Honors College Thesis

Submitted in partial fulfillment of the requirements for graduation from the Honors College

"A Streamlined DECT Analysis Tool for Pediatric Proton Therapy: Enhancing Stopping Power Estimation with Improved Calibration and Machine Learning"

Roya Parsa

Sixia Chen

Sukun Li, Matthew Wright

05/12/2025

Abstract

Proton therapy is a highly precise form of radiation treatment that is particularly advantageous in pediatric oncology because it minimizes damage to surrounding healthy tissue. A critical factor in proton therapy treatment planning is accurate estimation of the stopping power ratio (SPR), which determines proton range and dose deposition. Dual-energy computed tomography (DECT) is widely employed to derive electron density and effective atomic number for SPR estimation. However, existing calibration models often introduce uncertainties that can adversely affect dose planning.

This thesis presents the first streamlined application designed to consolidate DECT analysis into a single tool that enables clinicians to calculate SPR accurately in one place. The application leverages three distinct calibration methods – Saito, Hünemohr, and Tanaka – to refine and improve the accuracy of SPR estimations. In addition, a convolutional autoencoder network (CAN) is integrated into the workflow to effectively reduce noise and eliminate artifact interference that is typically associated with DECT imaging. This dual-pronged approach not only enhances calibration precision but also mitigates systematic errors, ultimately contributing to improved treatment planning in proton therapy.

Through the combination of established calibration techniques and advanced machine learning, this research provides a comprehensive computational tool that streamlines DECT data processing and SPR estimation while also addressing the challenges of noise and artifact management.

Contents

1. Introduction **4**
2. Background **5**
2.1 Proton Therapy 5
2.2 Dual-Energy Computed Tomography 8
2.3 Existing SPR Calibration Models 10
2.3.1 Saito Model 10
2.3.2 Hünemohr Model 13
2.3.3 Tanaka Model 15
2.4 Machine Learning in Medical Imaging 16
2.4.1 Overview of Machine Learning in Radiology 16
2.4.2 Convolutional Neural Networks (CNNs) 17
2.4.3 CNNs for Denoising 18
3. Methodology **19**
3.1 Overview of SPR-Net 19
3.1.1 Objectives and Functional Requirements 19
3.1.2 Workflow and User Interaction 21
3.2 CAN-Based Noise Reduction 22
3.2.1 Architecture of the CAN Model 22
3.2.2 Training and Optimization 24
4. Application Development **26**
4.1 Software Architecture 26
5. Evaluation and Validation **27**
5.1 Benchmarking Against Ground Truth Values 27
5.2 Model Accuracy and Performance Metrics 28
5.2.1 Evaluation of SPR Calculation Methods 28
5.2.2 CAN Performance in Noise Reduction 30
6. Discussion and Future Work **32**
6.1 Comparison of Model Performance 32
6.2 Limitations of Current Implementation 34
6.3 Future Directions 36
6.4 Potential Clinical Impact 37
7. Conclusion **38**
Tables **40**
Figures **47**
References **54**

1. Introduction

Proton therapy has emerged as a highly precise form of radiation treatment, particularly beneficial for pediatric oncology due to its ability to prevent common side effects that were present in previously used radiation therapy [17]. Accurate calculations of proton stopping power is essential in ensuring that the proton beam deposits the majority of its energy at the tumor site, minimizing damage to healthy tissue. In the 1970s, dual-energy computed tomography (DECT) began to be explored for clinical treatment of cancer. DECT utilizes two different X-ray spectra to differentiate between different tissue types, allowing for a more accurate identification of tumors [7]. DECT provides a key advantage over the traditional single-energy CT method (SECT) by allowing for the use of computer algorithms that use the two different energy data to evaluate tissue attenuation, allowing for more accurate and powerful tissue characterization [7].

The primary objective of this thesis is to provide a streamlined tool to analyze DECT images for the calculation of proton stopping power in pediatric patients. We propose a tool, SPR-Net, that uses uploaded files outputted by the scanner, DICOM files, to calculate electron density (ρ_e), effective atomic number (Z_{eff}), and stopping power ratio to water (SPR) using the equations outlined in Saito [22, 23], Hunehmohr [8], and Tanaka [27]. Then, we utilize curve fitting to optimize function parameters for the selected patient study to calculate ρ_e, Z_{eff}, and SPR. All uploaded scans are processed through the convolutional autoencoder network (CAN) to clean potential noise or artifact obstruction that could influence calculations.

The thesis will explore essential background information regarding proton therapy, DECT, machine learning – including Convolutional Neural Networks (CNNs) – and challenges being faced in the estimation of SPR for pediatric dose planning. A review of existing methods of calculation is presented. We then provide a detailed walkthrough of the tool's development, including programming languages, frameworks, and libraries used. The architecture of the CAN will also be explained. Finally, we will expand on the implications of such a tool in clinical practice as well as future work following this tool's development.

2. Background

2.1 Proton Therapy

Proton therapy is an advanced form of radiation therapy that utilizes protons to deliver precise doses of radiation to tumors while minimizing damage to surrounding healthy tissues. After originally being proposed by Robert Wilson in 1946, the first attempts of proton therapy were conducted by Lawerence et al. in the mid 1950s on the cyclotron at the California Lawrence Berkeley Laboratory [29]. Since then, proton therapy has become more popular in diagnostics with various clinical applications.

Clinical uses of proton therapy have been explored through various trials. The National Institute of Health (NIH) sponsored clinical trials focusing on cancers of the head, prostate, lung, gastrointestinal tract, and, particular to this study, pediatric [17]. About 10,000 children are diagnosed with cancer each year in the United States [17]. This provides great motivation to optimize proton therapy to

most effectively treat cancers while minimizing offset of the beam to healthy tissue in these smaller patients.

Proton therapy utilizes protons, which are large particles with a positive charge, that deposit their energy at the end of the beam. While other radiological techniques gradually deposit energy along their path, protons deposit minimal energy until reaching an optimal depth, known as the Bragg Peak, where they release the majority of their energy before sharply dropping off. The success and wide availability of this method provides an advantage over traditional radiology since the maximum amount of energy can be deposited through careful positioning of the beam in each direction [22]. Various implementations of proton beam therapy are utilized in practice. Specifically, DECT has been seen to contribute to more accurate proton range predictions in dose planning.

Particularly in younger patients, proton therapy poses risk of secondary radiation-induced tumors [5]. Reducing this risk is driving further research for advancements in proton therapy. As a result, proton therapy is often used as a means of reducing side effects in patients, especially in younger patients where there is a higher risk of mortality, thus requiring significantly lower dosages than in usual circumstances. Compared to traditional radiotherapy, a key advantage of proton therapy is its means of reducing total energy deposited outside of the tumor into healthy tissue in the patient [22]. However, there are many uncertainties that present themselves when dose planning for patients. These are due to a number of factors, including patient setup, patient geometry, and machine calibration [22].

Biological considerations may also contribute to uncertainties in the planned dose range.

Statistical analysis can evaluate the effect of range uncertainties on dose planning for proton therapy. The standard deviation of the Dose Volume Histogram (DVH) can be used to quantify dose variability within a tumor or organ, which may help to reflect the impact of uncertainties in dose delivery when the DVH deviates from its expected distribution [23]. DVHs show the relationship between radiation dose and the volume of tissues receiving at least that dosage, which help evaluate tumor coverage and the risk to healthy organs. An example DVH is provided in Figure 1. The x-axis explains the radiation dose and the y-axis displays the volume of tissue receiving at least that dose. Thresholds for organs at risk (OARs) are typically set by institutional protocol and physician judgement, which can vary between patients. When determining the tradeoff between tumor control and spillage risk, there are several factors that pose challenges. For instance, if the tumor is adjacent to an OAR, the chances of spillover will become much higher. The shape and size of a tumor may require more beam angles that could result in further opportunities for spillage to surrounding tissues. Additionally, if the patient loses weight or the tumor shrinks, the dose distributions may change and result in the need for further analysis. While analyses can be done to evaluate the potential harm of uncertain dose calculations, in the case of pediatric patients where they are more susceptible to side effects from proton therapy, accurate dose calculations are preferred to ensure proper treatment for the patient.

2.2 Dual-Energy Computed Tomography

DECT is an imaging modality commonly used to identify and diagnose tumors and cancerous growth in the bodies prior to treatment dose planning. DECT utilizes two different X-ray energy ranges, measured in kilovoltage peak (kVp), with an optimal difference between the energy pair to enhance tissue contrast [27]. The energy separation should be sufficient to differentiate materials, but not so extreme that it compromises image quality or increases noise.

There are two main interactions between the X-ray beams and the materials being scanned: The Compton effect, which depends on the electron density of the material, and the photoelectric effect, which depends on the effective atomic number [22]. The photoelectric effect in particular is exploited to differentiate the materials being scanned, as calcium and iodine are susceptible to it.

The linear attenuation coefficient (μ) describes how much of the intensity of X-ray photons is reduced as they pass through a material. This coefficient is material specific and energy-dependent. To quantify tissue response for CT imaging, μ is converted to a CT number or Hounsfield Unit (HU), which is a normalized measure representing the X-ray attenuation of a tissue relative to that of water. The HU is calculated as follows:

$$CT = 1000 \cdot \frac{\mu_m - \mu_w}{\mu_w},$$

where μ_m and μ_w are the linear attenuation coefficients of the material and water respectively [6].

Low-kVp beams result in higher image contrast but increased noise, whereas high-kVp beams penetrate more deeply and reduce image noise, but may compromise contrast. DECT takes advantage of this tradeoff by combining low and high-kVp scans to maximize both contrast resolution and noise reduction. This dual-spectrum approach provides superior material discrimination compared to Single-Energy CT, which uses only one energy setting.

To support accurate material property estimations, we computed the low-energy (HU_L) and high-energy (HU_H) Hounsfield Units for each material insert across all available slice thicknesses. This was achieved by iterating through DICOM series grouped by kilovoltage peak and slice thickness. For each given slice in a given DICOM series, circular regions of interest (ROIs) were defined using known insert locations from phantom metadata. Within each ROI, we computed the average pixel intensity value by collecting all pixels within the specified circular boundary. These average pixels were then aggregated across all slices of a given scan to obtain a representative mean for that insert.

To distinguish between low and high energy scans, the kVp metadata in the DICOM headers was used to classify each scan as either low (< 100 kVp) or high (≥ 100 kVp). For each insert, the average HU_L and HU_H were calculated by taking the mean pixel values from the corresponding low and high energy scan groups. Figure 2 provides an example of the resulting HU_L and HU_H values for each material from a scan series with 2.0 mm slice thickness used in our study.

Photon starvation, which is when an insufficient amount of photons reach the tissue from the beam, can result in an increase of noise. Patients with atypical geometry, such as those with larger BMI or children, may pose a need for further noise optimization to improve image quality [1]. Artifacts due to beam hardening are more likely to be present in calcium-rich materials and can interfere with SPR estimation [1]. Pre or post-processing of the reconstructed images may be paired with motion compensation algorithms to reduce artifacts in the image [2].

