Uncovering the Limits of Detection of Artificial Intelligence using Synthetic Lesions in Positron Emission Tomography

by

Quinn de Bourbon

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Physics

Ottawa-Carleton Institute for Physics
Department of Physics
Carleton University
Ottawa, Ontario

Abstract

The use of artificial intelligence (AI) for detection of lesions has been proposed to aid clinicians in detection tasks. However, before AI can be used clinically, the limits of detection must be studied to characterize AI performance. In this work, a library was constructed containing well-characterized spherical synthetic lesions in real positron emission tomography (PET) and x-ray computed tomography (CT) patient data, which had been previously reported free of lesions by expert physicians. These lesions were manually defined and automatically synthesized using the Lesion Synthesis Toolbox (LST). This library was used to study two FDG PET lesion-detection AI algorithms by their ability to detect lesions by size and intensity metrics, including lesion intensity, contrast, and contrast-to-noise ratio. The work demonstrates the utility of synthetic lesions for characterizing the limits of detection of AI and necessary tools available to other researchers.

Acknowledgements

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AI	Artificial Intelligence
β^+	Positron particle
Bq/cc	Becquerel per milliliter
CNR	Contrast to Noise Ratio
CSV	Comma-Separated Values
CT	Computed Tomography
D710	Discovery 710 PET/CT machine
DICOM	Digital Imaging and COmmunications in Medicine
FBP	Filtered Back Projection
FDG	Fluorodeoxyglucose
FOV	Field-of-View
GUI	Graphical User Interface
HU	Hounsfield Units
LOD	Limits of Detection
LOR	Line of Response
LST	Lesion Synthesis Toolbox
LYSO	Lutetium-Yttrium Orthosilicate
MLEM	Maximum-Likelihood Expectation-Maximization
MRI	Magnetic Resonance Imaging
OSEM	Ordered-Subset Expectation-Maximization
PET	Positron Emission Tomography
PMT	Photo-Multiplier Tube

PSMA   Prostate Specific Membrane Antigen

SNR   Signal to Noise Ratio

SPECT   Single Photon Emission Computed Tomography

SUV   Standard Uptake Values

TOF   Time of Flight

TOH   The Ottawa Hospital

Chapter 1: Introduction

Cancer is a leading cause of death in Canada and globally. Around 2 out of 5 Canadians are expected to develop cancer in their lifetime and 1 out of 4 Canadians are expected to pass away from cancer. [1] Although these numbers are high, the rate of deaths has been declining since 1988 due to developments in cancer detection and treatment. Detection techniques for cancer include blood tests, biopsies, and various imaging modalities. Nevertheless, cancer is notorious for being difficult to detect and can grow undetected for years before symptoms show. Once cancer metastasizes, different treatment modalities need to be applied for optimal outcomes. Therefore, detection of small lesions associated with early disease and metastases is critical.

Positron emission tomography (PET) has distinguished itself as a highly sensitive imaging modality for detection of many types of cancer due to high lesion to background contrast. High lesion contrast is enabled by radioactive tracers that target and accumulate in tissues with functions specific to each tracer. As PET technologies continue to evolve, smaller lesions can be detected due to increase in image spatial resolution, increase in signal sensitivity, improved image reconstruction algorithms, and recently the use of artificial intelligence (AI) to assist the physician in detecting lesions. However, comparing these technologies, and the performance of human observers, is challenging. Furthermore, a large library of ground truth lesion data, needed to complete this task, is difficult to acquire given current methodologies, either due to lack of realism or the large amount of time required for clinicians to allocate to this task.

In this work a new paradigm is proposed and explored for examining and characterizing the limits of detection of lesions. Using in-house developed software

named the Lesion Synthesis Toolbox (LST), well characterized artificial lesions were inserted into the raw PET data of real patients. The raw form data are then reconstructed to generate images in which the lesions are synthesized in a realistic manner. These images, in which the location, size, and intensity of the lesions are known, was then used to test the limits of lesion detection of two externally developed AI algorithms.

Chapter 2 delves into the background of PET imaging, including physics, image reconstruction, and clinical use. It also expands on lesion detection by human and AI observers.

Chapter 3 begins with a discussion on experimental models for characterizing lesion detection, follows with a justification of the motivation for generating well characterized synthetic lesions, and concludes with the description of the LST and its critical features.

Chapter 4 describes a perception study to characterize the limits of detection of two AI algorithms. It explains the methods and analysis involved in characterizing these AI algorithms using LST generated images to test the AI and the corresponding ground truth to score the AI responses.

Chapter 5 concludes with a discussion of the AI detection results and characterization, as well as possible future directions for this work.

1.1 Statement of Originality

In this thesis I built upon previous work by Hanif Gabrani-Juma and Ran Klein. I added functionality to the Lesion Synthesis Toolbox including:

- Creation of blobby spheres for realistic looking lesions.

- Addition of the 'retrieve lesions' tool to bring back simulations to the 'define lesion' sub-tab for modification, to enhance workflow when creating large lesion libraries.

- Creation of the toolbox user manual, included in the appendix.

- Testing, debugging, and updating of the toolbox, including creating a new logo and general bug fixes.

- Enhancement to the ground truth file generated by the toolbox to include more lesion and background activity parameters and patient information.

Throughout this process, Ran Klein supervised my work and contributed to some of the functionality updates explained in Chapter 3. The updated toolbox was made as an open-source repository on Github to be used by other researchers.

In addition to the functionality updates to the toolbox, I created the library of synthetic lesions by manually defining lesions in a variety of patients. This library was created for this project and to be shared as an open-source library. I was responsible for selecting patients that were clear of disease by reading the physicians reports and verified that the scans appeared normal using a clinical image viewer (HERMES Medical Solutions, Stockholm, Sweden) before downloading the raw image data. After the library was created, I was also responsible for verifying that the lesion insertion worked as intended and removed any studied containing lesion artifacts.

Finally, I handled all aspects of the AI limits of detection characterization (Chapter 4), including engaging participating sites, distributing the lesion library, aggregating the AI results, and analyzing the responses. The sole exception was the implementation of the psychophysical model which was assisted by Ran Klein.

This work led to an oral presentation at the Canadian Association of Nuclear Medicine conference, held in Ottawa in October 2023. [2] Furthermore, a manuscript was created to be submitted to peer-reviewed publication using sections of Chapters 3-5.

Chapter 2: Background

2.1 Principles of PET

Positron Emission Tomography (PET) is a medical imaging technique that uses positron-emitting isotopes to image physiologic function in the body. This method can be used in oncology for detecting and classifying cancer, cardiac imaging for myocardial perfusion, and brain imaging for detecting Alzheimer's disease and localizing seizures, to name a few applications. [3] Due to its proficiency, PET has been growing exponentially in its diagnostic use since the early 2000s.

2.1.1 Positron Emission

During positron emission, also known as β^+ decay, a proton from an unstable nucleus decays into a neutron, releasing a positron and a neutrino with respective kinetic energies. A positron (e^+) is the antiparticle of an electron, meaning it has the same mass as an electron but with a positive charge. When a positron is emitted during β^+ decay, the positron will travel through the neighboring medium, losing kinetic energy, until it encounters and combines with an electron, briefly forming positronium, before their mutual annihilation. During this annihilation event, the masses of the particles are converted into energy, according to conservation of energy, momentum, and charge, creating two antiparallel (nearly) colinear photons with energies of 511 keV each. The positron emission and annihilation processes are displayed in Figure 2.1. As an imaging device, PET is designed to detect these 511 keV photons in close coincidence.
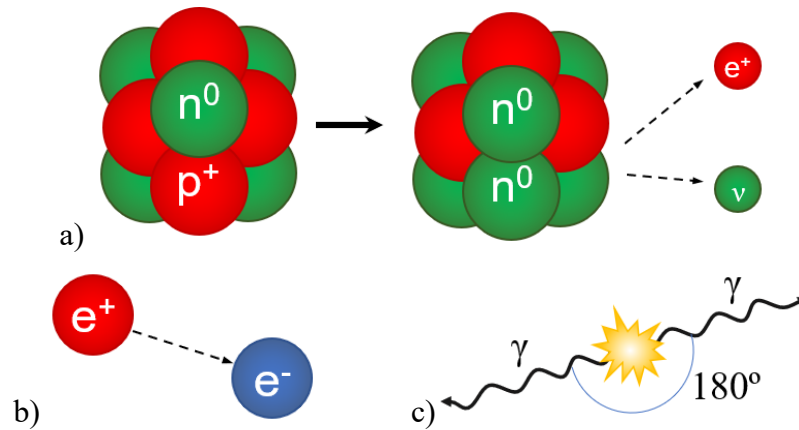
Figure 2.1: The nuclear processes involved in positron emission, a) positron emission from an unstable nucleus, and positron-electron annihilation b) before and c) after annihilation.

2.1.2   Radiotracers

For nuclear medicine testing, a patient is administered with a radioactive tracer that decays within the patient's body. These tracers can be administered by inhalation (e.g., for ventilation scans), taken orally (e.g., for gastric or thyroid testing), or most commonly by intravenous injection. Radioactive tracers can be a single element but are typically bound to another molecule by replacing a section of the molecule with a radioactive isotope. When these radioactive molecules enter the body of the patient, they are treated similarly to how the original molecule would be treated in the body. Consequently, these radiotracers accumulate in regions of the body with high uptake of the original molecule. There is a wide variety of radiotracers used in nuclear imaging, each targeting different organs or physiologic processes, depending on the type of disease that is being detected. Some common tracers used in PET imaging include ^{18}F-prostate specific membrane antigen (PSMA), ^{68}Ga-PSMA, and ^{11}C-Chlorine for prostate cancer, ^{82}RbCl and ^{15}O-water for myocardial perfusion, and ^{18}F-Fluorodeoxyglucose for cancer and glucose metabolism.

The most commonly used radioisotope in PET is Fluorine-18, which has a half-life of 110 minutes. This isotope may be bound to a glucose-like molecule, where one of the hydroxyl groups is replaced by the radioactive ^{18}F atom, creating a compound called Fluorodeoxyglucose (FDG). [4] When FDG enters the body, it is treated similarly to glucose. [5] However, when it enters the cell, it cannot be completely metabolized due to the missing hydroxyl group and therefore accumulates in the cell. This creates concentrated regions of positron emitting isotopes in regions of high glucose uptake. Regions of high uptake in the body include the brain, regions of inflammation, many types of tumors, and occasionally the heart, in cases of high insulin and low fatty acids in the body. There is moderate uptake in regions such as the liver, spleen, and thyroid. As blood is filtered through the kidneys, FDG is filtered out of the body through the urinary system, creating accumulation in the kidneys and bladder. Malignant tumors are FDG avid because their cells reproduce quickly; faster than vessels can develop to supply blood to match their oxygen need. Glucose provides an energy source to cancerous cells that are starved for oxygen. Figure 2.2 demonstrates FDG uptake in normal and abnormal scans, with a high intensity lesion in the upper abdomen of the abnormal scan. When FDG is administered to the patient, the patient must wait an hour for the isotope to accumulate in the target sites and wash out from background tissues to produce a high contrast image. The relatively long half-life of ^{18}F allows the tracer enough time to accumulate in the body while also maintaining enough activity to be detected. It also allows enough time to transport the tracer over a range of several hundred kilometers from the source of production. On the other hand, the radioisotope decays quickly enough

so that the patient isn't radioactive for too long, limiting the dose to both the patient and those around them.
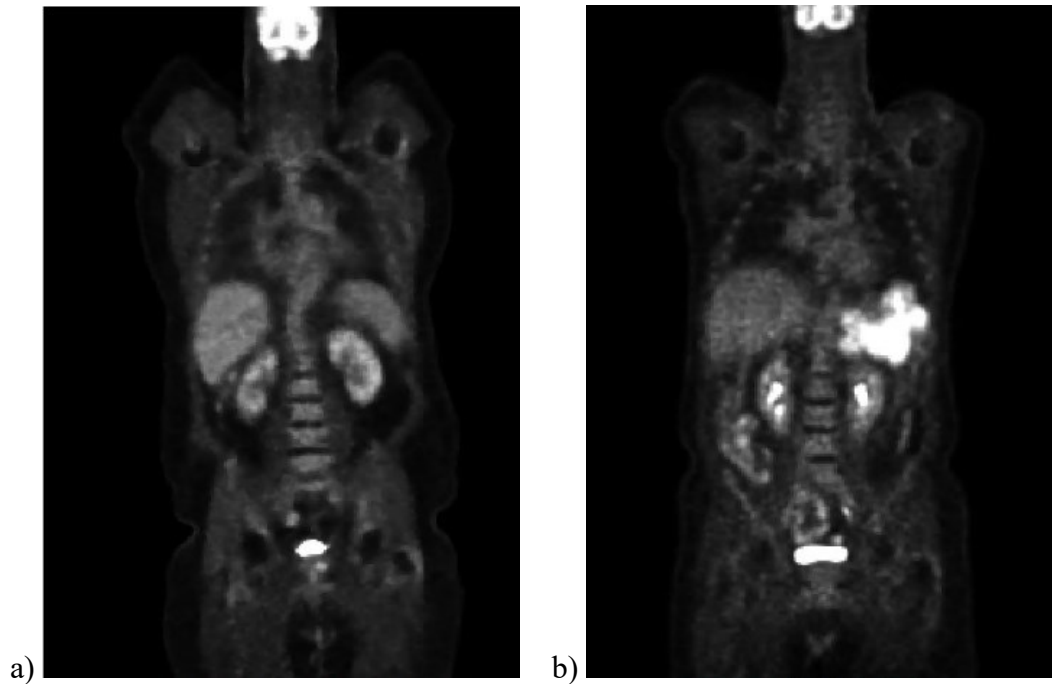
Figure 2.2: FDG PET scans showing a) normal and b) abnormal results. The abnormality is manifested as a high intensity (bright white) region in the upper abdomen, consistent with a cancerous growth.

The amount of radiotracer administered to the patient depends on several factors including patient size, type of tracer, sensitivity of the imaging equipment, and clinical considerations that include duration of image acquisition, radiation dose to the patient and image quality. FDG activity is typically adjusted by patient mass, on the order of 2 to 5 MBq per kilogram of patient mass. Variations in patient mass and administered dose makes quantitative comparison of FDG uptake between patients difficult. The standard uptake value (SUV) is used as a measure of FDG activity to compensate for variations between patients. SUV values are dimensionless with the assumption that 1 ml of tissue weighs 1 g. The equation to calculate the SUV is shown in Equation 2.1 below, where r is

the image sampled activity concentration (kBq/ml), a' is the injected FDG activity (kBq), and w is the patient weight (g). [6] Not shown in the equation, is that radioisotope decay between activity measurement and scan time must also be applied.

$$SUV = \frac{r}{(a'/w)}$$ **Equation 2.1**

2.1.3 PET Instrumentation

After the radiotracer has accumulated in the body, the patient is imaged. Only a small percent of the annihilation photons from the radiotracers in the body will make it out of the patient and be detected by the PET instrumentation, while the rest are attenuated and absorbed within the body or are released in a range outside of the PET detectors.

A typical PET scanner is comprised of stacked rings of specially designed detectors. The patient is located on a bed inside the ring of detectors that is moved through the ring over several minutes, to collect data for a given range, typically either 'eyes-to-thighs' or 'head-to-toe'. For a full body scan, the duration of the acquisition process is in the range of 10-30 minutes, depending on the scanner and time of detection per bed position. The typical time per bed position for FDG scan is in the range of 1-5 minutes. Some modern PET systems have long detectors that can perform a full body scan in one bed position, which can greatly reduce the time needed to image a patient. [7] These long detectors also increase the sensitivity of the scanner, so that less radiotracer activity needs to be administered to the patient, which reduces the radiation dose to the patient.

