs in generating sCTs lies in their ability to produce images that closely resemble ground-truth CTs, outlined by Liu et al. (2022). This approach has been valuable in mitigating issues related to data scarcity while offering a pathway to reduce reliance on multiple imaging modalities, such as the dual-simulation processes required for MRI and CT registration.

The technical versatility of CNNs and GANs has led to their widespread adoption in healthcare. CNNs, with their layered architecture, are the backbone of many medical imaging tasks, excelling in classification, detection, and segmentation. GANs, on the other hand, are revolutionising data generation and augmentation, particularly in situations where data scarcity is an issue. Models combining the two architectures have been proposed and tested, opening new possibilities for improving results (Nie et al., 2018).

Beyond imaging, deep learning extends to treatment planning, particularly in RT. For instance, GANs for automatically generating radiation therapy treatment plans, including dose prediction, based on patient-specific anatomical data for prostate patients (Babier et al., 2020). In RT planning, CNNs are used to automatically contour tumours and organ at risk (OAR)s, reducing the manual labor involved in treatment preparation while increasing precision (Wong et al., 2021).

Despite the impressive capabilities of CNNs and GANs in advancing healthcare technologies, their success in clinical practice depends on their acceptance by medical professionals and its implementation must be supervised by experts in this field (Alam et al., 2016). This brings us to a key concern, the need for Explainable AI (XAI). The medical sector requires XAI rather than "black box" AI models, particularly in real-world applications. Black box AI refers to systems where the internal workings of the model—how decisions or predictions are made—are hidden from the user (Tjoa and Guan, 2020). In deep learning models like CNNs and GANs, while we can observe the input and the output, the exact processes through which the model arrives at its predictions are often opaque.

The implementation of explainable XAI in healthcare is still in its early stages, despite the significant potential it holds for advancing AI applications in the field. This lack of transparency poses a challenge in the medical field, where clinicians need to understand and trust the decision-making process. Without clear explanations, medical professionals may be reluctant to adopt these technologies in practice, as errors or unexpected outcomes cannot be easily interpreted or corrected. XAI, on the other hand, aims to make these models more transparent by providing insights into how they make decisions. This level of understanding is crucial for real-world adoption, especially in life-or-death decisions like cancer diagnosis or treatment planning processes (Korica et al., 2021).

Despite these advancements, implementing deep learning in healthcare still faces challenges. The need for large, annotated datasets, concerns over data privacy, and the "black box" nature of many deep learning models pose significant hurdles to widespread clinical adoption. Nevertheless, ongoing research is focused on addressing these challenges, with innovations in explainability and synthetic data generation paving the way for more generalisable and scalable solutions.

In the specific context of creating reliable sCT images, deep learning models must undergo extensive validation to ensure that the generated sCT images accurately reflect the patient's anatomy.

2.5 | Existing Applications and Past Case Studies of sCT Generation

In this section, we review existing applications and past case studies publish in research papers for the generation of sCT. Focusing on how models such as CNNs and GANs have been implemented and tested in this area of study.

2.5.1 | Application of CNN for sCT Generation

The study by Han (2017) represents one of the earlier applications of CNNs for the generation of MRI-based sCT images. While previous approaches to sCT generation relied on atlas-based or segmentation-based methods, this paper demonstrates a learning-based method, where a CNN can learn to directly map MRI data to CT-equivalent images.

The study involved using paired MRI and CT datasets from 18 brain tumour patients. The model, comprising 27 convolutional layers, applied convolution to detect important features in the images, such as edges and textures. It employed pooling operations to reduce the data size while retaining essential information, making the process more efficient. Unpooling operations were then used to restore the spatial details, ensuring that important features were preserved during the MRI-to-CT conversion process. As illustrated in 2.2, this combination allowed the model to extract and preserve detailed spatial features during the conversion process.

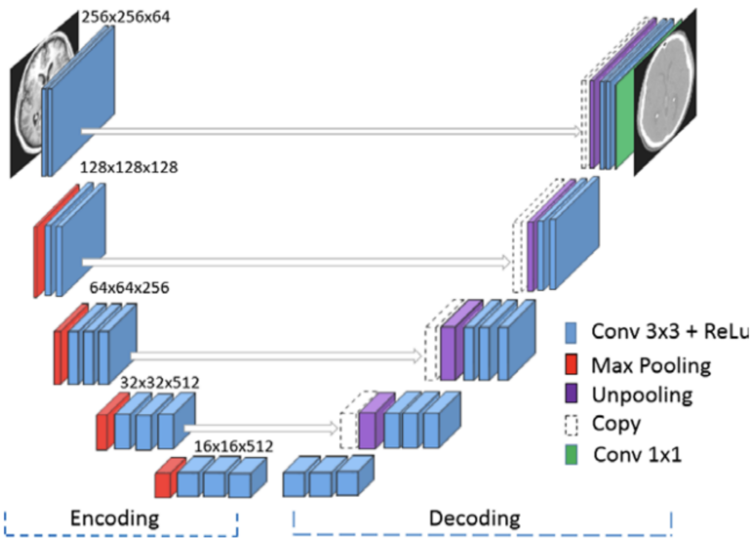


Figure 2.2: Overall architecture of the proposed sCT DCNN model (Han, 2017)

Despite it being published only 7 years ago, it is considered an old paper, being one of the first in this active area of study. The DCNN model developed "shown to be able to produce highly accurate sCT estimations from conventional, single-sequence MR images in near real time" (Han, 2017). One of the main contributions of the paper is its demonstration of improved accuracy and speed over traditional methods. The MAE of 84.8 Houndsfield Units (HU) for the CNN-based model represents a meaningful improvement compared to the 94.5 HU from atlas-based methods. Furthermore the CNN's ability to generate synthetic CT images in just 9 seconds is a substantial advancement in terms of computational efficiency.

Despite the results obtained, the limitations of the study must also be acknowledged. First, the study utilised a relatively small dataset, comprising only 18 brain tumor patients. Another limitation is that the model was trained and validated on images from a single institution, which may introduce bias due to uniform imaging protocols or scanner characteristics. These limitations may not fully capture the variability in patient anatomy and imaging conditions, which is crucial for the generalisation of the model to broader clinical settings.

Han's work underscored the ability of deep learning to automate and enhance the process of sCT generation. This advancement was a significant step toward MRI-only radiotherapy workflows, reducing registration errors and patient exposure to radiation.

2.5.2 | Application of GAN for sCT Generation

The study by Nie et al. (2018) addressed the generation of sCT scans, however, in this paper, the authors propose a deep learning-based approach that leverages GANs. The model consisted of a FCN that acts as the generator in a GAN framework. The FCN is tasked with learning the non-linear mapping from MRI to CT images, as illustrated in figures 2.3 and 2.4. Unlike traditional CNNs, the FCN preserves neighborhood information in the generated images, which is important for maintaining the spatial integrity of medical images. The discriminator network identifies whether the CT images are real or generated, while the generator is being trained to deceive the discriminator. To further improve the quality of the sCT images, the authors introduced a gradient difference loss function, ensuring that the generated images retain sharp details, particularly around edges and structures important for clinical interpretation. Additionally, the model uses an Auto-Context Model (ACM) to refine the predictions iteratively by incorporating contextual information from previously generated images.

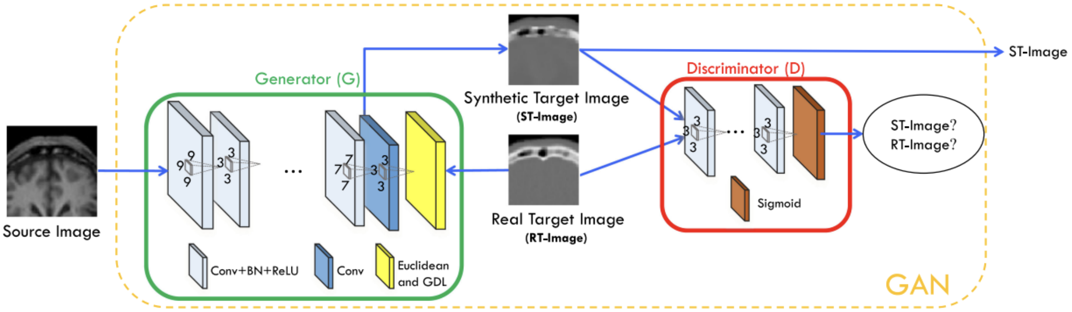


Figure 2.3: This figure illustrates the architecture used by Nie et al. (2018) in the deep convolutional adversarial setting for estimation of the synthetic target image

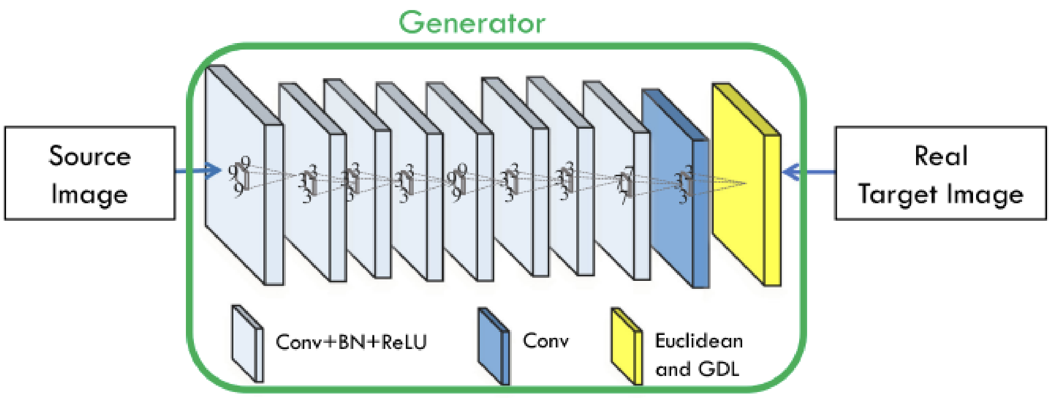


Figure 2.4: 3D FCN architecture for estimating a target image from a source image as outlined by Nie et al. (2018)

The proposed GAN model was evaluated on two datasets, a brain dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and a pelvic dataset. The model outperformed traditional atlas-based and random forest methods in terms of both accuracy and image quality. The MAE and PSNR for the GAN-based approach were significantly better than those of state-of-the-art methods. For the brain dataset, the proposed GAN model achieved a PSNR of 27.6, compared to 26.3 for the structured random forest method. On the pelvic dataset, the GAN-based method achieved an average PSNR of 34.1, outperforming the 32.1 PSNR of the best existing method.

This paper made several important contributions to the field of medical image synthesis. Firstly, it demonstrates the efficacy of GANs in generating high-quality MR-based sCT. Secondly, the introduction of the gradient difference loss function significantly improves the sharpness of generated images, addressing a common problem in deep learning-based image synthesis. Finally, the introduction of the ACM in this paper is designed to improve the quality and accuracy of generated images by taking into account the spatial relationships and context of the image data. In medical image synthesis, capturing long-range dependencies helps to maintain anatomical coherence and structural details of the generated images, particularly when dealing with medical images. The ACM refines the generated CT images by continuously incorporating feedback from the images generated in previous iterations. Essentially, it provides a way for the model to learn from the errors or imperfections in its earlier predictions and adjust the generated images accordingly.

Once again the scarcity of medical data is present and is highlighted as a limitation. The model was tested on relatively small datasets, with only 16 subjects in the brain dataset and 22 subjects in the pelvic dataset.

2.5.3 | SynthRAD2023 Challenge

The SynthRAD2023 Challenge Report, authored by Huijben et al. (2024), made a significant contribution to this active research area. This challenge facilitated the investigation of MR-based sCT generation, setting a benchmark and providing insights for developing MR-based approaches. Organised in conjunction with MICCAI 2023, it evaluates the latest deep learning-based sCT generation techniques using ground truth data from over 1,000 patients across multiple centers, focusing on MRI-to-CT and Cone Bean Computed Tomography (CBCT)-to-CT translation tasks.

This challenge fostered significant participation from 94 teams across the world, with 22 and 17 valid submissions for the two tasks, respectively. The accuracy of the generated sCT was evaluated using image similarity metrics together with photon and proton dose metrics.

Top-performing algorithms demonstrated impressive structural similarity indices as high as 0.90, and dose recalculations, showcasing the ability of deep learning models to produce high-quality synthetic CTs that match the groundtruth CT scans. Moreover, the sCTs generated for dose recalculation achieved excellent gamma pass rates, with photon and proton dose plans reaching up to 99.0%, showcasing the potential of these methods to accurately replicate CT-derived treatment plans. The deep learning architectures used in the challenge spanned various models, including CNNs, GANs, and newer approaches like vision transformers and diffusion models, demonstrating the breadth of techniques being explored in this field.

2.5.4 | Review Paper

The review by (Spadea et al., 2021), titled "Deep learning based sCT generation in radiotherapy and Positron Emission Tomography (PET): A review", presents a comprehensive review of recent advancements in deep learning-based methods for generating sCT. The focus is on three main clinical applications: (i) replacing CT in MRI-based treatment planning for RT, (ii) enhancing CBCT-based IMRT and (iii) generating attenuation maps for correcting PET data.

The review systematically examines studies published between 2014 and 2020, analysing the various deep learning architectures, such as CNN and GAN, and metrics used to evaluate the quality of sCTs. Key performance metrics include image metrics (such as MAE, and SSIM) as well as dose calculation accuracy metrics, such as Dose Difference (DD) and GPR

The paper concludes that deep learning- based sCT generation is a promising technology that could replace traditional methods for treatment planning in RT and PET correction, improving accuracy while reducing radiation exposure. However, challenges remain in terms of standardising protocols and improving generalisability across various patient populations.

2.6 | Clinical Acceptance criteria for sCT

Despite the close resemblance of sCTs generated by deep learning models to ground truth CT images, there is still no universally agreed-upon standard for determining the clinical acceptability of these sCTs (Vandewinckele et al., 2020).

Some works have investigated clinical acceptance criteria for sCTs. For example, Olberg et al. (2019) proposed that a photon GPR of greater than 98% using the 2 mm/2% criterion should be considered acceptable. This is a dose-related metric that assesses the quality of the sCT in terms of its ability to match the dose distribution of the original CT. On the other hand, Korsholm et al. (2014) proposes that treatments with a DVH difference of <2% are clinically acceptable. It is important to note that these criteria were proposed for breast, head-and-neck, and thorax sCT generation.

The suggested criteria was based on studies focused on specific anatomical regions, such as the breast, head-and-neck, and thorax. It is unclear whether this criteria is applicable to other anatomical regions, highlighting the need for further research and validation across different body areas.

Before addressing the clinical impact, it is imperative to consider the quality of the treatment plans adopted for the conducted studies. Treatment planning techniques may differ between institutes. Wieser et al. (2017) discusses the comparison of dose calculation accuracy between MatRad, an open-source treatment planning system created for research purposes, and clinically validated TPS. Despite these being generic models, the dose calculation results from MatRad were very close to those from clinically approved systems, with GPR deviating by no more than 0.5%. This indicates that matRad's dose calculation algorithms are highly effective and comparable to clinical TPS, regardless of the type of irradiation, whether photon, proton, or carbon ion therapy.

The SynthRAD2023 challenge set their own specific thresholds for this criterion: a GPR of over 99% for photon irradiation in regions receiving at least 10% of the prescribed dose. These values were proposed using a 2 mm/2% tolerance, which indicates how closely the sCT-generated dose distribution matches the reference distribution. Achieving these GPR ensures that the sCTs are clinically acceptable and can be relied upon for accurate dose calculations in radiotherapy (Huijben et al., 2024).

Several commercial solutions are already available for MR-only simulation is the Philips Magnetic Resonance for Calculating Attenuation (MRCAT) system as outlined by Köhler et al. (2015). This solution allows radiation dose calculations using only MRI data, eliminating the need for CT scans. MRCAT leverages specialised mDIXON imaging sequences to generate sCT images by classifying tissues into different categories and assigning appropriate HU values. However, many of these commercial algorithms require dedicated imaging protocols to generate accurate sCTs, potentially limiting their adaptability across different clinical settings. Moreover, its initial commercial application is limited to prostate cancer, with the potential to be adapted for other anatomical regions as MR-only simulation becomes more widespread in radiotherapy practices.

2.7 | Gaps in Literature

2.7.1 | The absence of standardised clinical criteria

The need for standardised clinical criteria for sCT acceptance is one of the most significant gaps in the current body of research on sCT generation for RT. The field still lacks a universally accepted set of guidelines or benchmarks that clearly define what constitutes a clinically acceptable sCT across different institutions and treatment modalities.

Regulatory bodies have not yet developed specific guidelines for the clinical validation and acceptance of sCTs. As sCT generation becomes more common in radiotherapy and diagnostic imaging, there is an urgent need for regulatory frameworks that define clear, validated acceptance criteria. This would not only ensure safety and efficacy but also pave the way for sCTs to be integrated into clinical workflows on a larger scale.

2.7.2 | Generalisation of sCT Algorithms

Generalisation of sCT algorithms is a persisting issue in the current research and clinical implementation in RT. Generalisation refers to the ability of an algorithm to perform accurately across different scenarios, such as varying patient anatomies, imaging protocols, scanner types, and clinical workflows, without the need for specialised adjustments (Maspero et al., 2020).

2.7.3 | The effects of diverse datasets

One of the key challenges in developing deep learning models, particularly in healthcare, is ensuring that they generalise well to new, unseen data, and that they perform equitably across diverse populations. Bias in data collection can result from imbalances in representation, such as over-representing certain demographic groups (e.g., male, female, white, or black). When models are trained on biased datasets, they may fail to perform well on underrepresented populations, which can have serious implications for clinical outcomes (Larrazabal et al., 2020).

2.8 | Conclusion

This chapter presented a detailed evaluation of the literature related to the study. In chapter three the research methodology used in the study will be discussed.

3

Research Methodology

3.1 | Introduction

The third chapter discusses the study's research methodology. This chapter covers the research approach, research strategy, data collection technique and execution used, as well as ethical considerations and study limitations.

3.2 | Research Approach

Given the objectives of this study, to assess a method for generating sCT images from MRI scans in the context of brain RT treatment planning, a quantitative research approach has been adopted. The nature of this study demands a high degree of objectivity that can be best achieved through quantitative methods. Numerical data analysis allows for the unbiased comparison of image qualities between the sCTs and ground truth CTs, providing clear, empirical evidence of the synthetic method's accuracy and efficacy.

The ultimate goal of this research is not only to generate sCT images but also to assess how the algorithm will perform on a local cohort of Maltese patients. With a population of just over 500,000 people, Malta is one of the smallest European countries in the world, making its population a minority group in the context of global research populations (Worldometer, 2024). This study thus contributes to the relatively limited data available on sCT generation for minority populations in RT.

The quantitative approach is inherently aligned with the study's objectives, providing a clear, measurable pathway to evaluating the quality of the sCTs generated from the local cohort, as well as their validity in RT application. This is achieved by collecting two sets of metrics; the image metrics and the dose metrics, further explained in section 3.7.2.5

The Image Metrics:

- MAE
- SSIM
- PSNR

The Dose Metrics:

- DVH Parameters
- GPR

3.3 | Research Strategy

The main research question addressed in this study is: How effectively can an algorithm developed for generating sCT images perform in the context of brain cancer RT treatment planning within a minority cohort of Maltese patients?

To answer this research question, an experimental and software testing strategy has been employed. This approach focuses on testing the algorithm on sCT images generated from MRI scans of 16 Maltese brain RT patients. The study involves a retrospective analysis, where pre-existing MRI scans are used to generate sCT images and CT scan data is used for comparison with the generated sCT scans. These images are then quantitatively compared with the corresponding ground-truth CT images to evaluate their accuracy and image quality.

The retrospective design of this study allowed for the evaluation of an existing algorithm, originally developed for ultrasound treatment planning, in the context of RT, without the need for real-time data collection or patient involvement. The dataset, sourced from Mater Dei Hospital, offered a valuable opportunity to assess the sCT generation algorithm on a minority cohort, providing unique insights into its performance in underrepresented populations.

While other research strategies, such as comparative or prospective studies, could provide additional insights, the chosen retrospective experimental approach is the most suitable for this study for several reasons:

3.3.1 | Focus on Algorithm Validation

The primary goal of this study is to assess the performance of a specific algorithm for generating sCT images from MRI scans in the context of radiotherapy treatment planning. An experimental approach allows for controlled testing of this algorithm on pre-exisiting MR images, providing direct insight into its effectiveness without the complexities of external comparisons or the need for new data collection in real-time.