Noisy scans have many adverse effects on SPR estimation. In the presence of gaussian noise, the distribution of SPR becomes asymmetrical – which contradicts the linear relationships between CT-ρ_e [15]. Noise is especially concerning to our study as it has been observed to produce the largest mean shift in soft tissue estimations for the Hünemohr and Saito methods that we will be using [15]. Therefore, the reduction of noise will be emphasized throughout development of the tool.

2.3 Existing SPR Calibration Models

2.3.1 Saito Model

Masatoshi Saito [25] proposed a simple conversion from the energy-subtracted CT (ΔHU) number to the relative electron density (ρ_e) through a singular linear relationship. The ΔHU-ρ_e conversion has since been the foundation of many additional DECT analyses. First, the energy-subtracted CT number, ΔHU, is defined as a weighted combination of high and low-kVp CT numbers:

$$\Delta HU = (1 + \alpha)HU_H - \alpha HU_L,$$

where HU_H and HU_L are the high and low CT numbers and α is the weighing factor.

To determine the optimal value for α, a range of possible values from 0 to 1 are defined. The following steps are repeated for each potential value. First, ΔHU is calculated using the previous equation with the current candidate for α. A linear model relating ΔHU and the true known electron densities, ρ_e^{true}, of the phantom materials, which is modeled as:

$$\rho_e^{cal} = a\frac{\Delta HU}{1000} + b,$$

where a and b are fitting parameters obtained from linear regression. For each candidate of α, the coefficient of determination r^2 is computed to assess how well the model fits the known ρ_e^{true} values. The value of α that yields the highest r^2 is selected as the optimal weighting factor for the scanner. In the ideal case, $a = 1$, $b = 0$, and $r^2 = 1$, indicating a perfect linear match between predicted and true electron density.

The ΔHU-ρ_e conversion was used to perform DECT image simulations that mimicked a second-generation dual-source CT scanner. A calibration curve was obtained for a study containing 16 inserts. The simulated line provided a predictable linear relationship over the ranges of ρ_e in the study, as displayed in Table 1. Between the ideal and simulated values, the maximum difference was -0.7%. The resultant curve's r^2, a, and b values were 0.99997, 0.999, and 1.001

respectively, which is close to unity and was expected for the ideal case. Due to its demonstrated accuracy in estimating electron density with minimal deviation from ideal values, the method proposed by Saito [25] is used in this study as a reliable calibration model.

Saito and Sagara [26] sought to expand on the ΔHU-ρ_e conversion to derive effective atomic numbers via electron density. This method, dubbed DEEDZ, utilized the equations presented in [25] to propose a reduced CT number, u_k:

$$u_k = \frac{HU_k}{1000} + 1,$$

where k = L or H depending on the energy spectra used. According to its definition, u_k may also be written as a linear attenuation coefficient relative to water:

$$u_k = \frac{\mu(E_k)}{\mu_w(E_k)},$$

which may be expanded to based on the definition of linear attenuation:

$$= \rho_e \left\{ P(E_k) + Q(E_k)\left(\frac{Z_{eff}}{Z_{eff,w}}\right)^m \right\},$$

where $P(E_k)$ and $Q(E_k)$ are the functions for the Compton Scattering and photoelectric absorption, respectively, E is the photon energy, $Z_{eff,w}$ is the effective atomic number of water, and m is quoted in the literature to be between 3 and 4. Considering the special case of water where $u_k = \rho_e = 1$ and $Z_{eff} = Z_{eff,w}$, the constraint $P(E_k) + Q(E_k) = 1$ is formed, allowing us to assert:

$$\left(\frac{Z_{eff}}{Z_{eff,w}}\right)^m - 1 = \lambda_k\left(\frac{u_k}{\rho_e} - 1\right),$$

where $\lambda_k = \frac{1}{Q(E_k)}$ and is a material-independent proportionality constant. Now, the reduced form of this relationship requires only two fitting parameters, m and λ_k, to be determined. For the sake of our study, we assume $m = 3.3$ and use this to calculate λ_k, which is the linear relationship $(Z_{eff,true}/Z_{eff,w})^m - 1$ and $\mu_L/\rho_{e,cal} - 1$. Similar to α, we take the value that provides the largest r^2.

The DEEDZ method was then used for analysis against tissue surrogates with similar chemical compositions and mass densities as commercially available tissue substitution phantoms. The determined ρ_e and Z_{eff} were then applied to human equivalent tissues. The resulting Z_{eff} from the simulations as seen in Table 2 were found to be within ±0.3% relative deviations, excluding the thyroid with a larger deviation of -3.1% due to the appearance of iodine in its concentration which poses a large atomic number ($Z = 53$). Thus, the DEEDZ method expands on the previous ΔHU-$\rho_e$conversion to produce highly accurate Z_{eff} formulations to be used in SPR calculations and will be used alongside the ΔHU-$\rho_e$conversion in this study.

2.3.2 Hünemohr Model

Hünemohr et al [9] proposed a method that utilizes DECT imaging techniques to predict SPR with respect to $\rho_e/\rho_{e,w}$ and Z_{eff} across a study consisting of twenty materials including tissue surrogates and artificial components. SPR was calculated by a simplified version of the Bethe equation:

$$SPR_w = \frac{\rho_e}{\rho_{e,w}} \cdot \frac{12.77-(a \cdot Z_{eff}+b)}{8.45},$$

where a and b are fitting parameters. The electron density ratios, $\rho_e/\rho_{e,w}$, were modeled as:

$$\frac{\rho_e}{\rho_{e,w}} = c_e \cdot \left(\frac{x_1}{1000} + 1\right) + (1 - c_e) \cdot \left(\frac{x_2}{1000} + 1\right),$$

where c_e is a fitting parameter and $x_1$and x_2 are the CT numbers for the study. Note that CT numbers and HU are used interchangeably. This formulation uses the reduced CT number, u_k, proposed by [25]. The true effective atomic number, Z_{eff}, for a given material was evaluated as:

$$Z_{eff} = \left(\frac{\Sigma_i n_i Z_i^{n+1}}{\Sigma_i n_i Z_i}\right)^{\frac{1}{n}},$$

where n_i is the number of atoms per unit volume of type i and Z_i is the atomic number of an element in a material's composition. The coefficient n is a CT-specific parameter. This true value is then used to determine the fitting parameter d_e to calculate the estimated Z_{eff} of a given material:

$$Z_{eff} = \left(\left(\frac{\rho_e}{\rho_{e,w}}\right)^{-1}\left(d_e\left(\frac{x_1}{1000HU} + 1\right) + \left(Z^2_{eff,w} - d_e\right)\left(\frac{x_2}{1000HU} + 1\right)\right)\right)^{\frac{1}{n}},$$

where n is the same parameter as the previous equation.

Findings for $\rho_e/\rho_{e,w}$ and Z_{eff} in the experimental analysis yielded a 0.4% and 1,7% mean accuracy for the tissue surrogates, respectively. SPR was calculated using the simplified Bethe equation and the calculated parameters using the proposed methodology with a yield of 0.6% mean accuracy for the experimental

values. The accuracy presented in this method demonstrates its strong predictive capabilities, thus making it a well suited option to use in this study.

2.3.3 Tanaka Model

Tanaka et al [30] compiled the works of Saito, Sagara and Hünemohr into a dose calculation plan that utilizes both papers in order to achieve an accurate SPR. Using the theoretical basis of the Bethe equation, SPR is approximated as:

$$SPR = \rho_e[1 - \frac{ln\frac{I}{I_w}}{ln[\frac{2m_e c^2\beta^2}{I_w(1-\beta^2)}\}-\beta^2}],$$

where I and I_w are the mean excitation energies of the material and water, m_e is the rest electron mass, c is the speed of light in a vacuum, and β is the speed of the projectile proton relative to light. The formulations for ρ_e proposed in [22] were used along with the formula for the ratio of effective atomic numbers. Instead of the reduced CT number proposed by Saito [25], Tanaka [30] utilized a more general form from Han et al [8]:

$$u_L = a_L\frac{HU_L}{1000} + b_L,$$

where a_L and b_L are also fitting parameters. To calculate $ln\frac{I}{I_w}$, an additional function proposed by Saito and Sagara [26] was used:

$$ln\frac{I}{I_w} = c_1\{(\frac{Z_{eff}}{Z_{eff,w}})^m - 1\} - c_0,$$

where c_0 and c_1 are human-tissue specific constants for the slope of intercept from the calibration line of the effective atomic number ratio.

The process proposed by Tanaka et al [30] reduces the need for the task of optimizing the calibration parameters for varying patient geometry and body size as the absolute difference in SPR between the high and low energy scans, ΔSPR, for tissue regions was only 0.3% of the reference value when using the DEEDZ conversion compared to the conventional stoichiometric SECT-SPR conversion as seen in Table 3. Due to its minimal reaction to variations in λ_L for the Z_{eff} calibration presented in [26], this method provides the advantage of not needing to be calibrated for every patient with a different body size in order to minimize the dose-calculation errors. Therefore, this method was utilized in our study due to its lack of limitations with smaller pediatric patients.

2.4 Machine Learning in Medical Imaging

2.4.1 Overview of Machine Learning in Radiology

Machine learning encompasses the study of algorithms which can learn relationships or patterns from data and make decisions based on them [18]. Due to the versatility in its applications, machine learning has a wide range of use in radiology. Machine learning is utilized across medical image segmentation, image registration, and content-based image retrieval – to name a few use cases [32]. Machine learning has also been used to automate the estimation of radiation doses from CT scans. Deep CNN classifiers trained on CT data sets have shown an accuracy upwards of 96% when estimating organ specific radiation dose estimations [4].

There are various barriers that stand between the widespread use of machine learning models in clinical practice. Supervised machine learning models require large datasets with annotated labels, which can be costly and time consuming for radiology professionals to produce [4]. There is also the assumption of blame, where physicians who rely on machine learning powered tools must take ownership of the results produced by the model [4]. Ethical and regulatory measures are necessary to ensure anything outputted by a machine learning system is being cross validated and approved by medical professionals before they are used to influence patient care.

2.4.2 Convolutional Neural Networks (CNNs)

Implementations of machine learning span across various fields, including neural networks. A neural network mimics the human brain, where the network identifies patterns in the data by adapting to changes in the input to generate the best possible result [18]. Neural networks are made up of a system of nodes called neurons, which are activated when the expected data has arrived [32].

Convolutional neural networks are a class of deep neural networks primarily used for image processing. They are composed of three foundational building blocks, dubbed convolution, pooling, and fully connected layers. The former extracts features from the image while the latter applies the extracted features to a final output [33]. A key benefit to CNNs is their ability to reduce the number of parameters by removing trivial features without having to extract them manually

[16]. CNNs provide a wide range of use cases including but not limited to: image classification, object detection, and image segmentation [16].

Due to their layered structure, CNNs are able to extract more meaningful high level features, which makes them well suited for tasks involving complex data structures. Due to their scalability, CNNs are a critical component in modern imaging workflows.