The detectors consist of a scintillation crystal, which converts high energy photons into lower energy light that can be detected by highly sensitive electronics, such

as a photomultiplier tube (PMT). [5] When light enters the PMT, it travels to and deposits its energy on a photocathode layer that subsequentially releases a few photoelectrons. If the electron has enough energy, it escapes the photocathode and travels through an electric field to the first dynode, which when hit will produce more secondary electrons, thus amplifying the electrical current. This effect cascades through a series of dynodes to create an amplification of electrons. At the final dynode (anode) the electrons are collected and output as an electrical signal to be passed on to additional circuitry before a digital signal is stored in a computer, encoding the photon detection event. Newer models of PET scanners are veering away from using PMTs and towards silicon photomultipliers, which work using photodiodes, producing higher quality signals, better system integration and the possibility of operating in strong magnetic fields as would be present in hybrid PET and magnetic resonance imaging (MRI) machines.

Moreover, the signal amplitude from the detector encodes the energy deposited in the scintillator by the incident photon. As the photon interacts with the scintillation material, it transfers its energy to the scintillator atom via the photoelectric effect scatter (~30 photons per keV absorbed using in LYSO crystal).[8] As electrons fill the ionization holes, they fluoresce lower energy photons in the blue light spectrum. Thus, the full absorption of a 511 keV photon produces tens of thousands of incident photons. PMTs and photodiodes produce and electrical signal proportional to the amount of light output from the scintillator. Electronic processing of the signal uses energy discrimination windows to eliminate background radiation, as well as photons that have lost a significant amount of energy from scatter, to limit noise in the collection process.

Since positron-electron annihilation creates two colinear photons, the detectors' signals are analyzed to detect coincident events within a narrow time window, to keep only the measurements from the same annihilation event and limit the number of random coincident events. Using the location of the two detectors that measured a coincident event within the given energy and time range, a line of response (LOR) can be drawn between the two detectors. The detection of two colinear photons is displayed in Figure 2.3.
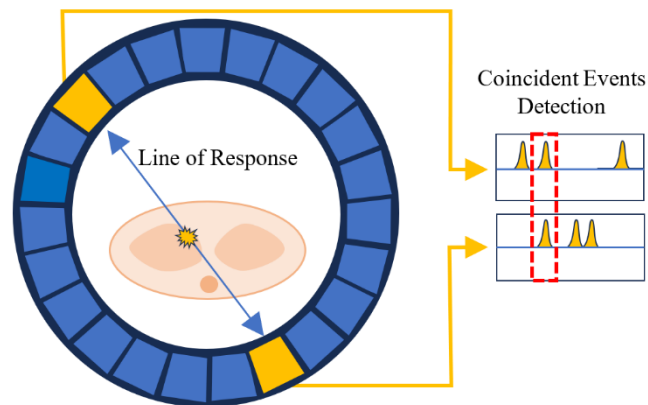
Figure 2.3: Schematics of a ring detector with coincidence detection demonstrating how the line of response is drawn between the activated detectors after an annihilation event.

On high-end PET systems, the time difference between the first photon and second photon hitting their respective detectors is known as the time of flight (TOF). [4] TOF can be used to measure approximately how far along the LOR the annihilation event occurred, creating more accurate measurements of the event location. The data collected by the PET scanner are stored in a sinogram, which contains projection histograms for each angle around the patient with the displacement from the center and can include TOF information.

The PET system used in the context of this thesis is the GE Healthcare Discovery 710 PET/CT scanner (D710), which was installed at The Ottawa Hospital (TOH) in 2014.

This scanner's detectors are comprised of lutetium-yttrium-orthosilicate (LYSO) crystals and PMTs enabling time of flight data acquisition, with timing resolution of 560 ps, coincidence window of 4.9 ns, and an energy window of 425-650 keV. [9] Reconstructed image spatial resolution is maximum ~4.5 mm. [10] The detector ring has a diameter of 81 cm, patient port of 70 cm, and axial field-of-view (FOV) of 15.7 cm. [11] With multiple bed acquisition the scanner range is up to 200 cm. At TOH, the time per bed position is 105 s, typically with 8 beds in an eyes-to-thighs scan, and 11 beds in a full body scan. For an FDG PET scan, the patient receives an effective dose of around 7 mSv. [12]

As with most modern PET systems, the D710 also contains a 64-slice diagnostic x-ray computed tomography (CT) machine. The CT is used for anatomical localization when co-registered with the PET image, and for applying corrections during the PET image reconstruction. An example image with registered PET and CT is shown in Figure 2.4, with the PET image in colour to indicate regions of high activity, and CT displaying anatomical features.
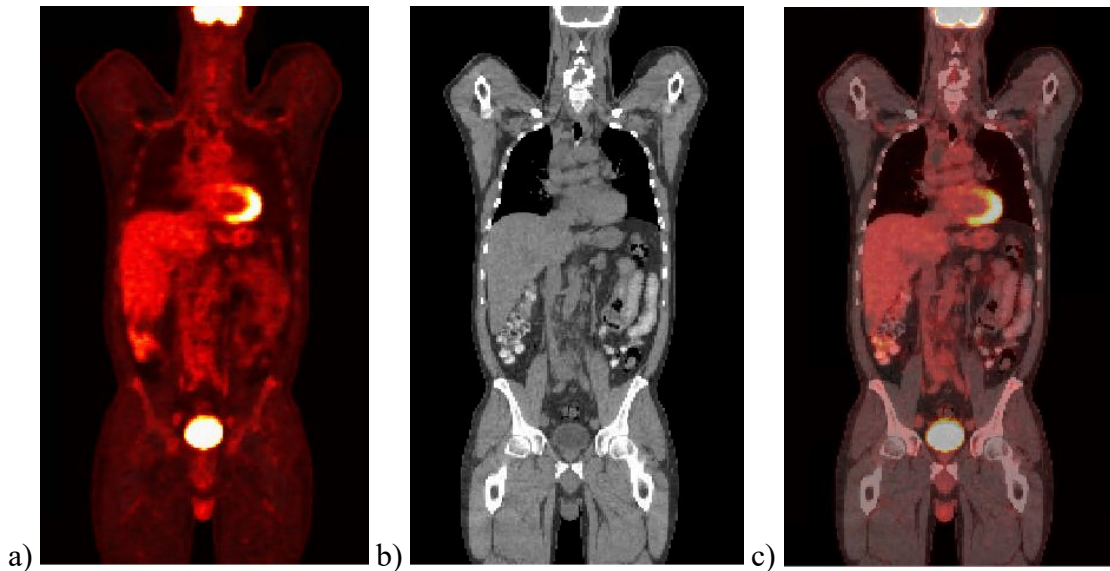
Figure 2.4: Patient images in the format of a) PET, b) CT, and c) co-registered PET/CT.

2.1.4   X-Ray Computed Tomography (CT)

CT imaging is an anatomical imaging modality. It uses a series of x-rays from different angles around the patient and scanned through the body to reconstruct a 3D attenuation map of different tissues in the patient. [13] When CT is used in the case of PET, it is used for attenuation correction of the PET image, as well as providing an anatomical map to localize physiologic features to the anatomy and better visualize the image. Since it is used as an accompanying image, the CT is usually acquired in a low dose mode (low x-ray tube current) and is not diagnostic, meaning that the quality of the image is lower than a regular CT scan and may not be used for all radiological interpretations. [14] This allows the radiation dose to be lower for the patient. For a chest scan, a diagnostic CT has an average effective dose of 7 mSv, whereas the low-dose CT has an average effective dose of 1.5 mSv. [15] On the Discovery 710 scanner, a whole-body low-dose CT scan has an effective dose of 3.2 mSv. [16]

2.1.5   Image Acquisition

Patients are required to follow a specific set of instructions to prepare for their PET study. In the case of FDG for PET imaging this includes fasting for 6 hours, no vigorous exercise for 48 hours prior to their appointment and especially in diabetics, ensuring that their blood glucose levels are sufficiently low - often using insulin. [12] At the time of their appointment, patients are administered FDG by venous injection, according to the patient's mass, with 4.99 MBq/kg in the case of TOH Regional Cancer Centre, to a maximum of 444 MBq. [12] Patients are to kept warm, and have limited movement, to prevent muscle uptake. Furthermore, patients are asked to hydrate and void prior to their study to encourage washing out of FDG through the urine to reduce patient dose and to reduce high intensity interference from the bladder. At 1-hour post-injection, patients are required to void and remove any metal objects before being imaged. The patient is imaged using CT according to the parameters in

Table 2.1. Immediately after the CT, the patient is imaged using PET according to the parameters in Table 2.2. After acquisition of the PET data, images are reconstructed on the console in several minutes, using the PET reconstruction parameters in Table 2.3. Scans are verified by technologists, checking for mis-registration of the CT and PET images (e.g., due to patient motion), artifacts, or data loss. On occasion, parts of the scan are repeated; for example, of the head, if patient motion has resulted in PET/CT misregistration.

Table 2.1: CT Acquisition parameters [12]

Parameter	Value
Mode	Helical
Rotation time	0.9 s
Rotation length	Full
Detector axial coverage	40 mm
Helical thickness	3.75 mm
Pitch	0.984:1
Bed axial speed	39.37 mm/rotation
kV	120
mA	Auto
Reconstruction algorithm	Q.Ac WideView
Recon mode	Full
ASiR (iterative reconstruction)	None
Reconstructed radial field of view	70 cm

Table 2.2: PET acquisition parameters [12]

Parameter	Value
Mode	Static
VPFX (Time of flight)	Yes
Bed motions	Step and shoot
Scan duration	1.75 min/bed position

Table 2.3: PET reconstruction parameters [12]

Parameter	Value
Reconstruction method	Q.Clear
Beta	550
Z-axis filter	Standard
Matrix size	192 x 192
Radial field of view	70 cm

2.1.6 Image Reconstruction

The simplest way to reconstruct an image from a sinogram is to simply

superimpose the projections from their respective angles. This analytical method, known

as back projection, is susceptible to blurring, and cannot incorporate accurate corrections of the imaging physics, so it is no longer used clinically. [5] Filtering can be applied when superimposing projections to limit blurring, known as filtered back projection (FBP), however, this simplistic technique produces low quality images as it cannot fully account for all aspects of the imaging process, including measurement noise, time-of-flight data, 3D PET geometry and motion. Contemporarily, iterative image reconstruction methods are used clinically with machine-learning approaches showing promise in the research setting.

Iterative image reconstruction methods use an estimate of the image, which is used to calculate projection data, and then compared with the measured projection data. After each iteration, the estimate is updated, the projection data are calculated, and the data are compared. During the projection calculation, physical effects can be modeled and therefore corrected for in the final reconstruction. Once the calculated and measured projection data converge (or after a certain number of iterations), the estimated image is now the reconstructed image.

2.1.6.1 Corrections

For accurate image reconstruction, corrections must be made to account for physical phenomena that may impact the quality of the signal measured by the device. These corrections include detector efficiency/normalization, attenuation of tissues, random coincident events, and photon scatter, to name the most common.

Detectors have inherent variations due to manufacturing variations and differences in geometry with respect to the activity distribution in the scanner FOV. The efficiency of the detector for measuring photons of different energies, angles, and

locations can be assessed by imaging a uniform source and/or by modelling. Efficiency is measured by the system during a calibration scan performed periodically (e.g., performed monthly or quarterly) or after any major repair to the system. The corrections are then applied during image reconstruction to compensate for these non-uniformities.

Another correction that must be made for the detector is the dead-time correction. [5] When a detector receives a signal, there is a brief time period afterwards when the detector is unable to detect another event. The correction can be done using system modelling. Dead-time correction becomes increasingly important with higher activity in the field of view, as the probability of near events (in time) can lead to greater event count loss.

The attenuation of photons in a medium depends not only on the material but also on the energy of the photon. For PET, photons have an energy of 511 keV, at which the most prominent type of interaction is Compton scattering. [17] Compton scattering causes photons to lose energy and changes their direction of travel through interactions with other particles in the body. The CT image is used to measure attenuation in the field of view and then derive a correction associated to the different tissues in the patient. It is worth noting that CT uses much lower energy (<150 keV) and polyenergetic photons, thus the attenuation coefficients must be scaled to 511 keV for PET attenuation correction. Since a PET measurement requires two photons detected coincidentally, the probability of both photons making it out of the patient depends on the whole length of the patient and isn't dependent on where on the LOR the event occurred. This makes it significantly easier to correct for attenuation in PET compared to single photon emission

computed tomography (SPECT), which has long been an important differentiator allowing for accurate quantification of activity concentrations with PET.

As mentioned above, photons can scatter when interacting with the patient. These scatter events cause the photon to lose energy, hence they can be accounted for by removing coincident events that are outside an energy window around 511 keV. [5] The upper limit of the energy window is larger than 511 keV (around 600 keV) to account for the limited energy resolution of the detectors (~10%), however the upper energy window is not much higher so as to reject more than one photon interacting with the detector at once. Energy discrimination, however, is not a solution to scatter, as the energy resolution of the detectors is limited, requiring relatively broad energy acceptance windows. Another method to correct for scatter is by creating simulated models using attenuation from the CT image and/or by analysis of the emission data.

Finally, random coincidence events create noise due to photons from separate decay events hitting detectors within the short time period of the coincidence time window. Random coincidence events are not from real annihilation events, and if left uncorrected, create a uniform blur in the reconstructed image. These events are purely random and are more difficult to account for. To correct for random events, delayed time windows or counting the number of single events can be used to estimate the rate of random events, which can then be removed from the histogram count rate prior to reconstruction.

Many of the above-described effects can be performed by incorporating models of them into the forward projection process during the image reconstruction process. This is a very powerful and convenient process, as almost any physical, instrumentation, or even

patient related effect can be corrected for, without having to devise an inverse model to correct for the phenomenon. Some vendors have used this approach to include corrections for system point-spread function (resolution correction), respiratory motion correction and cardiac motion correction. [18]–[20]

2.1.6.2 MLEM

In order to improve reconstruction of PET images, analytical methods like FBP have been replaced with iterative methods such as Maximum-Likelihood Expectation-Maximization (MLEM). MLEM uses statistical methods to limit image noise and artifacts, as well as apply the corrections mentioned above. [17] This technique can also be used with time-of-flight data, resulting in less noisy images.

Expectation-Maximization methods iterate to calculate a distribution of estimates of the activity in the patient. From there, Maximum-Likelihood methods are used to find which estimate is most likely by comparing with the projection data, and each iteration brings the image closer to this estimate and finally converges to a single image. The advantage of using this method, as opposed to analytical methods, is the reduction of streaking throughout the image and an increased signal-to-noise ratio (SNR). However, MLEM tends to have a slow convergence, taking many iterations to achieve an image with both high and low frequency features. [17] Also, this technique is susceptible to noise with too many iterations. To avoid this, a preset number of iterations is usually selected that balances high quality feature delineation without excessive image noise. However, MLEM is not frequently used due to the high computational time that makes it impractical for a busy clinical setting.

2.1.6.3 OSEM

Ordered-Subset Expectation-Maximization (OSEM) overcomes the limitation of MLEM by accelerating the reconstruction process, with little to no impact on the quality of the reconstructed image. [17] This technique uses the principles of MLEM but breaks up the data into S number of subsets of data, which are maximized separately, in sequence. This creates an accelerated convergence by a factor of $\leq S$, as fewer projections need to be forward projected at each step. While in MLEM, all projections of the estimated image are computed by the forward projector, in OSEM, only a subset of projections are computed each sub-iteration. The image update and forward projector remain otherwise the same as in MLEM, enabling respectively, to include all the physical correction in the system model and to ensure robust convergence towards a best estimate image. The advantages of MLEM still come through in this technique, such as the increased SNR, however, OSEM isn't guaranteed to fully converge to a single image. If noise is present in the image, then the iteration must be stopped before convergence, otherwise the noise is amplified. Another disadvantage to this technique is the distortion of small objects in the image. This is balanced by selecting a limited number of subsets (e.g. 8 or fewer).