3.3.2 | Efficiency and Feasibility

Conducting a retrospective study using a pre-existing dataset obtained from brain cancer patients who have already undergone scans for RT, minimises the logistical and time-related constraints associated with prospective studies. In a prospective design, data would need to be collected as patients progress through treatment, which would significantly extend the study duration. In contrast, the experimental approach with retrospective data allows for a more efficient evaluation of the algorithm's capabilities, delivering faster results without compromising the study's rigor.

3.3.3 | Applicability to Minority Populations

The cohort of Maltese patients represents a minority group in the global population, and this dataset provides a valuable opportunity to test the algorithm's performance in a population that is often underrepresented in global clinical studies. A comparative study involving multiple cohorts from different regions would be informative but also introduce variability that might obscure the algorithm's performance within this specific minority group. By focusing on this unique dataset, the experimental approach provides a clear, targeted assessment.

3.3.4 | Direct Quantitative Evaluation

The experimental method allows for a direct quantitative comparison between the generated sCT images and the ground truth CT images. This involves calculating keyimage metrics, as well as dosimetric evaluations. These quantitative measures offer clear, objective evidence regarding the algorithm's image accuracy and suitability for treatment planning, making the experimental approach ideal for achieving the research objectives.

3.4 | Data Collection Technique

The data for this study was collected retrospectively from Mater Dei Hospital in Malta. The dataset consists of pre-existing MRI and CT scans from 16 deceased patients diagnosed with brain cancer, which were originally acquired during the time of clinical procedures for RT treatment planning. MRI data were obtained using the Axial T1-weighted sequence with contrast, though specific MRI scanner model details were not accessible. CT imaging was performed using a Canon Aquilion Large Bore CT scanner, with scans acquired at 120 kV and a slice thickness of 2 mm, offering high-resolution data necessary for accurate treatment planning.

Since the study does not involve real-time data collection, there was no need for direct observation, patient consent, questionnaires or interviews. Instead, the research utilises secondary data, meaning that the scans were collected as part of the patients' standard clinical care and were then pseudoanonymised for use in this research. These images, stored in the hospital's medical imaging database, were accessed after securing the necessary ethical approvals.

The use of pre-existing clinical images was the most effective method. This approach allowed for a controlled comparison between the generated sCT images and the ground-truth CT images without the need to involve patients in new scanning procedures. Alternative data collection methods were considered but deemed less suitable for this study. Prospective data collection, involving the acquisition of new MRI and CT scans from patients, was rejected due to the logistical challenges and extended study timelines. Direct observation during RT planning sessions and qualitative methods, such as questionnaires or interviews with radiology professionals, were also considered. However, these approaches would have provided context-specific insights rather than the numerical image data necessary for algorithm validation.

3.5 | Data Collection Procedure

Data collection took place after obtaining ethical approval from Faculty Research Ethics Committee (FREC) to proceed with the research which commenced on the 23rd of November 2023 and was granted on the 27th of May 2024, the study was carried out over the course of 4 months, commencing in June 2024.

The collection of the dataset utilised for this study, followed institutional ethical guidelines, ensuring patient data confidentiality through the pseudoanonymisation of all personal identifiers by the intermediary of the study. The paired CT and MRI images were collected in Digital Imaging and Communications in Medicine (DICOM) format.

3.5.1 | Target Population

The target population for this study consisted of deceased Maltese patients diagnosed with brain cancer who had undergone both MRI and CT imaging for RT treatment planning. Given Malta's small population size, the cohort of 16 patients represents a unique minority group, underrepresented in global clinical studies.

The decision to use a purposive sampling technique was driven by the necessity to complete the study within a limited timeframe. By selecting deceased patients with pre-existing imaging data, the study was able to mitigate delays that could arise from collecting new data or obtaining consent from living patients.

This study is the first of its kind to specifically consider the Maltese population for sCT generation, contributing to the broader understanding of algorithm performance across diverse demographic groups. By focusing on this unique cohort, the research offers valuable insights into the applicability of sCT generation for populations that are often overlooked in global studies.

3.5.2 | Equipment Set Up

No new equipment was required for data collection, as the imaging data had already been collected as part of the patients' standard clinical care. The MRI and CT scans were stored within the hospital's Picture Archiving and Communication System (PACS) and retrieved for this study following ethical clearance.

3.6 | Data Collection Tools

The primary data collection tool employed in this study is the hospital's PACS at Mater Dei Hospital, which stores medical images taken within the institute. This medical imaging database provided access to the pre-existing imaging data necessary for the study. No physical instruments or direct patient involvement were necessary for this study, as all data were sourced from the hospital's internal records.

Alternative data collection tools, such as prospective imaging acquisition or direct observation methods, were deemed impractical for the study's objectives and timeline. Prospective data collection, involving the acquisition of new MRI and CT scans, would have required real-time patient involvement, extending the study period and introducing additional ethical considerations. Similarly, qualitative tools such as interviews with radiologists or clinicians would have provided contextual information but would not have been suitable for the quantitative nature of the research. Given the focus on algorithm validation, retrospective data from Mater Dei hospital's imaging archive was identified as the most appropriate choice.

3.6.1 | Validity and Reliability of the Data Collection Tool

The validity of the data collection tool was ensured through the hospital's standardised imaging protocols for MRI and CT acquisition. Since all the data was collected from the same institution, lack of standardisation was not a concern. All imaging data was collected using state-of-the-art radiological equipment, following clinical guidelines, ensuring the accuracy and consistency of the images.

3.7 | Data Analysis Technique

3.7.1 | Computer and GPU specifications

The computational analysis and sCT generation in this study were conducted on a Windows 10 system with a 64-bit operating environment. The system was equipped with an NVIDIA GeForce GTX 1080 Ti graphics processing unit (GPU), which provided the computational power required to process the MRI datasets and run deep learning model employed.

Core Specifications	
Component	**Specification**
CUDA Cores	3584
Base Clock Speed	1480MHz
Boost Clock Speed	1582MHz
Memory	11GB GDDR5X
Memory Speed	11Gbps
Memory Interface	352-bit
Memory Bandwidth	484 GB/s

Table 3.1: Core Specifications of the Computational Setup

The NVIDIA GeForce GTX 1080 Ti is based on the Pascal architecture, which is specifically designed to handle high-performance computing tasks, including deep learning applications. With 3584 CUDA cores and 11GB of GDDR5X video memory, the GTX 1080 Ti supports parallel processing of large data sets, accelerating inference times for deep learning models and allows for efficient handling of high-resolution medical imaging, such as MRI scans.

Furthermore, the GTX 1080 Ti is fully compatible with popular deep learning libraries, including TensorFlow and PyTorch. This compatibility made it an ideal choice for the study, as both TensorFlow and PyTorch were used extensively in running the sCT generation algorithm. These libraries offer optimised performance on NVIDIA GPUs by leveraging CUDA, a parallel computing platform and programming model developed by NVIDIA for general-purpose GPU computing, which enables the seamless execution of deep learning workflows (Heryadi and Hampton, 2019).

While there are newer GPUs available, the GTX 1080 Ti remains a reliable and efficient option for deep learning, offering a good balance between performance and cost. It was chosen for this study because its Pascal architecture and CUDA support make it well-suited to handle the computational demands of generating sCT images from MRI data, without introducing bottlenecks during the data processing and analysis phases.

To ensure compatibility with specialised medical imaging software, such as the FSL, the Windows Subsystem for Linux (WSL) was used. WSL enabled the execution of Linux-based applications within the Windows environment, allowing the study to leverage FSL for specific neuroimaging tasks and adequate visualisation of medical images without requiring a separate Linux machine. Using FSLeyes, the medical images were examined to ensure correct alignment and image quality during the analysis process.

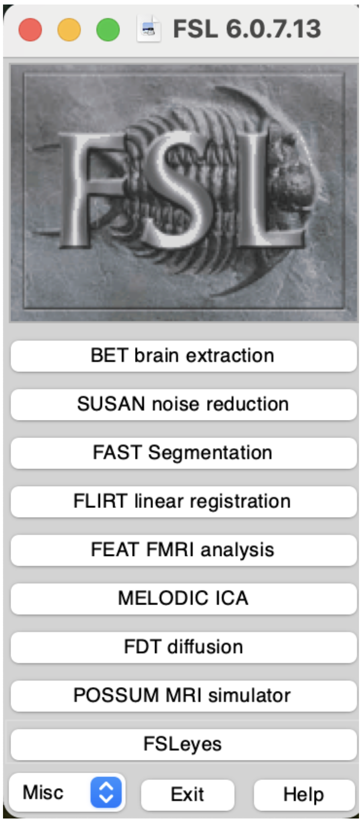


Figure 3.1: The FSL 6.0.7.13 graphical user interface, displaying various neuroimaging tools.

3.7.2 | Data Analysis Pipeline

The data analysis pipeline employed in this study comprises five key stages: Pre-processing, Algorithm Testing & sCT Generation, Post-Processing, Data Collection, and Metrics Evaluation. The schematic presented illustrates the flow of data throughout the study, from the initial preparation of MRI and CT datasets to the evaluation of the generated sCT images.

Pre-processing → Algorithm Testing & sCT Generation → Post-Processing → Data Collection → Metrics Evaluation

3.7.2.1 | Pre-processing

The preprocessing of CT and MRI images is an essential step in ensuring the quality and consistency of the data before and to ensure that the MRI images are in the correct format to be accepted by the algorithm. All the pre-processing steps were executed using Python scripts, which can be found in Appendix A section A.1.1, to ensure consistency and reproducibility across the dataset. The pre-processing workflow consisted of three key stages: file conversion, image registration and mask generation.

File Conversion → Image Registration → Mask Generation

File Conversion The original medical images were retrieved from the PACS in DICOM format, which is the standard format used for medical images. However, for compatibility with the deep learning framework, the files were converted to Neuroimaging Informatics Technology Initiative (NIfTI) format. The NIfTI format is widely used in neuroimaging due to its simplicity and ability to handle multi-dimensional image data efficiently (Li et al., 2016). The DICOM series was first converted into a SimpleITK image object and then saved as a NIfTI file at the specified path. This process was implemented using Python libraries such as pydicom for reading the DICOM files and SimpleITK for managing the conversion and handling image data.

Image Registration In the image registration step, the MRI scans were selected as the fixed images, while the CT scans were designated as the moving images. This approach ensured that the CT images were anatomically aligned to their respective MRI scans for each patient, maintaining consistency between the modalities. By carrying out image registration at this stage, it ensures that the sCT images generated from MRI data are correctly aligned with the ground-truth CT images.

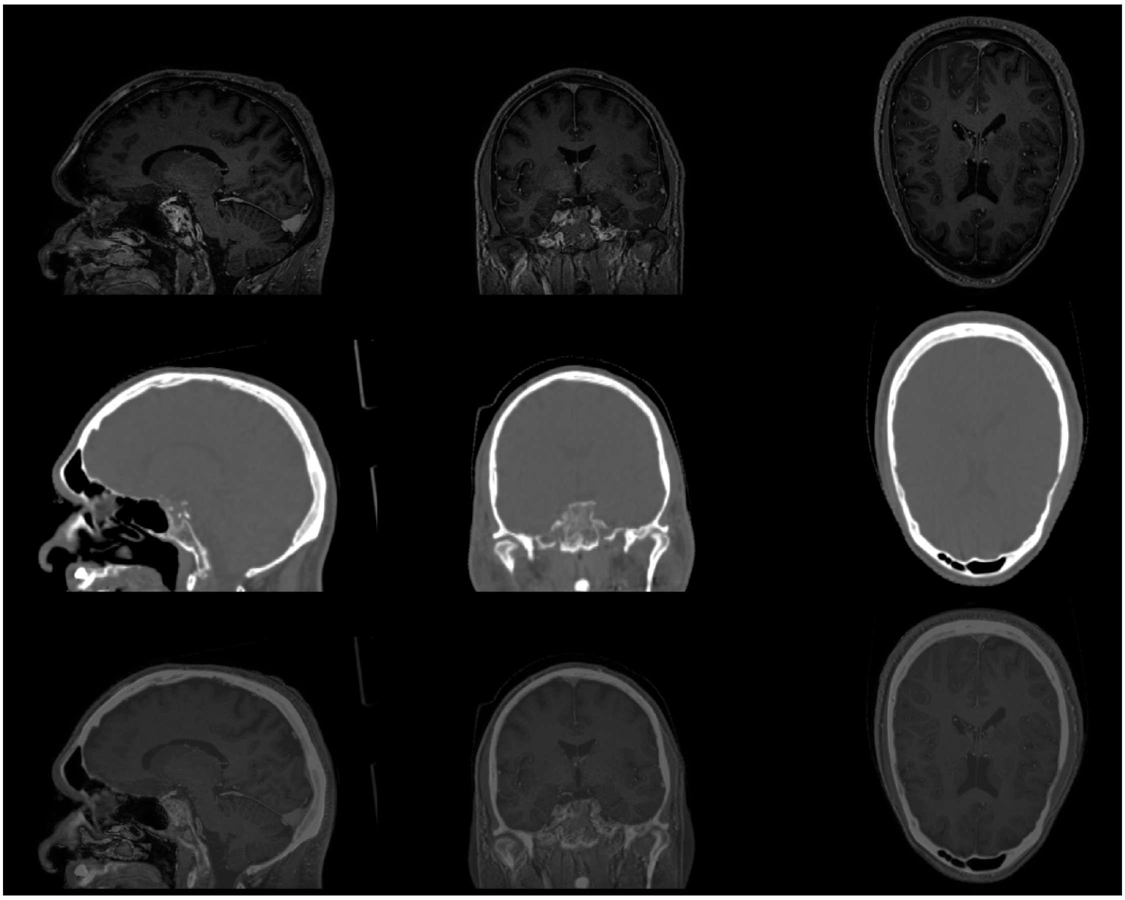


Figure 3.2: This figure visually represents the achievable result of CT and MRI co-registration, displayed as screenshots from the FSLeyes viewing software. The first row features the MRI scan in three orientations, while the subsequent rows show the CT scan and an overlay of the CT scan on the MRI scan, respectively.

To achieve this, rigid registration, as observed in figure 3.2, was performed using a predefined set of parameters from the rigid_parameter.txt file provided by Thummerer et al. (2023) in the SynthRad 2023 challenge. These parameters were developed for aligning CT images to MRI data during the pre-processing stage of SynthRAD2023 datasets, as outlined by Thummerer et al. (2023). The registration process was implemented using SimpleITK, applying rigid transformations to account for differences in position, orientation, and scale between the MRI and CT scans.

Mask Generation Following the registration process, mask generation was performed on the MRI images. Due to the presence of a stereotactic frame in some of the CT images, it was not possible to generate the required head masks from the CT scans. However, since the MRI and CT images were aligned for each patient, only one mask was needed to capture all anatomical structures within the skin boundary, this can be seen in figure 3.3. These masks will be utilised in section 3.7.2.5 to generate the image metrics, ensuring accurate evaluation of the sCT images based on the defined anatomical regions.

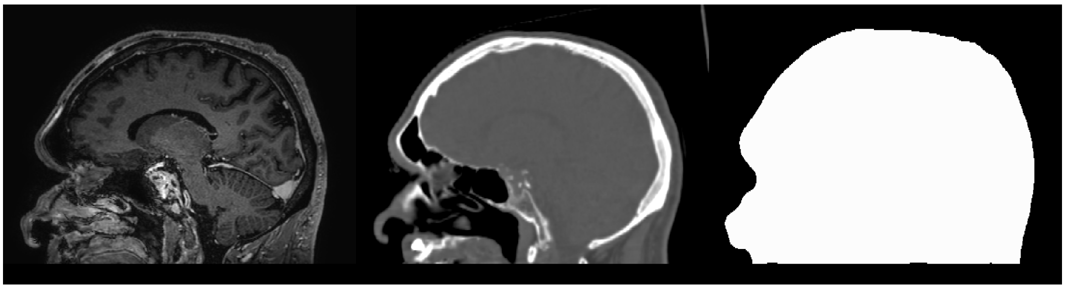


Figure 3.3: Figure displaying the registered MR and CT images along with the generated mask, visualised using FSLeyes. The MR image (left) and CT image (middle) demonstrate the anatomical alignment, while the mask (right) outlines the region of interest used in the analysis for metric evaluation.

3.7.2.2 | Algorithm Testing and sCT Generation

The sCT generation for this study was performed using a 3D patch-based cGAN, as outlined by Liu et al. (2022) in their work on generating sCT images from MRI data for transcranial MRI-guided focused ultrasound. This model was designed to synthesize high-quality CT images from routinely acquired MRI scans, minimising the need for actual CT scans during therapeutic planning. This is particularly beneficial for procedures such as transcranial MRI-guided focused ultrasound, where reducing radiation exposure is as minimal as possible. The conditional input from MRI scans allows the model to generate sCT images that are well-suited for medical applications, particularly in situations where CT data may be scarce or impractical to acquired.

A cGAN is a variant of the GAN designed to generate data conditioned on some input. Unlike a standard GAN, where the generator learns to produce data from random noise, a cGAN learns to generate data based on a specific input, such as, generating CT images conditioned on MRI scans . This makes it particularly suitable for applications such as image-to-image translation, where the goal is to map one domain to another domain (Goodfellow et al., 2014). The cGAN architecture employed in this study consists of two main components:

Generator The generator was responsible for synthesizing the sCT images and was based on a 3D ResNet architecture with 9 residual blocks. Residual blocks were chosen for their ability to improve the training of deeper networks by mitigating the vanishing gradient problem. This design enhances the generator's capacity to capture the complex anatomical structures present in MRI scans, which is critical for producing high-quality sCT images.

- Input: Single-channel 3D MRI patch
- Output: Corresponding sCT patch
- Final activation: Tanh function, followed by a scaling operation to map output to the appropriate HU range

The use of 3D patches enables the model to capture spatial relationships within volumetric data, improving the quality of the generated sCT images. Additionally, the overlapping patch strategy ensures smooth transitions when reconstructing full sCT volumes.

Discriminator The discriminator was implemented as a 3D convolutional PatchGAN classifier, which is tasked with distinguishing between real CT patches and sCT patches generated by the generator. Unlike a standard GAN discriminator that evaluates entire images, the PatchGAN operates on local patches, classifying each patch as real or synthetic. This strategy allows the discriminator to focus on high-frequency details, such as edges and textures, which are essential for generating realistic medical images.

- Patch-based Discriminator: The PatchGAN architecture is particularly suited for modeling high-frequency features, which are crucial for the accuracy and realism of medical images like CT scans.
- Loss Function: The discriminator is trained with an adversarial loss, which drives the generator to produce sCT images that are indistinguishable from original CT images.

The proposed model for generating MR-based sCT images involves several key steps, as illustrated in the schematic diagram figure 3.4. The workflow can be divided into three

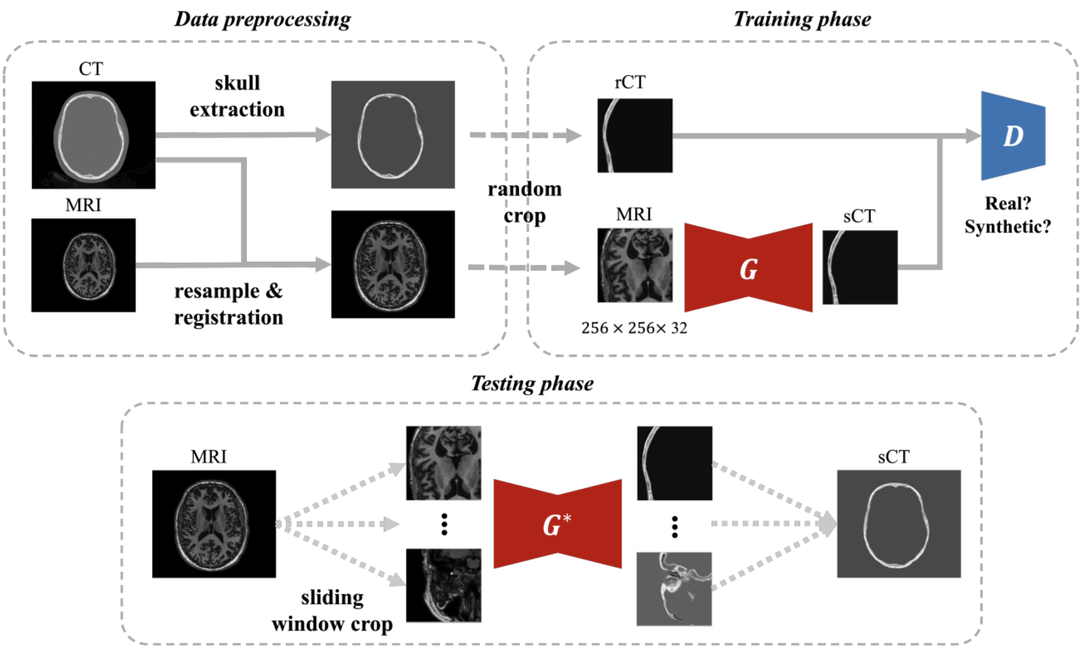


Figure 3.4: Workflow of the proposed method by Liu et al. (2022).

primary phases: data preprocessing, training, and testing. During data preprocessing, CT and MRI images are aligned through skull extraction, resampling, and registration. The training phase involves the generator (G) creating sCT images from random MRI patches, while the discriminator (D) determines if the output is real or synthetic. In the testing phase, the network generates full sCT images using sliding window crops from the input MRI images.