2.4.3 CNNs for Denoising

Medical images obtained by CT machines are subject to various degrees of noise that may be obtained during transmission and acquisition, which can complicate the analysis of DECT images, subsequently prolonging a diagnosis [11]. The most common sources of noise in CT are quantum noise and electronic noise and are increased by material separation [22]. Quantum noise refers to the natural statistical fluctuations present in X-ray signals due to its random nature. It is not reproducible or repeatable, making it challenging to mitigate when training a denoising model [31].

Classic auto encoder and decoder models fall short when working with more robust datasets. There have been various models designed for the purpose of denoising medical images. Residual learning, proposed by Jifara et al [11] provides a deep network that is ideal for small training data sets wherein it may learn the noise residual from the noisy image itself. This model, however, does not perform well against noise from image compression. Chen et al [3] propose a model that

utilizes learning-based noise reduction and patch encoding to circumnavigate the need for noise-specific methods to filter and restore clean images.

Beyond the architecture of the model, choosing an appropriate loss function can also have a staggering impact on its performance. Loss functions compare the output of the neural network with the ground truth values, wherein the model will then output a result that has a minimized loss [34]. Loss functions for image denoising tasks can be separated into pixel-level loss, perceptual loss, and adversarial loss [14]. Pixel-level loss measures the pixel-wise error between a denoised image and the clean image and is often calculated through mean absolute error (MAE) and mean square error (MSE). Perceptual loss is obtained from pixel-level loss and is used for feature maps. Adversarial loss is a measure of distance between distributions of a denoised image and clean image. When assuming the goals of a denoiser are to reduce noise while not introducing blur to the image, MAE and MSE proved to be the most optimal choices [14].

3. Methodology

3.1 Overview of SPR-Net

3.1.1 Objectives and Functional Requirements

The primary objective of SPR-Net is to streamline the analysis of DECT datasets for the purpose of estimating SPRs in pediatric cancer therapy planning. The tool was designed to assist researchers and clinicians by providing an end-to-end workflow: from DICOM input handling and image preprocessing to insert detection, *HU* extraction, and model-based SPR calculations. By integrating

multiple DECT-based models and offering an easy-to-use interface, the tool supports comparative evaluation and calibration of predicted values against known material properties, ultimately enhancing the reliability and efficiency of DECT data analysis.

Functionally, this tool is required to accept raw DICOM folders containing both a high and low kVp image series, regardless of file naming. It classifies and separates the series using metadata present in the DICOM files and prepares them for analysis through basic preprocessing including rescaling, normalization, and spatial alignment. The tool automatically detects the positions of known insert configurations for both head and body sized phantoms and extracts region-specific HU values for both energy levels. Users are given the ability to adjust the insert boundary radius to assess how changes in region size affect HU measurements and model outputs. Users also have the ability to denoise their images using a trained CNN to clean any noise or artifacts that may be present on the images before proceeding with analysis.

The tool incorporates three SPR estimation models from existing literature, as described in Section 2.3. These models are calibrated against the truth tables present in their respective literature, which contain validated physical properties for each phantom insert material. Model outputs are then compared against these true values to evaluate prediction accuracy. Finally, the tool includes a React-based graphical user interface that allows users to upload data, visualize DECT images, inspect insert overlays, select SPR models, and view output results in real time.

Results include identified materials, model based ρ_e and Z_{eff} predictions, and error metrics relative to the ground truth.

3.1.2 Workflow and User Interaction

The tool was developed with a user-centered design approach. Upon launching the application, users are prompted to upload a folder containing DICOM series. These series must include a folder for the high kVp and the low kVp due to the nature of the imaging modality. The backend will parse the DICOM metadata to classify each series and pair them accordingly.

Once the data is processed, the user is presented with a preview of the images taken by the machine. From this view, they can iterate through them and select their image pairing. Users may also elect to utilize the denoising feature to clean any noise on their selected images before continuing,

The tool automatically detects and overlays insert locations onto the images. A slider allows the user to adjust the radius around each insert to fine-tune the region of interest in HU calculations. The user must also input information about the scan parameters, such as high and low kVp levels and slice thickness. The user will then select one of the models to use with their data.

A reference table is then generated that contains the list of identified materials from the scans, model outputs for ρ_e, Z_{eff} and SPR. To evaluate the accuracy of ρ_e and Z_{eff}, two statistical metrics were reported: the Root Mean Square Error ($RMSE$) and the coefficient of determination (R^2). $RMSE$ measures the average

magnitude of the errors between predicted and true values, where lower values signify better performance. R^2 indicates the proportion of variance in the true values that is predictable from the model outputs, where a score closer to 1 suggests a stronger correlation and reliable predictions. The user also has the option of comparing the results of the study with another study of their choice by either selecting a new model to benchmark against or by adjusting the current selection through the radius size or denoising feature. The full stage of user interactions are summarized in the workflow diagram shown in Figure 3. This diagram outlines the sequence from DICOM input to model output.

3.2 CAN-Based Noise Reduction

3.2.1 Architecture of the CAN Model

To effectively reduce noise in DECT images while preserving the structural details of phantom substitutes, we implemented a convolutional autoencoder neural network. The autoencoder architecture consists of two primary components: an encoder that compresses the input image into a lower-dimensional representation (latent space), and a decoder that reconstructs the image from this compact encoding. This process allows the model to learn the most salient features of the input data while disregarding irrelevant noise.

Figure 2 shows the overall workflow of our denoising pipeline. The original clean DECT image (128x128 pixels) is synthetically degraded by adding Gaussian noise, simulating a low-dose acquisition:

$$\hat{x} = x + \eta,$$

where x is the clean input image, $\widehat{x}$ is the noisy input, and η represents the added noise from a Gaussian distribution. Although Gaussian noise is not the predominant form of noise found in DECT imaging, at higher photon flux conditions, quantum noise converges towards a Gaussian distribution, making Gaussian noise a reasonable benchmark for this denoising experiment [12]. The noisy image is then passed through the encoder, which progressively compresses the input through the encoder network f_θ, which maps it to a latent representation z:

$$z = f_\theta(\widehat{x}).$$

The latent representation of the input data is essential to understanding relationships of the features that are observable in the input data. Reducing the dimensionality of the input data while maintaining its essential features makes for easier handling and processing of complex input data. This representation is then passed into the decoder.

After reaching the bottleneck layer and obtaining the latent representation, the decoder g_ϕ reconstructs the denoised image $\bar{x}$ from the latent space:

$$\bar{x} = g_\phi(z) = g_\phi(f_\theta(\widehat{x})),$$

where $\bar{x}$ is constructed by minimizing the reconstruction error of the latent data.

The detailed architecture of the CAN is shown in Figure 3. The encoder comprises three layers with decreasing dimensionality: 700, 500, and 300 neurons, respectively. These layers extract increasingly abstract features while reducing spatial resolution. The bottleneck is the center of our network and consists of 100

neurons that serve as the compressed latent representation of the input image. The decoder mirrors the encoder architecture, with three symmetric layers in increasing dimensionality, reconstructing the denoised output image from the bottleneck features.

Each layer uses ReLU (Rectified Linear Unit) activation to introduce non-linearity. ReLU is defined as $f(x) = max(0, x)$, which means it outputs 0 for all negative inputs and passes positive values unchanged. This activation function is widely used in convolutional neural networks due to its simplicity and computational efficiency, allowing the network to learn sparse and efficient representations. These properties are particularly beneficial in denoising applications, where the model must emphasize important features while suppressing irrelevant noise.

The output layer, however, uses a sigmoid activation function, defined as $\sigma(x) = \frac{1}{1+e^{-x}}$. This function maps input values to a range between 0 and 1, making it suitable for constraining pixel intensity values in normalized grayscale images. In the context of image denoising, sigmoid ensures that the reconstructed pixel values remain within a realistic and interpretable range, which is crucial for visual accuracy and quantitative evaluation later.

3.2.2 Training and Optimization

The dataset consisted of 348 128x128 pixel grayscale DECT images of phantoms with clearly defined inserts. Clean images x were artificially corrupted

using Gaussian noise to generate corresponding noisy images $\widehat{x}$. This synthetic augmentation allowed the model to learn an array of noise characteristics without the trial and error of manually taking scans that would result in a diverse set of noise.

Each training pair $(\widehat{x}, x)$ was normalized to have pixel intensities within the range $[0, 1]$. The dataset was then split into 80% for training, 10% for validation, and 10% for testing.

The network was trained by minimizing the Mean Squared Error (MSE) loss function:

$$MSE = \frac{1}{n}\sum i = 1^{n}(x_{i} - \widehat{x}_{i})^{2}$$

where x_{i} and $\widehat{x}_{i}$ denote the ground truth and reconstructed pixel values, respectively, and n is the total number of pixels per image. This encourages the model to produce denoised outputs that are as close as possible to the clean reference images in a pixel-wise sense.

The model was trained using the Adam optimizer, a variant of stochastic gradient. This optimizer was chosen due to its dynamic learning rate, which allows each parameter to learn on past gradients and get past local minima more quickly, and that it uses fewer hyperparameters compared to standard SGD. Adam converges much faster compared to standard SGD and requires minimal tuning, which made it an effective option for this project. Training was conducted for 50 epochs with a batch size of 16. Early stopping was employed by monitoring the

validation loss. Training would halt if the validation loss failed to improve after 10 consecutive epochs, thereby reducing the risk of overfitting.

4. Application Development.

4.1 Software Architecture

The software architecture consists of a modular client-server architecture consisting of two components: a React frontend for user interaction and a FastAPI backend for data processing, computation and model execution . This ensures separation between user interface logic and computational functionality.

The React frontend manages file uploads, parameter selections, and real-time visualization of results. It communicates with the backend through HTTPs requests. The user interface is designed to be intuitive and responsive, allowing users to preview DICOM images, adjust insert boundaries, clean images, select analysis models, and view outputs including calculated HU, ρ_e, Z_{eff} and SPR. interactive elements such as sliders are implemented to provide fine control over analysis parameters. Figures 4, 5, 6, 7, and 8 showcase the frontend in order of user flow.

The backend is responsible for the core computational tasks. Upon receiving a HTTP request from the frontend, it will access the associated API to perform the requested task. The order of operations begin with the backend parsing and organizing the DICOM files using the `pydicom` library and performing preprocessing such as image rescaling and alignment using `NumPy` and `OpenCV`. If the user chooses to clean the image of any noise, this will be done next before any

calculations occur. Insert detection is guided by predefined phantom layouts for different phantom sizes and insert specific HU are calculated for both the high and low kVp scans. The backend then applies the selected SPR model and calculates the calibrated parameters. The calculated results, ground truth, identified materials, and error metrics results are then returned.

The application enables modular enhancements, specifically for the addition of further models for SPR calculation. The backend contains individual Python files that contain the code for each model. Any additional model from new literature must simply be translated to Python code and added to its own file that can be called upon the main FastAPI application within the backend. This allows for the tool to grow as research in DECT grows as well, ensuring long term support for clinicians, researchers, and patients.