2.1.6.4   Q.Clear

The GE Discovery scanners, one of which was used to collect data for this thesis, were the first to use a reconstruction algorithm called Q.Clear, that was developed by GE Healthcare. Q.Clear uses Bayesian penalized likelihood to suppress noise where images are uniform, using a relative difference penalty term, and accentuate contrast where edges exist. [21] The increased contrast along edges may create oscillations, called the Gibbs phenomenon, however, this can be minimized by increasing the image matrix

dimensions. [22] A beta factor is selected by the user to trade off between image smoothness and edge preservation. A beta value of zero ignores the Bayesian penalty and produces an OSEM like reconstruction. A beta value of 550 was used in this thesis, corresponding TOH clinical practice for FDG-PET. This algorithm is able to reach full convergence while maintaining image quality, and multiple studies have reported improved resolution, contrast, and SNR, with decreased noise, and even improved lesion detection. [23]–[25] Q.Clear is seen as a reconstruction method that effectively improves the system's overall sensitivity. This added sensitivity can be leveraged to improve image quality, reduce the amount of tracer activity used, and/or to reduce imaging time, according to the clinic's needs. In the case of TOH, Q.Clear was adopted to reduce image acquisition times, with the goal of increasing clinic throughput to match the growing demand for PET imaging in the region.

2.1.6.5   Image Reconstruction in Practice

Regardless of the method, reconstruction of PET images remains a computationally demanding task that can take minutes to hours on high-end dedicated computer hardware costing hundreds of thousands of dollars. On modern desktop computers, reconstruction takes several hours and up to a day. At TOH, in a clinical setting, reconstruction is done on the console using OSEM to produce a quick image to check for artifacts before the patient leaves the department and then Q.Clear reconstruction is queued to be completed overnight. More modern PET/CT systems are equipped with computing hardware that enables Q.Clear reconstruction within minutes, negating the need for preliminary reconstruction. As reconstruction times continue to

decrease, more elaborate corrections, such as respiratory motion, are increasingly being applied in the clinic.

2.1.7   Artifacts

In addition to the corrections of physical phenomena mentioned previously, there are other errors that can affect the quality and accuracy of PET images. Image artifacts are abnormalities that occur in images that are not in the original object.

Patient movement between the PET and CT images can cause misalignment and attenuation correction artifacts. [26], [27] In addition, breathing can cause blurring and misregistration of the CT image in the chest region of the image. This results in artifacts such as the banana artifact, which is banana shaped region of low activity on the liver-lung boundary. For breathing related artifacts, gated imaging can be used to reduce these effects. However, for gross patient motion artifacts, the technologist will often repeat the PET and CT image acquisition for anatomical regions that have been severely misregistered.

Furthermore, CT beam hardening and photon starvation can cause streaking and excess noise due to high density regions. [28] This can be caused by wide regions of the body with boney structures, such as through the shoulders and hips, but also from metal implants, calcifications, or x-ray contrast media. Consequently, the PET image will contain artifacts from overcorrection of high attenuation objects in the CT. Filters and current modulation can be used to prevent beam hardening and reduce the effect of these artifacts. Alternatively, iterative reconstruction software containing beam-hardening correction can be used to limit beam-hardening artifacts.

2.2   Image Visualization and Lesion Detection

A lesion is an area of damaged tissue, which can be from disease or an injury. This includes wounds, abscess, ulcers, and benign or malignant tumors. [29], [30] Since FDG is an analogue of glucose, lesions that are visible on FDG PET are typically malignant tumors, but benign lesions may also be seen. Tumors vary in size, from being a single cell to being larger than a human head. [31] Typically, larger objects are easier to perceive than smaller objects, however, that greatly depends on their contrast with the surroundings. FDG PET is no exception to this pattern. The relation between size and intensity relative to the ability to detect the lesion is studied in this work.

2.2.1   Perception Versus Detection

Clinically, lesion detection is done visually by clinicians. Ideally, clinicians use dedicated visualization software, using well calibrated displays, in specially designed reporting rooms. These are intended to create an optimal viewing environment to maximize the detection of disease. [32] Nevertheless, even the most well-trained clinician is prone to making mistakes. These errors can occur at multiple points throughout the disease detection process, including during perception and detection of the disease. Although perception and detection are often used interchangeably, they are two different processes. Perception is a subconscious process during which the brain takes in sensory information, visual in this case, and processes this information to make sense of it. Detection is a cognitive process during which the observer eliminates other options and decides that the object is in fact what they are looking for. Cognitive errors are estimated to account for around 15-28% of all errors in radiology, some of which are caused by biases. [33]

2.2.2   Human Errors

During detection tasks, errors can occur during perception of objects or the cognition process. When scanning an image, the observer may not consciously become aware of an object by either completely missing the object or by perceiving it but subconsciously disregarding it. Errors during this phase can be studied by tracking eye movement of the observer and measuring the amount of time spent on each region to tell whether the observer perceived the object. [33] Once the observer is consciously aware of the object, the observer must consciously decide to detect or disregard it. Human perception studies have shown that there are many different types of biases that are prevalent in this phase, such as confirmation bias, alliterative errors, anchoring bias, and satisfaction of search. [33]

Confirmation bias can occur when prior diagnostic information about the patient is known, causing observers to search for evidence supporting this information, and potentially missing other information in the process. Alliterative errors can occur if there were errors from a previous report that the observer takes as truth, without reviewing the findings, causing a continuous error. Furthermore, anchoring bias may occur when new information is received after an initial impression was made, but the observer does not adjust their prior impression based on the new information. Another common error is termed "satisfaction of search" in which the observer stops looking for additional abnormalities after initial findings. To mitigate these types of errors, readers are trained to systematically search the entire image space. Some have even advocated for the use of checklists, to guide the reader through an exhaustive search. These are just a few of the many types of biases when it comes to human detection.

2.2.3   Human Versus AI

Although there are techniques to reduce human bias, including structured reporting and error review, these strategies require more human resources, which are already stretched thin in most clinical practices. [33] Artificial intelligence has been proposed as a detection tool to allow for a reduction in bias and more accurate detection. During detection tasks, human observers rely on experience to produce accurate detection, whereas AI algorithms are trained from imaging datasets. If these datasets are robust, machine learning algorithms have been shown capable of learning to detect a wide variety of diseases. However, some rare conditions may not be included in the training dataset, causing AI to miss these diseases, whereas a human observer might not miss them through the ability to rationalize. Evidently, the training dataset is crucial to producing a sophisticated AI algorithm.

Moreover, test/retest discrepancies for human observers can be quite significant, even for highly trained clinicians, due to guesses and lapses in judgement. [34] The advantage of AI is that they are reproduceable and will retest with the same results.

2.2.4   Characterization of Detection Performance

Most commonly, detection accuracy is characterized using a validation dataset that has been labelled with reference ground truth of the detection outcome. The observer (whether human or machine) results are then compared against the ground truth. The results can then be summarized in a confusion matrix (see Figure 2.5) from which summary statistics can be calculated. The most common statistics include: [35]

- Sensitivity: probability of testing positive given that the ground truth is positive, equal to the number of true positives divided by the number of actual positives. Sensitivity is also referred to as recall.

- Specificity: probability of testing negative given that the ground truth is negative, equal to the number of true negatives divided by the number of actual negatives.

- Precision: probability of the ground truth being positive given that the test is positive, equal to the number of true positives divided by the number of predicted positives. Precision is also referred to as the positive predictive value.

		Actual Value	
		Positive	Negative
Predicted Value	Positive	**True Positive**	**False Positive**
	Negative	**False Negative**	**True Negative**

Figure 2.5: Confusion matrix for detection tasks

In the context of this work, sensitivity looks at an observer's performance from the point of view of the known ground truth lesions that were detected correctly. Precision, in contrast, looks at the lesions that were reported by the observer and evaluates which of these were actually ground-truth lesions (as opposed to mistakenly identified lesions).

In the context of lesion detection, for a given observer, these summary statistics may vary depending on how obvious the lesions are to detect. These statistics do not inform on the limits of detection with regards to the lesion characteristics. Hence, they make it difficult to infer how an observer will perform under conditions that are not represented by the validation set. For example if an observer is good at detecting advanced, large tumors, it is not known how well they can detect early stage, small tumors.

2.2.5   Perception Models

The ability to perceive a signal, such as lesion detection, may depend on numerous factors including lesion intensity, lesion size, background intensity and pattern, method of image display, observer proficiency, and lapses by the observer. Accounting for all these factors separately is challenging, thus a perception model can relate the likelihood of lesion detection to various lesion characteristics. When lesion characteristics are known, such as size and intensity, modelling can be used to extrapolate detection performance for values outside of tested values. The benefit of perception models is that they can succinctly summarize complex relationships of lesion properties to perception likelihood using a few parameters.[36]–[38] These parameters may then be used to benchmark or grade the lesion detection performance of observers.

This concept is well studied in the psychophysical discipline where human test subjects are tested for their ability to detect a physical stimulus (e.g., sound intensity, weight, colour difference or light intensity). Low intensity stimuli may not be perceived by the observer, but as the stimuli increases in intensity the observer transitions towards increased likelihood of perception. High intensity stimuli are rarely missed. This

relationship is typically modelled using a sigmoid function. To our knowledge psychometric modeling has not been applied to characterize AI's limits of detection with respect to lesion properties.

The following chapters describe a method in which well characterized synthetic lesions are embedded into real clinical image data, which are then used to test AI observers. The observer reported lesions are then compared to the ground-truth data to characterize the limits of detection of the AI algorithm with regards to the lesion properties using a perception model.

Chapter 3: Lesion Synthesis Toolbox

A major challenge to evaluating performance on the task of lesion detection using clinical images is the absence of reference ground truth data. Various workarounds have been applied to this problem. These are listed in Figure 3.1 in the context of the imaging process, and therefore illustrate the downstream imaging steps that can be probed for their impact on lesion detection. For example, Correlative imaging (e.g. magnetic resonance imaging) can be used to verify the presence of detected lesions, but may be unwarranted, as certain types of lesions may not be discernible on different imaging modalities. Moreover, availability of multiple imaging modalities (in close time proximity) is usually limited by consideration for costs, patient convenience and reducing the radiation dose to the patient. Another approach is to have PET images reviewed by a panel of expert human readers to mark lesions by consensus. This has the limitations of being extremely time demanding and expensive, while being limited by the performance of the human readers.

Histopathology, the sampling of suspect tissues (biopsy) for laboratory testing, is often considered to be a definitive ground truth. However, biopsies are limited by their invasiveness as lesions cannot always be extracted safely and needle biopsies often miss or under-sample the lesion, even while using image guidance. Perhaps most importantly, biopsies are taken only of suspect diseases sites, which are usually defined by the very same imaging we want to test for limits of detection. While many biopsies are taken during surgery, these do not qualify for ground-truth for subsequent imaging, as at that point the lesion site has been excised.

As an alternative, ground truths can be known by using a phantom. Phantoms are human-made patients, usually built with plastic and other materials, mimicking the properties of tissues. They have known dimensions and can be injected with radionuclides to perform various tasks in nuclear medicine, such as quality control testing of imaging equipment and methods. However, they are not realistic enough to be useful for most lesion detection studies, as they cannot represent the wide variety of patients, specifically their anatomy, physiology, and differing disease manifestations. Consequently, lesion placement in physical phantoms tends to be predictable.

Another method for acquiring a ground truth is by using digital phantoms, which are digitally modelled patients that can be fed through a computer model to create realistic looking image scans. These models can be simulated with noise and patient variability, as well as simulated instrumentation effects, to better reflect the complexities and uncertainties of real patient data. However, these types of simulations include many pitfalls. Synthetic patients are tedious, computationally demanding, lack image artifacts, and still do not account for all instrumentational and patient variability encountered in real data. Furthermore, uniform tracer concentrations in the phantom "organs" create unrealistic results, and pixel level realism phantoms are not readily available.

Thus the use of real patient data is advantageous to allow for realism and variability in anatomy and physiology. When it comes to how to insert the lesions, there are several possible approaches that can be taken. In the case of animal studies, it may be possible to do so surgically, but this method is expensive, inhumane and does not accurately represent the human subject. Artificial lesions may also be placed superficially on human volunteers, but these don't emulate deep tissue lesions. To overcome these

limitations, we sought an approach to insert synthetic lesions into real PET data in silico. Rather than adding them in the reconstructed images, where the effects of image reconstruction are difficult to model, we sought to insert the lesions into the raw data measured by the PET instrumentation. This allows the lesions to undergo reconstruction and obtain characteristics of the reconstruction parameters used.
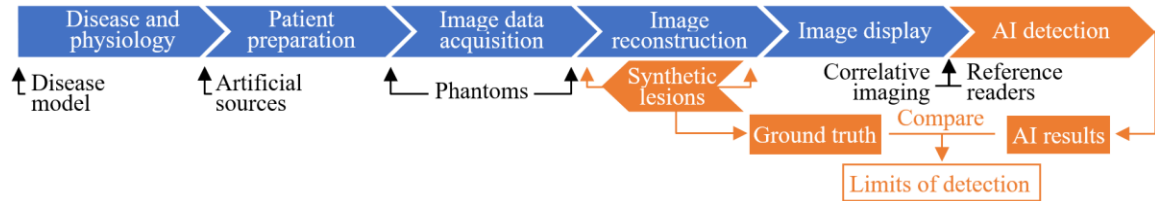
Figure 3.1: Workflow to display a medical image, with blue indicating the clinical path, orange indicating the path added in this project, and black indicating other methods of creating a known ground truth.

This chapter describes the methods by which lesions were synthesized in PET and CT data and follows with a description of an in-house developed software tool for lesion insertion by non-expert users – the Lesion Synthesis Toolbox (LST). This chapter concludes by distinguishing the incremental developments performed as part of this thesis project to enable efficient creation of large libraries of images with synthetic lesions to empower researchers.

3.1 The Lesion Synthesis Toolbox (LST)

The LST was designed to be a standalone application that could be used by non-technical experts to efficiently build libraries of PET/CT images with synthetic lesions and their corresponding ground-truth values. When the LST was created, the following key requirements were identified to be included in the toolbox:

1. <u>Importing data:</u> the program should be able to easily import all relevant data directly from the imaging modality. Because the raw PET data are relatively large

(~1 GB per study), these data are rarely archived and are only kept on the modality for a limited time (several days).

2. Data anonymization: to be able to use patient data in research, removal of sensitive patient identifiers is paramount. The tool must be able to remove all patient identifier information easily and robustly, then replace it with convenient study identifiers.

3. Visualizing data: the program should allow users to easily visualize data for quality control of reconstructed images and to intuitively define lesions within the image.

4. Realistic lesion insertion: the program should be able to insert realistic-looking and accurately characterized (e.g., location, size, and intensity) lesions into the PET data and allow lesions to undergo reconstruction with the same parameters as the image. Also, the reconstruction should be the same as those used clinically.

5. Batch data: the program should be able to batch process data from a queue to eliminate the need for the user to supervise the program and manually start reconstruction jobs.

6. Reviewing results: the program should allow users to review the results after lesions have been inserted and view the properties of each lesion.

7. Organized database: the program should create an organized database for raw data, reconstructed images, and lesion-simulated images.

8. Reconstruction parameters: the program should allow users to modify the image reconstruction parameters and allow for saved parameter options. The

reconstruction methods should be representative of those available in a clinical setting.

9. <u>Easy to use Graphical User Interface (GUI):</u> the program should be intuitive for users to navigate without previous experience using the program. The program should include separate sections for data retrieval, image reconstruction, and lesion insertion tasks.

10. <u>Exporting data:</u> the program should allow users to export the reconstructed and simulated images, as well as the corresponding ground truth values for the simulated lesions.

11. <u>Multiple users:</u> the program should be able to have different users with their own saved data folders and project options. Since image reconstruction requires long processing times, in some lab settings, it may be preferable to host the program on a shared compute server where multiple users can queue jobs. The software should isolate users from interrupting each other but manage the shared computing resources.

The LST is a tool created at the Ottawa Hospital by Dr. Ran Klein and Hanif Gabrani-Juma to simulate synthetic lesions in real PET/CT data, based on the above stated requirements. The requirements that were not met in the original creation of this toolbox were added as part of this project, as listed in the incremental developments in Section 3.3, by either the author or Dr. Ran Klein. The LST is developed in Matlab (MathWorks, Natick, Massachusetts), with the most current version tested on Matlab version 2023a. The toolbox is divided into 5 tabs, including the login, data retrieval, image reconstruction, lesion simulation, and configuration tabs. These tabs can be seen in

the user manual in Appendix A along with descriptions of each component of the user interface. The toolbox GUI is organized to be intuitive and easy to use, making use of common GUI elements and a top-left to bottom-right workflow common in many modern applications. With the further accompaniment of the user manual, the eighth requirement in the list above is satisfied.