The model is trained and validated on a dataset consisting of 86 paired CT and T1-weighted MRI images. The dataset was randomly split into 66 images for training, 10 for validation, and 10 for testing.

The cGAN was tested on a Maltese dataset consisting of 16 patients. To ensure smooth transitions and continuity between the sCT patches generated by the model, the MRI data was divided into overlapping 3D patches, with a 0.9 overlap between consecutive patches. The entire testing process was performed within a Docker container, ensuring reproducibility and consistency across different computational environments.

3.7.2.3 | Post-Processing

The post-processing steps were performed to ensure the sCT images could be successfully imported into Monaco, the TPS provided by Elekta, currently used at the radiotherapy department at SAMOC, Mater Dei Hospital. All the post-processing steps were executed using Python scripts, which can be found in A in sections A.1.2 and A.1.3, to ensure consistency and reproducibility. The post-processing workflow consisted of three key stages: Image re-orientation, file conversion and the use of an intermediary software.

Image Re-Orientation → File Conversion → Intermediary Software

Image Re-Orientation All generated sCT images were reoriented to match the standard orientation matrix $\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}$, ensuring compatibility with the imaging standards required by the Monaco system. This orientation ensures that the images are reconstructed in an axial plane, which is the expected format for proper interpretation by the TPS. Without this reorientation step, oblique reconstructions were produced, which restricted importation into Monaco TPS. The re-orientation was implemented using Python, ensuring that the images were standardised and correctly aligned to meet the axial orientation required by the system. The developed code can be located in Appendix A, section A.1.2.

File Conversion Given that the generated sCT images output in NIfTI format, they had to be converted back to the DICOM format to be compatible with the Monaco system. This step was more complex than converting DICOM to NIfTI because it involved rewriting specific DICOM tags required for proper importation into the TPS. The NIfTI images are converted to DICOM format using the 3D Slicer software. This conversion was carried out within the DICOM module as shown in figure 3.5, the DICOM tags offered as inputs by the software are illustrated in the screenshot.

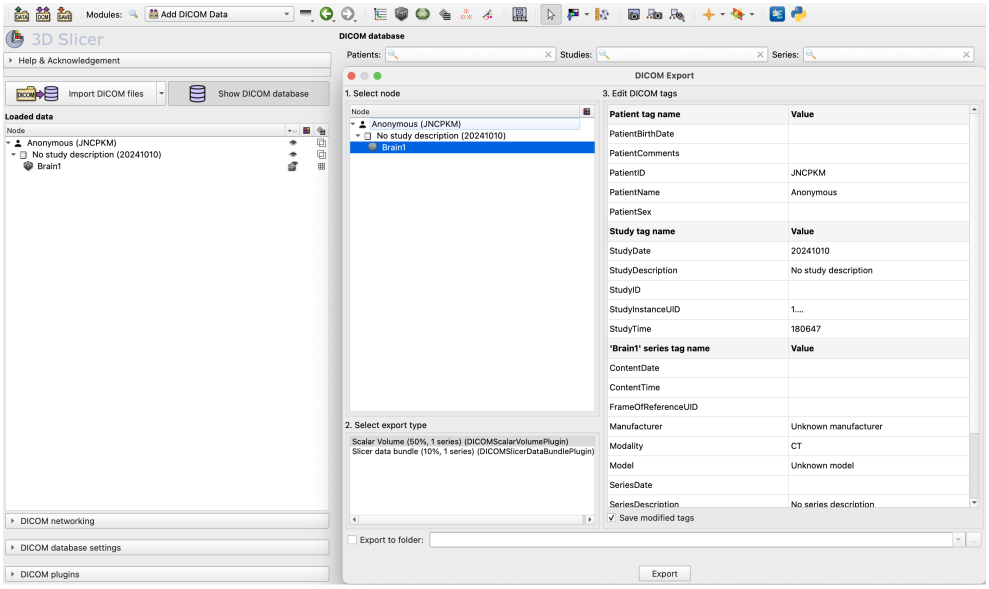


Figure 3.5: The figure shows a screenshot of 3D Slicer's software interface, specifically the DICOM module.

After the conversion, additional required DICOM tags were added or modified using Python and the pydicom library, the developed code can be located in Appendix A, section A.1.3. The required DICOM tags, also outlined in Elekta Solutions AB (2024), are listed in table 3.2.

DICOM Tag	**Tag Code**
Patient Name	(0010,0010)
Patient ID	(0010,0020)
Patient Sex	(0010,0040)
Referring Physician's Name	(0008,0090)
Study ID	(0020,0010)
Accession Number	(0008,0050)
Study Description	(0008,1030)
Patient Instance UID	(0010,0021)
Study Instance UID	(0020,000D)
Pixel Spacing	(0028,0030)
Image Orientation	(0020,0037)
Image Position	(0020,0032)
Rows	(0028,0010)
Columns	(0028,0011)
Bits Allocated	(0028,0100)
Bits Stored	(0028,0101)
Pixel Representation	(0028,0103)
Pixel Data	(7FE0,0010)
Smallest Image Pixel Value	(0028,0106)
Rescale Intercept	(0028,1052)
Rescale Slope	(0028,1053)
Window Centre	(0028,1050)
Window Width	(0028,1051)
Image Type	(0008,0008)

Table 3.2: DICOM Tags and Corresponding Codes required for Monaco TPS Importation

Intermediary Software ProKnow DS is a medical software developed by Elekta Solutions designed for RT treatment planning and analysis. It serves as an intermediary tool, allowing for the organisation, evaluation, and management of treatment data (ProKnow, 2024). In this study, ProKnow DS version 2.0.2.0 was used to facilitate the importation of medical imaging data into the MONACO TPS. By aiding in data transfer and preparation, ProKnow DS ensures seamless integration with MONACO TPS, as seen in figure 3.6.

Figure 3.6: The figure shows a screenshot the ProKnow DS used as the intermediary software, showing an imported sCT image.

3.7.2.4 | Data Collection

In order to comprehensively evaluate the quality and performance of the sCT images, a range of image and dose metrics were collected and analysed. These metrics include both image and dose metrics, which together provide a thorough assessment of how well the sCT images compare to the original CT images and how accurately they reproduce the dose distribution for RT treatment planning.

A series of additional steps were undertaken to ensure the accurate extraction of numerical values from the sCT images within the Monaco treatment planning system, as outlined in the workflow diagram.

Registration → Skin Contour Creation → sCT Treatment Plan Application

First, the sCT images underwent a meticulous registration process to ensure precise alignment with the original anonymised CT images, a required step in facilitating an accurate comparison between both image quality and dose metrics retrieved from Monaco. This registration step, as illustrated in figure 3.7, visually represents the alignment process, where the sCT and corresponding original CT images are color-coded with a green and magenta overlay, respectively. The registration process was repeated three times for each patient to ensure consistency and reliability across all comparisons.

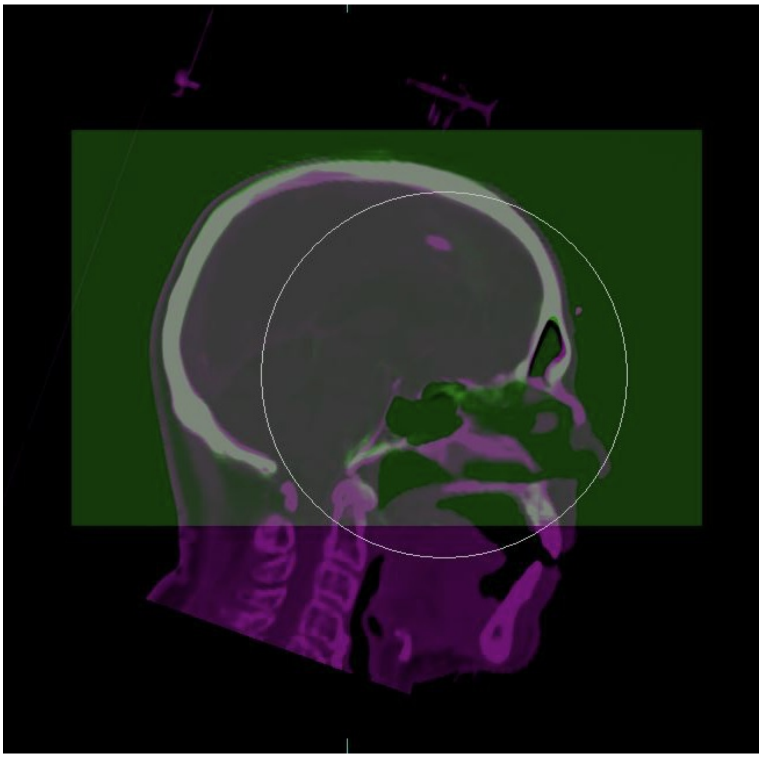


Figure 3.7: The figure shows a screenshot of the registration visual located within MONACO TPS

Following this, the skin contour creation step was performed, where the patient's skin contour was delineated for the CT excluding any immobilisation devices. A few of the anonymised CTs in the dataset used for this studies include patients that were images while wearing the stereotactic frame designed for CT imaging. The skin contours were copied to their corresponding sCT images using functions within the structure edit module using the 'Copy Structure' in the 'Structure Edit' section as shown in figure 3.8. This contour defined the boundaries of the patient's skin, ensuring consistency in dose calculations across both sets of images.

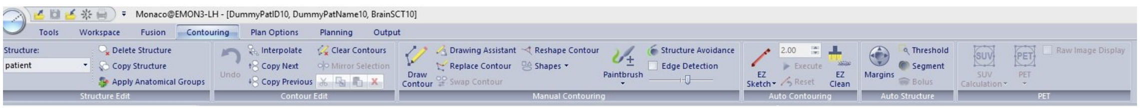


Figure 3.8: The figure shows a screenshot of the contouring section located within Monaco TPS

Finally, the sCT treatment plan application was carried out, where the original treatment plan created for the CT images was applied to the sCT images, allowing for a direct comparison of the dose distribution metrics, such as DVH and GPR. These steps were taken in ensuring the validity of the evaluation process and the accuracy of the extracted numerical values for both image and dose metrics collected from Monaco TPS.

1. Image Metrics Collection
The image metrics collected include MAE, PSNR, and SSIM, each of which plays a role in evaluating different aspects of the structural accuracy and quality of the sCT images compared to the original CT images.

Mean Absolute Error (MAE) Collection The MAE metric was calculated by importing both the generated sCT and the original CT images into the Monaco treatment planning system (version 5.11). The 'Image Statistics' tool, in the 'Tools' section, as depicted in figure 3.9, was utilised to compute statistical values such as the minimum, maximum, mean, and median intensity levels in HU. These values were specifically calculated for both the skin and brain contours, providing an in-depth evaluation of the consistency and accuracy of the sCT images in replicating the original CT data. The results of this analysis are shown in the Image Statistics window, in figure 3.10, where the calculated metrics help validate the reliability of the sCT images for radiotherapy treatment planning.

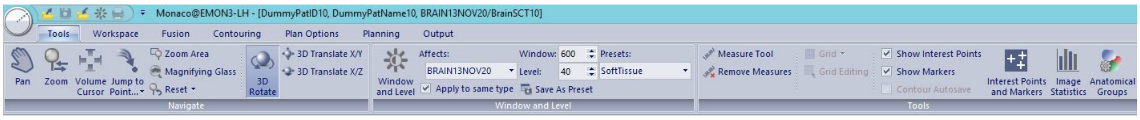


Figure 3.9: The figure shows a screenshot of the tools section located within Monaco TPS

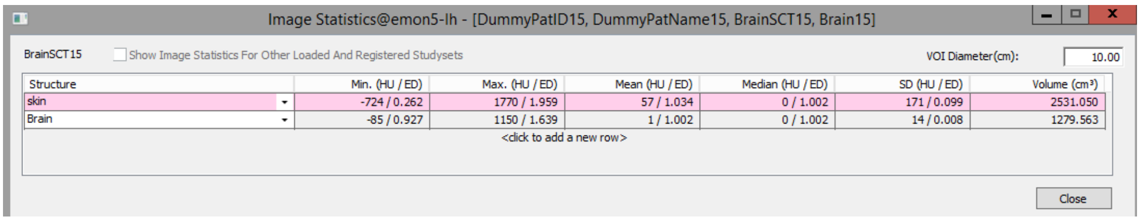
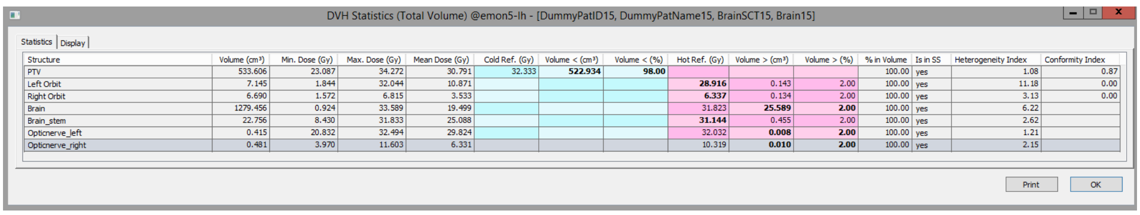


Figure 3.10: The figure shows a screenshot taken from Monaco TPS of the Image Statistics window

PSNR and SSIM Collection The remaining image metrics, PSNR and SSIM, were calculated by developing Python scripts, which utilised the skimage and numpy libraries for advanced image processing. These metrics offered valuable additional insights into the image quality and structural similarity between the sCT and CT images generated from the Maltese dataset. The scripts, located in Appendix A section A.1.4, systematically processed the registered images and computed precise quantitative values for both PSNR and SSIM, which further validated the accuracy and reliability of the sCT images in replicating key structural features present in the original CT scans. The detailed Python code used to calculate these metrics, which can be found in Appendix A, section A.1.4, allowing for reproducibility of the analysis performed in this study.

2. Dose Metrics The dose metrics collected include DVH parameters and GPR, which are insightful metrics for assessing how accurately the sCT images replicate the dose distribution observed in the original CT images. Accurate dose distribution is imperative for ensuring that the treatment plan effectively delivers the prescribed radiation dose to the tumour while minimising exposure to surrounding healthy tissues and OARs.

DVH Parameters Collection The DVH parameters were extracted from the Volumetric Modulated Arc Therapy (VMAT) plan for both the CT and sCT images. The DVH parameters collected, including V95%, D98, D2, and Dmean, represent objective functions and dose constraints that are commonly used to evaluate the quality and effectiveness of RT treatment plans. These parameters provide critical insights into how well the treatment plan meets the prescribed dose requirements for both the target volume and surrounding OARs. Specifically, V95% reflects the volume of the target receiving at least 95% of the prescribed dose, D98 indicates the dose received by 98% of the target volume, D2 represents the dose received by the most exposed 2% of the volume, and Dmean indicates the average dose to the structure. This can be observed in the DVH statistics window in figure 3.11


DVH Statistics (Total Volume) @emon5-lh - [DummyPatID15, DummyPatName15, BrainSCT15, Brain15]

Statistics | Display

Structure	Volume (cm^3)	Min. Dose (Gy)	Max. Dose (Gy)	Mean Dose (Gy)	Cold Ref. (Gy)	Volume < (cm^3)	Volume < (%)	Hot Ref. (Gy)	Volume > (cm^3)	Volume > (%)	% in Volume	Is in SS	Heterogeneity Index	Conformity Index
PTV	533.606	23.087	34.272	30.791	32.333	**522.934**	**98.00**				100.00	yes	1.08	0.87
Left Orbit	7.145	1.844	32.044	10.871				**28.916**	0.143	2.00	100.00	yes	11.18	0.00
Right Orbit	6.690	1.572	6.815	3.533				**6.337**	0.134	2.00	100.00	yes	3.13	0.00
Brain	1279.456	0.924	33.589	19.499				31.823	**25.589**	**2.00**	100.00	yes	6.22	
Brain_stem	22.756	8.430	31.833	25.088				**31.144**	0.455	2.00	100.00	yes	2.62	
Opticnerve_left	0.415	20.832	32.494	29.824				32.032	**0.008**	**2.00**	100.00	yes	1.21	
Opticnerve_right	0.481	3.970	11.603	6.331				10.319	**0.010**	**2.00**	100.00	yes	2.15	

Print OK

Figure 3.11: Screenshot taken from Monaco TPS of the DVH Statistics window.

These DVH parameters help assess whether the treatment plan effectively delivers the required therapeutic dose to the tumour while minimising radiation exposure to healthy tissues. The dose distribution itself, illustrated in 3.12, is graphically represented on the x-axis as the percentage of the prescribed dose, while the y-axis depicts the percentage

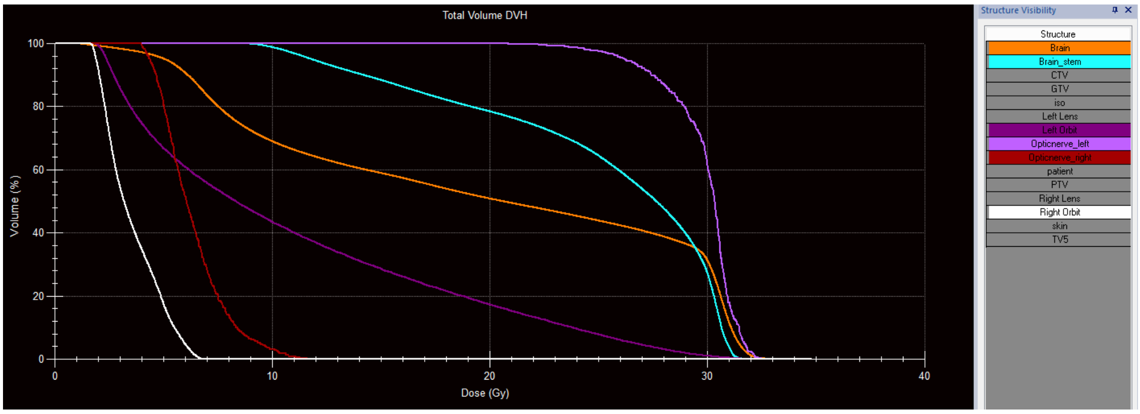


Figure 3.12: Screenshot of Total Volume DVH Graph from Monaco TPS, featuring a color-coordinated legend.

of tissue volume receiving that dose. Steeper curves represent OARs receiving minimal radiation exposure, while flatter curves correspond to the tumour target volume, which requires a higher dose for effective treatment. The graph allows for a detailed evaluation of the plan's effectiveness in delivering the optimal dose to the target volume while protecting the surrounding healthy tissues.

GPR Calculation The final dose metric, GPR, was calculated using the Sun Nuclear Arc-check system, a 3D dosimetry system designed for verifying dose delivery in rotational therapies like VMAT and IMRT. The system was employed to verify the dose accuracy of the sCT images by comparing them to the original VMAT plans. A screenshot of the GPR calculation within the Arc-check system.

Figure 3.13 presents a screenshot taken from the Arc-check system, showing the GPR calculation in action. In the upper panels, the color wash indicates the dose distribution overlaid on the patient anatomy, where red represents the areas receiving the highest dose, and blue indicates the lowest dose regions. In the lower-left panel, a gamma analysis map shows regions where the dose agreement is acceptable in blue and regions where the dose discrepancies exceed the tolerance limits are highlighted in red. The lower-right panel shows the dose profile, plotting the dose across a cross-section of the patient, further validating the dose consistency between the planned and delivered treatments.