5. Evaluation and Validation

5.1 Benchmarking Against Ground Truth Values

To measure the effectiveness of the SPR models, we compared the calculated values from our data with the values presented in the literature for each study. These served as our ground truth values and were used in all error metrics. For ρ_e and Z_{eff}, the values for both parameters were obtained from the literature for each material used in our study.

While these models are meant to be calibrated to work for any study setup, it is important to note differences in our experimental phantoms and those presented in the literature. Particularly for the Hünemohr model which calculates Z_{eff} based

on the elemental composition of the phantom insert, there will be a slightly higher error than that for ρ_e. This is due to the slight and subtle differences in elemental composition between the phantom substitutes used in this study and that of Hünemohr. Thus, those using the application must familiarize themselves with the contents of their tissue substitutes and how they differ from those presented in the literature of the model they are utilizing.

5.2 Model Accuracy and Performance Metrics

5.2.1 Evaluation of SPR Calculation Methods

The three models presented in this application must have their respective hyperparameters fit to the data. Drawing mathematical relationships between HU, ρ_e, Z_{eff} and SPR is complex – even if the physics behind these subjects are known (e.g., the Bethe equation, photoelectric effect, Compton scattering). Analytically, this conversion is difficult due to various systemic factors. HU values are influenced by both material composition and scanner-specific factors, which may differ from those used in the existing literature. Beam hardening, noise, and other real-world effects also influence the image. To compensate, these models are empirical and must be fitted to adapt the model to our setup.

Error metrics were calculated for ρ_e and Z_{eff} as they are direct model outputs that are measurable and validated individually. SPR, however, is a derived value. Any error in ρ_e and Z_{eff} propagates through the Bethe formula nonlinearly, thus

masking issues in the underlying model. For this reason, only the aforementioned parameters were used for error calculations.

$RMSE$ and R^2 were calculated. $RMSE$ measures the average magnitude of the error between our predicted parameters and the ground truth values present in the literature. Because ρ_e and Z_{eff} are continuous and quantitative variables, it gives us an absolute sense of error while penalizing larger errors more heavily. This is especially important in clinical settings where even small inaccuracies can impact SPR. R^2 was calculated to explain how well the calibrated models are fitting for our range of parameters. This is used to track how reliable and explanatory our model is across scans.

To evaluate the accuracy of our application and validate its underlying methodology, we conducted an experimental study using a DECT scan pair at 80/120 kVp. The study included 8 tissue-equivalent materials with known physical properties. For each insert, the application calculated the relative ρ_e, Z_{eff}, and SPR. The results of this analysis are displayed in Tables 5, 6, and 7, which summarizes the calculated values for each material along with the fitted model parameters and corresponding error metrics across the three models implemented in the application.

The purpose of this comparison was to determine which method yields the most accurate results within the phantom dataset. The Saito method (Table 6) yielded the strongest performance with the lowest $RMSE$ for Z_{eff} and the highest R^2

for Z_{eff} as well. The Tanaka method (Table 5) yielded identical results for ρ_e – which makes sense considering it uses the same equations as presented in Saito [25] and [26]. Hünemohr presented the worst results, with a staggeringly low R^2 of 0.22917 and 0.68743 for ρ_e and Z_{eff}, respectively.

These results suggest that the Saito model provides the most balanced and reliable estimates of physical properties in this DECT study. The resulting accuracy in ρ_e and Z_{eff} directly enables reliable estimation of SPR, which is critical for dose planning.

However, it is important to note that these findings are specific to the materials, scanning conditions, and experimental setup used in this study. Different scanners, kVp pairings, or tissue compositions may yield different outcomes, and other studies may find alternative models more appropriate. The goal of this tool is not to universally prescribe one method, but to allow researchers and clinicians to test multiple models on their own data and identify which approach provides the most accurate predictions for their specific use case. This flexibility is especially valuable when using DECT to inform SPR calculations for pediatric patient dose planning, where even small deviations could prove to be fatal.

5.2.2 CAN Performance in Noise Reduction

The goal of the CAN denoiser was to reduce variations in HU values across noisy images in our dataset. To measure this, we used three different error metrics:

PSNR, SSIM, and LPIPS. The results of these metrics are detailed in Table 4 and explained in more detail below.

PSNR, or peak signal-to-noise ratio, is the measure of the maximum possible pixel value in an image (e.g., 255 in 8-bit images). This measures how close a denoised image is to the original image, pixel by pixel. However, this does not correlate well with perceived image quality. Changes in brightness, contrast, or other visual quality metrics can affect PSNR. PSNR is calculated as:

$$PSNR = 20 \times log_{10}(\frac{MAX_i}{\sqrt{MSE}}),$$

where MAX_i is the maximum possible pixel value. We obtained a PSNR of 35.62 dB, which is well within the range of a good score for 8-bit images (30 dB to 50 dB).

SSIM is a structural similarity index measure, which calculates the perceived quality of the image based on luminance, contrast, and structure. SSIM compares local patterns of pixel intensities that have been normalized for luminance and contrast. Scores range from -1 to 1, where 1 is perfect similarity. SSIM also accounts for structural information and is more aligned with human visual perception. Our SSIM score was 0.9058, which indicates that the noisy and clean images are very similar to each other.

LPIPS is the learned perceptual image path similarity. Deep features extracted from images using pre-trained neural networks compute the distance between image feature maps at multiple layers to capture high-level perceptual differences. A lower score indicates that the images are similar. Compared to PSNR,

LPIPS can capture semantic differences that aren't noticed by pixel wise metrics. Using AlexNet, we obtained a score of 0.0635.

According to the presented metrics, our CAN is effective at reducing noise while maintaining pixel values to ensure that HU values are preserved for our analyses purposes.

6. Discussion and Future Work

6.1 Comparison of Model Performance

An important feature of SPR-Net is its ability to facilitate direct comparison between different DECT models for a given image pair. Beyond a single evaluation, SPR-Net allows users to test the same image pair against multiple models, providing a side-by-side comparison of predicted values. This comparative approach helps identify which model yields results that best align with the known ground truth under specific imaging conditions.

To assess the reliability and flexibility of SPR-Net, we conducted a controlled validation study designed to evaluate how changes in model selection and processing parameters influence the calculated values of ρ_e, Z_{eff}, and SPR. The goal of this validation was not only to verify output accuracy, but also demonstrate the extent to which different DECT models and user-defined parameters can affect material property estimation, an important consideration for dose planning in proton therapy.

This experimental setup utilized a 70/140 kVp pair of eight known tissue-equivalent inserts taken at a slice thickness of 4.0 mm. We performed a

series of tests across multiple conditions, namely switching between the Tanaka, Saito, and Hünemohr models and adjusting the insert radius used for averaging. For each configuration, we recorded the resulting ρ_e, Z_{eff}, and SPR.

Tables 8, 9, and 10 present the physical properties calculated using the Tanaka, Saito, and Hünemohr methods under three different radius settings: 100%, 50%, and 25% of the original insert size. Each table presents side-by-side outputs from all three models for direct comparison. In general, Tanaka consistently produced the most clinically plausible SPR values, maintaining relative stability across all three radii. For example, for the inner bone, Tanaka predicted SPRs of 1.28746 (100%), 1.28611 (50%), and 1.28868 (25%), suggesting strong resistance to ROI size fluctuations. This resistance is particularly important when identifying an appropriate model for pediatric patients, whose smaller body size and localized tumors often result in smaller ROIs. In such cases, the ability to derive a consistent and accurate SPR, regardless of insert or lesion size, is essential for ensuring proper proton range calculations and avoiding treatment inaccuracies due to scaling errors.

The Saito model, while reasonably consistent across radii, tended to underpredict Z_{eff} and showed sensitivity to radius changes. The LN-350 SPR values varied from 0.27250 (100%) to 0.31006 (50%) and 1.30882 (25%), indicating that reduced spatial averaging of the HU shifted predictions in low-density regions. This is not ideal for the case of pediatrics, where fluctuations due to body size should not be a concern when planning for patients.

In contrast, the Hünemohr model yielded consistently low SPR values across all material and radii, which may not reflect realistic stopping powers for human tissue. For cortical bone, a material that requires a high stopping power due to its density, the model predicted SPRS of 0.03368, 0.03514, and 0.03517 across the three radii, which were nearly two orders of magnitude lower than those from Tanaka or Saito. This consistent underestimation highlights a potential limitation of the model's calibration for this dataset or kVp pairing.

These results reinforce the importance of customizable comparison in SPR-Net. The ability to vary both model and preprocessing parameters within a single platform allows users to test multiple hypotheses, observe their effect on clinical metrics like SPR, and identify the best-fitting approach for their data. With full export of configuration and results, SPR-Net supports transparent, reproducible research that enables evidence-based model selection for both academic and clinical environments.

6.2 Limitations of Current Implementation

Presently, this study utilizes tissue substitute phantoms in our analysis and application workflow instead of actual patient scans. There are several reasons for this decision. Phantoms contain tissue substitutes with known physical properties, which allows researchers to compare predicted results from model outputs to known values provided by the manufacturer. This way, we are able to evaluate model accuracy without the variability of human tissue like hydration, tissue composition, or movement. This makes testing reproducible and eliminates ethical concerns

surrounding prolonged radiation exposure, particularly for pediatric patients. Utilizing phantoms also introduces important limitations. Tissue substitutes cannot fully replicate the physiological complexity of real human tissues. Biological variability in organ shape, texture, disease presentation, and tumor composition is absent from phantom studies. The limited range of materials in phantom inserts may also fail to capture overlapping tissue properties seen in patients.

This study is also limited in the models it chooses to implement. Presently, we use three models. The Tanaka model especially combines the efforts of both Saito and Hünemohr into a single directed workflow. This introduces dependencies between the models where any inaccuracies present in one may propagate through the combined framework. As such, understanding the individual strengths and limitations of each model is essential to interpreting results.

Additionally, limitations exist in the current noise reduction strategy. The denoiser was trained on synthetic Gaussian noise added to clean phantom images. While this provides a controlled and reproducible training environment, it does not fully reflect the noise characteristics typical of clinical CT imaging, where quantum noise is more prevalent. The use of Gaussian noise was intended as an initial benchmark to assess the feasibility of automated denoising in the DECT workflow. However, before clinical use, it will be necessary to retrain or further tune the denoiser using more realistic noise models derived from scan simulations to capture the stochastic properties of quantum noise. Expanding the denoiser in this way will

ensure better preservation of anatomical structures and improve confidence in the clinical applicability of the generated outputs.

6.3 Future Directions

Moving forward, the project offers several key opportunities for expansion and refinement. One of the most immediate directions involves the implementations of additional DECT models for SPR estimation. While this study focused on three core models, incorporating alternative or emerging models could offer deeper insights into model performance across different imaging protocols and tissue types. Comparative analysis among a broader set of models would allow for more nuanced validation and the potential to develop approaches that combine model outputs to reduce uncertainty and improve overall prediction accuracy.