The Data Retrieval tab allows users to import data directly from a PET/CT scanner and select which patients to download, satisfying the first requirement for an efficient toolbox. The PET/CT scanner connection can be set up in the "Configure" tab. Alternatively, patient data can be saved from external sources into the same directory as the downloaded patients. Immediately after downloading, patient data can be anonymized to protect patient confidentiality, satisfying the second requirement.

On the Image Reconstruction tab, downloaded patients can be reconstructed with user-defined reconstruction parameters, using OSEM and Q.Clear reconstruction algorithms. Reconstruction on a modern desktop computer takes several hours, depending on computer resources and number of PET scan length (number of bed positions).

The Lesion Simulation tab consists of 4 sub-tabs, namely "Select Patient", "Lesion Insertion", "Job Queue", and "Review Simulation". The "Select Patient" tab allows the user to view the reconstructed patient images and select one for lesion insertion. The "Lesion Insertion" tab (see Figure 3.2) is where lesions can be defined by their location, size, shape, intensity for the PET image (in either Becquerels per ml (Bq/cc), SUV, or relative to another reference region in the image), and intensity for the CT image in Hounsfield Units (HU). Reference objects can be added as well, typically in the liver. Once the lesions are added, the reconstruction parameters can be chosen from a

list of preset ones or can be manually input. These tabs satisfy the third and fourth

requirements for an efficient toolbox.

Figure 3.2: Screenshot of the lesion insertion tab from the Lesion Synthesis Toolbox

The "Job Queue" sub-tab allows the user to view and remove simulations that are

in the queue to be processed for generating lesion images. After simulation, the images

can be reviewed and downloaded in the "Review Simulation" tab. The ground truth data

can also be downloaded in this tab. In addition, lesions can be retrieved and brought back

to the "Lesion Insertion" tab, which allows the user to repeat the same reconstruction

using different parameters or allows for adding and removing lesions. The "Job Queue"

and "Review Simulation" tabs satisfy the fifth, sixth, and tenth requirements for the

toolbox.

The Configure tab allows the user to customize settings of the toolbox.

Administrators can configure common settings such as connections to scanner consoles

for data retrieval and location of job queues. Individual users can configure projects that

include locations of their image libraries, image naming conventions, default image

reconstruction methods, where to send reconstructed images, and push notification of

processing progress. The "Configure" tab satisfies the seventh, eighth, and eleventh requirement.

3.2   Lesion Insertion

As mentioned previously, the "Lesion Insertion" sub-tab allows users to add synthetic lesions into previously reconstructed PET/CT data. The previously reconstructed data are only used to visualize the image, to sample activities in reference regions of interest and to sample background activities that inserted lesion activities must complement to achieve the target lesion activity. For this reason, it is preferable to first reconstruct the baseline image with the exact same reconstruction parameters as the simulation. In fact, if this is not manually done during image reconstruction, reconstruction with the same parameters as the simulation, will be automatically performed during the simulation process.

In this work, all reconstructions were performed using the GE Healthcare, DUETTO package, which performs image reconstructions that are numerically identical to the clinical reconstructions performed by the scanner. However, the DUETTO package is limited to researchers in collaboration with GE Healthcare. In theory, the software is designed to accommodate reconstruction packages by other vendors, but this was out of scope for this thesis.

Lesions are forward projected also using the DUETTO package, which estimates the resulting sinogram from the added lesion activity distribution. It does so using the same forward projector that DUETTO uses during iterative image reconstruction. [39] Next, the lesion sinogram is added to the image sinogram before reconstruction by DUETTO. This process is illustrated in Figure 3.3.
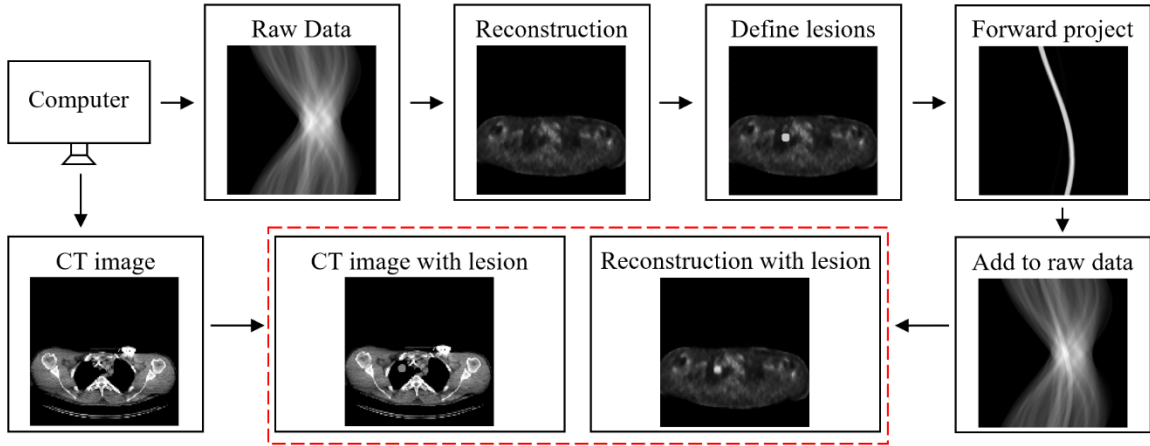
Figure 3.3: Flow diagram of lesion insertion into raw sinogram data and the direct addition of the lesion to the CT image, for realistic lesion simulation.

The DUETTO package is numerically equivalent to the reconstructions that are done on the PET/CT machines, although the speed of reconstruction is typically slower on a computer than on the console. By inserting lesions into the sinogram space, lesions undergo the same reconstruction and experience the same distortion or enhancement effects as the measured patient data, rather than simply inserting a blurred lesion directly into the PET image.

3.3   Incremental Developments

For the Lesion Synthesis Toolbox to be used for this project and to be ready for distribution to researchers outside our group, the toolbox needed updated functionality and to undergo thorough testing. The following incremental developments were performed in the context of this thesis by the author and Dr. Ran Klein:

3.3.1   Increasing Lesion Realism

Although spherical lesions are easily characterizable with only a radius parameter, real tumors are rarely perfectly spherical. Therefore, blobby spheres were created as a way to create more realistic lesions to reflect actual patient data, as seen in Figure 3.4.

These lesions are defined using two parameters, the lesion size and the amount of 'blobby-ness'. Blobby spheres were created using a sphere with varying radius values determined by random numbers within a given range. The range is defined by the second parameter input from the user, where the input ranges from 1 to 10, where 0 would represent a regular sphere, and blobby spheres of 1 being the least irregular and 10 being the most irregular (i.e., the radii values would be: radius $\pm$ range$\times$rand$\times$radius/10, where rand is a uniformly distributed random number between 0 and 1).

Figure 3.4: Examples of spheres with diameters of 15 mm, with a regular sphere on the left, followed by blobby spheres of parameters 1, 4, 7, and 10, respectively.

However, as the aim of this thesis was to study the detectability of lesions, only spherical lesions were used in this thesis, since characterizing lesions by one shape parameter is ideal for this initial study of lesion perception by AI.

3.3.2   Lesion Synthesis in CT

As with PET, lesion synthesis in CT was also desirable to enhance realism. Unlike PET, for the CT, the lesions are appended directly into the reconstructed image in one of two modes. The first (homogenous) mode inserts the lesion with the user defined activity, overwriting the original voxel values. In the second (maintain texture) the original voxel values are sampled from the image and a complimentary, uniform, intensity is added to achieve the target mean value, while preserving the difference between voxels. The second method is preferred for realism, as it maintains realistic image noise characteristics. Neither method accounts for partial volume effects of the

lesion, hence a voxel is either in the lesion or not. In addition, a button was added to redo the lesion insertion in the CT images, which can be used for previously simulated files that do not have CT lesions or if any future adjustments are made.

To clarify, synthetic lesions are added to the CT for visualization and analysis purposes, but not for PET image reconstruction corrections. Because the CT is used to describe the physics (i.e., attenuation and scatter) during PET acquisition, modifying the CT by lesion insertion could produce mismatches between these physical effects during image acquisition and their correction during PET image reconstruction, resulting in image artifacts. A copy of the unaltered, original CT must be preserved to enable future reconstructions and simulations.

3.3.3   Functionality for Generating Large Datasets

In order to create large sets of data, batch processing was added to allow users to add simulations to a queue and have the toolbox complete the tasks consecutively, instead of manually starting each simulation. Also, a retrieve lesions button was added as a way to bring the original file and lesion parameters of previous simulations back into the lesion insertion window. This allows the user to redo the simulation with different reconstruction parameters or edit and add new lesions. The ground truth table was updated to include anonymized patient information, as is described in Appendix A.

3.3.4   Integration with Clinical Image Visualization Systems

In order to integrate the toolbox with clinical systems such as HERMES (HERMES Medical Solutions, Stockholm, Sweden), a DICOM send function was added as a way to send reconstructed data to an external server. This is useful for visualizing

data using the same tools as clinicians. Files are sent in a universally used DICOM format, which is explained below.

3.3.4.1   DICOM Image Format

The PET data from the LST are saved in a DICOM format, which stands for Digital Imaging and Communications in Medicine. This format is the standard for transferring medical information and can be used across many vendors and equipment. It includes information structures, which in this case contains patient, image, and reconstruction information. [40]

3.3.5   Knowledge Translation

In an effort to share the LST with the scientific community, several components were created and updated. First, a user manual was created to allow users to have a reference for the functionality of the toolbox, which was updated as new functionalities were added. A GitHub repository containing the toolbox was updated as the toolbox was developed, and the user manual was added to it. Once the toolbox is ready to share, this repository will be used to share the toolbox and associated documentation. In addition to the toolbox, an open-source lesion database is in development using the lesion library, which is discussed in Chapter 4.

Chapter 4: Limits of Detection in AI

4.1   Background

Using the Lesion Synthesis Toolbox, a library of synthetic lesions and their associated ground truth values was generated. This library was then used to characterize the performance of two externally developed lesion-detection AI based systems in terms of their limits of detection (LOD). Detectability was characterized in terms of the size and intensity of the lesions at which the AI algorithm can reliably detect them. The hypothesis is that smaller and fainter lesions are more difficult to detect than larger and brighter lesions, with each AI algorithm having different LOD characteristics that can be used to objectively benchmark their performance for the clinical task of lesion detection. Furthermore, the AI algorithms were evaluated for the rates of falsely reported lesions, which were also characterized in terms of reported lesion size and local image activity.

4.2   Methods

This study was approved by The Ottawa Hospital Research Ethics Board (Protocol number 20180722-01H). The study was exempt from needing to obtain patient consent since it made retrospective, secondary use of patient data and the data were fully anonymized.

4.2.1   Patient Data

Routine clinical PET/CT scans from The Ottawa Hospital were selected if they were clinically reported as free of disease by the nuclear medicine physicians in definitive terms. Studies were checked on the department clinical viewing system (HybridViewer version 6.1.4, HERMES Medical Solutions, Stockholm, Sweden) before downloading data, to ensure the images were free of severe image artifacts or distortions

and appeared free of lesions. All images were acquired at TOH using the GE Healthcare

Discovery 710 PET/CT scanner (Waukesha, Wisconsin, USA). All scan relevant data

were imported to the LST directly from the PET/CT system console and immediately

anonymized before reconstructing the images. The downloaded files included the patient

specific raw histogram data (with time-of-flight) and reconstructed CT scans, and the

system detector normalization, geometry correction, and calibration files. These files are

required to perform reconstruction of the PET data using the same methods as are used

clinically.

All patient demographics data (e.g., age, sex, weight, height, study type) were

extracted from the raw data header metadata. Patient gender and ethnicity were not

captured in these as per clinical practice, nor were they considered significant for this

research, as these are not known to be important differentiators of anatomy and

physiology.

4.2.2   Lesion Library

Anonymized patient files were reconstructed in the LST using the same algorithm

and parameters used in the clinic (see Table 2.3 in Section 2.1.5). Briefly, reconstruction

used the Q.Clear algorithm including time-of-flight, attenuation, scatter, dead-time,

normalization, and decay corrections. Using the reconstructed PET image, registered with

the CT for anatomical and physiology references, synthetic lesions were manually

defined on these images to create a library of simulated PET/CT data. A variety of well-

defined synthetic lesions were placed in anatomically realistic sites. The author reviewed

medical images and journals to gather a list of common locations for soft tissue tumours

before creating the library. Tumours in bone were avoided, as they had not yet been

validated using the LST. The goal of the library is to create an open-source database of images that can be used for training, testing, and studying the limits of detection for both human observers and AI. The library itself doesn't contain any information on the location or number of the lesions to preserve objectivity of the AI developers, but the ground truth data were downloaded from the LST for internal use to score the AI reported lesions.

A perception study had been performed previously using the LST to study limits of detection in human observers with varying levels of experience. [41] The experience from that study was used to create a baseline estimation for how small and dim lesions could be while still being detected by human readers of varying levels of experience. Using the results from that study as a guideline, the lesion library was built.

This library consisted of 56 patients, 114 simulations, and 565 lesions. More than one simulation was performed for most patients using different lesions, which is why there is a higher number of simulations than patients. Lesions were defined with intensity values between 0.5 and 6 times the intensity of the liver and sizes varying between 1 and 15 mm. Liver intensity was determined by manually placing a 20 mm radius sphere within a uniform region of the liver. Lesions were defined manually in the lesion synthesis tab of the toolbox by their location, size, shape, and PET/CT intensities. The process to insert reference regions and synthetic lesions is depicted in Figure 4.1.
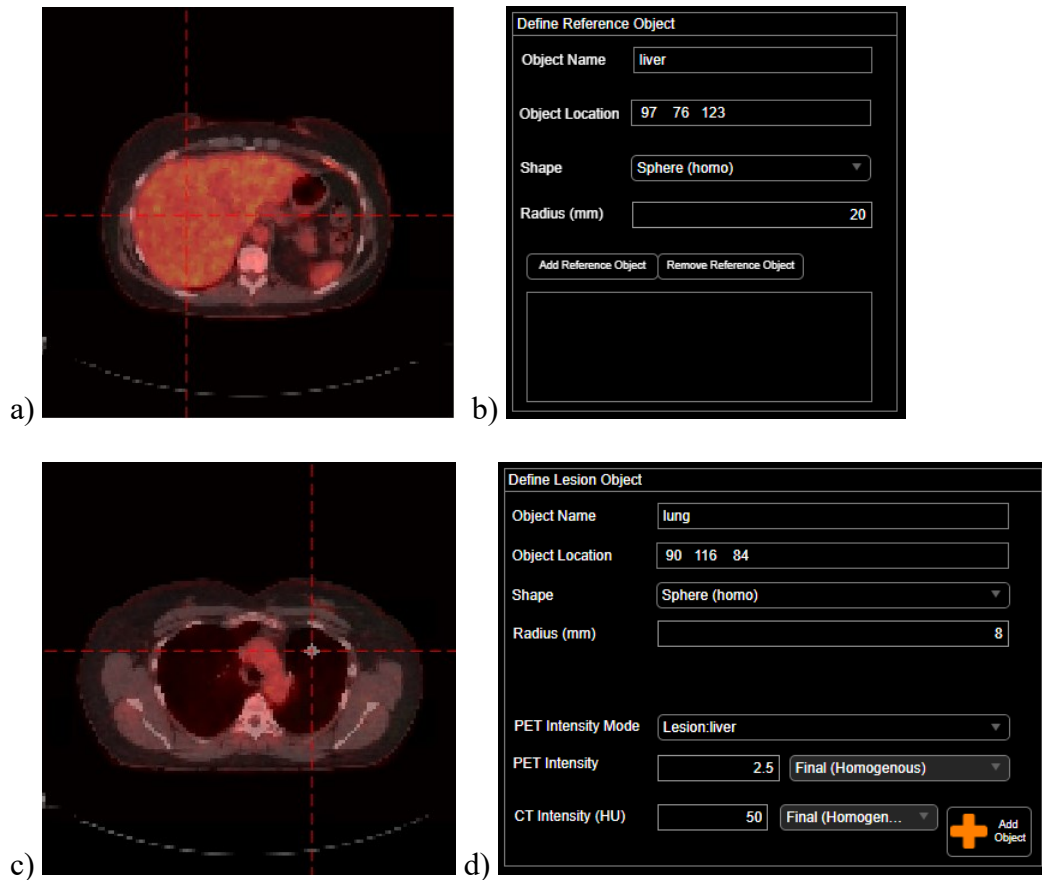
Figure 4.1: Steps to insert a synthetic lesion into PET/CT data in the LST, a) the location of the reference region is set in a uniform region of the liver, b) reference object parameters are defined, c) lesion is set in a realistic region of the body, and d) lesion parameters are defined.