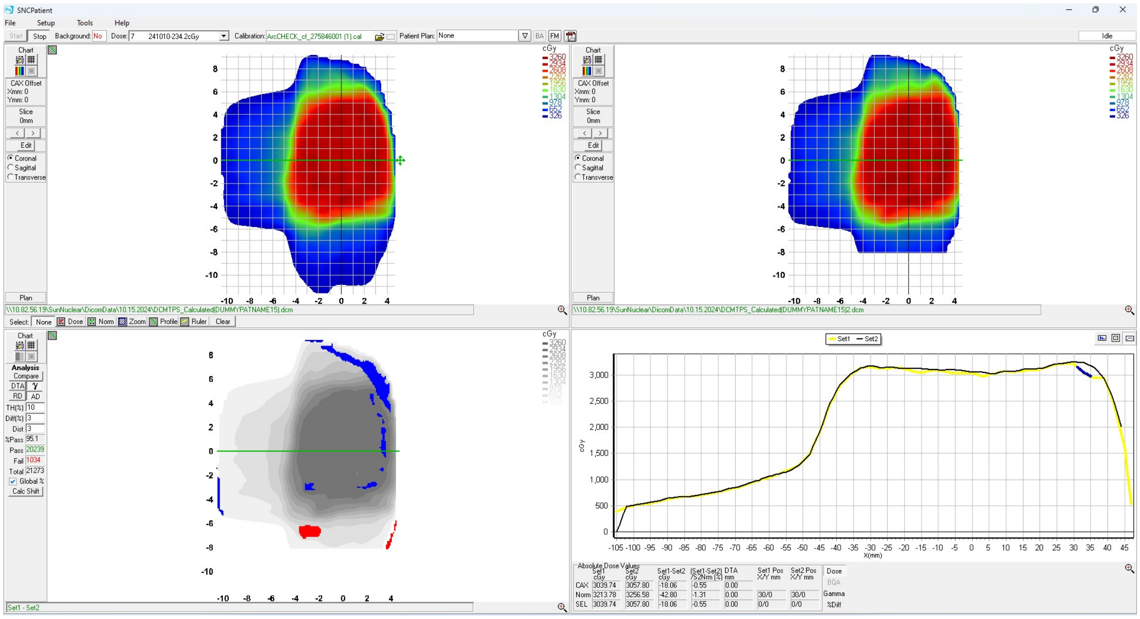


Figure 3.13: Screenshot from the Sun Nuclear Arc-check system showing the GPR calculation.

The numerical values for all the aforementioned metrics were systematically tabulated and analysed using Microsoft Excel. This step allowed for efficient data organisation and detailed comparison between the sCT and CT images, providing insights into the accuracy and clinical viability of the generated sCTs. The comprehensive results and their corresponding analysis, presented in formats, are discussed in detail in section 4.2.

3.7.2.5 | Metrics Evaluation

By evaluating the metrics this study aims to assess how closely the sCTs resembled the groundtruth CTs. Two sets of metrics were collected to evaluate the accuracy and clinical applicability of the sCT images, these are the image metrics and the dose metrics. The choice of these metrics was guided by those used by Huijben et al. (2024) in the SynthRAD 2023 Challenge.

Image Metrics

The accuracy of the generated sCT images was evaluated using the following standard image similarity metrics, which were computed within the body contour mask generated in the pre-processing phase and the skin contour created within Monaco, outlined in sections 3.7.2.1 and 3.7.2.4 respectively. Together, these regions enabled a thorough assessment of the sCT image quality, allowing for an in-depth comparison of both structural accuracy and dose distribution fidelity.

1. MAE

This measures the average absolute differences between the sCT and the original CT images, providing a quantitative assessment of how closely the sCT images resemble the ground-truth CTs. The final masked MAE value is then obtained by calculating:

$$\text{MAE}(\text{CT}, s\text{CT}) = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} |\text{CT}_i - s\text{CT}_i|$$

Where B represents the set of voxels within the body contour, and the sum is over all voxels in this region. It excludes areas outside the mask, such as air or regions not relevant for clinical analysis, making the metric more focused on the anatomical region of interest.

A lower MAE indicates a better agreement between the synthetic and ground-truth CTs, meaning the sCT accurately replicates the intensity values of the groundtruth CT. Intuitively, a higher MAE suggests larger discrepancies between the synthetic and grountruth CT images, indicating less accurate image generation.

2. PSNR

PSNR quantifies the ratio between the maximum possible pixel value and the noise present in the sCT images. It provides insight into the image quality by comparing the dynamic range of pixel intensities between the sCT and CT images. The final masked PSNR value is then obtained by calculating:

$$\text{PSNR}(\text{CT}, s\text{CT}) = 10 \log_{10} \left(\frac{Q^2}{\frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} (\text{CT}_i - s\text{CT}_i)^2} \right),$$

Where Q is the dynamic range of the voxel intensities ([-1024, 3000] HU). The CT and sCT were clipped to the dynamic range to calculate the masked PSNR The higher the PSNR value, the closer the sCT is to the CT.

A higher PSNR values indicate better image quality, with less noise and greater similarity between the sCT and original CT images. In general, a high PSNR suggests that the synthetic image accurately replicates the pixel intensity values of the original CT, meaning the structural details and dynamic range are well preserved. Conversely, lower PSNR values indicate increased noise and greater discrepancies between the sCT and CT, implying a loss of detail or significant deviation in pixel intensity. In the context of medical imaging, a higher PSNR is imperative as it reflects the ability of the sCT to closely mimic the original CT data, which is essential for accurate radiotherapy planning and other clinical applications. However, while PSNR provides useful information about the overall image quality, it should be interpreted alongside other metrics, such as SSIM, to gain a more comprehensive understanding of both structural and perceptual image fidelity.

3. SSIM

This metric evaluates the perceptual quality of the images by comparing structural information, including luminance, contrast and texture, between the sCT and CT images. The final masked SSIM value is then obtained by computing:

$$\text{SSIM}(\text{CT}, s\text{CT}) = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \text{SSIM}(\text{CT}_i, s\text{CT}_i),$$

It ranges from -1 to 1, where higher values indicate greater similarity. A value of 1 implies perfect similarity, meaning the two images are structurally identical, while lower values indicate less similarity.

Dose Metrics

In addition to evaluating the accuracy of the sCT images through image-based metrics, the clinical viability of the sCT for RT treatment planning was assessed using dose metrics. These metrics assess how well the synthetic images preserve the accuracy of the radiation dose distributions when compared to ground-truth CT images. Two primary dose metrics were utilised in this study: the DVH parameters and the GPR.

4. DVH Parameters

The DVH parameters was employed to evaluate the differences in radiation dose distributions between the sCT and CT images. These metrics provide a comprehensive analysis of how closely the dose calculated on the sCT images resembles that of the original treatment plan based on the CT. The DVH parameter metrics evaluate key dose characteristics for both the planning treatment volume (PTV) and the OARs. The use of the near-minimum and near-maximum was suggested by International Commission on Radiation Units and Measurements (2010). Four key terms were calculated and then combined to generate a singular overall DVH metric for each patient, as follows:

D98 for the PTV

This term represents the near-minimum dose that covers 98% of the PTV. A lower D98 difference indicates better agreement between the two dose distributions at the PTV. It is computed as the relative absolute difference between the D98 value from the CT and the sCT:

$$D_{98,\text{PTV,CT}} = \frac{|D_{98,\text{PTV,CT}} - D_{98,\text{PTV,sCT}}| + \epsilon}{D_{98,\text{PTV,CT}} + \epsilon}$$

V95 for the PTV

This parameter evaluates the percentage of the PTV that receives at least 95% of the prescribed dose. The relative difference between the V95 values from the CT and sCT is computed as:

$$V_{95,\text{PTV,CT}} = \frac{|V_{95,\text{PTV,CT}} - V_{95,\text{PTV,sCT}}| + \epsilon}{V_{95,\text{PTV,CT}} + \epsilon}$$

D2 for the OARs

This metric is used to ensure the sCT does not result in a higher dose to critical organs. This term represents the near-maximum dose received by the OARs, calculated as the average relative difference for all considered OARs:

$$D_{2,\text{OARs}} = \frac{1}{n_{\text{OARs}}} \sum_{\text{OAR}} \frac{|D_{2,\text{OAR,CT}} - D_{2,\text{OAR,sCT}}| + \epsilon}{D_{2,\text{OAR,CT}} + \epsilon}$$

Dmean for the OARs

A low difference in Dmean ensures that the sCT maintains dose accuracy to critical structures. This evaluates the average dose received by the OARs, computed as the mean relative difference in the dose to the OARs between the original CT and sCT:

$$D_{\text{mean,OARs}} = \frac{1}{n_{\text{OARs}}} \sum_{\text{OAR}} \frac{|D_{\text{mean,OAR,CT}} - D_{\text{mean,OAR,sCT}}| + \epsilon}{D_{\text{mean,OAR,CT}} + \epsilon}$$

Where $\epsilon = 1 \times 10^{-12}$ is added to avoid division by zero and n_{OARs} represents the number of OARs included in the analysis. The above computed DVH parameters are summed up to generate a single, overall DVH metric for each patient. This thorough DVH parameters assessment ensures that the sCT images are clinically viable by closely matching the dose distribution of the ground-truth CT.

A lower difference in parameters such as D98 and V95% for the PTV indicates that the sCT closely matches the intended treatment plan by delivering the prescribed dose to the target volume. A high D98 or V95% discrepancy, on the other hand, suggests that the sCT might underdose or overdose portions of the PTV, which could compromise treatment effectiveness and possibly increase the risk of radiation-induced side effects. For the OARs, metrics such as D2 and Dmean reflect how well the sCT spares healthy tissues from excessive radiation exposure. A low value in D2 and Dmean ensures that the sCT is successfully limiting the dose to critical organs, thus reducing the likelihood of complications. However, larger differences in these parameters could indicate potential risks to healthy tissue, which would necessitate further refinement of the sCT generation process.

5. GPR

The GPR metric evaluates the agreement between two sets of data. The gamma index is calculated for each point in the image, comparing the values at corresponding locations. The metric is defined by two key criteria points; Distance to Agreement (DTA) and DD. The approach suggested by Low et al. (1998) recommends using a DTA of 2mm and a DD of 2%. Additionally, an alternative set of thresholds, with 3mm for DTA and 3% for DD, was applied based to assess the metric on less strict criteria. A lower GPR under stricter criteria may highlight minor inaccuracies in the sCT, whereas a higher GPR with more lenient criteria can still indicate acceptable clinical performance. Therefore, a careful balance of these values is necessary to ensure both precision and clinical applicability. A point is considered a pass if both criteria are satisfied, and a higher percentage value indicates better agreement between the paired images. The GPR metric is evaluated within regions receiving doses greater than 10% of the prescribed dose.

The interpretation of the Gamma Passing Rate (GPR) will determine the level of agreement between the sCT and CT dose distributions. A higher GPR percentage indicates that the majority of the points within the dose distribution meet the established criteria for spatial and dose agreement, reflecting a strong correlation between the planned and delivered doses. Conversely, a lower GPR value suggests discrepancies between the dose distributions, particularly in regions where spatial alignment or dose delivery may differ. This may indicate that the sCT images require further refinement to ensure accurate dose replication, especially in high-dose regions.

3.7.3 | Baseline Algorithms

Baseline models are relevant for establishing a point of comparison in the evaluation of more advanced sCT generation algorithms. They help demonstrate the extent to which the advanced algorithm enhances image quality and clinical applicability compared to simpler, bulk-assignment approaches. These models provide a simple yet effective reference that allows researchers to quantify the improvements made by more complex methods. The metrics obtained in this study are compared to the two baseline models, tabulated in tables 3.3 and 3.4, published by Huijben et al. (2024) in the SynthRAD Challenge report was used as a standard.

Team	MAE (HU, ↓)	PSNR (dB, ↑)	SSIM (↑)
Water Baseline	332.93 ± 89.53	17.95 ± 1.73	0.552 ± 0.127
Stratified Baseline	69.45 ± 16.33	28.42 ± 1.92	0.854 ± 0.027

Table 3.3: Baseline Model Image Metrics

Team	DVH (↓)	2%/2mm (↑)
Water Baseline	0.0972 ± 0.1057	96.70 ± 5.36
Stratified Baseline	0.0347 ± 0.0504	98.21 ± 5.17

Table 3.4: Baseline Model Dose Metrics (Photon)

3.7.3.1 | Water Baseline Model

The water baseline model offers a basic approach for sCT by assigning uniform HU values. All voxels within the body contour are assigned a value of 0 HU, representing water, while voxels outside the contour (air) are assigned -1000 HU. This model, although simplistic, serves as a lower-bound reference for evaluating the generated sCT images. It allows for a comparison that highlights the improvements achieved by more sophisticated algorithms.

3.7.3.2 | Stratified Baseline Model

The stratified baseline model, introduced by Maspero et al. (2017), refines the bulk-assignment approach by categorizing voxels based on specific HU ranges, thereby providing a more anatomically accurate representation. The ground-truth CT is segmented into five categories:

- **Air**: $(-\infty, -210)$ HU $\rightarrow$ -968 HU
- **Adipose Tissue**: $[-210, -20)$ HU $\rightarrow$ -86 HU
- **Soft Tissue**: $[-20, 120)$ HU $\rightarrow$ 42 HU
- **Bone Marrow**: $[120, 555)$ HU $\rightarrow$ 198 HU
- **Cortical Bone**: $[555, \infty)$ HU $\rightarrow$ 949 HU

This stratified approach provides a more realistic comparison for the synthetic CT images by better approximating the varying densities of different anatomical structures. To further refine the bone segmentation, a binary hole-filling algorithm was applied, ensuring that no air or soft tissue voxels were erroneously assigned within bone structures.

3.8 | Ethical Considerations

This study adhered to strict ethical guidelines to ensure the protection of patient data and compliance with relevant regulations in medical research.

Obtaining medical images from a hospital, such as paired MRI and CT scans, involves several ethical considerations and protocols to ensure patient privacy, confidentiality, and compliance with relevant regulations. The following must be adhered to when opting for this route.

Given the study timeline, the medical images required were retrieved from patients who were already deceased. Ethical approval was sought and granted for the retrospective use of this data. This ensured that no living patients were involved in the study, minimising potential ethical concerns associated with ongoing patient care. Since the data used in this study came from deceased patients, the standard informed consent process was modified. Ethical approval from the hospital's review board allowed for the use of these retrospective datasets without direct patient or family consent.

All patient data used in the study was pseudoanonymised to protect the identity and personal information of the individuals involved. Identifying details such as patient names, ID numbers, and other personal identifiers were removed before data processing. This step was carried out by the intermediary of the study to maintain patient confidentiality in compliance with the GDPR and other relevant local and European data privacy laws.

The study was reviewed and approved by the Data Protection Officer (DPO) at Mater Dei Hospital, Malta, which oversees research related to data or medical devices at the institution. The study is also approved by FREC to ensure that the University of Malta's Research Code of Practice is adhered to. The ethical approval ensured that the research complied with local regulations and international standards for ethical medical research.

To safeguard patient data utilised in this study, secure storage and processing measures were employed. All data were stored on password-protected storage devices, and access was limited to authorised personnel only.

3.9 | Limitations of the Research Methodology

Despite the thorough design and execution of this study, several limitations in the research methodology should be acknowledged. These limitations may have affected the outcomes of the study and should be considered when interpreting the results.

- **Retrospective Data Collection**
  The study utilised retrospective data from patients who were already deceased. While this allowed for ethical use without direct patient intervention, it introduced some limitations in terms of controlling the quality and consistency of the imaging data. Although the MRI and CT scans were not originally acquired for this study, they were collected for the same purpose, for RT treatment planning. This being said there may have been variations in scan parameters that could influence the accuracy of the generated sCT images.

- **Dependence on Data Quality**
  One of the primary limitations stems from the reliance on the quality of MRI and CT scan data. The efficacy of sCT generation is heavily contingent upon the resolution, contrast, and overall quality of the input MRI scans. Variabilities in scan quality, attributable to differences in imaging protocols, scanner hardware, and patient movement, can introduce inconsistencies in the sCT images generated. While efforts were made to standardise and preprocess the images, the potential for residual variability remains, which could affect the generalisability of the sCT generation technique.

- **Training Data for the AI Model**
  The AI model used for sCTgeneration in this study was trained using a dataset that was divided into 66 images for training, 10 for validation, and 10 for testing. This relatively small dataset, compared to the SynthRAD2023 dataset, which included data from 1,080 patients undergoing radiotherapy treatment, may limit the generalisability and performance of the sCT generation model. The smaller dataset may have introduced variability or errors in the results, as deep learning models typically benefit from larger datasets.

- **Differences in Training Datasets and Baseline Reference**
  Despite the significant difference in the size of the training datasets, this study will still refer to the baseline algorithm outlined in the SynthRAD 2023 competition. While this provides valuable context for comparison, the discrepancy between the datasets introduces limitations in the accuracy of direct comparisons. The larger

training set in the SynthRAD challenge could lead to more robust results, and referencing this baseline may not fully reflect the performance limitations introduced by the smaller dataset used in this study.

3.10 | Conclusion

This chapter laid out the study's research methodology. The results will be presented, analysed, and discussed in chapter four.

4

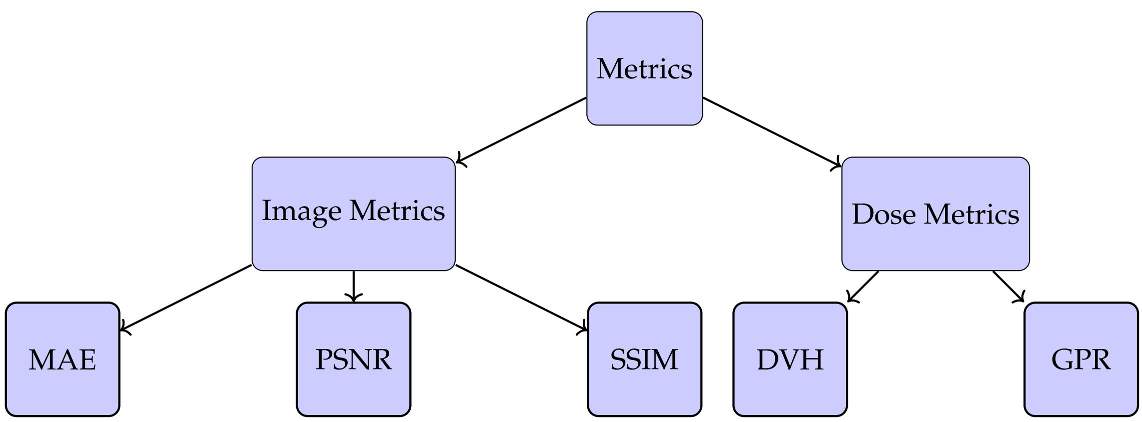
Results

4.1 | Introduction

This chapter presents the results of the image and dose metrics collected from the groundtruth CT and the generated sCT images. The full set of results, the Python code created to calculate specific metrics and the uncertainties can be located in Appendix A, in sections A.1.4 and A.1.5.

4.2 | Data

This section presents the results and are organised into two primary metric groups: image metrics and dose metrics, as illustated in the hierarchical diagram below. The tables below summarise the results obtained for each metric, offering a comprehensive comparison between the sCT and CT images to evaluate the accuracy and clinical viability of the synthetic images for RT planning.