Another critical future step is the integration of patient scan data into the analysis pipeline. While phantoms provide a reliable baseline for controlled testing, they cannot fully capture the anatomic complexity and biological variability present in clinical settings. Expanding this work to include patient datasets will allow for more rigorous testing of model generalizability and will be essential for clinical translations.

Additionally, future efforts will focus on improving the software's user interface and integrating it more seamlessly with research workflows, including automated preprocessing and batch processing for offline and asynchronous processing of big data. Future work will also focus on improving the performance of the CAN used in this application. Future improvements to the denoiser must be

done so on a training set that includes quantum noise to increase the clinical applicability of this tool. While the current CAN supports denoising of images through the simple autoencoder structure, further learning of contextual features may occur through the use of more involved architectures, like Unet or residual learning. These enhancements will make the model more robust and increase confidence in its use for accuracy improvements.

6.4 Potential Clinical Impact

Accurate estimation of SPR is essential for effective proton therapy, particularly in pediatric patients where minimizing radiation exposure to healthy tissue is critical. Our data-driven approach integrates machine learning with DECT analysis to improve the precision and reliability of stopping power estimates. By benchmarking model predictions against tissue-substitute phantoms with known properties, the framework ensures foundational accuracies before eventual clinical translation.

The proposed workflow has the potential to streamline pre-treatment planning in proton therapy by reducing the uncertainties in dose calculations. For pediatric patients, where treatment margins must be minimized to avoid damage to developing organs and tissues, even small improvements in SPR can lead to more targeted radiation delivery and better long-term outcomes. Moreover, the integration of the CAN allows researchers to offset the noise typical of DECT scanning, improving confidence in treatment plans and reducing the need for additional imaging.

Looking ahead, this project represents a collaborative effort between the New York Proton Center (NYPC), where the developed application will be directly integrated into ongoing research within their imaging and treatment planning laboratories. By leveraging this tool, NYPC researchers aim to streamline DECT analysis and improve SPR estimation for pediatric patients. The ability to automate insert detection, material identification, and parameter calculation will support more efficient and reproducible experiments, while also reducing the manual workload on medical physicists and research staff. This project bridges the gap between research innovation and practical implementation in a clinical research environment.

7. Conclusion

This study presents a machine learning-assisted workflow for analyzing DECT images to improve SPR estimation. By utilizing tissue substitutes phantoms with known material properties, the analysis ensures a controlled and reproducible testing environment for evaluating the accuracy and reliability of various DECT models. The integration of these models has enabled the automatic calculation of ρ_e and Z_{eff}. The embedded CAN allows for denoising of DECT images for improved accuracy in HU calculation.

However, the study is not without limitations. Phantoms, while valuable for validation, do not capture the full biological variability of real patients, which may affect generalizability. Additionally, the dependency of the Tanaka model on the assumptions of the Saito and Hünemohr models introduces compound sources of

error that must be carefully considered when interpreting results. Future work will address these challenges by incorporating clinical datasets and expanding the scope of models used.

This work was developed in collaboration with the NYPC, where the application will be adopted into their research workflow to support ongoing DECT-based studies. With continued development, the methods and tools presented in this study represent a meaningful contribution toward safer, more accurate, and personalized proton therapy planning.

Tables

Material	ρ_e (ideal)	HU_H	HU_L	ΔHU ($\alpha = 0.459$)	ρ_e^{cal}	$\rho_e^{cal} - \rho_e$(%)
Lung	0.258	− 729.7	− 702.9	− 742.1	0.260	0.2
Adipose tissue	0.951	− 61.4	− 87.0	− 49.6	0.952	0.1
Yellow Marrow	0.982	− 31.6	− 59.6	− 18.7	0.982	0.0
Solid Water	0.990	− 4.2	9.6	− 10.5	0.991	0.01
Water	1.000	1.1	5.5	− 0.9	1.000	0.0
Breast	1.014	8.7	− 1.0	13.2	1.014	0.0
Brain	1.035	36.9	43.1	34.1	1.035	0.0
Eye lens	1.055	53.7	53.3	53.9	1.055	0.0
Cartilage	1.083	92.2	111.4	83.4	1.084	0.1
Spongiosa	1.150	208.6	338.5	148.9	1.150	0.0
PVC	1.246	510.2	1094.2	242.2	1.243	− 0.3
Femur	1.278	401.0	685.4	270.5	1.271	− 0.7
Ribs	1.441	612.5	988.2	440.0	1.440	− 0.1
Mandible	1.577	809.0	1326.1	571.6	1.572	− 0.5
Cortical bone	1.781	1086.8	1740.2	786.9	1.787	0.6
PTFE	1.817	835.9	877.5	816.8	1.817	0.0
Aluminum	2.347	1689.8	2427.3	1351.2	2.350	0.3

Table 1: Electron density values and CT numbers based on Saito's calibration model using $\alpha = 0.459$. Table adapted from [25].

Material	$\rho_{e,true}$	$Z_{eff,true}$	HU_H	HU_L	$p_{e,cal}$	Dev (%)	$Z_{eff,cal}$	Dev (%)
Lung	0.258	7.60	−741.8	−740.9	0.258	0.0	7.60	0.0
Adipose	0.933	6.23	−80.8	−107.7	0.931	−0.2	6.22	−0.2
Breast	0.962	6.79	−46.1	−62.8	0.961	−0.1	6.79	−0.1
Brain	1.035	7.58	35.9	38.9	1.035	0.0	7.58	0.0
Eye lens	1.055	7.30	52.0	47.1	1.054	0.0	7.31	0.0
Spongiosa	1.150	10.23	207.6	339.7	1.149	−0.1	10.23	0.0
Ribs	1.441	12.32	610.6	994.1	1.441	0.0	12.32	0.0
Cortical Bone	1.780	13.63	1094.6	1801.4	1.782	0.1	13.63	0.0
Bone mineral-hydroxyapatite	2.891	16.11	2838.3	4966.3	2.898	0.2	16.11	0.0
Thyroid	1.042	8.41	52.9	74.6	1.043	0.1	8.15	−3.1

Table 2: Electron density values, CT numbers, and Effective Atomic Numbers based on Saito's DEEDZ conversion using $\alpha = 0.442$. Table adapted from [26].

Calibration	SECT-SPR	DEEDZ-SPR
18 cm phantom	94.9%	95.1%
33 cm phantom	90.3%	95.4%
Difference	−4.6	+0.3

Table 3: Comparison of the SECT-SPR and DEEDZ-SPR conversion. Table adapted from [30].

PSNR	SSIM	LPIPS
$35.62\ dB$	0.9058	0.0635

Table 4: Quantitative Evaluation Metrics for CAN Image Denoising Performance

Material	ρ_e	Z_{eff}	SPR
Liver	0.972	8.20	1.07168
LN-450	0.531	9.02	0.58457
Breast	0.903	7.82	0.99560
LN-350	0.410	7.86	0.45085
Cortical Bone	1.882	12.10	2.08301
Adipose	0.891	7.07	0.98268
Brain	0.949	7.94	1.04641
Inner Bone	1.133	10.27	1.25145
$RMSE$	0.00674	1.30481	-
R^2	0.96224	0.75358	-

Table 5: Calculated Physical Properties for Tissue-Equivalent Materials Using the Tanaka Method with 80/120 kVp DECT

Material	ρ_e	Z_{eff}	SPR
Liver	0.972	7.33	1.07003
LN-450	0.531	6.34	0.58541
Breast	0.903	6.41	0.99341
LN-350	0.410	9.70	0.45228
Cortical Bone	1.882	12.01	2.08864
Adipose	0.891	5.19	0.98050
Brain	0.949	6.90	1.04509
Inner Bone	1.133	9.80	1.25133
$RMSE$	0.00674	0.79768	-
R^2	0.96224	0.84936	-

Table 6: Calculated Physical Properties for Tissue-Equivalent Materials Using the Saito Method with 80/120 kVp DECT

Material	ρ_e	Z_{eff}	SPR
Liver	1.064	7.97	0.01673
LN-450	0.500	10.35	0.00735
Breast	0.958	7.29	0.01572
LN-350	0.312	9.25	0.00458
Cortical Bone	2.791	12.72	0.03135
Adipose	0.918	5.22	0.01527
Brain	1.023	7.47	0.01600
Inner Bone	1.413	10.98	0.01964
$RMSE$	0.13762	1.65512	-
R^2	0.22917	0.68743	-

Table 7: Calculated Physical Properties for Tissue-Equivalent Materials Using the Hünemohr Method with 80/120 kVp DECT

Material	ρ_e (T)	Z_{eff} (T)	SPR (T)	ρ_e (S)	Z_{eff} (S)	SPR (S)	ρ_e (H)	Z_{eff} (H)	SPR (H)
Liver	1.043	8.14	1.14896	1.043	4.10	1.14923	1.058	8.43	0.01665
LN-450	0.511	8.59	0.56293	0.511	7.28	0.56342	0.558	9.14	0.00821
Breast	0.967	7.56	1.06373	0.967	4.76	1.06431	0.944	6.65	0.01550
LN-350	0.247	9.32	0.27276	0.247	9.82	0.27250	0.309	9.99	0.00453
Cortical Bone	1.773	13.25	1.96761	1.773	13.38	1.96780	2.999	15.12	0.03368
Adipose	0.943	7.45	1.03824	0.943	5.10	1.03769	0.916	6.20	0.01523
Brain	1.020	8.16	1.12322	1.020	4.32	1.12359	1.037	8.47	0.01623
Inner Bone	1.166	10.58	1.28746	1.166	9.91	1.28776	1.465	12.57	0.02037

Table 8: Calculated Physical Properties for Tissue-Equivalent Materials with 70/140 kVp DECT and 100% Radius

Material	ρ_e (T)	Z_{eff} (T)	SPR (T)	ρ_e (S)	Z_{eff} (S)	SPR (S)	ρ_e (H)	Z_{eff} (H)	SPR (H)
Liver	1.053	8.11	1.16028	1.053	4.13	1.16023	1.070	8.40	0.01684
LN-450	0.545	8.48	0.50600	0.459	6.36	0.50651	0.489	9.27	0.00719
Breast	0.970	7.62	1.06827	0.970	4.71	1.06766	0.947	6.59	0.01553
LN-350	0.281	8.74	0.30988	0.281	7.41	0.31006	0.312	9.74	0.00457
Cortical Bone	1.773	12.87	1.96903	1.773	13.38	1.96857	3.128	15.23	0.03514
Adipose	0.946	7.56	1.04119	0.946	4.96	1.04033	0.918	6.28	0.01526
Brain	1.022	8.14	1.12576	1.022	4.37	1.12582	1.042	8.50	0.01630
Inner Bone	1.165	10.28	1.28611	1.165	9.78	1.28619	1.481	12.61	0.02059

Table 9: Calculated Physical Properties for Tissue-Equivalent Materials with 70/140 kVp DECT and 50% Radius