The reconstruction parameters were selected to match those used clinically; then the studies were queued for synthesis. After reconstruction, the images were proofed using both the LST and HERMES. After processing and reviewing the library, three studies were removed from the library due to 1) corruption of data during information transfer and 2) image artifacts created from lesion insertion during image reconstruction. Note that the library values listed above do not include the excluded studies.

The ground truth data were exported by the LST in a Microsoft Excel (Redmond, Washington) file and included table columns for the patient and simulation parameters as listed in Table 4.1.

Table 4.1: Exported Ground Truth Parameters

Parameters	Details
Patient level	
Patient ID	
Age (years)	
Sex	-
Height (m)	
Weight (kg)	
Study level	
Injected Activity (MBq)	-
Simulation level	
Simulation Name	
Reconstruction Profile	
Series Description	-
Simulation Date	
Lesion level	
Lesion Name	e.g., lung, liver, etc.
Location Coronal (units)	Units include pixels, mm from center, mm from origin, and relative fraction. Selected by user before ground truth download
Location Sagittal (units)	
Location Transaxial (units)	
PET Intensity Mode	e.g., Lesion: Background, Lesion:ROI, SUV, Bq/cc
PET Intensity Value (Bq/cc, SUV or unitless)	As per PET Intensity Mode
CT Intensity Value (HU)	-
Shape	e.g., sphere, blobby sphere
Shape Size (mm)	-
Reference Lesion Activity (Bq/cc)	As per reference ROI, or Baseline Background Activity
Baseline Background Activity (Bq/cc)	Sample from the reconstructed image (same reconstruction parameters as simulation) prior to lesion insertion
Target Lesion Activity (Bq/cc)	Calculated
Surrounding Margin (mm)	Surrounding region is defined as a 2 cm margin around the defined lesion. These values are sampled from the reconstructed image (same reconstruction parameters as simulation) prior to lesion insertion.
Surrounding Mean PET Intensity (Bq/cc)	
Surrounding SD PET Intensity (Bq/cc)	
Surrounding Min PET Intensity (Bq/cc)	
Surrounding Max PET Intensity (Bq/cc)	

4.2.3   AI Processing

Two independent research groups that were developing AI for lesion detection in PET were invited to participate in this study. Both of these groups used real tumours to train, test, and validate their AI.

PAIRE is a Paris based startup company using a 3D deep convolutional neural net-based AI to detect lesions in FDG PET/CT scans. [42] The PAIRE AI was trained using real lesions from two lymphoma clinical trial datasets. The lesions were semi-automatically segmented using an external software and reviewed by two physicians to form the ground truth on which PAIRE AI was trained and validated.

QURIT is a lab based at the University of British Columbia which has developed a PET/CT lesion detection AI using 3D residual U-nets. [43] This AI was trained using real lesions from a combination of 4 external datasets containing a variety of lymphoma cases from Canada, South Korea, and Germany. Three of these datasets were segmented by five physicians separately. The fourth dataset was taken from the Cancer Imaging Archive, which included ground truth segmentation results from two radiologists.

In both cases, the complete image library was sent for automated analysis. Reports of detected lesions were returned in NIFTI segmented files for objective scoring of lesion detection.

For the QURIT data, all the lesions were segmented using a single region of interest. Therefore, in-house analysis was used to separate these into unique disjoint regions. In the PAIRE data, each lesion was already labelled as a unique region. Thus, the results of both methods were converted to a consistent format with each lesion indicated as a spatially connected segment in the image pixel matrix.

Herein, the methods are arbitrarily referred to as Algorithm A and Algorithm B to obscure performance of each AI.

The center of mass of each segmented region was calculated and used to encode the center coordinate of each lesion. To measure the lesion size, the average of two methods was used. The first measurement was done by averaging the difference between the minimum and maximum coordinate values in each Cartesian direction. If the lesion is not a round shape, or has protrusions, this method is susceptible to overestimating the lesion size. The second measurement was done by calculating the volume (number of voxels in the segment multiplied by the voxel volume) and solving for the diameter from the volume of a sphere equation. This method assumes that the segmented region is spherical. Because neither approximation of the lesion diameter is completely accurate, their average was used, to avoid extreme outlier values. Lesion diameter was used, as opposed to radius or volume, since the lesion diameter was varied in intervals from 1 to 15 mm, making the x-axis of graphs simpler and more readable. Furthermore, segmentation metrics such as dice coefficients were not used, since this study deals with the detection of lesions, not segmentation.

In addition, the lesion intensity, background intensity, and background standard deviation were measured from each image. These values were used to calculate the contrast and contrast-to-noise ratio, as they are defined in Section 4.2.4. After reading in each data set and calculating the location and size values, the patient ID, study number, Cartesian coordinates, size, intensity, contrast, and contrast-to-noise ratio were exported as a comma-separated values (CSV) table for both AI.

4.2.4   Characterization of AI Performance

Lesions reported by the AI were compared with ground truth data by relative position. Reported lesions that were within 2 pixels (~ 0.7 cm) of the center of ground truth lesions were scored as true positives and a hit (i.e., these were deemed lesions that were correctly detected by the AI). Reported lesions that were not within 2 pixels of ground truth lesions were deemed false positives (i.e., lesions that were reported, but were not synthesized). Finally, lesions that were synthesized but did not have a corresponding reported lesion within 2 pixels distance to the center were scored as false negative and a miss (i.e., a true lesion that was not detected). True negatives are regions that had no lesion inserted or detected and were not considered in the analysis. Figure 4.2 summarizes these scoring combinations.

		Actual Value	
		Positive	Negative
Predicted Value	Positive	**True Positive** **(Hit)** Synthetic lesion is detected	**False Positive** **(False Finding)** Non-lesion is detected
	Negative	**False Negative** **(Miss)** Synthetic lesion is not detected	**True Negative** - Non-lesion is not detected

Figure 4.2: Lesion scoring grid.

After scoring, true positive lesions were plotted on a bar graph by their simulated anatomical location for each algorithm, alongside the total number of simulated lesions in said location.

Moreover, true positives (hits), false negatives (misses), and false positives (false findings) were plotted with respect to lesion size (diameter) and intensity. Several intensity metrics were evaluated including:

- Lesion intensity relative to the liver

- Lesion to background contrast

- Contrast-to-Noise Ratio (CNR)

As mentioned previously, the lesion intensity relative to the liver was defined in the LST during lesion creation using a reference region in a uniform region of the liver. Lesion intensity was also measured from the image post-reconstruction and compared to the defined lesion intensity (i.e., measured versus expected). When measuring the values from the image, the mean lesion intensity (I_{les}) was measured within the detected region, for AI lesions, or defined region, for ground truth lesions. The background mean intensity (I_{bkg}) and standard deviation (σ_{bkg}) were measured using a 20 mm margin around the lesions. Lesion contrast was calculated using the mean background and mean lesion activity, as seen in Equation 4.1. [44] The CNR was calculated using the mean background activity, mean lesion activity, and the standard deviation of the background, as seen in Equation 4.2.

$$\boldsymbol{Contrast} = \frac{|I_{les} - I_{bkg}|}{I_{bkg}} \qquad\qquad \textbf{Equation 4.1}$$

$$\boldsymbol{CNR} = \frac{|I_{les} - I_{bkg}|}{\sigma_{bkg}} \qquad\qquad \textbf{Equation 4.2}$$

The contrast and CNR values were calculated during both the download of ground truth data, for true positives and false negatives, and during analysis of each AI measurement, for true and false positives. The expected contrast and CNR values were

calculated using the expected lesion intensity (I_{exp}) calculated from the lesion intensity

relative to the liver ($(I/I)_{rel}$) defined by the user in the toolbox,

$$I_{exp} = \left(\frac{I}{I}\right)_{rel} I_{liver} \qquad \textbf{Equation 4.3}$$

where I_{liver} is the intensity of the liver.

The measured contrast and CNR were calculated using the mean lesion intensity

measured from the image. Equation 4.3 was also used to convert the measured lesion

intensity (by replacing the expected intensity with the measured intensity) to the intensity

relative to the liver by dividing by the intensity of the liver reference region, for detected

lesions. The expected and measured values were compared to influence of the lesion

insertion and reconstruction on the final contrast and CNR values.

Likewise, the expected and measured lesion sizes were plotted for each AI

algorithm. A line of best fit (with 0 intercept) was fitted on these graphs and subsequently

used to calibrate the measured size for false positives. This was done to model what the

lesion size would have been, if the false positive lesions were true positives, allowing for

a more similar comparison of data of the false positive lesions with true lesions. The

calibrated false positive values were plotted along with the hit and miss lesions with

respect to lesion size and the measured intensity, contrast, or CNR values. By using the

measured values for all three lesion types, the values are in comparable terms.

4.2.5   Psychophysical Response Model

The hits and miss data were used to fit a perception model of lesion detection

probability as a function of lesion size and lesion contrast. First, a power-law model was

used as a measure of the intensity of the stimulus (S) as previously proposed in,

$$S = A \cdot d^D \cdot c^C \qquad \textbf{Equation 4.4}$$

consisting of three free parameters: D the power of the lesion diameter, d; C the power of the lesion contrast or CNR, c; and A an amplitude normalization factor. [45], [46]

The signal perception psychometric response was modeled using a Weibull function, as shown in Equation 4.5. [36] The model determines the probability (P) of detecting a lesion of a given size and contrast.

$$P = (1 - \gamma - \lambda)\left[1 - e^{S^{\beta}}\right] \qquad \textbf{Equation 4.5}$$

Both the guess and lapse rates, γ and λ respectively, were set to zero. This was done under the assumptions that the experiment consisted of lesion localization and the odds of guessing the lesion location was very low ($\gamma = 0$). Furthermore, AI, unlike human observers, are expected to be reproducible and untiring, hence lapses were assumed unlikely ($\lambda = 0$). The remaining parameter, β, is redundant with parameters A, D, and C and therefore was set to 1. Consequently, parameter A is correlated with the breadth of the transition zone between where signal intensity shifts from a high degree of lesion detection to high degree of lesion miss. Model fitting was performed using an iterative (without a differential estimation) cost minimization function, as shown in Equation 4.6. The first term penalizes for difference between lesion score (where 0 is a miss and 1 is a hit) and the model predicted probability of lesion detection. The second term strongly penalizes for negative model parameter (A, D, and C) values. W was set at 500.

$$Cost(x) = \sum_{i=1}^{n}(P(x, d_i, c_i) - Score_i)^2 + W \cdot \sum_{j=1}^{3}\begin{cases} 0 & x_j \geq 0 \\ 1 & x_j < 0 \end{cases} \qquad \textbf{Equation 4.6}$$

The model response was plotted as a function of diameter and contrast, with the responses superimposed. Furthermore, the 80% and 95% probabilities of lesion detection contours were arbitrarily selected to represent fair and good levels of performance respectively and were emphasized graphically on the model response plots. Limits of

detection were extrapolated from the model outside the range of experimental values as approximations for subsequent research.

4.2.6   Statistical Analysis

Population summary statistics (including age, weight, injected activity, etc.) were presented as mean values (μ), standard deviations (σ), and range in the form of $\mu \pm \sigma$ [min, max]. Sensitivity and precision were measured using Equation 4.7 and 4.8. Due to a lack of true negatives, specificity was not measured. [35]

$$sensitivity = \frac{true\ positives}{true\ positives + false\ negatives} \qquad \textbf{Equation 4.7}$$

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \qquad \textbf{Equation 4.8}$$

To compare the psychophysical model parameters of the AI algorithms, model fitting was performed with 1000 bootstrap sampling. The parameters were reported as median value and interquartile ranges, and the medians were compared using a Wilcoxon rank-sum test and p-values were reported, with $p<0.05$ considered statistically significant.

4.3   Results

The lesion library's patient demographics can be seen in Table 4.2, with a wide distribution of weight, height, and age, and a balanced distribution of males and females.

Table 4.2: Demographics of the lesion library

Demographics [n = 56 patients]	$\mu \pm \sigma$ [min-max]
Injected Activity (MBq)	253 ± 45 [133, 322]
Weight (kg)	80 ± 19 [41, 128]
Height (m)	1.69 ± 0.12 [1.47, 1.92]
Age (years)	63 ± 14 [27, 85]
Females (%)	46%
Number of Lesions per Study	5.0 ± 1.1 [3, 7]

4.3.1   AI Reported Lesion Scores

Table 4.3 summarizes the number of simulations comprising the lesion database and the number of lesions reported by each AI. In both cases many more lesions were reported by the AI than were actually simulated (1381 and 1098 for Algorithms A and B, respectively, versus 565).

Table 4.4 summarizes the scoring of these lesions against the ground truth data with Algorithm B having more true positives, fewer false positives, and fewer false negatives, indicating superior performance over Algorithm A in all metrics. Both algorithms have low precision, whereas only Algorithm A has low sensitivity.

Figure 4.3 shows an example image of the detection of lesions for a patient for both algorithms. Some detected regions are matching in both AI, but Algorithm A seems to pick up large regions in the abdomen.

Table 4.3: Lesion simulation and detection information

Number of Patients	Number of Studies	Number of Simulated Lesions	Number of Detected Lesions
56	114	565	Algorithm A: 1381
			Algorithm B: 1098

Table 4.4: Lesion scores for each algorithm

AI	True Positive (Hit)	False Positive	False Negative (Miss)	Sensitivity	Precision
Algorithm A	119	1182	366	24.5%	9.1%
Algorithm B	374	724	191	66.2%	34%

a)   b)
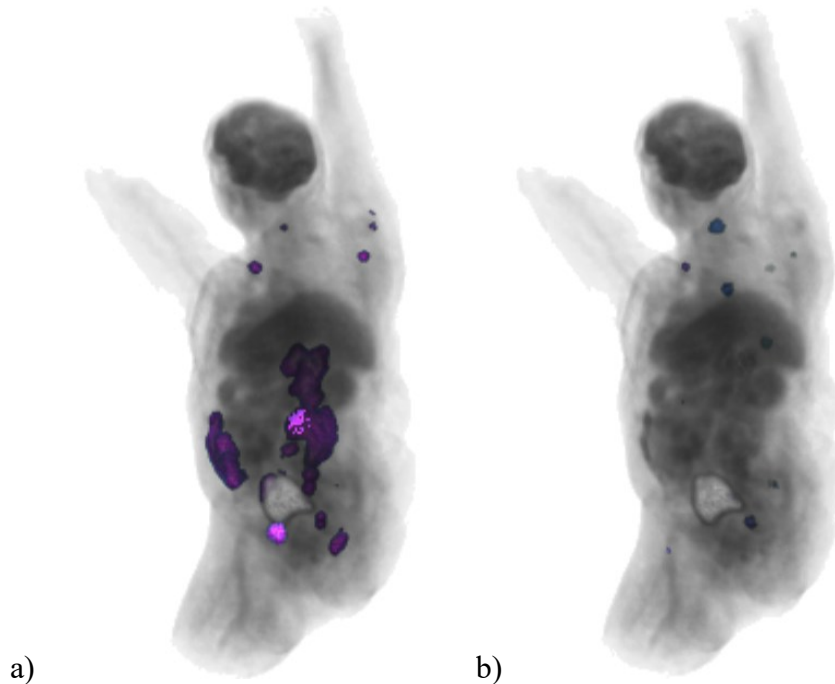
Figure 4.3: Detected lesions (blue and purple) in Volume Viewer on Matlab, overlayed with the respective image (greyscale) for a) Algorithm A and b) Algorithm B.