4.2.1 | Image Metrics

4.2.1.1 | MAE

Patient ID	**Original CT Mean (HU)**	**SynCT Mean (HU)**	**Difference Mean (HU)**
Brain1	134	67	67
Brain2	164	76	88
Brain3	143	121	22
Brain4	153	142	11
Brain5	143	76	67
Brain5	168	155	13
Brain7	139	128	11
Brain8	177	166	11
Brain9	179	59	120
Brain10	181	198	17
Brain11	245	99	146
Brain12	159	57	102
Brain13	126	102	24
Brain14	123	123	0
Brain15	177	169	8
Brain16	126	2	124
Average			**51.94** $\pm$12.42

Table 4.1: The mean intensity values of sCT and groundtruth CT to calculate the averaged MAE metric for the skin contour

Patient ID	Original CT Mean (HU)	SynCT Mean (HU)	Difference Mean (HU)
Brain1	32	2	30
Brain2	34	1	33
Brain3	33	4	29
Brain4	45	14	31
Brain5	39	5	34
Brain6	35	6	29
Brain7	40	8	32
Brain8	37	9	28
Brain9	36	2	34
Brain10	41	6	35
Brain11	46	1	45
Brain12	37	1	36
Brain13	40	6	34
Brain14	41	5	36
Brain15	35	10	25
Brain16	126	2	124
Average			**38.44** ± 5.81

Table 4.2: The mean intensity values of sCT and groundtruth CT to calculate the averaged MAE metric for the brain contour

4.2.1.2 | PSNR

Patient ID	PSNR (dB)
Brain1	20.92
Brain2	21.51
Brain3	23.13
Brain4	22.85
Brain5	20.09
Brain6	22.88
Brain7	22.74
Brain8	22.14
Brain9	20.07
Brain10	23.28
Brain11	19.39
Brain12	20.24
Brain13	22.21
Brain14	22.74
Brain15	22.06
Brain16	19.98
Average	**21.64** ± 0.33

Table 4.3: The PSNR values of the sCT and groundtruth CT to calculate the average PSNR metric value.

4.2.1.3 | SSIM

Patient ID	**SSIM**
Brain1	0.774
Brain2	0.767
Brain3	0.801
Brain4	0.792
Brain5	0.733
Brain6	0.800
Brain7	0.796
Brain8	0.792
Brain9	0.798
Brain10	0.799
Brain11	0.741
Brain12	0.785
Brain13	0.720
Brain14	0.797
Brain15	0.788
Brain16	0.788
Average	**0.779** $\pm$0.006

Table 4.4: The SSIM values of the sCT and groundtruth CT to calculate the average SSIM metric value.

4.2.2 | Dose Metrics

4.2.2.1 | DVH

Patient ID	V95 PTV, CT	D98 PTV, CT
Brain1	0.0047	0.0635
Brain2	0.0989	0.0018
Brain3	0.2477	0.0347
Brain4	0.0838	0.0201
Brain5	0.0127	0.0084
Brain6	0.0358	0.0010
Brain7	0.0912	0.0157
Brain8	0.0367	0.0056
Brain9	0.0358	0.0025
Brain10	0.0341	0.0012
Brain11	0.0037	0.0262
Brain12	0.0061	0.0135
Brain13	0.0139	0.0039
Brain14	0.0641	0.0136
Brain15	0.0300	0.0010
Brain16	0.0353	0.0137
Average	**0.0522** $\pm$0.0150	**0.0141** $\pm$0.0041

Table 4.5: The V95PTV,CT and D98PTV,CT values of the sCT and groundtruth CT to calculate the average DVH metric value.

	D2 OARs			
Patient ID	**brainstem**	**optic nerves**	**orbits**	**brain**
Brain1	0.0493	0.5893	0.1511	0.0607
Brain2	0.0020	0.0215	0.1977	0.0048
Brain3	0.0245	0.2002	0.3632	0.0269
Brain4	0.0108	0.2842	0.6159	0.0194
Brain5	0.0025	0.0191	0.1117	0.0089
Brain6	0.0273	0.0002	0.1016	0.0127
Brain7	0.0525	0.1101	0.0495	0.0175
Brain8	0.0061	0.0011	0.3260	0.0000
Brain9	0.0006	0.0097	0.0388	0.0045
Brain10	0.0013	0.0940	0.0664	0.0094
Brain11	0.0069	0.0164	0.0434	0.0105
Brain12	0.0089	0.0478	0.7051	0.0065
Brain13	0.0105	0.0075	0.1044	0.0044
Brain14	0.0121	0.0175	0.2969	0.4232
Brain15	0.1274	0.0298	0.4909	0.0222
Brain16	0.0040	0.0012	0.2304	0.0104
Average	**0.0217** $\pm$0.0081	**0.0906** $\pm$0.0388	**0.2433** $\pm$0.0523	**0.0388** $\pm$0.0259

Table 4.6: The D2 OAR values of the sCT and groundtruth CT to calculate the average DVH metric value.

	Dmean OARs			
Patient ID	**brainstem**	**optic nerves**	**orbits**	**brain**
Brain1	0.0554	0.3124	0.0434	0.0644
Brain2	0.0809	0.0554	0.0372	0.0124
Brain3	0.1264	0.0485	0.1691	0.0068
Brain4	0.0168	0.2940	0.2258	0.0137
Brain5	0.0136	0.0881	0.1960	0.0142
Brain6	0.0015	0.0289	0.0417	0.0304
Brain7	0.3980	0.2144	0.0732	0.0613
Brain8	0.0073	0.0892	0.0659	0.0138
Brain9	0.0083	0.1355	0.2473	0.0004
Brain10	0.1202	0.0887	0.1870	0.0057
Brain11	0.0119	0.0211	0.0200	0.0108
Brain12	0.0151	0.1738	0.5931	0.0128
Brain13	0.0215	0.0070	0.0036	0.0100
Brain14	0.0014	0.0009	0.0481	0.0002
Brain15	0.0759	0.0665	0.7011	0.0278
Brain16	0.1133	0.0877	0.1448	0.0212
Average	**0.0667** $\pm$0.0248	**0.1070** $\pm$0.0239	**0.1748** $\pm$0.0503	**0.0191** $\pm$0.0047

Table 4.7: The Dmean OAR values of the sCT and groundtruth CT to calculate the average DVH metric value.

V95 PTV,CT	**D98 PTV,CT**	**D2 OAR**	**Dmean OAR**	**DVH**
0.0522	0.0142	0.0986	0.0919	**0.2568** $\pm$0.0195

Table 4.8: Final average values of all four DVH parameters and the final summed DVH metric value.

4.2.2.2 | GPR

Patient ID	Criteria 2mm, 2%	Criteria 3mm, 3%
Brain 1	49.60%	60.40%
Brain 2	79.10%	88.00%
Brain 3	65.40%	77.10%
Brain 4	92.30%	95.60%
Brain 5	81.70%	92.40%
Brain 6	94.10%	99.10%
Brain 7	64.30%	80.40%
Brain 8	92.70%	95.80%
Brain 9	87.70%	93.10%
Brain 10	79.70%	90.40%
Brain 11	93.30%	98.20%
Brain 12	89.50%	95.10%
Brain 13	97.10%	99.60%
Brain 14	94.30%	98.40%
Brain 15	93.60%	97.20%
Brain 16	92.90%	96.20%
Average	**84.21%** ± 3.41	**91.06%** ± 2.60

Table 4.9: Final averaged percentages for two different sets of criteria and the final averaged GPR metric value.

4.3 | Data Analysis and results

In this section, the results obtained after testing the algorithm on a Maltese cohort in the context of RT treatment planning are analysed. These results, which are comprehensively tabulated in Section 4.2, are compared to the water and stratified baseline models published by Huijben et al. (2024) in SynthRAD 2023. SynthRAD 2023, published this year, serves as a valuable reference, facilitating the investigation and benchmarking of sCT generation techniques. The necessary uncertainties were computed using a python script located in Appendix A, section A.1.5, to ensure the reliability and validity of the reported metrics. The inclusion of uncertainties provides a more comprehensive understanding of the algorithm's performance and allows for more accurate benchmarking against baseline models and competing methods.

The first set of metrics analysed are the image metrics: MAE, PSNR, and SSIM. These metrics are commonly used in medical image synthesis to assess the performance of algorithms in generating synthetic medical images (Spadea et al., 2021).

Following the analysis of image metrics, the dose metrics, DVH and GPR measures will be evaluated. These will provide insights into the accuracy of RT treatment plans based on sCT images.

The MAE was calculated to evaluate the accuracy of the sCT generation. Both the skin and brain contour values were published in tables 4.1 and 4.2 respectively, however, only the skin contour is valid for comparison purposes. The skin contour encloses all contents within, covering the same region as the metric collected during the SynthRAD 2023 competition.

The MAE value obtained for the skin contour was 51.94 ± 12.416 HU. The results obtained by the participating teams in SynthRAD 2023 varied from 58.83 ± 13.41 HU to 117.88 ± 45.08 HU, with a lower MAE indicating better agreement between the synthetic and ground-truth CTs. This means that the sCT produced by the tested algorithm replicates the intensity values of the ground-truth CT to an acceptable degree, surpassing both water and stratified baseline models outlined in table 3.3.

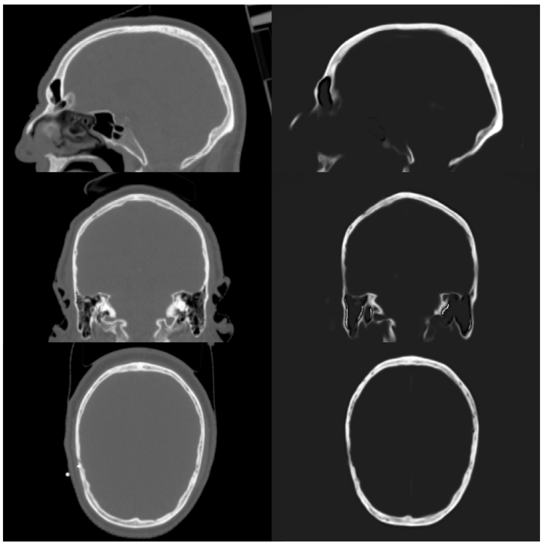


Figure 4.1: Brain14

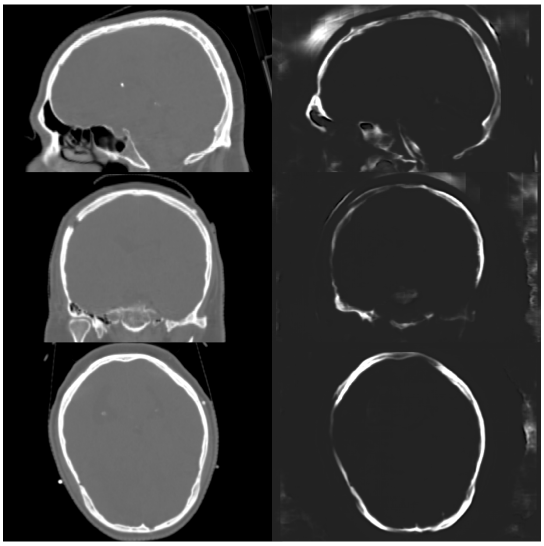


Figure 4.2: Brain11

Figure 4.3: This figure presents a side-by-side comparison of the cases with the lowest and highest MAE values. Each image illustrates the original CT on the left and the corresponding generated sCT on the right, highlighting the variation in accuracy between the two cases.

The patients with the best-performing results include Brain14, Brain15, Brain4, Brain7, and Brain8, all of whom displayed a low MAE, ranging between 0 and 11 HU. These results indicate a strong agreement between the intensity values of the original and synthetic CT images, suggesting that the algorithm was particularly effective at replicating the anatomical structures in these cases. Specifically, observed in figure 4.1, Brain14 exhibited the lowest MAE of 0 HU, demonstrating exceptional performance in replicating the skin contour accurately.

On the other hand, outliers like Brain9, Brain11, Brain12, and Brain16, with MAE values of 120, 146, 102, and 124 HU, respectively, highlight the algorithm's struggle in certain cases. These higher MAE values could be indicative of more complex anatomical features or inconsistencies in image registration for these patients, however, this is unlikely the case, based on the side-by-side comparison observed in figure 4.3. Brain11, observed in figure 4.2, showed the highest deviation with an MAE of 146 HU and is visually observed to be of lesser quality when compared to the lowest achieved MAE of Brain14 in figure 4.1. This suggests that the synthetic CT had significant difficulty in accurately capturing the intensity values for this patient, indicating lack of generalisability.

This range of results, with an average MAE of 51.94 ± 12.42 HU, indicates that while the algorithm performs well in many cases, further refinement is needed to improve consistency, particularly for complex cases where intensity values significantly deviate from the ground-truth CT. These findings underscore the importance of continuous optimization, especially for challenging anatomical regions or patients with atypical structures.

The average difference in intensity values between the sCT and ground-truth CT for the brain contour across all patients was 38.44 ±5.81 HU, which is lower than the value obtained for the skin contour, 51.94 ±12.42 HU. This suggests that the deviations between the sCT and ground-truth CT are smaller in the brain region, indicating a better agreement between the synthetic and actual CT values for the brain contour.

Though the brain contour is not directly comparable to the SynthRAD 2023 dataset due to differences in the volumes considered, these results provide additional insights into the performance of the algorithm for more complex internal structures. The larger difference observed in some cases, such as Brain16 having a difference mean of 124 HU in both volumes, points to potential areas where the algorithm may require further refinement to better capture the intensity values in these regions.

The PSNR values obtained for the sCT and ground-truth CT are presented in Table 4.3 with an average value of 21.64 ±0.33 dB. This result is compared against the baseline models from SynthRAD 2023. The water baseline model scored 17.95 ± 1.73 dB, while the stratified baseline model achieved a higher PSNR of 28.42 ± 1.92 dB.

The participating teams in SynthRAD 2023 had scores ranging from 23.69 ± 0.94 dB to 29.61 ± 1.79 dB. It is evident that while the average PSNR value of 21.64 ±0.33 dB falls short of the highest-performing models in SynthRAD 2023, it still represents a notable performance compared to the water baseline model. The highest-performing brain scan in this study was Brain10, which achieved a PSNR of 23.28 dB, indicating the strongest agreement between the synthetic and ground-truth CT images. On the other hand, Brain11 had the lowest PSNR value of 19.39 dB, suggesting the least accurate reconstruction in terms of image quality for that specific case.

This highlights the need for refining the algorithm to better accommodate a wider range of anatomical variations, ensuring more consistent image quality across different cases. Brain11, which scored the lowest in both MAE and PSNR, reinforces the observation that the algorithm had difficulty generalising to this specific case, resulting in a lower-quality sCT.

The SSIM values obtained for each patient are presented in Table 4.4, with an average SSIM score of 0.779 $\pm$0.006. This result falls between the water and stratified baseline models from SynthRAD 2023, which scored 0.552 ± 0.127 and 0.854 ± 0.027, respectively.

Participants in the SynthRAD 2023 challenge achieved SSIM values ranging from 0.756 ± 0.028 to 0.876 ± 0.030. The highest SSIM values achieved in this study were for Brain3 and Brain6, with scores of 0.801 and 0.800, respectively. These results show that while the algorithm performs reasonably well, reaching SSIM scores close to the upper bounds achieved by the SynthRAD 2023 participants, there is still some variability across different patients, with lower values observed for Brain13 having an SSIM value of 0.720.

Overall, the SSIM result suggest that the sCTs generated by the algorithm preserves a significant amount of structural similarity, but further improvements could help close the gap between this study's results and the higher-performing models in the competition.

Moving on to the dose metrics, the DVH value obtained in this study is 0.2568 ±0.0195, as shown in Table 4.8. This value is notably higher compared to the baseline models and the results obtained by participants in the SynthRAD 2023 competition. The water baseline model scored 0.0972 ± 0.1057, while the stratified baseline model achieved a lower DVH of 0.0347 ± 0.0504. Results from the participating teams ranged from 0.0265 ± 0.0382 to 0.0972 ± 0.1057.

One possible reason for the higher DVH result may lie in the limitations of the sCT images, particularly in their ability to capture the fine anatomical details necessary for precise dose calculations. The discrepancies observed between the values for D2 OARs and Dmean OARs, observed in tables 4.6 and 4.7, which were significantly higher compared to the other DVH parameters, V95% PTV,CT and D98 PTV,CT, suggest that the sCT may not possess the high level of anatomical accuracy required to match the dosimetric plans generated from the original CT images. This could be especially true for sensitive and complex structures like the optic nerves and orbits, which overall resulted in the highest values.

The inability of the sCT to faithfully replicate these intricate regions likely impacts the algorithm's performance in delivering accurate dose metrics, emphasising the need for further refinement. Improving the structural fidelity of the sCT images, particularly in these complex areas, is critical to achieving better dosimetric outcomes.

The final averaged percentages representing the GPR values for the different criteria considered are shown in Table 4.9. For the 2mm/2% criteria, the average GPR was 84.21 ±3.41%, while for the 3mm/3% criteria, the average was 91.06 ±2.60%. The results obtained following 2mm/2% criteria are comparable to the baseline models and participants' results from SynthRAD 2023. The water baseline model achieved a GPR of 96.70% ± 5.36, while the stratified baseline model scored 98.21% ± 5.17. Participants' results ranged from 96.42% ± 5.00 to 99% ± 1.98. The less strict criteria considering 3mm/3% passing criteria yielded higher GPR percentages, reflecting the current practices in RT departments, and suggesting that the algorithm's performance improves under these more relaxed criteria.

Among the best performers for both sets of criteria are Brain13, which achieved a GPR of 97.10% for the 2mm/2% criteria and 99.60% for the 3mm/3% criteria, and Brain6 follows, with scores of 94.10% and 99.10%, respectively. These cases demonstrate excellent agreement between the sCT and CT images, indicating that the algorithm performed well in replicating the dose distributions in these patients.

Conversely, Brain1 and Brain7 were the worst performers, particularly for the 2mm/2% criteria. Brain1 had the lowest GPR value of 49.60%, and Brain7 followed with a score of 64.30%. Even under the more lenient 3mm/3% criteria, Brain1 still scored the lowest at 60.40%, indicating significant discrepancies between the sCT and CT for these cases. These outliers highlight cases where the algorithm struggles.

In summary, the evaluation of the algorithm's performance across both image and dose metrics has provided valuable insights. The image similarity metrics, MAE, PSNR, and SSIM, showed reasonable performance, with results closely aligning with the baseline models and some of the participant results from SynthRAD 2023.

On the other hand, the dose metrics DVH and GPR revealed larger deviations, especially in the case of the DVH, where the values obtained were significantly higher than those reported in SynthRAD 2023. The performance under both conditions for the GPR metric fell below the both water and stratified baseline model values.

Overall, better agreement and performance were observed in the image metrics compared to the dose metrics, suggesting that the algorithm excels in replicating structural and intensity values but may require further refinement to achieve higher accuracy in dosimetric consistency for RT treatment planning.

4.4 | Conclusion

This chapter presented the results that were collected in the study together with a brief summary of how the data was analysed. The data will be discussed in Chapter 5.

5

Discussion

5.1 | Introduction

The discussion chapter provides an interpretation and analysis of the results presented in the thesis, including comparisons to relevant literature and suggestions for future research. It addresses the strengths and limitations of the study. The chapter offers a critical reflection on the research process and the implications of the results, contributing to the advancement of knowledge in the field.

5.2 | Discussion

The primary aim of this study was to evaluate the effectiveness of an algorithm designed to generate MR-based sCT images and assess their potential application in RT treatment planning. By focusing on an underrepresented cohort of Maltese patients diagnosed with brain cancer, this study not only explores the technical feasibility of the sCT generation process but also its clinical relevance in a real-world setting where patient-specific data can often be limited. The dual emphasis on both image quality and dosimetric accuracy metrics offers a comprehensive evaluation of the algorithm's capabilities, setting the stage for a deeper analysis of its strengths and limitations within the context of modern RT workflows.

While the field of sCT research often relies on image similarity metrics to gauge performance, such as MAE, PSNR, and SSIM, these metrics, while valuable, do not provide the full picture when it comes to RT planning. The ultimate measure of success for sCT in this context lies in its ability to replicate dose distributions accurately, a far more critical factor for ensuring effective and safe radiation treatment. Indeed, as highlighted in the literature, there is often little to no significant correlation between image similarity metrics and dose accuracy (Kieselmann et al., 2018). This disconnect was clearly demonstrated in the present study, where strong performance in image metrics did not consistently result in accurate dose distributions for the Maltese cohort. Despite the algorithm's ability to generate visually accurate sCT images, the disparities in dose metrics reveal that good image quality alone does not guarantee dosimetric precision.

Across the image similarity metrics, the results indicate that the algorithm performs reasonably well, particularly in replicating structural and intensity values. Despite the model being trained on a small data set in comparison to the SynthRAD algorithms the MAE value of 51.94 ± 12.42 for the skin contour exhibits a lower MAE value when compared to values obtained by the participants of the challenge, indicating better agreement between the groundtruth and synthetic CT. This demonstrates that the algorithm can produce sCT images with a high degree of intensity accuracy compared to groundtruth CT. The PSNR value of 21.64 ± 0.33, while lower than the highest-performing models, suggests that image quality is within a reasonable range, though improvements are necessary to match the top benchmarks. Similarly, the SSIM value of 0.779 ± 0.006 shows that the algorithm preserves a substantial amount of structural similarity. Although the SSIM score is lower than that of the stratified baseline model and the highest scores achieved by SynthRAD 2023 participants, it falls within an acceptable range, indicating that the algorithm captures structural information well, though there remains room for improvement in consistency across different patients.