Material	ρ_e (T)	Z_{eff} (T)	SPR (T)	ρ_e (S)	Z_{eff} (S)	SPR (S)	ρ_e (H)	Z_{eff} (H)	SPR (H)
Liver	1.052	8.11	1.15912	1.052	4.17	1.15903	1.070	8.41	0.01683
LN-450	0.461	8.47	0.50773	0.461	6.27	0.50824	0.490	9.26	0.00720
Breast	0.971	7.62	1.06849	0.971	4.73	1.06785	0.947	6.55	0.01553
LN-350	0.280	8.65	0.30861	0.280	7.12	0.30882	0.307	9.57	0.00450
Cortical Bone	1.773	12.88	1.96839	1.773	13.38	1.96791	3.131	15.27	0.03517
Adipose	0.944	7.56	1.03886	0.944	4.94	1.03797	0.917	6.27	0.01523
Brain	1.022	8.13	1.12674	1.023	4.27	1.12679	1.041	8.44	0.01629
Inner Bone	1.167	10.27	1.28868	1.167	9.75	1.28877	1.482	12.61	0.02060

Table 10: Calculated Physical Properties for Tissue-Equivalent Materials with 70/140 kVp DECT and 25% Radius

Figures

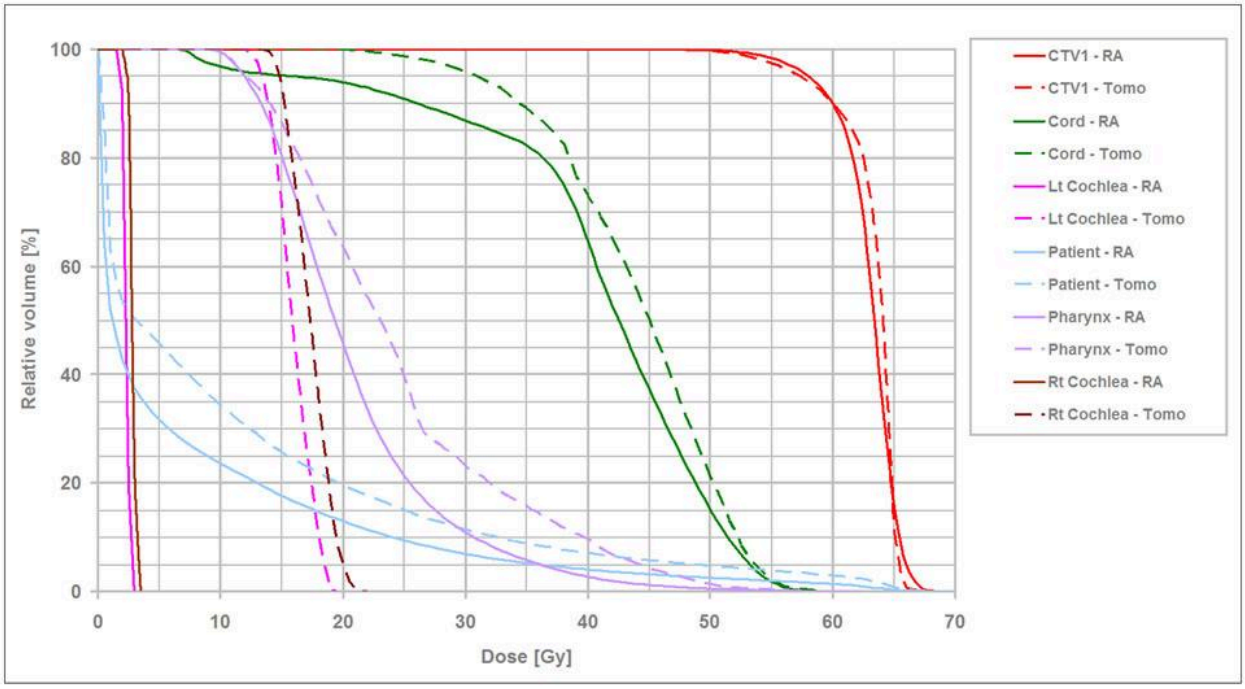

Figure 1: Example of Dose Volume Histogram. CC BY 3.0,

https://en.wikipedia.org/w/index.php?curid=32800049

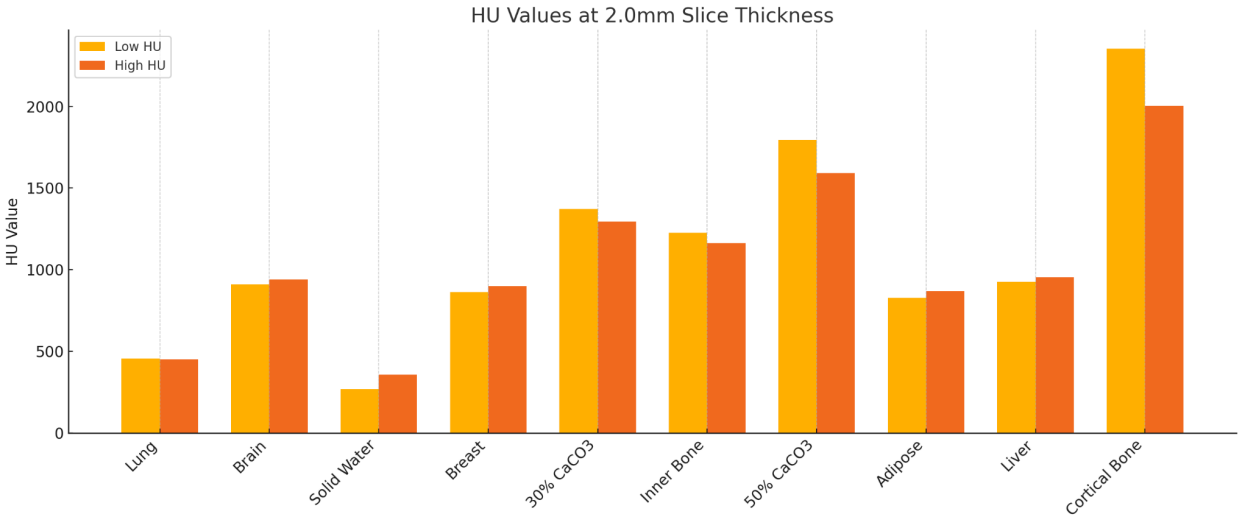

Figure 2: Comparison of Low and High *HU* Values Across Materials at 2.0mm Slice Thickness