4.3.1.1   Anatomical Lesion Locations

The hit and miss lesions categorized by anatomical region are listed in Table 4.5 and Figure 4.4 for Algorithms A and B. Algorithm B outperformed Algorithm A for 16 of the 22 lesion locations. Three of the locations had equal detection between algorithms, while the remaining 3 locations had no hits for either algorithm. By design, a large number of lesions were placed in the lungs, since the lungs are the only region in the body with both low CT and low PET values, allowing for high contrast lesions. Of the locations with more than 15 total lesions, only the axilla and breast had detection of 50% or more of their respective lesion number for both AI.

Table 4.5: Total Lesions and Hits for Algorithms A and B by Anatomical Location

Anatomical Location	Total Lesions	Algorithm A Hits	Algorithm B Hits
Abdomen	84	28	57
Aorta	2	0	2
Arm	1	1	1
Axilla	74	41	62
Breast	16	8	14
Bronchi	26	9	17
Esophagus	1	0	0
Groin	2	1	2
Hip	12	6	9
Jaw	1	1	1
Knee	1	0	1
Leg	1	0	0
Liver	23	6	19
Lung	197	67	117
Neck	8	2	4
Ribs	12	3	10
Shoulder	1	1	1
Stomach	5	0	3
Thigh	46	15	32
Throat	11	2	3
Trachea	28	8	19
Uterus	3	0	0
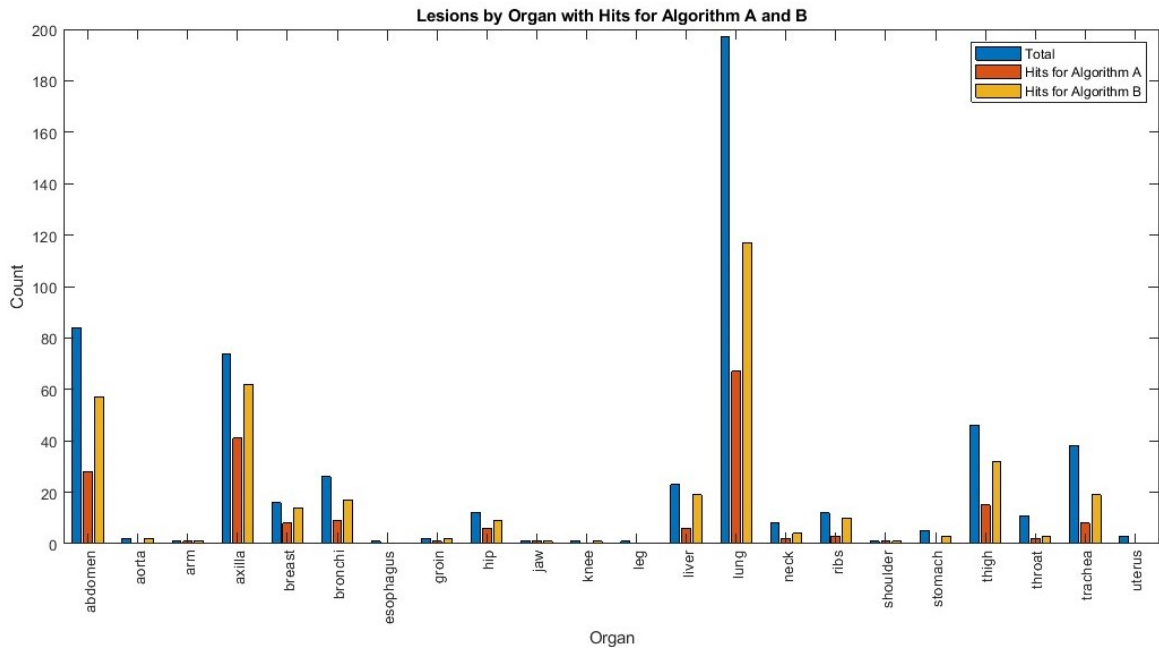
Figure 4.4: Distribution of lesion anatomical location and corresponding number of hits for Algorithms A and B. Number of misses are the differences between the total and hit numbers.

4.3.1.2   Limits of Detection Modelling

Figure 4.5 displays the relationship between the size and intensity of lesion in terms of hit and missed lesions. There appears to be a significant amount of overlap in the mid region, preventing a well fitted model from matching the data. Using other lesion features, such as contrast and CNR, allows the data to be better separated, as seen in Figure 4.6 and Figure 4.7, respectively. Regardless, complete separation between hit and miss regions is not achieved with either.
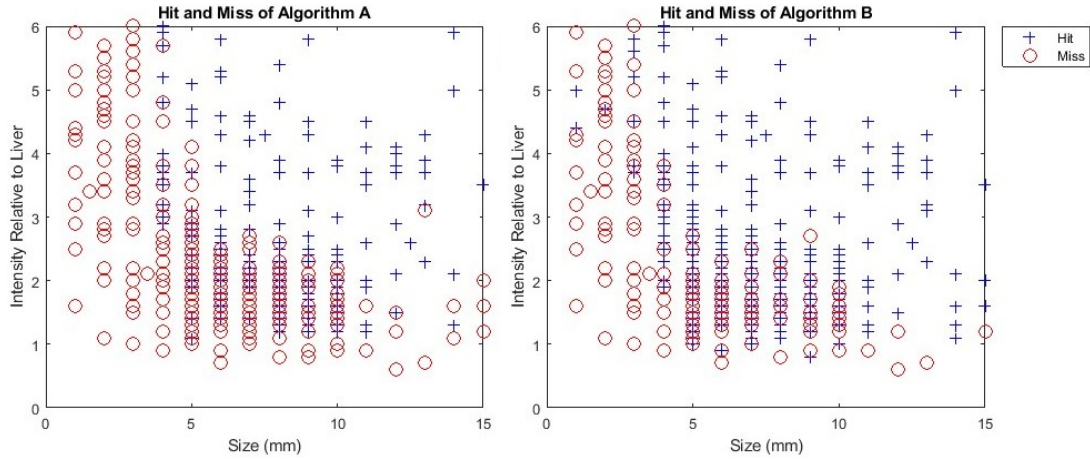
Figure 4.5: Detectability of lesions for Algorithms A and B with expected intensity relative to the liver.
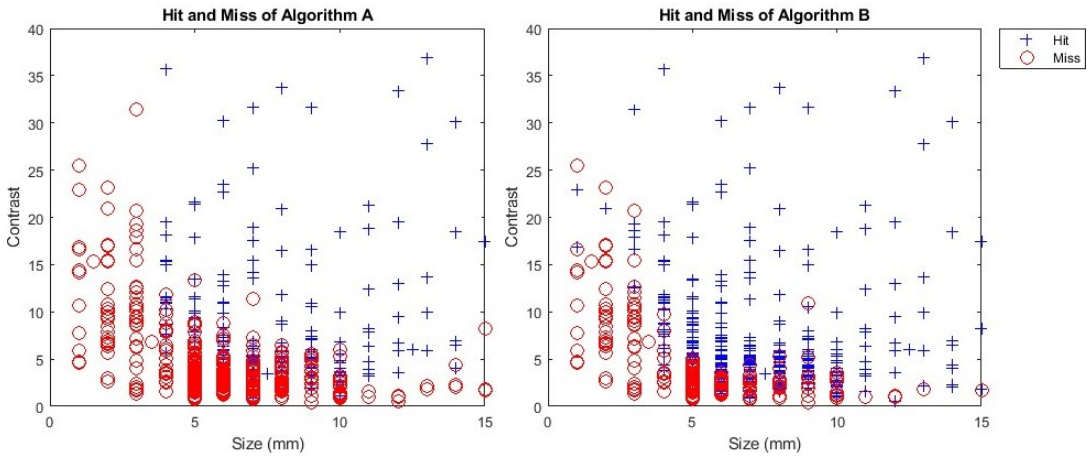
Figure 4.6: Detectability of lesions for Algorithms A and B with expected contrast.
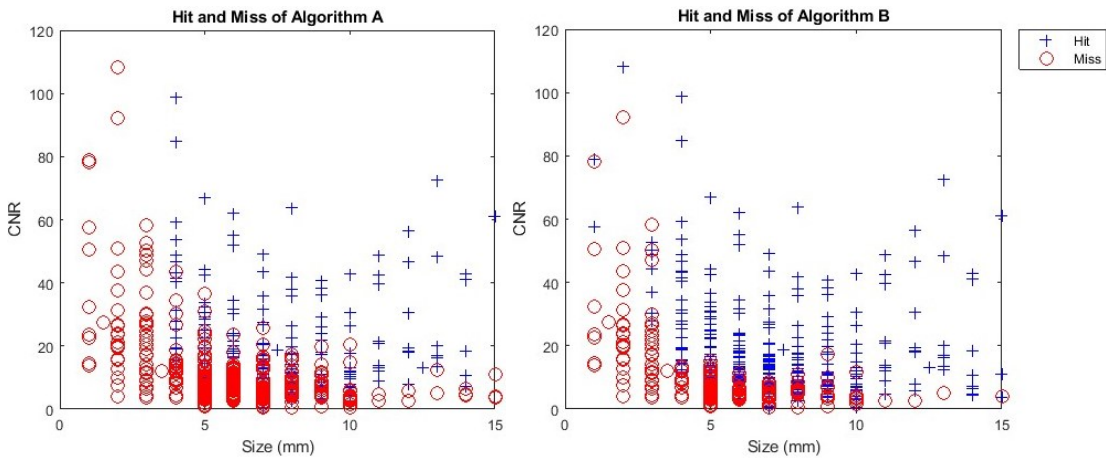
Figure 4.7: Detectability of lesions for Algorithms A and B with expected contrast-to-noise ratio (CNR)

Corresponding graphs of the probability of lesion detection for any given size and contrast for each AI algorithm are shown in Figure 4.8 and using CNR is shown in Figure 4.9. Model fit to the data was judged to be good. Visually, agreement of the model and the overlayed responses (dots) is apparent. Numerically, the model predicted responses (hit: probably > 50%; miss: probability ≤50%) agreed with the true responses in 80% of lesions for both Algorithms using contrast. When combining responses from both Algorithms response agreement dropped to 75% indicating a divergence in their responses. For CNR, predicted response agreement maintained the same, except for Algorithm B, which increased to 83%.

The parameters for each psychophysical response model can be seen in the top right of each figure. The larger blue region in the perception model of Algorithm B, for both contrast and CNR, corresponds to the higher hit (true positive) and lower miss (false negative) rates in Table 4.4. Furthermore, the smaller detectable lesion size and lower lesion contrast for Algorithm B are reflected by the higher model parameters D and C respectively. Finally, the transition from red to blue zones, associated with imperceptible and obvious lesion zones, respectively, is depicted by the white strip region (roughly 40% to 60% probability). The width of this region reflects a range of uncertainty, which is visually narrower in Algorithm B and corresponds to a larger A value.
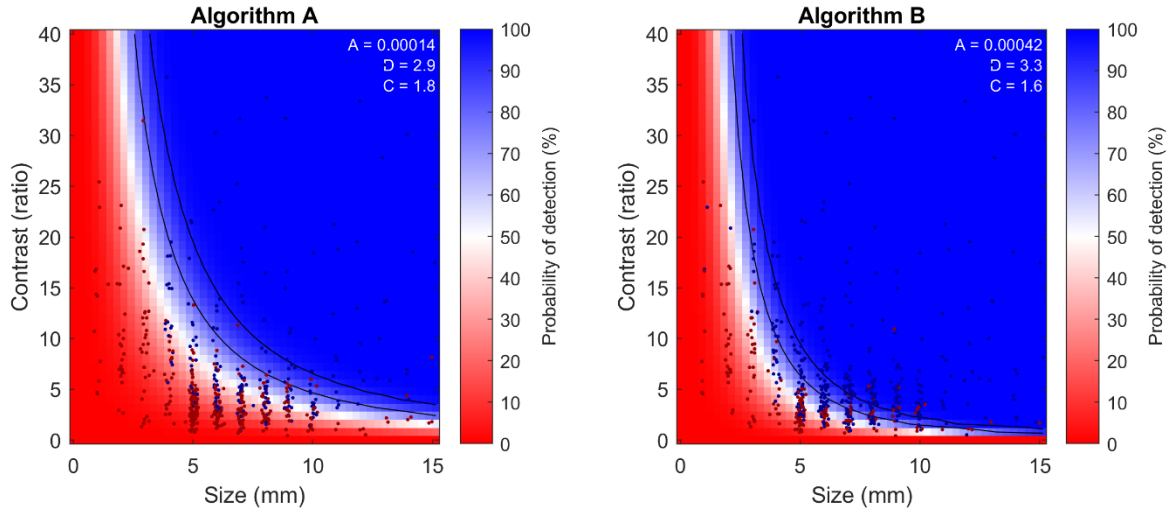
Figure 4.8: Perception model fit to hit and miss scores (dark red and dark blue dots, respectively) indicating the probability of lesion detection by simulated lesion contrast and size. Black lines indicating 80% and 95% probability of detection represent the limits of detection.
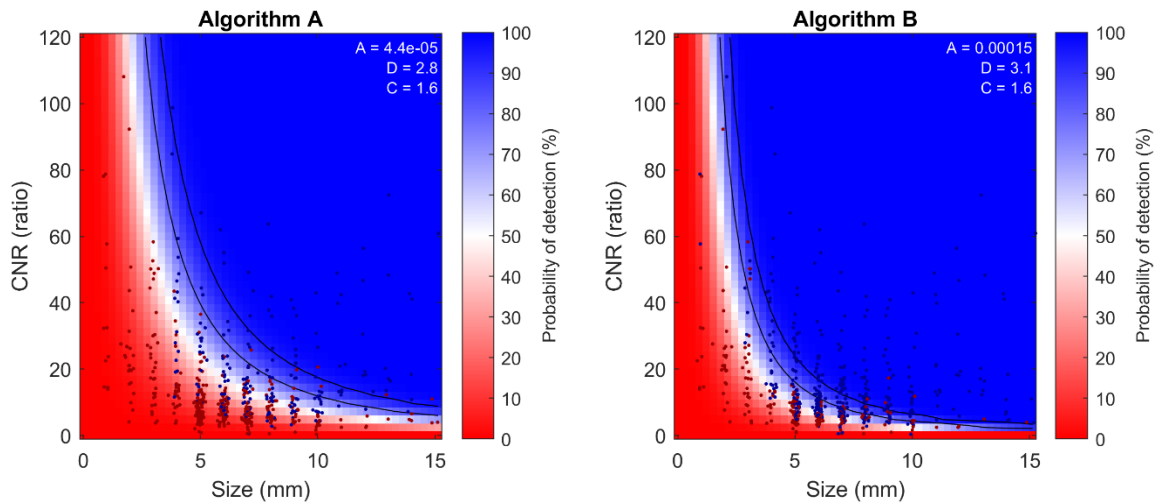
Figure 4.9: Perception model fit to hit and miss scores (dark red and dark blue dots, respectively) indicating the probability of lesion detection by simulated lesion contrast-to-noise ratio and size. Black lines indicating 80% and 95% probability of detection represent the limits of detection.

In bootstrap analysis of the fitted perception models, all parameters were significantly different (p<0.001) between the two algorithms both using contrast and CNR as a lesion parameter as shown in Table 4.6.

Table 4.6: Perception model parameter (median ± interquartile-range) using 1000 bootstrap samples and test for significant difference.