However, the dose metrics, particularly DVH and GPR, revealed significant deviations. The DVH value of 0.2568 ± 0.0195 was considerably higher than the baselines and participant results in SynthRAD 2023, indicating potential challenges in accurately replicating dose distributions. The GPR values, $84.21 \pm 3.41\%$ for 2mm/2% and $91.06 \pm 2.60\%$ for 3mm/3%, were also lower than those of the baseline models, though the algorithm performed better under the more relaxed 3mm/3% criteria. Overall, while the algorithm shows promise in generating accurate sCT images for RT treatment planning, further refinements are required, particularly in the area of dosimetric accuracy.

In this study, the treatment plans utilised were designed for photon therapy using a 6flattening filter-free (FFF) beam, which is commonly employed in stereotactic RT due to its higher dose rate and reduced treatment time (Yan et al., 2016). The Monaco TPS uses the Monte Carlo method for dose calculations, it is known for its high accuracy in simulating particle interactions and dose deposition within heterogeneous tissues. This algorithm is particularly beneficial in capturing the complex anatomical details required for accurate dosimetric calculations, as it accounts for scattering effects and tissue density variations more precisely than conventional algorithms (Tuğrul, 2021).

The prescription doses (Rx), observed in table 5.1, varied across the treatment plans between 30Gy and 60Gy, depending on the clinical needs of each patient. This range reflects the typical doses required for treating brain tumours, where higher doses may be necessary to achieve local control of more aggressive tumour types, while lower doses are prescribed to minimise damage to surrounding healthy tissues and critical OARs.

Patient	**Rx (Gy)**
Brain1	30
Brain2	60
Brain3	40.05
Brain4	60
Brain5	30
Brain6	54
Brain7	30
Brain8	60
Brain9	30
Brain10	60
Brain11	30
Brain12	60
Brain13	30
Brain14	60
Brain15	60
Brain16	30

Table 5.1: Patient Labels and Corresponding Prescription Doses (Rx in Gy)

These parameters highlight the importance of ensuring that the sCT images generated from MRI are able to replicate the dosimetric accuracy required for clinical applications. Ensuring that the sCT images can reliably support such treatment plans is essential for their integration into routine clinical workflows.

In relation to the other study conducted using this algorithm by Liu et al. (2022), while both studies employed the algorithm in different contexts, the only directly comparable metric between them is the MAE. This comparison is waranted as the algorithm used in both studies remained unchanged, allowing for a direct evaluation of its performance across different datasets and clinical contexts. The MAE values obtained for the Maltese cohort were 51.94 HU for the skin contour as seen in table 4.1, significantly outperforming the MAE of 190.94 ± 22.40 HU reported for the skull region in their sCT generation for Transcranial MRI-guided Focused Ultrasound interventions. One potential reason for this discrepancy is the possible difference in anatomical regions considered or the different pre-processing steps taken, such as, skull extraction and resampling. The lower MAE achieved in this study highlights the strong performance of the algorithm on the Maltese dataset. The study by Liu et al. (2022) does not explicitly mention the nationality or geographical origin of the dataset, however, the paper states that the data was collected from Vanderbilt University, suggesting that the dataset is likely to be derived from a population in the United States. This raises questions about potential biases in different populations, such as the Maltese cohort, and its effects on the quality of sCT achieved. This underscores the need for further investigation into how algorithms perform across diverse populations to ensure unbiased and equitable outcomes.

The findings of this study suggest that the algorithm demonstrates strong potential for generating sCT images from MRI, particularly in replicating structural and intensity values, as indicated by the image metrics. These results show that the algorithm can produce images of acceptable quality, which is a critical prerequisite for its application in RT treatment planning. However, discrepancies observed in the dose metrics highlight the need for further research to refine the algorithm and improve its accuracy in replicating dose distributions, especially in areas involving complex anatomical structures

In the context of radiation protection, this emerging technology has the potential to significantly enhance patient care by reducing the number of required scans, minimising patient fatigue, lowering radiation exposure and streamlining the RT treatment planning process. These benefits further underscore the critical importance of continuously benchmarking new algorithms against established models, such as those in SynthRAD 2023, to ensure they meet the high standards needed for safe and effective clinical implementation.

This study has several implications for RT treatment planning, particularly when dealing with minority populations like the Maltese cohort. Minority populations face a unique challenge in clinical settings where there is limited access to patient-specific data. This means there are fewer patients from whom data can be collected. Consequently, the availability patient data is limited, which can make it harder to develop highly personalised treatment plans or train algorithms on diverse data, locally. The results obtained highlight the need for large datasets and suggest the need for further investigation.

A number of methodological issues emerged during the course of the study, likely influencing the results obtained. These challenges relate to both the dataset used and the trained algorithm used.

One of the primary limitations was the restricted size of the dataset, which included only 66 training images, in contrast to the SynthRAD 2023 dataset, which contained data from 1,080 patients. This discrepancy limits the algorithm's ability to generalise effectively. While the algorithm performed well on the training data, it struggled when applied to new or unseen data. In this study, the smaller dataset likely contributed to the slightly lower PSNR values observed, as the algorithm may not have been exposed to a sufficiently wide range of anatomical and intensity variations present in larger cohorts. Additionally, the limited dataset may have impacted the dose metric results, as the algorithm's capacity to generate sCT images with accurate dose distributions could have been constrained by the lack of diverse training examples.

Another significant methodological issue was the high DVH value computed. The higher DVH values observed for the oribts and the optic nerves in this study are indicative of the algorithm's limitations in generating sCT images with sufficient anatomical detail for accurate dose planning. The complexity of the optic nerves and orbits, which require a high degree of precision in treatment planning, suggests that the sCT images lacked the necessary structural detail to replicate the fine variations seen in the original CT images. This deficiency likely contributed to the observed discrepancies in dose distribution, emphasizing the need for further refinement of the algorithm to ensure it can achieve the same level of dosimetric accuracy as traditional CT-based plans.

Variability in the algorithm's performance across different patients was another notable issue. This variability was particularly evident in the MAE results, as presented in Tables 4.2 and 4.1. For instance, Brain16 exhibited significantly higher error values, with an MAE difference of 124, while other patients scored as low as zero. This suggests that the algorithm may have difficulty handling more complex or outlier cases where the anatomical or intensity characteristics of the patient differ markedly from the majority of the training data. This limitation could be eliminated with standardised protocols for medical datasets used for research, and more specifically, algorithm development and testing.

The use of a 3D patch-based cGAN in this study introduced several implications that should be considered. First, the patch-based approach, while effective in reducing memory usage, can potentially result in inconsistencies at the boundaries between adjacent patches. This is particularly relevant when generating sCT images from MRI data, where anatomical structures may span across multiple patches. Although overlapping patches were used to address this issue, some minor inconsistencies could still arise.

Secondly, processing 3D volumetric data adds to the computational complexity, requiring substantial resources and resulting in longer training times, with an average generation time of 4909.46 ± 383.97 seconds. Additionally, the memory requirements are higher compared to 2D approaches. Nonetheless, the ability to capture spatial relationships in three dimensions is a non-negotiable for accurate image synthesis in medical applications.

Additionally, it is important to acknowledge that the algorithm used in this study was originally designed for Transcranial MR Imaging–Guided Focused Ultrasound interventions, rather than for RT treatment planning. This may explain the strong performance in image similarity metrics, as high image quality and structural definition are critical for ultrasound interventions. This may account for the strong performance in image similarity metrics, as high image quality and structural definition are critical in ultrasound procedures. Furthermore, by conditioning the generation of sCT images on MRI scans, the model was able to produce outputs that closely aligned with the anatomical structures captured in the MRI data, thereby improving the accuracy of the generated sCT images.

A significant strength of this study is its focus on a minority population, specifically a cohort of Maltese patients with brain cancer. By examining the performance of the algorithm in a smaller, less-represented population, the study contributes valuable insights into how sCT generation algorithms perform in a population with limited patient-specific data. These findings may inform future research and clinical practice, particularly for smaller or minority populations where data availability is restricted.

Another strength is the use of SynthRAD 2023 as a benchmarking reference. Benchmarking against such a recent and validated challenge results allows for a clear comparison of the algorithm's performance relative to established models. This approach not only highlights areas where the algorithm performs well, such as image similarity metrics, but also helps identify aspects that require improvement, such as dose metric accuracy. This comparison with SynthRAD 2023 enhances the study's credibility and provides a framework for continuous improvement of the algorithm.

This study effectively addressed its primary objectives, providing valuable insights into the algorithm's performance in generating sCT images from MRI, specifically within the context of a minority population, such as the Maltese cohort. One of the main objectives was to evaluate the quality of sCT generation achieved by the algorithm. Although the results obtained suggests that although the algorithm performs well in generating images that replicate anatomical structures, further refinement is necessary to improve its accuracy in replicating dose distributions. Meeting this objective more fully would require additional work to enhance the algorithm's ability to predict accurate dose metrics, especially in critical regions such as OARs.

Another key objective of the study was to address the registration errors that commonly arise from dual-simulation processes involving both CT and MRI scans. The study demonstrated that sCT generation from MRI can reduce these registration errors by mapping from image-to-image translations, thereby improving alignment and consistency in RT treatment planning.

Despite these successes, the study also found areas that need improvement. The study did not achieve outcomes comparable to the best-performing teams in SynthRAD 2023, particularly in terms of dose accuracy. While the image similarity metrics were strong, with the final MAE value surpassing the best-performing team in SynthRad23, the discrepancies in dose metrics indicate that further refinement is needed before it can match top-tier models. The algorithm struggles to handle complex anatomical changes, particularly in regions with detailed structures like the orbits and optic nerves. Further improvements are needed to ensure the algorithm consistently captures these variations and provides both accurate images and precise dose calculations.

With continued development and optimisation, particularly focusing on enhancing dose accuracy and improving the handling of diverse anatomical structures, MRI-based sCT generation could become an integral part of RT treatment planning. This technology has the potential to eliminate registration error, reduce patient radiation exposure, streamline clinical workflows and offer high quality sCT for treatment planning. This area of research shows great potential, and with continued development, MRI-based sCT generation could be transformative in the future of RT treatment planning.

5.3 | Conclusion

This chapter discussed the main results of this research whilst making reference to published literature. The significance of this study was highlighted together with the strengths, limitations and methodology issues encountered during the conduct of the study. The next chapter will include the concluding remarks.

6

Conclusion and recommendations

6.1 | Introduction

This chapter will include the concluding remarks, recommendations for professional practice and future research.

6.2 | Summary of Conclusions from Study

- The algorithm demonstrated strong potential in generating structurally accurate sCT images from MRI, as evidenced by favourable image similarity metrics (MAE, PSNR, SSIM). This suggests that the algorithm can produce high-quality images suitable for RT planning.

- Despite the success in image generation, discrepancies were observed in dose metrics, such as DVH and GPR, particularly in regions with complex anatomical structures. These results indicate that while the algorithm performs well in structural replication, further refinement is needed to achieve accurate dose distributions for safe clinical application.

- The study highlighted the value of testing the algorithm on a minority cohort, such as Maltese brain cancer patients. Although the algorithm performed well overall, the variability in results suggests the need for algorithms to be tailored to diverse populations to ensure consistent accuracy across different demographic groups.

- The algorithm addressed the registration errors that arise from dual-simulation processes involving CT and MRI scans. By successfully generating sCT images directly from MRI, the image-to-image translation process effectively eliminates these misalignment issues.

6.3 | Recommendations for Professional Practice

Based on the findings of this study and the observed potential of sCT generation from MRI, several key recommendations can be made to enhance its integration into clinical RT workflows:

- **Inclusion of sCT in Clinical Practice**: One of the most immediate benefits of incorporating sCT into clinical practice is the elimination of the dual imaging process, where both MRI and CT are required for treatment planning. The use of sCT derived from MRI can help avoid registration errors, which often arise from anatomical or positional discrepancies between separate CT and MRI scans, ultimately streamlining the workflow to rely solely on MRI for both imaging and treatment planning.
- **Validation Using Larger Datasets**: To further improve the generalisability and reliability of sCT algorithms, it is essential to validate them using larger and more diverse datasets. The current study highlighted certain limitations due to the small sample size. Expanding the dataset would allow the algorithm to learn and generalise better, particularly when applied to different anatomical regions and patient demographics, ensuring it is optimised for clinical use.
- **Staff Training and Expertise Development**: Should sCT become more integrated into clinical workflows, it is vital that RT professionals receive specialised training to operate these algorithms effectively. Staff should be proficient in running the sCT generation process, troubleshooting, and identifying areas for improvement in the algorithm. Additionally, as different anatomical regions present unique challenges, continued training will enable staff to fine-tune the algorithm for various treatment sites, improving accuracy and treatment outcomes.
- **Automation of sCT Generation**: For seamless integration into clinical settings, a dedicated system should be established where all MRI scans of RT patients are automatically processed through the sCT generation algorithm. This automation will standardise the generation of sCTs, reducing manual intervention and ensuring that sCT images are consistently available for treatment planning. Hosting such an environment would support the widespread adoption of sCT.

6.4 | Recommendations for Future Research

- **Expanding Dataset Size and Diversity**: Future studies should involve larger and more diverse datasets to train and test the algorithm. This would help improve its generalisability across different populations and reduce variability in performance.
- **Refinement for RT Applications**: Given that the algorithm was originally designed for Transcranial MR Imaging–Guided Focused Ultrasound interventions, additional research should focus on optimising it specifically for RT applications. This includes enhancing its ability to predict accurate dose metrics in radiotherapy treatment planning.
- **Alternative Algorithmic Techniques**: Future research could explore the incorporation of alternative, advanced techniques such to improve the accuracy of dose distributions in complex anatomical regions.
- **Anonymisation and Open Data Practices**: Investigate how SynthRAD23 managed to anonymise and share medical imaging data as open datasets, in contrast to current practices where such data is still kept behind closed doors. Understanding their approach to data anonymisation and sharing could provide insights into improving accessibility while maintaining patient confidentiality.
- **Expanding to More Anatomical Regions**: Explore the potential of applying sCT generation and analysis to additional anatomical regions beyond the brain. Expanding the application of sCT to other areas could provide broader insights into the algorithm's generalisability and clinical relevance in different radiotherapy contexts.
- **Phantom and Model-Based Studies:** Utilising anthropomorphic phantoms and computational models for RT treatment planning can simulate the use of sCT images, allowing researchers to evaluate the precision and reliability of sCT in predicting dose distributions. This method provides a controlled environment to test the clinical applicability of sCT generation without the need for live patient data.

6.5 | Conclusion

In conclusion, the objectives of this study were successfully met to varying extents. The algorithm's ability to generate structurally accurate sCT images from MRI was thoroughly evaluated, with strong performance in replicating anatomical structures as demonstrated by the image metrics. However, the results revealed significant challenges in dose accuracy, particularly in complex anatomical regions, indicating that further refinement is necessary.

The study also achieved its goal of addressing registration errors by reducing the need for dual-simulation processes. Additionally, the algorithm's performance in a minority cohort highlighted both its potential and the importance of tailoring algorithms to diverse populations.

Overall, while the study met its objectives in evaluating the algorithm's performance, further research is needed to optimise its clinical feasibility, particularly in terms of dose accuracy, before it can be a possibility to implement into the RT treatment planning pipeline.

A

Appendix A

A.1 | Python Code

A.1.1 | Python Code developed for image pre-processing

```
# PREPROCESSING STEPS

#Import libraries

import argparse
import SimpleITK as sitk
import numpy as np
import matplotlib.pyplot as plt
import os
from operator import mul
import nibabel as nib

# DIMENSIONS OF TEST IMAGE PROVIDED FOR ALGORITHM - TEST_MRI.NII
#Check the shape of the test image

# Load the NIfTI file
nifti_img = nib.load('/home/matthew/preprocessing/Brain1_Data/NiftiData/
    Brain1.nii')

# Get the image data as a NumPy array
img_data = nifti_img.get_fdata()

# Print the dimensions of the image
print(f"Image dimensions: {img_data.shape}")

# 1. CONVERTING DICOM INTO NIFTI FILE FORMATS
```

```
#Function facilitating dicom to nifTI conversion

def convert_dicom_to_nifti(input, output):
    # reads a series of DICOM files as a single volume
    reader = sitk.ImageSeriesReader()
    # get DICOM series file names
    dicom_names = reader.GetGDCMSeriesFileNames(input)
    # set file names to reader
    reader.SetFileNames(dicom_names)
    # execute the reader - convert into SimpleITK image object
    image = reader.Execute()
    # SimpleITK image object is specified to a specific NifTI path
    sitk.WriteImage(image, output)

#Paths for MRI images

input_path_MRI = '/home/matthew/preprocessing/Brain1_Data/DCMData'
output_path_MRI = '/home/matthew/preprocessing/Brain1_Data/NiftiData/
    Brain1.nii'

#Paths for CT images

input_path_CT = 'Anonymisation/DICOMS/CTs/Brain1'
output_path_CT = 'Anonymisation/NifTI/Nifti_CTs/Brain1.nii'

#Execute function

convert_dicom_to_nifti(input_path_MRI, output_path_MRI)

def limit_pixel_values(input_fn, output_fn,low_val=-1024):
    input_im = sitk.ReadImage(input_fn)
    input_np = sitk.GetArrayFromImage(input_im)
    input_np[input_np<low_val]=low_val
    output_im = sitk.GetImageFromArray(input_np)
    output_im.CopyInformation(input_im)
    sitk.WriteImage(output_im,output_fn)

limit_pixel_values("/home/matthew/preprocessing/sCT/Brain1_sCT.nii", "/
    home/matthew/preprocessing/sCT/sCT_aligned/Brain1_lim.nii")

# 2. REGISTRATION

def register(fixed, moving, parameter, output):
    # Load images
    fixed_image = sitk.ReadImage(fixed)
```

```
    moving_image = sitk.ReadImage(moving)

    # Read the parameter map from the file
    parameter_map = sitk.ReadParameterFile(parameter)

    # Perform registration based on parameter file
    elastixImageFilter = sitk.ElastixImageFilter()
    elastixImageFilter.SetParameterMap(parameter_map)
    elastixImageFilter.PrintParameterMap()
    elastixImageFilter.SetFixedImage(moving_image)  # due to FOV
        differences CT first registered to MR an inverted in the end
    elastixImageFilter.SetMovingImage(fixed_image)
    elastixImageFilter.LogToConsoleOn()
    elastixImageFilter.LogToFileOff()
    elastixImageFilter.Execute()

    # convert to itk transform format
    transform = elastixImageFilter.GetTransformParameterMap(0)
    x = transform.values()
    center = np.array((x[0])).astype(np.float64)
    rigid = np.array((x[22])).astype(np.float64)
    transform_itk = sitk.Euler3DTransform()
    transform_itk.SetParameters(rigid)
    transform_itk.SetCenter(center)
    transform_itk.SetComputeZYX(False)

    # save itk transform to correct MR mask later
    output = str(output)
    #transform_itk.WriteTransform(str(output.split('.')[:-2][0]) + '
        _parameters.txt')
    transform_itk.WriteTransform(str('registration_parameters.txt'))

    ##invert transform to get MR registered to CT
    inverse = transform_itk.GetInverse()

    ## check if moving image is an mr or cbct
    min_moving = np.amin(sitk.GetArrayFromImage(moving_image))
    if min_moving <-500:
        background = -1000
    else:
        background = 0

    ##transform MR image
    resample = sitk.ResampleImageFilter()
    resample.SetReferenceImage(fixed_image)
```

```
    resample.SetTransform(inverse)
    resample.SetInterpolator(sitk.sitkLinear)
    resample.SetDefaultPixelValue(background)
    output_image = resample.Execute(moving_image)

    # write output image
    sitk.WriteImage(output_image, output)

register(
    fixed='/home/matthew/preprocessing/Anonymisation/Registered_MRIs/Brain1.nii',
    moving='/home/matthew/preprocessing/sCT/Brain1_sCT.nii',
    parameter='param_files/parameters_MR.txt',
    output='/home/matthew/preprocessing/sCT/sCT_aligned/Brain1_mri.nii')

# 3. IMAGE MASK GENERATION
#segmented with only (1,1,1) dilation

def segment_no_dilation(input_image, output_mask=None, radius=(12, 12, 12), return_sitk=False):
    # Read and preprocess the image
    image = sitk.InvertIntensity(sitk.Cast(sitk.ReadImage(input_image), sitk.sitkFloat32))

    # Apply Otsu thresholding to create a binary mask
    mask = sitk.OtsuThreshold(image)


    # Option 2: Minimal dilation if some dilation
    minimal_dil_mask = sitk.BinaryDilate(mask, (1, 1, 1))  # Reduced dilation for less expansion

    # Perform connected component analysis
    component_image = sitk.ConnectedComponent(minimal_dil_mask)  # Using minimal_dil_mask

    # Sort by object size and extract the largest connected component
    sorted_component_image = sitk.RelabelComponent(component_image, sortByObjectSize=True)
    largest_component_binary_image = sorted_component_image == 1

    # Apply morphological closing to fill small gaps (keep the same radius here)
    mask_closed = sitk.BinaryMorphologicalClosing(largest_component_binary_image, radius)
```

```

    # Fill holes inside the mask (no additional dilation)
    filled_mask = sitk.BinaryFillhole(mask_closed)

    # Return or save the resulting mask
    if return_sitk:
        return filled_mask
    else:
        sitk.WriteImage(filled_mask, output_mask)


# CHECKING IMAGE SHAPE
#Check the shape of the test image

# Load the NIfTI file
nifti_img = nib.load('/home/matthew/preprocessing/Brain1_Data/NiftiData/
    Brain1.nii')

# Get the image data as a NumPy array
img_data = nifti_img.get_fdata()

# Print the dimensions of the image
print(f"Image dimensions: {img_data.shape}")

# CHECKING FILE PATH

file_path = input_path_CT

if os.path.exists(file_path):
    print(f"The file '{file_path}' exists.")
else:
    print(f"The file '{file_path}' does not exist or the path is incorrect
        .")
```