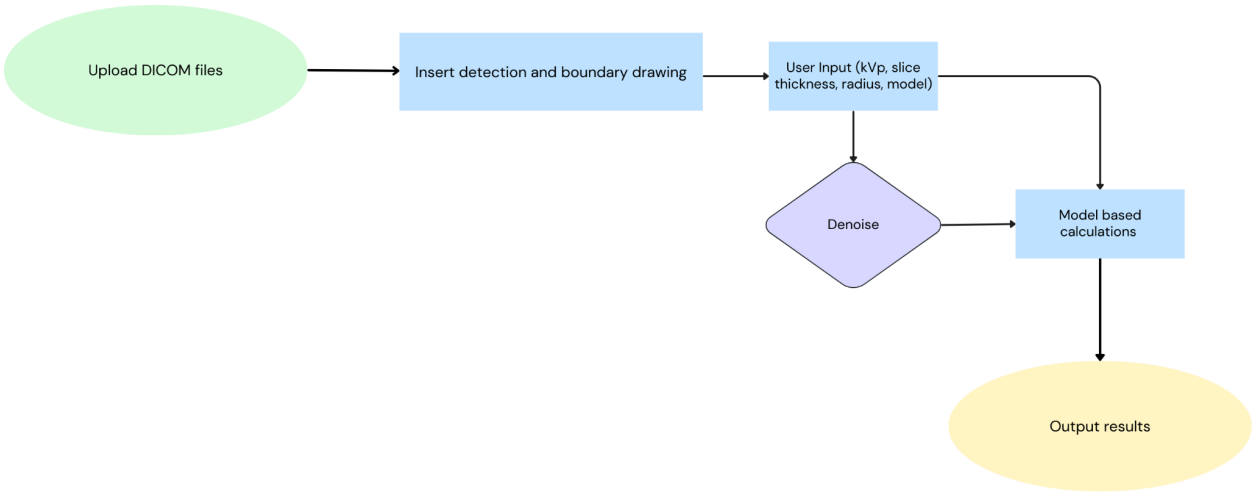

Figure 3: SPR-Net User Workflow

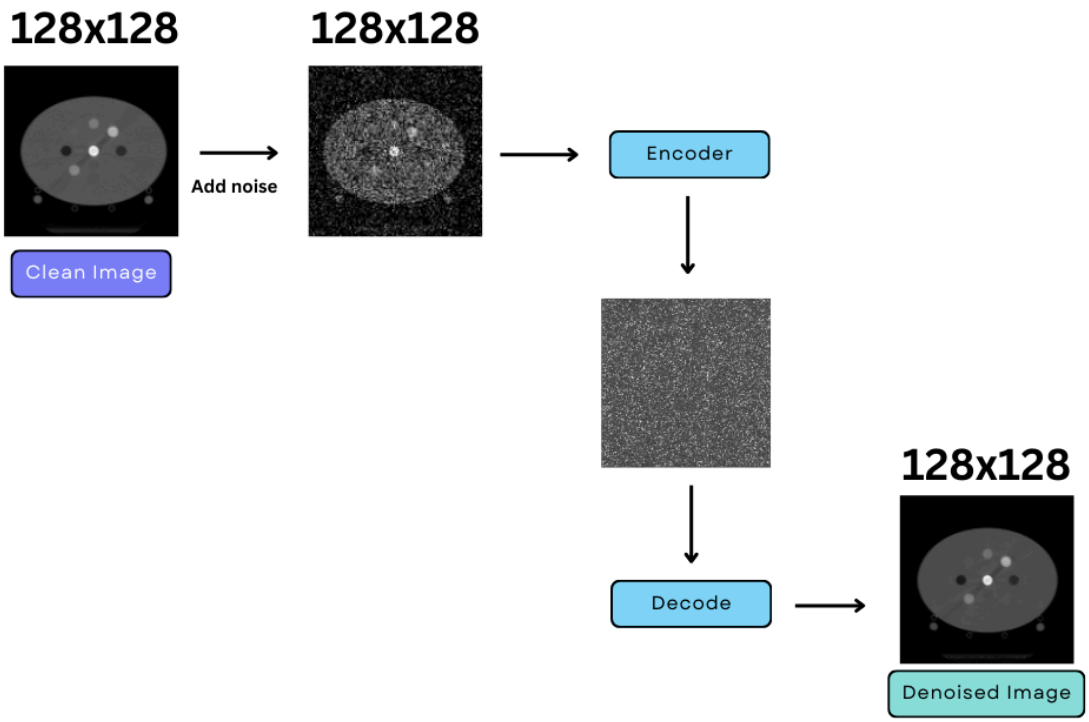

Figure 4: Autoencoder Flow Chart

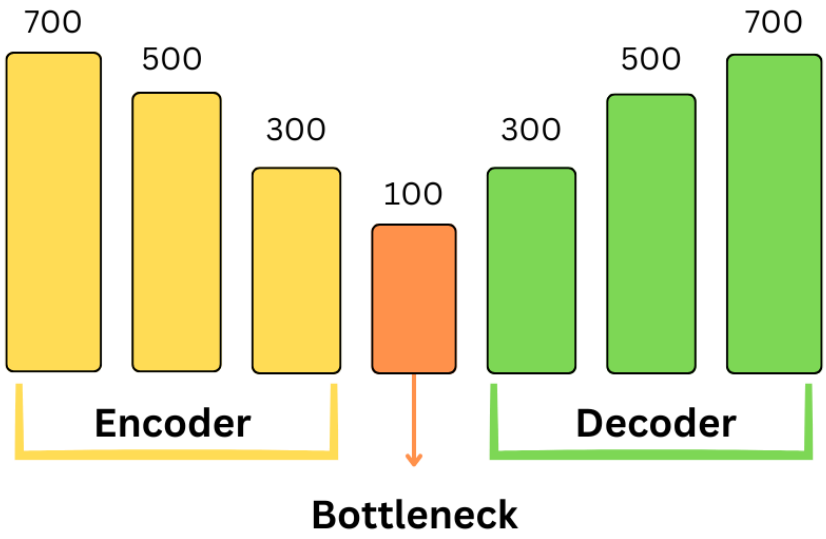

Figure 5: Autoencoder Architecture

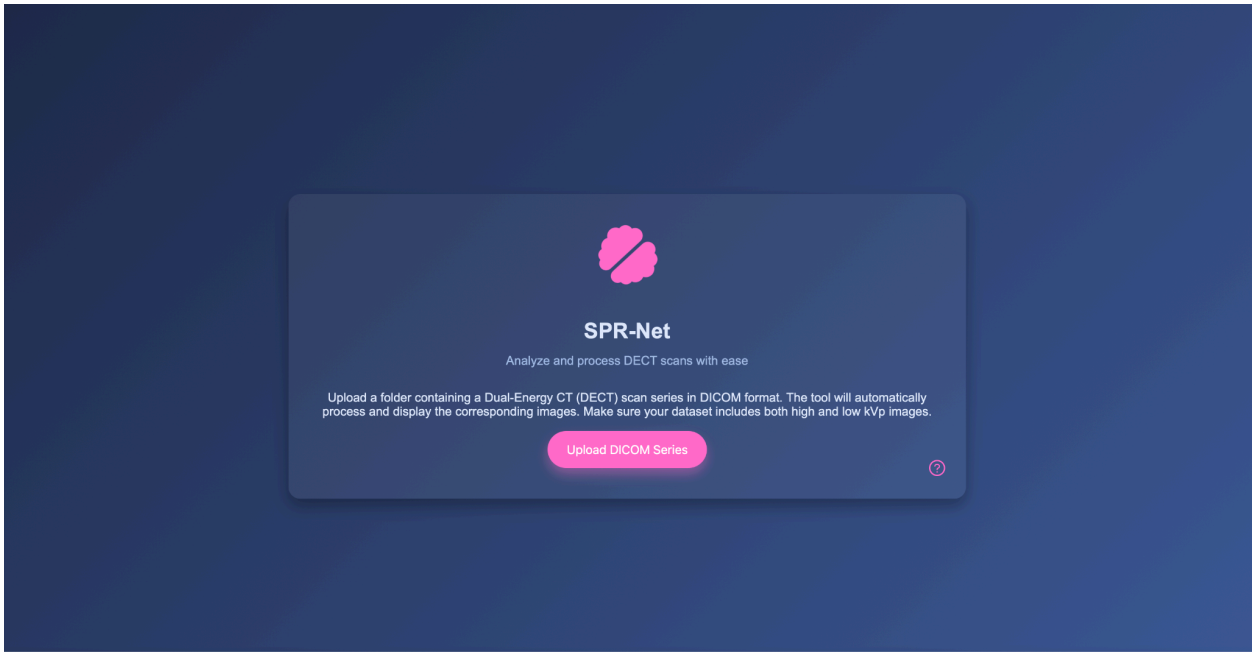

Figure 6: SPR-Net Landing Page

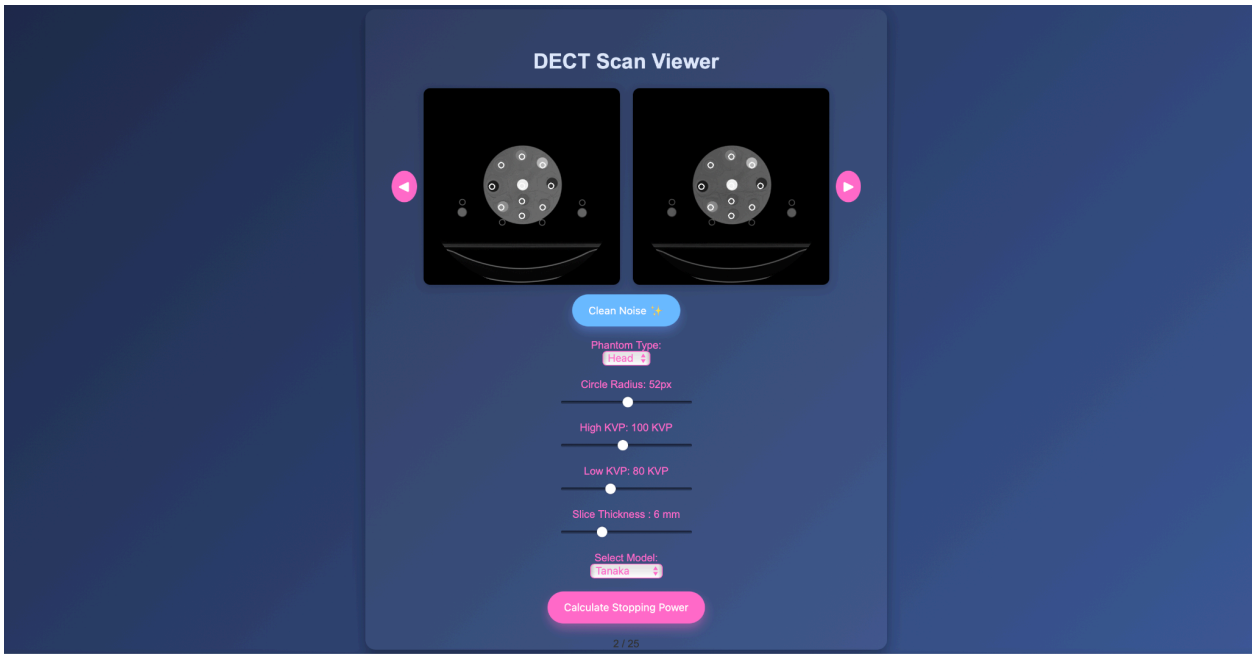

Figure 7: SPR-Net Configuration Page

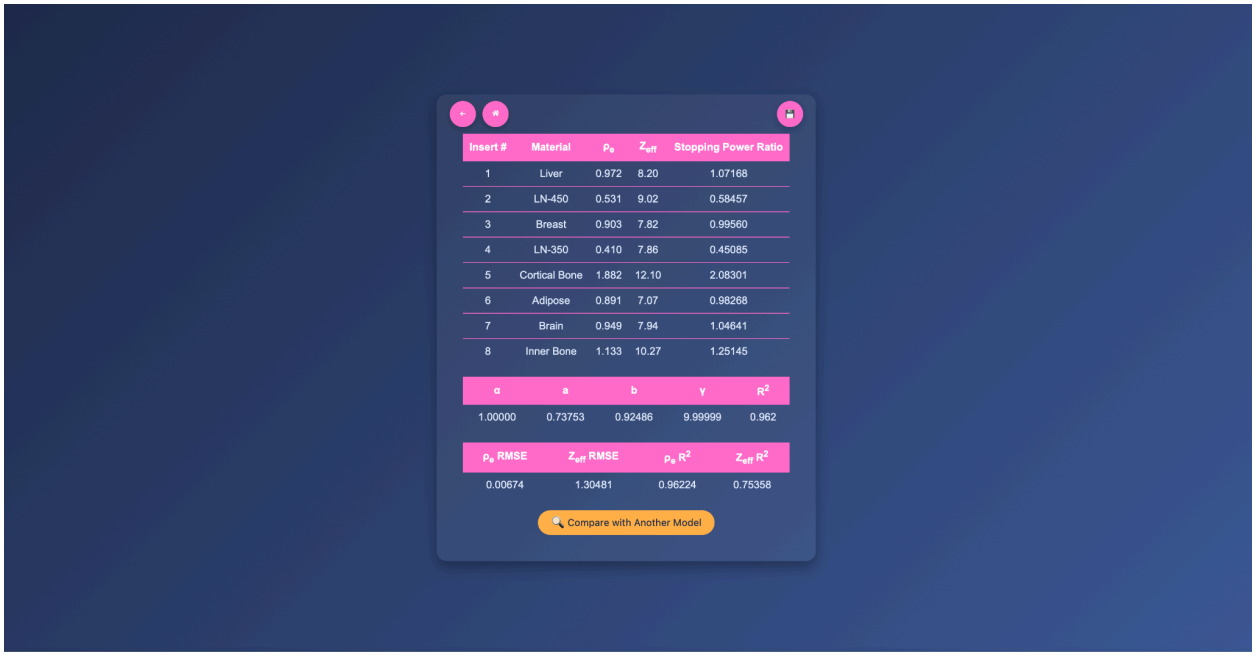

Figure 8: SPR-Net Results Page

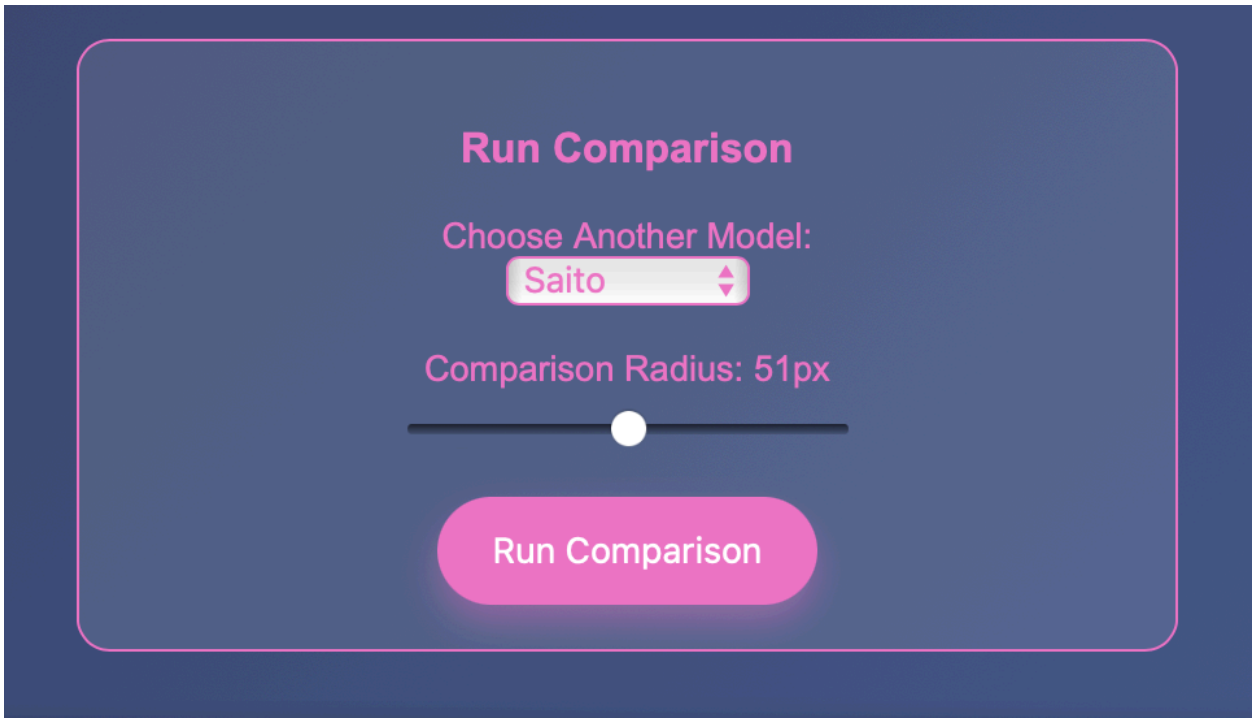

Figure 9: SPR-Net Comparison Tab

Comparison (Saito)