Parameter	Algorithm A	Algorithm B	p-value (Wilcoxon rank-sum)
Contrast			
A	$(1.3 \pm 1.8) \times 10^{-4}$	$(3.8 \pm 6.4) \times 10^{-4}$	2.3239×10^{-86}
D	2.90 ± 0.49	3.39 ± 0.82	1.5752×10^{-97}
C	1.83 ± 0.31	1.58 ± 0.29	2.006×10^{-107}
CNR			
A	$(3.96 \pm 6.3) \times 10^{-5}$	$(1.3 \pm 2.2) \times 10^{-4}$	4.8569×10^{-82}
D	2.84 ± 0.53	3.14 ± 0.62	4.8769×10^{-47}
C	1.64 ± 0.25	1.58 ± 0.33	3.1937×10^{-7}

4.3.2 Lesion Size

When comparing the size of the ground truth lesion to the size of the detected lesion, the detected lesion size is larger on average than the simulated lesion size, as seen in Figure 4.10, and as is reflected by the lines of best fit for both algorithms. The line of best fit was set to have an intercept at y = 0, to ensure the size calibration didn't result in negative size values.
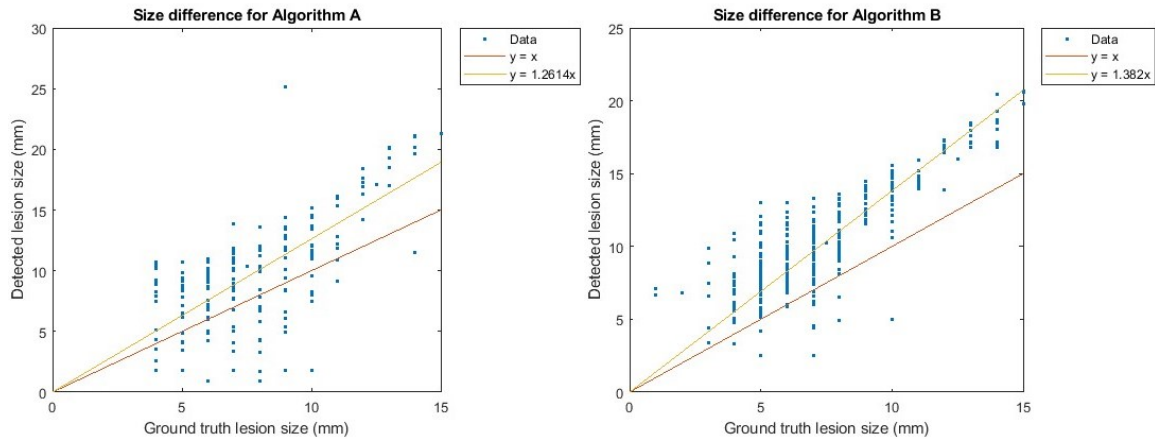
Figure 4.10: Variation in size for detected lesions compared to their ground truth size for Algorithms A and B.

4.3.3 Measured versus Expected Lesion Intensity Metrics

As mentioned previously, the values for intensity were taken from the LST and measured from the image and used to calculate both expected and measured contrast and CNR. The difference in the expected and measured values can be seen in Figure 4.11, where the size of the lesion in encoded by its colour. It can be seen that larger lesions appear closer to the line of identity, whereas smaller lesions appear further away and closer to the x-axis.
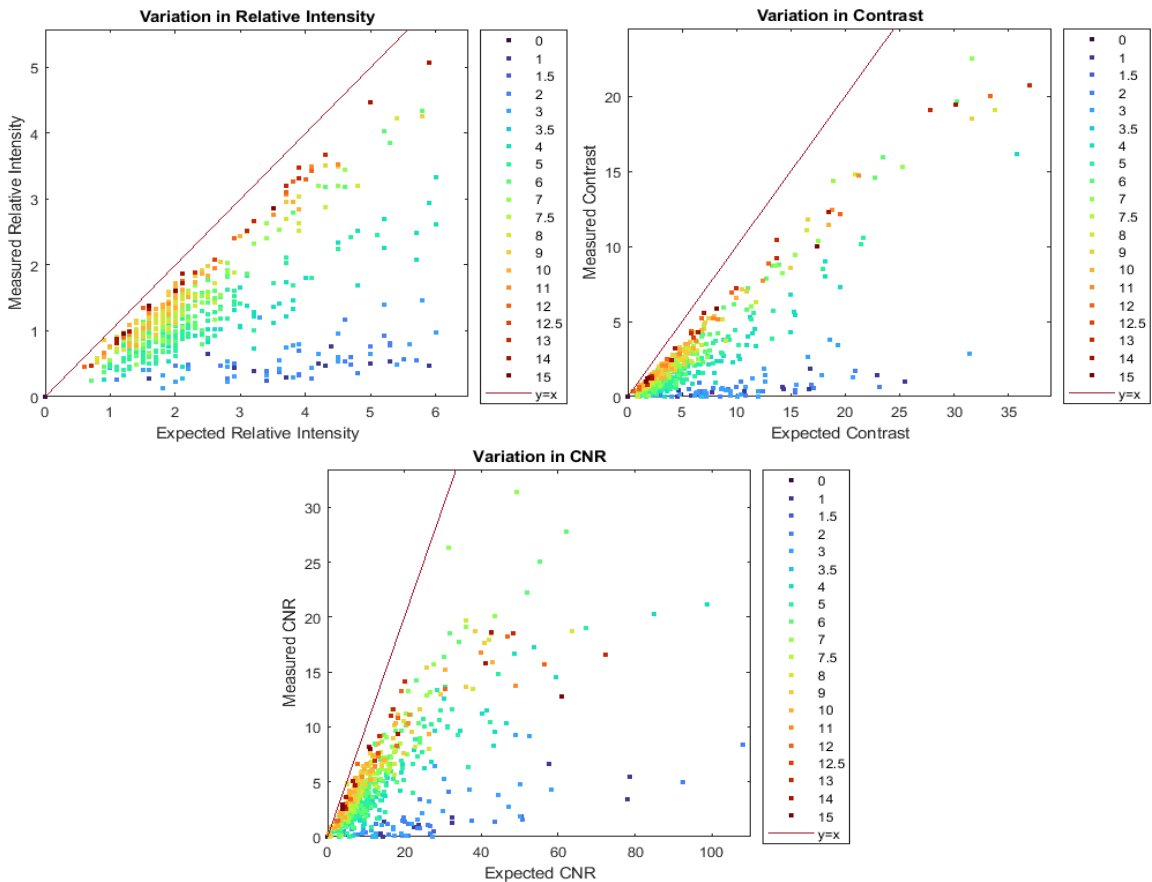
Figure 4.11: Variation between expected and measured metrics of lesion intensity including: intensity relative to the liver, contrast with background (ratio), and contrast-to-noise ratio (CNR), for ground truth lesions.

4.3.4   False positives

For false positive results, no ground truth data exists. Regardless, these were plotted along with the hit and miss data with the measured intensity, contrast, and CNR

values, as seen in Figure 4.12, with lesion size calibrated from the true positives data

shown in Figure 4.10. The false positives appear to be mostly associated with low

contrast and low CNR compared to the hit and miss lesions. Furthermore, even after size

calibration, Algorithm A has a very large range of sizes, extending up to 73 mm.

Algorithm B tends to have most false positive lesions between 0 and 10 mm, with a

maximum size of 18 mm.
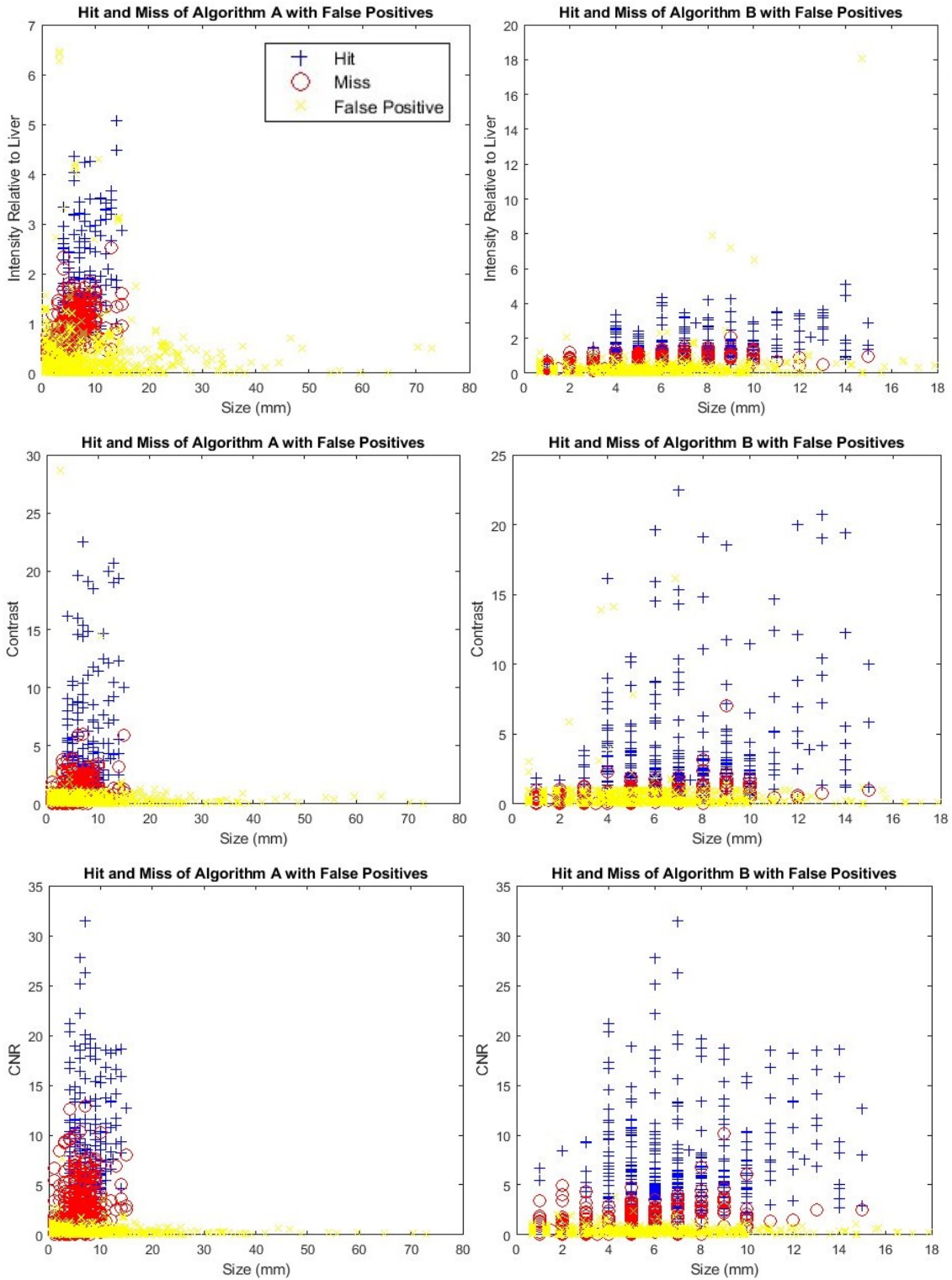
Figure 4.12: Hit (blue) and miss (red) graphs against image measured metrics, including false positives (yellow), where the sizes of the false positives are calibrated using the line of best fit from Figure 4.10.

Chapter 5: Discussion and Conclusion

Using real patient data and the LST, a relatively large image library was constructed with well characterized synthetic lesions. The lesion characteristics, including location, size, and intensity, served as a ground truth reference for objective evaluation of lesion detection by observers. The image data were distributed to two independent laboratories developing AI for lesion detection from which their lesion prediction results were objectively analyzed against the reference truth. Novel analysis was performed that includes not only the overall lesion detection rates, but breaks these down by the lesion characteristics, including size and contrast, or size and CNR. Lastly, a parameterised psychophysical based model was created to briefly summarize the lesion detection performance of an observer.

5.1 Lesion Synthesis Toolbox

The Lesion Synthesis Toolbox was an essential tool in this research, as it allowed for the creation of synthetic lesions in real PET/CT data for the lesion library. The incremental developments to the toolbox were vital to achieve a productive lesion insertion software for creating a large library of synthetic lesions. The most significant developments include integrating the toolbox with clinical software to visualize completed simulations to allow for verification using clinical software and the industry standard DICOM format. Furthermore, the download of ground truth lesion values and image characteristics enabled the analysis of detected lesions.

With the software development completed, the bottleneck for creating synthetic images was the simulation and reconstruction time. On a desktop computer with i7-4790 CPU at 3.6 GHz reconstruction of a single PET bed position was about 2 hours. With 4

physical cores (8 logical cores) and 16 GB of RAM up to 3 bed positions could be processed simultaneously. Thus, for a whole-body PET study with 9 bed positions, a typical simulation required approximately 6 hours. A more modern desktop with more memory and more cores can likely perform the same reconstruction within 2 hours. The GE DUETTO toolbox does not yet support GPU acceleration but is expected to be available soon and could further improve reconstruction times. Likewise, even a modest computer cluster could be leveraged to highly increase simulation throughput by parallelizing the operation across multiple simulations.

The time to define lesions in an image using the LST only amounts to several minutes per simulation, and thus dozens of simulations can be configured and queued in a workday by a single user. Future versions may benefit of functionality to define simulations with multiple variations of the same lesion (e.g., varying intensity and/or sizes). Presently, a trained user is still required to define anatomically realistic lesion locations.

5.2   Characterization of AI Performance

This study characterized two lesion-detection AI algorithms. Both AI detected significantly more lesions than had been inserted, with Algorithm A detecting more than Algorithm B. Furthermore, Algorithm B was shown to be more sensitive and precise than Algorithm A. Lesion detection did vary by anatomical location but not significantly, with Algorithm B consistently finding a higher number of lesions per location. Thus, overall, it appeared that Algorithm B outperformed Algorithm A. Although from the summary statistics it appears as if these AI did not perform well, this is not necessarily the case. The lesion library contained a large number of lesions below the typical resolution of

PET images, resulting in these lesions being missed. The lesion characteristics, including size and intensity, heavily impact the ability of AI to detect such lesions.

To better understand what types of lesions each algorithm was able to detect, deeper analysis was performed by lesion size and intensity metrics (Figure 4.5, Figure 4.6, and Figure 4.7). These confirmed the expectation that larger and brighter lesions could be more reliably detected than faint and/or smaller lesions. Lesion parameters were compared to determine the best intensity metric to distinguish between detected and missed lesions. Lesion to background contrast and CNR were better suited metrics than lesion intensity, as demonstrated by better separation of the hits and misses in these figures. This is anticipated, since the detectability of a lesion is dependent on how well it can be distinguished from surrounding tissue, which is essentially the definition of contrast.

Contrast and CNR performed similarly and produced similar patterns on the score charts (Figure 4.6 and Figure 4.7), however, CNR had less data occupying the upper right region of the graph. This indicates that larger lesions have lower CNR, even while having high contrast values. The likely explanation for this trend is due to the fixed margin around the lesions when measuring intensity values from the image. The margin causes larger lesions to have a larger sample size than smaller lesions, due to the cubic term from volume measurements. The larger sample size could result in more variation and more noise, causing the CNR to be lower for larger lesions. This is especially true for large lesions that are defined on the boundary or an organ (e.g., mediastinum), thus part of the lesion may have moderate background from the organ and the other part of the lesion background may have low background signal from the surrounding tissue. This

scenario increases the variability in the background region, and hence the noise level in the CNR denominator. To resolve this inconsistency, a fixed region of interest should be used around the lesion, instead of a fixed margin.

A potential limitation to the detection analysis was the method of characterizing detected and missed lesions. Using the proximity (3 cm) of centers of mass as a criterion for lesion detection May have mis-scored large, overlaying lesions as a miss if they were off-centred. Alternatively, using the area of overlap between lesions may have been a more accurate measure of detection. However, this approach is also not without caveats. For example, Algorithm A segmented very large regions, up to 73 mm across, and if there was a small lesion placed inside this large region, such as a 5mm lesion, it may be labelled as detected when it should not be.

5.2.1   Perception Model

While visual comparison of the score charts for each AI algorithm (Figure 4.6 and Figure 4.7) clearly showed that Algorithm B could reliably detect smaller and fainter lesions than Algorithm A, a quantitative comparison of performance was sought out. Hence, a lesion detection prediction model was fit from the domain of psychophysics. The model was simplified to consist of three parameters that encoded the sensitivity to lesion size, D, sensitivity to lesion contrast, C, and a third parameter, A, that both normalized the perception signal intensity (determined by D and C) and encoded the rate of transition from undetectable to robust lesion detection. Quality of fit metrics (percent agreement between model predicted and actual lesion score) indicated that the model was suitable to characterize the AI performance in these data. Although it is worth noting that other commonly used quality of fit metrics (e.g., coefficient of determinations,

correlation, sum-squared-error) were not investigated due to the difference between the model response (continuous probability of detection in the range of 0 to 1) and the measures (dichotomous hit score as 0 or 1).