A.1.2 | Python Code developed for Image Post-Processing - Image Re-Orientation

```
#Import libraries

import argparse
import SimpleITK as sitk
import numpy as np
import matplotlib.pyplot as plt
```

```
import os
from operator import mul
import nibabel as nib

# CT alignment

# Load the volume from a DICOM directory

image = sitk.ReadImage("/home/matthew/preprocessing/Anonymisation/NifTI/
    Nifti_MRIs/Brain16.nii")

# Get the direction matrix
direction_matrix = image.GetDirection()

# Reshape the direction matrix to 3x3
direction_matrix_reshaped = [
    direction_matrix[0:3],
    direction_matrix[3:6],
    direction_matrix[6:9]
]

# Print the direction matrix
for row in direction_matrix_reshaped:
    print(row)

import SimpleITK as sitk

# Load the volume from a NIfTI file
image = sitk.ReadImage("/home/matthew/preprocessing/Anonymisation/NifTI/
    Nifti_MRIs/Brain16.nii")

# Set the direction to the identity matrix (1, 0, 0, 0, 1, 0, 0, 0, 1)
identity_direction = (1.0, 0.0, 0.0,
                      0.0, 1.0, 0.0,
                      0.0, 0.0, 1.0)

# Apply the new direction matrix to the image
image.SetDirection(identity_direction)

# Save the modified image back to a NIfTI file
sitk.WriteImage(image, "/home/matthew/preprocessing/Anonymisation/
    Aligned_MRIs/Brain16_aligned.nii")

# Print the updated direction matrix to confirm the change
updated_direction_matrix = image.GetDirection()
```

```
print(updated_direction_matrix)

# Load the volume from a DICOM directory
image = sitk.ReadImage("/home/matthew/preprocessing/Anonymisation/
    Aligned_MRIs/Brain16_aligned.nii")

# Get the direction matrix
direction_matrix = image.GetDirection()

# Reshape the direction matrix to 3x3
direction_matrix_reshaped = [
    direction_matrix[0:3],
    direction_matrix[3:6],
    direction_matrix[6:9]
]

# Print the direction matrix
for row in direction_matrix_reshaped:
    print(row)

# sCT alignment

# Load the volume from a DICOM directory
image = sitk.ReadImage("/home/matthew/preprocessing/sCT/Brain3_sCT.nii")

# Get the direction matrix
direction_matrix = image.GetDirection()

# Reshape the direction matrix to 3x3
direction_matrix_reshaped = [
    direction_matrix[0:3],
    direction_matrix[3:6],
    direction_matrix[6:9]
]

# Print the direction matrix
for row in direction_matrix_reshaped:
    print(row)

import SimpleITK as sitk

# Load the volume from a NIfTI file
image = sitk.ReadImage("/home/matthew/preprocessing/sCT/Brain16_sCT.nii")

# Set the direction to the identity matrix (1, 0, 0, 0, 1, 0, 0, 0, 1)
```

```
identity_direction = (1.0, 0.0, 0.0,
                      0.0, 1.0, 0.0,
                      0.0, 0.0, 1.0)

# Apply the new direction matrix to the image
image.SetDirection(identity_direction)

# Save the modified image back to a NIfTI file
sitk.WriteImage(image, "/home/matthew/preprocessing/sCT/Brain16_aligned.
    nii")

# Print the updated direction matrix to confirm the change
updated_direction_matrix = image.GetDirection()

print(updated_direction_matrix)

# Success

# Load the volume from a DICOM directory
image = sitk.ReadImage("/home/matthew/preprocessing/sCT/Brain16_aligned.
    nii")

# Get the direction matrix
direction_matrix = image.GetDirection()

# Reshape the direction matrix to 3x3
direction_matrix_reshaped = [
    direction_matrix[0:3],
    direction_matrix[3:6],
    direction_matrix[6:9]
]

# Print the direction matrix
for row in direction_matrix_reshaped:
    print(row)
```

A.1.3 | Python Code developed for Image Post-Processing - DICOM Tags

```

#Import libraries

import pydicom
```

```
import os
import numpy as np

#PRINT ALL DICOM TAGS

def print_all_dicom_tags(filepath):
    # Read the DICOM file
    ds = pydicom.dcmread(filepath)

    # Iterate through all DICOM tags and print them
    for elem in ds:
        print(f"Tag: {elem.tag}, Name: {elem.name}, VR: {elem.VR}, Value:
            {elem.value}")

print_all_dicom_tags("/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN16/
    BrainSCT16/IMG0001.dcm")

def print_smallest_pixel_values(directory):
    files = [f for f in os.listdir(directory) if f.endswith('.dcm')]
    for filename in sorted(files):
        filepath = os.path.join(directory, filename)
        ds = pydicom.dcmread(filepath)
        print(f"{filename}: Study Date= {ds.StudyDate}")

print_smallest_pixel_values("/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN16
    /BrainSCT16")

def assign_dates(input_directory, output_directory, study_date,
    series_date, content_date, study_time):
    if not os.path.exists(output_directory):
        os.makedirs(output_directory)

    files = [f for f in os.listdir(input_directory) if f.endswith('.dcm')]

    for filename in sorted(files):
        filepath = os.path.join(input_directory, filename)
        ds = pydicom.dcmread(filepath)

        # Assign Study Date
        ds.StudyDate = study_date

        # Assign Series Date
        ds.SeriesDate = series_date

        # Assign Content Date
```

```
        ds.ContentDate = content_date

        # Assign Study Time
        ds.StudyTime= study_time

        # Save the modified DICOM file in the output directory
        new_filepath = os.path.join(output_directory, filename)
        ds.save_as(new_filepath)
        print(f"{filename}: Study Date set to {study_date}, Series Date
            set to {series_date}, Content Date set to {content_date} and
            saved in {output_directory}")

assign_dates(
    "/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN21/BrainSCT21",
    "/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN21/BrainSCT21",
    "20230321",
    "20230321",
    "20230321",
    "103628")


def assign_physician_and_desc(input_directory, output_directory,
    physician_name, series_desc):
    if not os.path.exists(output_directory):
        os.makedirs(output_directory)

    files = [f for f in os.listdir(input_directory) if f.endswith('.dcm')]

    for filename in sorted(files):
        filepath = os.path.join(input_directory, filename)
        ds = pydicom.dcmread(filepath)

        # Assign Referring Physician's Name
        ds.ReferringPhysicianName = physician_name

        # Assign Accession Number
        ds.SeriesDescription = series_desc

        # Save the modified DICOM file in the output directory
        new_filepath = os.path.join(output_directory, filename)
        ds.save_as(new_filepath)
        print(f"{filename}: Referring Physician's Name set to {
            physician_name}, Series Description set to {series_desc}, and
            saved in {output_directory}")

```

```
assign_physician_and_desc(
    "/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN21/BrainSCT21",
    "/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN21/BrainSCT21",
    "DummyRefPhys!",
    "DummySeriesDesc!")

# Define the folder containing DICOM files
folder_path = "/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN21/BrainSCT21"


# Iterate through all files in the folder
for filename in os.listdir(folder_path):
    if filename.endswith(".dcm"):  # Only process DICOM files
        dicom_file = os.path.join(folder_path, filename)

        # Load the DICOM file
        ds = pydicom.dcmread(dicom_file)

        # Delete the tag (0008,0050) Accession Number if it exists
        if (0x0008, 0x0050) in ds:
            del ds[(0x0008, 0x0050)]

        # Save the updated DICOM file, overwriting the original file
        ds.save_as(dicom_file)

print("Tag (0008,0050) removed from all DICOM files in the folder.")
```

A.1.4 | Python Code developed for metrics computation

```

#Import libraries

import numpy as np
import nibabel as nib
from skimage.metrics import structural_similarity as ssim
import matplotlib.pyplot as plt
import SimpleITK as sitk

# GENERATE MASKS FROM MRI SCANS

#segmented with only (1,1,1) dilation

def segment_no_dilation(input_image, output_mask=None, radius=(12, 12, 12)
    , return_sitk=False):
```

```
    # Read and preprocess the image
    image = sitk.InvertIntensity(sitk.Cast(sitk.ReadImage(input_image),
        sitk.sitkFloat32))

    # Apply Otsu thresholding to create a binary mask
    mask = sitk.OtsuThreshold(image)

    # Remove or reduce dilation to prevent expanding beyond the skin
    # Option 1: Skip dilation entirely
    # dil_mask = sitk.BinaryDilate(mask, (10, 10, 1))  # Original dilation
         step - removed

    # Option 2: Minimal dilation if some dilation is still needed (reduce
        the dilation size)
    minimal_dil_mask = sitk.BinaryDilate(mask, (1, 1, 1))  # Reduced
        dilation for less expansion

    # Perform connected component analysis
    component_image = sitk.ConnectedComponent(minimal_dil_mask)  # Using
        minimal_dil_mask

    # Sort by object size and extract the largest connected component
    sorted_component_image = sitk.RelabelComponent(component_image,
        sortByObjectSize=True)
    largest_component_binary_image = sorted_component_image == 1

    # Apply morphological closing to fill small gaps (keep the same radius
         here)
    mask_closed = sitk.BinaryMorphologicalClosing(
        largest_component_binary_image, radius)

    # Fill holes inside the mask (no additional dilation)
    filled_mask = sitk.BinaryFillhole(mask_closed)

    # Return or save the resulting mask
    if return_sitk:
        return filled_mask
    else:
        sitk.WriteImage(filled_mask, output_mask)

input_image_path = '/Users/martina/Desktop/DICOMRT/MRIs_Aligned/
    Brain21_aligned.nii'
output_mask_path = '/Users/martina/Desktop/DICOMRT/MASKS/Mask21_skin.nii'

segment_no_dilation(input_image_path, output_mask = output_mask_path)
```

```

# PLOT THE IMAGE BEFORE PERFORMING METRICS

# Load NIfTI images

ct_nifti = nib.load('/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN1/
    Brain1_reg_CT_aligned.nii')
sct_nifti = nib.load('/Users/martina/Desktop/DICOMRT/BRAINS/BRAIN1/
    Brain1_aligned.nii')
mask_nifti = nib.load('/Users/martina/Desktop/DICOMRT/MASKS/Mask1_skin.nii
    ')

# Convert to NumPy arrays

ct_image = ct_nifti.get_fdata()
sct_image = sct_nifti.get_fdata()
mask_image = mask_nifti.get_fdata()

# Select a slice to analyze (e.g., slice 0)
slice_index = 120  # Change this index as needed
ct_image_slice = ct_image[:, :, slice_index]
sct_image_slice = sct_image[:, :, slice_index]
mask_image_slice = mask_image[:, :, slice_index]


# Visualize the slices
plt.figure(figsize=(12, 6))

# CT image slice
plt.subplot(1, 3, 1)
plt.imshow(ct_image_slice, cmap='gray')
plt.title('CT Image Slice')
plt.axis('off')

# Synthetic CT image slice
plt.subplot(1, 3, 2)
plt.imshow(sct_image_slice, cmap='gray')
plt.title('Synthetic CT Image Slice')
plt.axis('off')

# Synthetic CT image slice
plt.subplot(1, 3, 3)
plt.imshow(mask_image_slice, cmap='gray')
plt.title('Mask Image Slice')
plt.axis('off')
```

```

plt.show()

# CLIPPING CTs & sCTs

def load_nifti_image(file_path):

    nifti_img = nib.load(file_path)
    return nifti_img.get_fdata(), nifti_img.affine  # Return image data and affine transformation


def save_nifti_image(data, affine, output_path):

    nifti_img = nib.Nifti1Image(data, affine)
    nib.save(nifti_img, output_path)


def clip_image(image_data, min_value, max_value):

    return np.clip(image_data, min_value, max_value)

#CLIP SCT

# Paths to your NIfTI files
image_path = "/Users/martina/Desktop/DICOMRT/BRAINS-NIFTIS/NifTI_sCT/Brain21_sCT.nii"
output_path = "/Users/martina/Desktop/DICOMRT/Clipped_Images/BRAIN 21/brain21_clipped_SCT.nii"

# Load the NIfTI image
image_data, affine = load_nifti_image(image_path)

# Clip the image to the desired dynamic range (-1024 to 3000 HU)
clipped_image = clip_image(image_data, -1024, 3000)

# Save the clipped image as a new NIfTI file
save_nifti_image(clipped_image, affine, output_path)

print(f"Clipped image saved to {output_path}")

#CLIP CT

# Paths to your NIfTI files
```

```
image_path = "/Users/martina/Desktop/DICOMRT/BRAINS-NIFTIS/Nifti_CT/
    Brain21.nii"
output_path = "/Users/martina/Desktop/DICOMRT/Clipped_Images/BRAIN 21/
    brain21_clipped_CT.nii"

# Load the NIfTI image
image_data, affine = load_nifti_image(image_path)

# Clip the image to the desired dynamic range (-1024 to 3000 HU)
clipped_image = clip_image(image_data, -1024, 3000)

# Save the clipped image as a new NIfTI file
save_nifti_image(clipped_image, affine, output_path)

print(f"Clipped image saved to {output_path}")

# METRIC COLLECTION

#Import Libararies

import nibabel as nib
import numpy as np

# Load NIfTI files for CT, sCT, and the mask
ct_nifti = nib.load('/Users/martina/Desktop/DICOMRT/Clipped_Images/BRAIN1/
    brain1_clipped_CT.nii')
sct_nifti = nib.load('/Users/martina/Desktop/DICOMRT/Clipped_Images/BRAIN1
    /brain1_clipped_SCT.nii')
mask_nifti = nib.load('/Users/martina/Desktop/DICOMRT/MASKS/Mask1_skin.nii
    ')

# Convert NIfTI images to NumPy arrays
ct_data = ct_nifti.get_fdata()
sct_data = sct_nifti.get_fdata()
mask_data = mask_nifti.get_fdata()

# Ensure that CT, sCT, and mask have the same dimensions
assert ct_data.shape == sct_data.shape == mask_data.shape, "Dimensions do
    not match!"

# Apply the mask to select only the voxels within the body contour
# Voxels where mask == 1 are selected
masked_ct = ct_data[mask_data == 1]
masked_sct = sct_data[mask_data == 1]

```

```
# Calculate the Mean Absolute Error (MAE)
mae = np.mean(np.abs(masked_ct - masked_sct))

print(f"Mean Absolute Error (MAE): {mae}")

import numpy as np

def calculate_mae(ct_image, sct_image, mask):
    """
    Calculate the MAE between CT and synthetic CT images within a mask.

    :param ct_image: 3D numpy array of reference CT scan
    :param sct_image: 3D numpy array of synthetic CT scan
    :param mask: 3D binary numpy array representing the masked region
    :return: MAE value
    """


    # Convert NIfTI images to NumPy arrays
    ct_data = ct_nifti.get_fdata()
    sct_data = sct_nifti.get_fdata()
    mask_data = mask_nifti.get_fdata()

    # Apply the mask
    ct_masked = ct_image[mask == 1]
    sct_masked = sct_image[mask == 1]

    # Calculate absolute differences
    absolute_differences = np.abs(ct_masked - sct_masked)

    # Compute the mean absolute error
    mae = np.mean(absolute_differences)

    return mae

# Example usage
# ct_image = np.load("path_to_ct_image.npy")  # Load your CT image
# sct_image = np.load("path_to_sct_image.npy")  # Load your synthetic CT
    image
# mask = np.load("path_to_mask.npy")  # Load your mask

    mae_value = calculate_mae(ct_image, sct_image, mask)
    print(f"MAE: {mae_value}")

import nibabel as nib
```

```
import numpy as np

def load_nifti_image(file_path):
    """
    Load a NIfTI image from a file and return the image data as a NumPy
        array.

    :param file_path: Path to the NIfTI file
    :return: 3D NumPy array of the image
    """
    nifti_img = nib.load(file_path)
    return nifti_img.get_fdata()  # Convert to NumPy array

def calculate_mae(ct_image, sct_image, mask):
    """
    Calculate the MAE between CT and synthetic CT images within a mask.

    :param ct_image: 3D numpy array of reference CT scan
    :param sct_image: 3D numpy array of synthetic CT scan
    :param mask: 3D binary numpy array representing the masked region
    :return: MAE value
    """
    # Apply the mask
    ct_masked = ct_image[mask == 1]
    sct_masked = sct_image[mask == 1]

    # Calculate absolute differences
    absolute_differences = np.abs(ct_masked - sct_masked)
    print('the absolute differences are:', absolute_differences)

    # Compute the mean absolute error
    mae = np.mean(absolute_differences)

    return mae

# Load your NIfTI files
ct_image = load_nifti_image("/Users/martina/Desktop/DICOMRT/Clipped_Images
    /BRAIN1/brain1_clipped_CT.nii")
sct_image = load_nifti_image("/Users/martina/Desktop/DICOMRT/
    Clipped_Images/BRAIN1/brain1_clipped_SCT.nii")
mask = load_nifti_image("/Users/martina/Desktop/DICOMRT/MASKS/Mask1_skin.
    nii")

# Calculate MAE
mae_value = calculate_mae(ct_image, sct_image, mask)
```

```
print(f"MAE: {mae_value}")

# MASKED PSNR

import nibabel as nib
import numpy as np

def load_nifti_image(file_path):
    """
    Load a NIfTI image from a file and return the image data as a NumPy
        array.

    :param file_path: Path to the NIfTI file
    :return: 3D NumPy array of the image
    """
    nifti_img = nib.load(file_path)
    return nifti_img.get_fdata()  # Return only image data as a NumPy
        array

def calculate_psnr(ct_image, sct_image, mask, dynamic_range):
    """
    Calculate PSNR between CT and synthetic CT images within a mask.

    :param ct_image: 3D numpy array of reference CT scan
    :param sct_image: 3D numpy array of synthetic CT scan
    :param mask: 3D binary numpy array representing the masked region
    :param dynamic_range: Dynamic range of the voxel intensities (Q)
    :return: PSNR value
    """
    # Apply the mask
    ct_masked = ct_image[mask == 1]
    sct_masked = sct_image[mask == 1]

    # Calculate the Mean Squared Error (MSE)
    mse = np.mean((ct_masked - sct_masked) ** 2)

    # Calculate PSNR
    if mse == 0:  # Avoid division by zero
        return float('inf')

    psnr = 10 * np.log10(dynamic_range ** 2 / mse)

    return psnr

# Paths to your NIfTI files
```

```
ct_image_path = "/Users/martina/Desktop/DICOMRT/Clipped_Images/BRAIN 21/
    brain21_clipped_CT.nii"
sct_image_path = "/Users/martina/Desktop/DICOMRT/Clipped_Images/BRAIN 21/
    brain21_clipped_SCT.nii"
mask_path = "/Users/martina/Desktop/DICOMRT/MASKS/Mask21_skin.nii"

# Load the NIfTI images and mask
ct_image = load_nifti_image(ct_image_path)
sct_image = load_nifti_image(sct_image_path)
mask = load_nifti_image(mask_path)

# Define the dynamic range Q (clipped to [-1024, 3000] HU)
dynamic_range = 3000 - (-1024)

# Calculate PSNR
psnr_value = calculate_psnr(ct_image, sct_image, mask, dynamic_range)
print(f"PSNR: {psnr_value} dB")

# MASKED SSIM

import nibabel as nib
import numpy as np
from skimage.metrics import structural_similarity as ssim

def clip_and_adjust_image(image, min_value, max_value):
    """
    Clip the image to the specified HU range and adjust values to be non-
        negative.

    :param image: 3D numpy array of the image
    :param min_value: Minimum clipping value (HU)
    :param max_value: Maximum clipping value (HU)
    :return: Clipped and adjusted image
    """
    # Clip image to the specified range
    clipped_image = np.clip(image, min_value, max_value)

    # Adjust the clipped image to be non-negative by adding 1024
    adjusted_image = clipped_image + 1024

    return adjusted_image

def calculate_masked_ssim(ct_image, sct_image, mask, window_size=7,
    dynamic_range=4024):
    """
```

```
    Calculate the Masked SSIM between CT and synthetic CT images within a
        mask.

    :param ct_image: 3D numpy array of reference CT scan
    :param sct_image: 3D numpy array of synthetic CT scan
    :param mask: 3D binary numpy array representing the masked region
    :param window_size: The size of the window (N) for local SSIM
        calculation
    :param dynamic_range: The dynamic range of voxel intensities (L)
    :return: Mean SSIM value for the masked region
    """
    # Apply the mask
    ct_masked = ct_image[mask == 1]
    sct_masked = sct_image[mask == 1]

    # Calculate SSIM over the masked region
    ssim_value, _ = ssim(ct_masked, sct_masked, win_size=window_size,
        data_range=dynamic_range, full=True)

    return np.mean(ssim_value)

# Paths to your NIfTI files
ct_image_path = "/Users/martina/Desktop/DICOMRT/Clipped_Images/BRAIN 21/
    brain21_clipped_CT.nii"
sct_image_path = "/Users/martina/Desktop/DICOMRT/Clipped_Images/BRAIN 21/
    brain21_clipped_SCT.nii"
mask_path = "/Users/martina/Desktop/DICOMRT/MASKS/Mask21_skin.nii"

# Load the NIfTI images and mask
ct_image = load_nifti_image(ct_image_path)
sct_image = load_nifti_image(sct_image_path)
mask = load_nifti_image(mask_path)

# Clip images to [-1024, 3000] HU and adjust to non-negative range
ct_image_clipped = clip_and_adjust_image(ct_image, -1024, 3000)
sct_image_clipped = clip_and_adjust_image(sct_image, -1024, 3000)

# Calculate Masked SSIM
masked_ssim_value = calculate_masked_ssim(ct_image_clipped,
    sct_image_clipped, mask)
print(f"Masked SSIM: {masked_ssim_value}")
```