Insert #	Material	ρ_e	Z_{eff}	Stopping Power Ratio
1	Liver	0.972	7.33	1.07003
2	LN-450	0.531	6.34	0.58541
3	Breast	0.903	6.41	0.99341
4	LN-350	0.410	9.70	0.45228
5	Cortical Bone	1.882	12.01	2.08864
6	Adipose	0.891	5.19	0.98050
7	Brain	0.949	6.90	1.04509
8	Inner Bone	1.133	9.80	1.25133

Figure 10: SPR-Net Comparison Result View

References

[1] E. S. Armentia, “Strengths, weaknesses, opportunities, and threat analysis of dual-energy CT in head and neck imaging,” 2022.

[2] A. P. Borges, C. Antunes, and L. Curvo-Semedo, “Pros and Cons of Dual-Energy CT Systems: ‘One Does Not Fit All,’” *Tomography*, vol. 9, no. 1, pp. 195–216, Jan. 2023, doi: 10.3390/tomography9010017.

[3] H. Chen *et al.*, “Low-dose CT denoising with convolutional neural network,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, Australia: IEEE, Apr. 2017, pp. 143–146. doi: 10.1109/ISBI.2017.7950488.

[4] G. Choy *et al.*, “Current Applications and Future Impact of Machine Learning in Radiology,” *Radiology*, vol. 288, no. 2, pp. 318–328, Aug. 2018, doi: 10.1148/radiol.2018171820.

[5] A. M. Crellin and N. G. Burnet, “Proton Beam Therapy: The Context, Future Direction and Challenges Become Clearer,” *Clinical Oncology*, vol. 26, no. 12, pp. 736–738, Dec. 2014, doi: 10.1016/j.clon.2014.10.009.

[6] C. N. De Cecco, A. Laghi, U. J. Schoepf, and F. G. Meinel, Eds., *Dual Energy CT in Oncology*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-19563-6.

[7] R. Forghani, B. De Man, and R. Gupta, “Dual-Energy Computed Tomography,” *Neuroimaging Clinics of North America*, vol. 27, no. 3, pp. 371–384, Aug. 2017, doi: 10.1016/j.nic.2017.03.002.

[8] D. Han, J. V. Siebers, and J. F. Williamson, “A linear, separable two-parameter model for dual energy CT imaging of proton stopping power computation: Dual-energy CT, proton stopping power,” *Med. Phys.*, vol. 43, no. 1, pp. 600–612, Jan. 2016, doi: 10.1118/1.4939082.

[9] N. Hünemohr, B. Krauss, C. Tremmel, B. Ackermann, O. Jäkel, and S. Greilich, "Experimental verification of ion stopping power prediction from dual energy CT data in tissue surrogates," *Phys. Med. Biol.*, vol. 59, no. 1, pp. 83–96, Jan. 2014, doi: 10.1088/0031-9155/59/1/83.

[10] A. E. Ilesanmi and T. O. Ilesanmi, "Methods for image denoising using convolutional neural network: a review," *Complex Intell. Syst.*, vol. 7, no. 5, pp. 2179–2198, Oct. 2021, doi: 10.1007/s40747-021-00428-4.

[11] W. Jifara, F. Jiang, S. Rho, M. Cheng, and S. Liu, "Medical image denoising using convolutional neural network: a residual learning approach," *J Supercomput*, vol. 75, no. 2, pp. 704–718, Feb. 2019, doi: 10.1007/s11227-017-2080-0.

[12] C. Kaethner, J. Müller, and T. M. Buzug, "Determining Noise Distribution in Computed Tomography – A Simple Phantom Based Approach," *Biomedical Engineering / Biomedizinische Technik*, vol. 57, no. SI-1-Track-B, p. 000010151520124050, Aug. 2012, doi: 10.1515/bmt-2012-4050.

[13] M. Karcaaltincaba and A. Aykut, "Dual-energy CT revisited by multidetector ct: review of principles and clinical applications," *Diagn Interv Radiol*, 2010, doi: 10.4261/1305-3825.DIR.3860-10.0.

[14] B. Kim, M. Han, H. Shim, and J. Baek, "A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose CT images," *Medical Physics*, vol. 46, no. 9, pp. 3906–3923, Sep. 2019, doi: 10.1002/mp.13713.

[15] H. H. C. Lee, B. Li, X. Duan, L. Zhou, X. Jia, and M. Yang, "Systematic analysis of the impact of imaging noise on dual-energy CT -based proton stopping power ratio estimation," *Medical Physics*, vol. 46, no. 5, pp. 2251–2263, May 2019, doi: 10.1002/mp.13493.

[16] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[17] H. Liu and Y.J. Chang, "Proton Therapy in Clinical Practice," *Chinese Journal of Cancer.*, vol. 30, no.5, pp. 315-326, doi: 0.5732/cjc.010.10529

[18] B. Mahesh, "Machine Learning Algorithms - A Review," *IJSR*, vol. 9, no. 1, pp. 381–386, Jan. 2020, doi: 10.21275/ART20203995.

[19] R. S. Maia, C. Jacob, A. K. Hara, A. C. Silva, W. Pavlicek, and M. J. Ross, "An algorithm for noise correction of dual-energy computed tomography material density images," *Int J CARS*, vol. 10, no. 1, pp. 87–100, Jan. 2015, doi: 10.1007/s11548-014-1006-z.

[20] T. E. Merchant, "Clinical Controversies: Proton Therapy for Pediatric Tumors," *Seminars in Radiation Oncology*, vol. 23, no. 2, pp. 97–108, Apr. 2013, doi: 10.1016/j.semradonc.2012.11.008.

[21] P. Omoumi, F. Becce, D. Racine, J. Ott, G. Andreisek, and F. Verdun, "Dual-Energy CT: Basic Principles, Technical Approaches, and Applications in Musculoskeletal Imaging (Part 1)," *Semin Musculoskelet Radiol*, vol. 19, no. 05, pp. 431–437, Dec. 2015, doi: 10.1055/s-0035-1569253.

[22] H. Paganetti, "Range uncertainties in proton therapy and the role of Monte Carlo simulations," *Phys. Med. Biol.*, vol. 57, no. 11, pp. R99–R117, Jun. 2012, doi: 10.1088/0031-9155/57/11/R99.

[23] P. C. Park *et al.*, "Statistical Assessment of Proton Treatment Plans Under Setup and Range Uncertainties," *International Journal of Radiation*

*Oncology*Biology*Physics*, vol. 86, no. 5, pp. 1007–1013, Aug. 2013, doi: 10.1016/j.ijrobp.2013.04.009.

[24] D. M. C. Poon, S. Wu, L. Ho, K. Y. Cheung, and B. Yu, "Proton Therapy for Prostate Cancer: Challenges and Opportunities," *Cancers*, vol. 14, no. 4, p. 925, Feb. 2022, doi: 10.3390/cancers14040925.

[25] M. Saito, "Potential of dual-energy subtraction for converting CT numbers to electron density based on a single linear relationship," *Medical Physics*, vol. 39, no. 4, pp. 2021–2030, Apr. 2012, doi: 10.1118/1.3694111.

[26] M. Saito and S. Sagara, "A simple formulation for deriving effective atomic numbers via electron density calibration from dual-energy CT data in the human body," *Med. Phys.*, vol. 44, no. 6, pp. 2293–2303, Jun. 2017, doi: 10.1002/mp.12176.

[27] S. Sajja *et al.*, "Technical Principles of Dual-Energy Cone Beam Computed Tomography and Clinical Applications for Radiation Therapy," *Advances in Radiation Oncology*, vol. 5, no. 1, pp. 1–16, Jan. 2020, doi: 10.1016/j.adro.2019.07.013.

[28] E. Santos Armentia, T. Martín Noguerol, N. Silva Priegue, C. Delgado Sánchez-Gracián, C. Trinidad López, and R. Prada González, "Strengths, weaknesses, opportunities, and threat analysis of dual-energy CT in head and neck imaging," *Radiología (English Edition)*, vol. 64, no. 4, pp. 333–347, Jul. 2022, doi: 10.1016/j.rxeng.2022.05.003.

[29] A. R. Smith, "Proton therapy," *Phys. Med. Biol.*, vol. 51, no. 13, pp. R491–R504, Jul. 2006, doi: 10.1088/0031-9155/51/13/R26.

[30] S. Tanaka, Y. Noto, S. Utsunomiya, T. Yoshimura, T. Matsuura, and M. Saito, "Proton dose calculation based on converting dual-energy CT data to stopping power ratio (DEEDZ-SPR): a beam-hardening assessment," *Phys.*

Med. Biol., vol. 65, no. 23, p. 235046, Dec. 2020, doi: 10.1088/1361-6560/abae09.

[31] S.-J. Tu, W.-Y. Chen, and C.-T. Wu, "Uncertainty measurement of radiomics features against inherent quantum noise in computed tomography imaging," *Eur Radiol*, vol. 31, no. 10, pp. 7865–7875, Oct. 2021, doi: 10.1007/s00330-021-07943-5.

[32] S. Wang and R. M. Summers, "Machine learning and radiology," *Medical Image Analysis*, vol. 16, no. 5, pp. 933–951, Jul. 2012, doi: 10.1016/j.media.2012.02.005.

[33] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/s13244-018-0639-9.

[34] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Neural Networks for Image Processing," Apr. 20, 2018, *arXiv*: arXiv:1511.08861. doi: 10.48550/arXiv.1511.08861.