A.1.5 | Python Code developed for uncertainties calculation

```

#Import libraries

import numpy as np

#Define Function
def calculate_uncertainty(measurements):
    # Number of measurements
    n = len(measurements)

    # Calculate the mean of the measurements
    mean_value = np.mean(measurements)

    # Calculate the standard deviation of the measurements
    standard_deviation = np.std(measurements, ddof=1)  # Using ddof=1 for sample standard deviation

    # Calculate the uncertainty of the average (standard error of the mean)
    uncertainty = standard_deviation / np.sqrt(n)

    return mean_value, uncertainty

#Use Function
measurements = []
mean, uncertainty = calculate_uncertainty(measurements)

#Print
print(f"Mean value: {mean}")
print(f"Uncertainty of the average: {uncertainty}")
```

References

AAPM Task Group 111. Tissue inhomogeneity correction for megavoltage photon beams, 2021. URL `https://www.aapm.org/pubs/reports/rpt_85.pdf`. The accurate delivery of a prescribed dose to a well-defined target volume is dependent firstly on the accuracy with which the radiation beam can be calibrated.

Martin D Abeloff, James O Armitage, John E Niederhuber, Michael B Kastan, and W Gillies McKenna. *Abeloff's Clinical Oncology E-Book*. Elsevier Health Sciences, 2008.

Dorothy Abshire and Matthew K Lang. The evolution of radiation therapy in treating cancer. In *Seminars in oncology nursing*, volume 34, pages 151–157. Elsevier, 2018.

Yuichi Akino, Naoki Tohyama, Kazuhiko Akita, Masayori Ishikawa, Ryu Kawamorita, Masahiko Kurooka, Iori Sumida, Yuji Nakaguchi, Kazunori Miyaura, and Mitsuhiro Nakamura. Modalities and techniques used for stereotactic radiotherapy, intensity-modulated radiotherapy, and image-guided radiotherapy: A 2018 survey by the japan society of medical physics. *Physica Medica*, 64:182–187, 2019.

Fakhre Alam, Sami Ur Rahman, Sehat Ullah, Adnan Khalil, and Aziz Uddin. A review on extrinsic registration methods for medical images. *Tech J Univ Eng Technol Taxila*, 21:110–9, 2016.

Asraa Faris Aldoghachi, Ahmed Faris Aldoghachi, Koen Breyne, King-Hwa Ling, and Pike-See Cheah. Recent advances in the therapeutic strategies of glioblastoma multiforme. *Neuroscience*, 491:240–270, 2022.

Damien Autret, Camille Guillerminet, Alban Roussel, Erwan Cossec-Kerloc'h, and Stéphane Dufreneix. Comparison of four synthetic ct generators for brain and prostate mr-only workflow in radiotherapy. *Radiation Oncology*, 18(1):146, 2023.

Aaron Babier, Rafid Mahmood, Andrea L McNiven, Adam Diamant, and Timothy CY Chan. Knowledge-based automated planning with three-dimensional generative adversarial networks. *Medical Physics*, 47 (2):297–306, 2020.

Ana Barragán-Montero, Umair Javaid, Gilmer Valdés, Dan Nguyen, Paul Desbordes, Benoit Macq, Siri Willems, Liesbeth Vandewinckele, Mats Holmström, Fredrik Löfman, et al. Artificial intelligence and machine learning for medical imaging: A technology review. *Physica Medica*, 83:242–256, 2021.

Laura Beaton, Steve Bandula, Mark N Gaze, and Ricky A Sharma. How rapid advances in imaging are defining the future of precision radiation oncology. *British journal of cancer*, 120(8):779–790, 2019.

Mario Caccese, Fabio Busato, Angela Guerriero, Marta Padovan, Giulia Cerretti, Marina Paola Gardiman, Vittorina Zagonel, and Giuseppe Lombardi. The role of radiation therapy and systemic treatments in meningioma: The present and the future. *Cancer Medicine*, 12(15):16041–16053, 2023.

Hersh Chandarana, Hesheng Wang, RHN Tijssen, and Indra J Das. Emerging role of mri in radiation therapy. *Journal of Magnetic Resonance Imaging*, 48(6):1468–1478, 2018.

Sunan Cui, Huan-Hsin Tseng, Julia Pakela, Randall K Ten Haken, and Issam El Naqa. Introduction to machine and deep learning for medical physicists. *Medical physics*, 47(5):e127–e147, 2020.

Fida K Dankar, Marton Gergely, and Samar K Dankar. Informed consent in biomedical research. *Computational and structural biotechnology journal*, 17:463–474, 2019.

Jia Deng, Olga Russakovsky, Jonathan Krause, Michael S Bernstein, Alex Berg, and Li Fei-Fei. Scalable multi-label annotation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3099–3102, 2014.

Oliver Diaz, Kaisar Kushibar, Richard Osuala, Akis Linardos, Lidia Garrucho, Laura Igual, Petia Radeva, Fred Prior, Polyxeni Gkontra, and Karim Lekadir. Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools. *Physica medica*, 83:25–37, 2021.

Hyun Do Huh and Seonghoon Kim. History of radiation therapy technology. *Progress in Medical Physic*, 31 (3):124–134, 2020.

Pronnoy Dutta, Pradumn Upadhyay, Madhurima De, and RG Khalkar. Medical image analysis using deep convolutional neural networks: Cnn architectures and transfer learning. In *2020 International Conference on Inventive Computation Technologies (ICICT)*, pages 175–180. IEEE, 2020.

Elekta. Mr-linac & mrgrt cancer treatment, 2024. URL `https://www.elekta.com/patients/mr-linac/`. Accessed: 2024-08-03.

Elekta Solutions AB. *Monaco 6.2 DICOM Conformance Statement (Rev04)*. Elekta Solutions AB, Stockholm, Sweden, April 2024. URL `https://www.elekta.com`. Document ID: E059389/04.

Jie Fu, Kamal Singhrao, Minsong Cao, Victoria Yu, Anand P Santhanam, Yingli Yang, Minghao Guo, Ann C Raldow, Dan Ruan, and John H Lewis. Generation of abdominal synthetic cts from 0.35 t mr images using generative adversarial networks for mr-only liver radiotherapy. *Biomedical Physics & Engineering Express*, 6(1):015033, 2020.

Ian Goodfellow. Deep learning, 2016.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Xiao Han. Mr-based synthetic ct generation using a deep convolutional neural network method. *Medical physics*, 44(4):1408–1419, 2017.

Dodi Heryadi and Scott Hampton. Characterizing performance improvement of gpus. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning)*, pages 1–5. 2019.

Evi MC Huijben, Maarten L Terpstra, Suraj Pai, Adrian Thummerer, Peter Koopmans, Manya Afonso, Maureen van Eijnatten, Oliver Gurney-Champion, Zeli Chen, Yiwen Zhang, et al. Generating synthetic computed tomography for radiotherapy: Synthrad2023 challenge report. *Medical image analysis*, page 103276, 2024.

International Atomic Energy Agency. Malta upgrades radiotherapy services following iaea support, 2024. URL `https://www.iaea.org/newscenter/news/malta-upgrades-radiotherapy-services-following-iaea-support`. Accessed: 2024-11-02.

International Commission on Radiation Units and Measurements. *ICRU Report 83: Prescribing, Recording, and Reporting Intensity-Modulated Radiation Therapy (IMRT)*. ICRU, Bethesda, MD, 2010. URL `https://www.fnkv.cz/soubory/216/icru-83.pdf`.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

Kyulee Jeon, Woo Yeon Park, Charles E Kahn Jr, Paul Nagy, Seng Chan You, and Soon Ho Yoon. Advancing medical imaging research through standardization: The path to rapid development, rigorous validation, and robust reproducibility. *Investigative radiology*, pages 10–1097, 2023.

Wan Jeon, Hyun Joon An, Jung-in Kim, Jong Min Park, Hyoungnyoun Kim, Kyung Hwan Shin, and Eui Kyu Chie. Preliminary application of synthetic computed tomography image generation from magnetic resonance image using deep-learning in breast cancer patients. *Journal of Radiation Protection and Research*, 44(4):149–155, 2019.

Cheng-Bin Jin, Hakil Kim, Mingjie Liu, Wonmo Jung, Seongsu Joo, Eunsik Park, Young Saem Ahn, In Ho Han, Jae Il Lee, and Xuenan Cui. Deep ct to mr synthesis using paired and unpaired data. *Sensors*, 19 (10):2361, 2019.

Dakai Jin, Adam P Harrison, Ling Zhang, Ke Yan, Yirui Wang, Jinzheng Cai, Shun Miao, and Le Lu. Artificial intelligence in radiology. In *Artificial Intelligence in Medicine*, pages 265–289. Elsevier, 2021.

Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim. Deep learning applications in medical image analysis. *Ieee Access*, 6:9375–9389, 2017.

Nazish Khalid, Adnan Qayyum, Muhammad Bilal, Ala Al-Fuqaha, and Junaid Qadir. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, 158: 106848, 2023.

Mohamed Khalifa and Mona Albadawy. Ai in diagnostic imaging: Revolutionising accuracy and efficiency. *Computer Methods and Programs in Biomedicine Update*, page 100146, 2024.

Jennifer Petra Kieselmann, Cornelis Philippus Kamerling, Ninon Burgos, Martin J Menten, Clifton David Fuller, Simeon Nill, MJ Cardoso, and Uwe Oelfke. Geometric and dosimetric evaluations of atlas-based segmentation methods of mr images in the head and neck region. *Physics in Medicine & Biology*, 63(14): 145007, 2018.

M Köhler, T Vaara, MV Grootel, R Hoogeveen, R Kemppainen, and S Renisch. Mr-only simulation for radiotherapy planning. *Philips White Paper*, 2015.

Petra Korica, Neamat El Gayar, and Wei Pang. Explainable artificial intelligence in healthcare: Opportunities, gaps and challenges and a novel way to look at the problem space. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 333–342. Springer, 2021.

Marie E Korsholm, Line W Waring, and Jens M Edmund. A criterion for the reliable use of mri-only radiotherapy. *Radiation Oncology*, 9:1–7, 2014.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

Agostina J Larrazabal, Nicolás Nieto, Victoria Peterson, Diego H Milone, and Enzo Ferrante. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences*, 117(23):12592–12594, 2020.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo, and Namkug Kim. Deep learning in medical imaging: general overview. *Korean journal of radiology*, 18(4):570–584, 2017.

Xiangrui Li, Paul S Morgan, John Ashburner, Jolinda Smith, and Christopher Rorden. The first step for neuroimaging data analysis: Dicom to nifti conversion. *Journal of neuroscience methods*, 264:47–56, 2016.

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

Han Liu, Michelle K Sigona, Thomas J Manuel, Li Min Chen, Charles F Caskey, and Benoit M Dawant. Synthetic ct skull generation for transcranial mr imaging–guided focused ultrasound interventions with conditional adversarial networks. In *Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 12034, pages 135–143. SPIE, 2022.

Daniel A Low, William B Harms, Sasa Mutic, and James A Purdy. A technique for the quantitative evaluation of dose distributions. *Medical physics*, 25(5):656–661, 1998.

Wisam Hayder Mahdi, Shaimaa Khamees Ahmed, Zobeda Hatif Naji AL-Azzawi, and Raghda Salam Al Mahdawi. A comparative study of edge detection technique with mri (magnetic resonance imaging) images. *Webology (ISSN: 1735-188X)*, 19(1), 2022.

Matteo Maspero, Cornelis AT Van den Berg, Guillaume Landry, Claus Belka, Katia Parodi, Peter R Seevinck, Bas W Raaymakers, and Christopher Kurz. Feasibility of mr-only proton dose calculations for prostate cancer radiotherapy using a commercial pseudo-ct generation method. *Physics in Medicine & Biology*, 62(24):9159, 2017.

Matteo Maspero, Laura G Bentvelzen, Mark HF Savenije, Filipa Guerreiro, Enrica Seravalli, Geert O Janssens, Cornelis AT van den Berg, and Marielle EP Philippens. Deep learning-based synthetic ct generation for paediatric brain mr-only photon and proton radiotherapy. *Radiotherapy and Oncology*, 153: 197–204, 2020.

Ministry for Health, Malta. A national health systems strategy for malta 2023 - 2030: Investing successfully for a healthy future, 2023. URL `https://health.gov.mt/wp-content/uploads/2023/04/A_National_Health_Systems_Strategy_for_Malta_2023_-_2030_Investing_Successfully_for_a_Healthy_Future_EN.pdf`. Accessed: 2024-08-03.

Mehdi Mirza. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

Dong Nie, Roger Trullo, Jun Lian, Li Wang, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, 65(12):2720–2730, 2018.

Sven Olberg, Hao Zhang, William R Kennedy, Jaehee Chun, Vivian Rodriguez, Imran Zoberi, Maria A Thomas, Jin Sung Kim, Sasa Mutic, Olga L Green, et al. Synthetic ct reconstruction using a deep spatial pyramid convolutional framework for mr-only breast radiotherapy. *Medical physics*, 46(9):4135–4147, 2019.

ProKnow. *ProKnow DS v1.31.0 User Guide*, 2024. URL `https://cdn.proknow.com/docs/v1.31.0/ProKnow%20DS%20v1.31.0%20User%20Guide.pdf`. Version 1.31.0.

Y Zhai Jiajin Fan E Wang R Croce S Asbell Tamara LaCouture G Kubicek Quanyue Xu, Yao Chen. Internal brain motion between ct and mr scanning. 2023.

Chenyang Shen, Dan Nguyen, Zhiguo Zhou, Steve B Jiang, Bin Dong, and Xun Jia. An introduction to deep learning in medical physics: advantages, potential, and challenges. *Physics in Medicine & Biology*, 65(5):05TR01, 2020.

Maria Francesca Spadea, Matteo Maspero, Paolo Zaffino, and Joao Seco. Deep learning based synthetic-ct generation in radiotherapy and pet: a review. *Medical physics*, 48(11):6537–6566, 2021.

Abhishek Thakur and Gopal Kumar Thakur. Developing gans for synthetic medical imaging data: Enhancing training and research. *Int. J. Adv. Multidiscip. Res*, 11(1):70–82, 2024.

Adrian Thummerer, Erik van der Bijl, Arthur Galapon Jr, Joost JC Verhoeff, Johannes A Langendijk, Stefan Both, Cornelis (Nico) AT van den Berg, and Matteo Maspero. Synthrad2023 grand challenge dataset: Generating synthetic ct for radiotherapy. *Medical physics*, 50(7):4664–4674, 2023.

Erico Tjoa and Cuntai Guan. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems*, 32(11):4793–4813, 2020.

Taylan Tuğrul. The effect of algorithms on dose distribution in inhomogeneous phantom: Monaco treatment planning system versus monte carlo simulation. *Journal of Medical Physics*, 46(2):111–115, 2021.

TVM News. Cancer patients soon to receive new radiotherapy which limits organ damage, 2024. URL `https://tvmnews.mt/en/news/cancer-patients-soon-to-receive-new-radiotherapy-which-limits-organ-damage/`. Accessed: 2024-08-03.

Aäron Van Den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *International conference on machine learning*, pages 1747–1756. PMLR, 2016.

Liesbeth Vandewinckele, Michaël Claessens, Anna Dinkla, Charlotte Brouwer, Wouter Crijns, Dirk Verellen, and Wouter van Elmpt. Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. *Radiotherapy and Oncology*, 153:55–66, 2020.

Henry O Velesaca, Gisel Bastidas, Mohammad Rouhani, and Angel D Sappa. Multimodal image registration techniques: a comprehensive survey. *Multimedia Tools and Applications*, pages 1–29, 2024.

Catherine Westbrook and John Talbot. *MRI in Practice*. John Wiley & Sons, 2018.

Hans-Peter Wieser, Eduardo Cisternas, Niklas Wahl, Silke Ulrich, Alexander Stadler, Henning Mescher, Lucas-Raphael Müller, Thomas Klinge, Hubert Gabrys, Lucas Burigo, et al. Development of the open-source dose calculation and optimization toolkit matrad. *Medical physics*, 44(6):2556–2568, 2017.

Jordan Wong, Vicky Huang, Derek Wells, Joshua Giambattista, Jonathan Giambattista, Carter Kolbeck, Karl Otto, Elantholi P Saibishkumar, and Abraham Alexander. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. *Radiation Oncology*, 16(1):101, 2021.

Worldometer. Malta population (live), 2024. URL `https://www.worldometers.info/world-population/malta-population/`. Accessed: October 3, 2024.

Ryuya Yamanaka, Eisuke Abe, Toshiteru Sato, Azusa Hayano, and Yasuo Takashima. Secondary intracranial tumors following radiotherapy for pituitary adenomas: A systematic review. *Cancers*, 9(8):103, 2017.

Yue Yan, Poonam Yadav, Michael Bassetti, Kaifang Du, Daniel Saenz, Paul Harari, and Bhudatt R Paliwal. Dosimetric differences in flattened and flattening filter-free beam treatment plans. *Journal of Medical Physics*, 41(2):92–99, 2016.

Orlando Zapata-Cortes, Martin Darío Arango-Serna, Julian Andres Zapata-Cortes, and Jaime Alonso Restrepo-Carmona. Machine learning models and applications for early detection. *Sensors (Basel, Switzerland)*, 24(14), 2024.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.