Through bootstrap analysis (Table 4.6) it was demonstrated that the variance of these parameters (represented by the interquartile ranges) were sufficiently small that the median model parameters were significantly different from one another. Therefore, these parameters may be useful to easily benchmark alternative observers on the quality of their limits of detection. Furthermore, these parameters may be used to compare between each intensity metric to differentiate the best metric for characterizing lesion detectability, however, further understanding of the model parameters is needed before this can be done.

Regardless, the use of the psychophysical model and its parameters is preliminary, and more work is required to fully understand how this model may be best applied. It is not completely clear how the model parameters should be interpreted or whether better summary statistics may be derived from these parameters.

5.3 False Positives

While an observers' sensitivity to detect mild lesions is desired, its important to be cautious of a model that also produces many false-positive lesions. In medicine, it is a common concern that with excessive testing, patients are being increasingly exposed to unnecessary medical workup with the associated risks and costs they entail. Therefore, characterizing where false lesions may be reported by an observer was studied. The results demonstrated (Figure 4.12) that most false-positive lesions have similar sizes to

real lesions and contrast similar to missed lesions. Hence many of these lesions may be best to not report (either by automated exclusion or by physicians reviewing the study).

In few cases, reported lesions had high contrast and/or size consistent with true-positive lesions. While these false-positive lesions were not studied exhaustively, they may represent several scenarios. One possibility is the detection of anatomical structures with high intensity that are not lesions (e.g., kidney, urinary bladder, brain, or heart). Another possibility is the detection of actual lesions that were not synthesized, but were present in the original patient data, despite best efforts to exclude such studies.

5.4 Expected Versus Measured Lesion Characteristics

For both AI algorithms, the detected size of lesions was larger on average than the simulated size (Figure 4.10). The variation in size between simulated and detected lesions is likely due to blurring during image reconstruction. In addition, the AI detection could include a margin around lesions, which would explain the different slopes on the figures.

Contrarily, the measured values for intensity, contrast, and CNR were smaller than expected values for all points. The distribution of sizes from these graphs (Figure 4.11) illustrated a clear pattern, with larger lesions reproducing their defined characteristics better than smaller lesions. During lesion simulation and image reconstruction, lesions undergo blurring, which partially limits the spatial resolution of PET (~6 mm in our case). For smaller lesions this blurring is more significant, resulting in their intensity being reduced significantly. For example, a lesion of size 1 mm and relative intensity of 5.9 times the intensity of the liver underwent blurring and had a resulting value of 0.5 times the intensity of the liver, which is a reduction of about 92%. A lesion of size 9 mm, however, decreased only from 5.8 to 4.3 times the intensity of the

liver which only is a reduction of 26%. These results are consistent with earlier literature on Q.Clear reconstruction. [39], [41]

5.4.1 Characterization of Human Observer Performance

Much like the analysis performed for AI, the same methods could be applied to characterize the performance of human observers. In fact, the perception models in this work were similar in shape and magnitude of model parameters to those previously published for a small subset of human observers. [41] However, several considerations should be accounted for. Unlike AI, humans may not be completely reproducible due to distractions, mistakes, ambiguity of the task, satisfaction of search or willingness of participants to guess when they struggle to find lesions. In psychophysical models guess and lapse rates are typically included in the psychometric model to account for this. In this work, the guess and lapse rates were set to zero for AI. For human observers, these parameters are either fit to the response data or are estimated. In previous work in humans for image-based detection guess rates are typically set low (e.g., <1%) and the lapse rate is set at ~5%. Another consideration when working with human observers is that they become tired of the task and can only produce a limited amount of results data and/or quality data. Furthermore, expert observers (e.g., physicians) may be difficult to obtain and are costly to hire. With AI it is possible to generate many more data points than with human observers, especially when large image libraries with ground truth data exist.

5.5 Other Application of Synthetic Lesions

The potential applications of synthetic lesions go beyond those explored in this work. Other applications can be envisioned including the following:

- <u>Evaluation of segmentation quality</u> (whether by humans or AI). Image segmentation metrics typically include the dice coefficient for quality of overlap of the segmentation with the ground truth lesion. Other commonly reported measures are of the distance of the segmentation boundaries from those of the ground truth. These are often reported as the mean or as the maximum deviation (Hausdorff distance).

- <u>Evaluating the impact of technologies</u> on lesion detection and segmentation performance. As implied in Figure 3.1, after lesion synthesis downstream steps include image reconstruction and display. Likewise, and image processing could be included. Synthetic lesions may be used to evaluate the impact of alternative technologies on the clinical task. This was clearly demonstrated in prior work evaluating PET reconstruction algorithms.[47], [48]

- <u>Evaluating human reader performance.</u> The last step in the imaging PET imaging cascade is interpretation by the clinician. Synthetic lesions could be applied to test human reader performance with training and experience, however given the tediousness of this task and the need to generate clinically relevant lesions it is doubtful that this will ever be applied broadly. More practically, synthetic lesions may be useful in experiments seeking to guide optimal image viewing conditions (e.g., lit/dark room, desktop versus mobile device, colour rendering) and for evaluating search strategies such as the use of checklists, standardized reporting, computer aided reporting and quality improvement initiatives.[48]

- <u>Training of AI</u> is another clear application for synthetic lesions. In the same way that it is difficult to label data for AI and human performance validation, it is

more so applicable for generating ground truth data to train AI, where even more data are required. Synthetic lesions can be applied to better control the training process to teach the AI to detect subtle lesions and learn to recognize lesions in uncommon anatomical sites that may be under-represented in clinically derived datasets. A possible pitfall is that the synthetic lesions may lack necessary realism to represents clinical patterns. Validation with clinical lesions is therefore paramount.

5.6   Application to Other Pathologies

This work was performed in the context of oncology PET where lesions are manifested as regions with elevated tracer activity. Cancer imaging represents that vast majority of PET use internationally. Nevertheless, other PET applications have been growing.

In cardiac PET, positive pathology findings are typically associated with decreased tracer uptake (e.g., reduced myocardial blood flow).[49] Thus, in that context synthetic lesions would have to incur negative activity (for a count reduction), which we have not attempted with LST, but theoretically should be feasible.

In neurology PET, positive PET findings may be associated with either increased or decreased regional activity depending on the disease in question and the physiologic state (e.g., ictal versus interictal epilepsy). While synthetic lesions may be useful in these applications, it would require lesions consistent with anatomical and physiologic structures which exceed the simple shapes (i.e., spheres) demonstrated in this work. Defining such shapes would require significant enhancements to the lesion drawing tools in the LST.

PET for detection of infection of unknown origin in most aspects is very similar to oncology PET, and in this domain the LST in its current form may be suitable.

5.7   Future Work

There are many possible directions in which this work can be used in future research. Firstly, this work can be expanded to include more AI algorithms. To date, other labs have been in communication, but have yet to submit their results. Increasing the number of AI algorithms would allow for a deeper understanding of characterization of AI detection, and would be less of a simple comparison between two AI.

In addition, the lesion library can be used to train AI for lesion detection tasks, as the ground truth values are known from the toolbox. However, the realism of artificial lesions may heavily impact the AI's ability to detect real lesions. Further testing would be required to implement artificial lesions for the purpose of training AI for lesion detection.

Moreover, a larger library of lesions would be beneficial to increasing the accuracy of the results. Specifically, more lesions in different anatomical regions would allow to study the likelihood of AI detecting lesions in a variety of locations. For example, we did not evaluate lesion detection in the context of prostate cancer. Since the prostate is in proximity of the urinary bladder, which accumulates radioactive urine, so lesion detection performance may be very different than the other anatomical sites. Likewise for tumours in the proximity of the brain. Lesions could also be placed along boundaries of tissues, to study the effects of edges on detection of lesions.

Furthermore, a library could be created using identical lesions, but with different reconstruction parameters. This would be made easy by using the 'retrieve lesions' button to bring back the original image and defined lesions, hence allowing for a comparison

between the detectability for different image reconstructions. In addition, this would allow to compare contrast and CNR measurements from different reconstruction algorithms or parameters.

Clearly it would be beneficial to extend the LST to be used with other vendors, as currently it is only functional with GE Healthcare scanners. The LST software was designed with that in mind, but because file types and reconstruction methods are not standardized in the industry, collaboration with expert researchers and industry partners is pertinent. Furthermore, extending to other pathologies, such as MRI or ultrasound, would also be beneficial.

An up-and-coming issue with medical imaging and advances in AI technology is the use of medical images to identify faces. [50] This would break anonymity and causes privacy and safety concerns for patients. Fortunately, de-identification techniques have been developed to anonymize patient faces by modifying facial features without the loss of important medical information. [51] Integrating this technology with the LST would ensure open-source libraries maintain patient privacy. Currently, this applies more to the CT portion of the PET/CT study which contains exquisite anatomical data. But even for PET with its lesser anatomical data, generative AI has been demonstrated to be moderately capable of reconstructing patient faces. [52] Another malicious use of this technology includes falsifying patient data. In the wrong hands, this technology could be used to insert false lesions into healthy patients and cause unnecessary medical intervention.

Lesion detection using the CT component of a PET/CT study is surely to arrive soon. Currently, lesion detection AI uses the PET component of the study. While the LST

can insert lesions into reconstructed CT data, these have yet to be used for any research application. Regardless, it is likely that lesion insertion in both PET and CT will be required to characterize the performance of AI for lesion detection.

5.8   Conclusion

In conclusion, synthetic lesions are a useful tool to characterize the limits of detection in AI. Visual assessment and psychophysical models were effective at characterizing each AI's detectability performance by size versus contrast and CNR, and proved each method to be statistically different using bootstrapping methods. The Lesion Synthesis Toolbox is a feature rich tool that may aid researchers in training and characterizing AI, and thus may contribute to the delivery of precision healthcare.

References

[1] "Canadian Cancer Statistics Advisory Committee in collaboration with the Canadian Cancer Society." Statistics Canada and the Public Health Agency of Canada, 2023.

[2] "Book of Abstracts." Canadian Association of Nuclear Medicine, Oct. 2023. [Online]. Available: https://canm-acmn.ca/resources/Documents/Book%20of%20Abstracts_2023.pdf

[3] S. Anand, H. Singh, and A. Dash, "Clinical Applications of PET and PET-CT," *Med. J. Armed Forces India*, vol. 65, no. 4, pp. 353–358, Oct. 2009, doi: 10.1016/S0377-1237(09)80099-3.

[4] J. L. Prince and J. M. Links, *Medical imaging signals and systems*, 2nd ed. Boston: Pearson, 2015.

[5] D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, Eds., *Positron emission tomography: basic sciences*. New York: Springer, 2005.

[6] P. E. Kinahan and J. W. Fletcher, "Positron Emission Tomography-Computed Tomography Standardized Uptake Values in Clinical Practice and Assessing Response to Therapy," *Semin. Ultrasound CT MRI*, vol. 31, no. 6, pp. 496–505, Dec. 2010, doi: 10.1053/j.sult.2010.10.001.

[7] X. Zhang *et al.*, "Total-Body Dynamic Reconstruction and Parametric Imaging on the uEXPLORER," *J. Nucl. Med.*, vol. 61, no. 2, pp. 285–291, Feb. 2020, doi: 10.2967/jnumed.119.230565.

[8] F. E. Enríquez-Mier-y-Terán, A. S. Ortega-Galindo, T. Murrieta-Rodríguez, M. Rodríguez-Villafuerte, A. Martínez-Dávalos, and H. Alva-Sánchez, "Coincidence energy spectra due to the intrinsic radioactivity of LYSO scintillation crystals," *EJNMMI Phys.*, vol. 7, no. 1, p. 21, Dec. 2020, doi: 10.1186/s40658-020-00291-1.

[9] H. J. Yoon, Y. J. Jeong, H. J. Son, D.-Y. Kang, K.-Y. Hyun, and M.-K. Lee, "Optimization of the spatial resolution for the GE discovery PET/CT 710 by using NEMA NU 2-2007 standards," *J. Korean Phys. Soc.*, vol. 66, no. 2, pp. 287–294, Jan. 2015, doi: 10.3938/jkps.66.287.

[10] "Molecular Imaging Equipment." University of Wisconsin School of Medicine and Public Health. [Online]. Available: https://radiology.wisc.edu/research/modalities/molecular-imaging/equipment/

[11] K. Wagatsuma *et al.*, "Comparison between new-generation SiPM-based and conventional PMT-based TOF-PET/CT," *Phys. Med.*, vol. 42, pp. 203–210, Oct. 2017, doi: 10.1016/j.ejmp.2017.09.124.

[12] E. Alie, "FDG and NaF PET/CT Imaging, NM SOP 015." The Ottawa Hospital, 2023.

[13] N. J. Pelc, "Recent and Future Directions in CT Imaging," *Ann. Biomed. Eng.*, vol. 42, no. 2, pp. 260–268, Feb. 2014, doi: 10.1007/s10439-014-0974-z.

[14] R. Boellaard *et al.*, "FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 42, no. 2, pp. 328–354, Feb. 2015, doi: 10.1007/s00259-014-2961-x.

[15] "A Guide to CT Radiation Dose Management." GE Healthcare, 2012.

[16] S. Alameen, N. Tamam, S. Awadain, A. Sulieman, L. Alkhaldi, and A. B. Hmed, "Radiobiological risks in terms of effective dose and organ dose from 18F-FDG

whole-body PET/CT procedures," *Saudi J. Biol. Sci.*, vol. 28, no. 10, pp. 5947–5951, Oct. 2021, doi: 10.1016/j.sjbs.2021.06.055.

[17]   S. Tong, A. M. Alessio, and P. E. Kinahan, "Image reconstruction for PET/CT scanners: past achievements and future challenges," *Imaging Med.*, vol. 2, no. 5, pp. 529–545, Oct. 2010, doi: 10.2217/iim.10.49.

[18]   I. S. Armstrong, C. Hayden, M. J. Memmott, and P. Arumugam, "A preliminary evaluation of a high temporal resolution data-driven motion correction algorithm for rubidium-82 on a SiPM PET-CT system," *J. Nucl. Cardiol.*, vol. 29, no. 1, pp. 56–68, Feb. 2022, doi: 10.1007/s12350-020-02177-2.

[19]   S. Manwell, R. Klein, T. Xu, and R. A. deKemp, "Clinical comparison of the positron emission tracking (PeTrack) algorithm with the real-time position management system for respiratory gating in cardiac positron emission tomography," *Med. Phys.*, vol. 47, no. 4, pp. 1713–1726, Apr. 2020, doi: 10.1002/mp.14052.

[20]   J. Schaefferkoetter, M. Casey, D. Townsend, and G. El Fakhri, "Clinical impact of time-of-flight and point response modeling in PET reconstructions: a lesion detection study," *Phys. Med. Biol.*, vol. 58, no. 5, pp. 1465–1478, Mar. 2013, doi: 10.1088/0031-9155/58/5/1465.

[21]   M. Bjöersdorff *et al.*, "Impact of penalizing factor in a block-sequential regularized expectation maximization reconstruction algorithm for 18F-fluorocholine PET-CT regarding image quality and interpretation," *EJNMMI Phys.*, vol. 6, no. 5, Dec. 2019, doi: 10.1186/s40658-019-0242-2.

[22]   "Discovery(TM) MI Digital Ready, Discovery 710 and Discovery 710 Clarity User Manual." General Electric Company, 2021.

[23]   E. Texte *et al.*, "Impact of the Bayesian penalized likelihood algorithm (Q.Clear®) in comparison with the OSEM reconstruction on low contrast PET hypoxic images," *EJNMMI Phys.*, vol. 7, no. 28, Dec. 2020, doi: 10.1186/s40658-020-00300-3.

[24]   D. Tian, H. Yang, Y. Li, B. Cui, and J. Lu, "The effect of Q.Clear reconstruction on quantification and spatial resolution of 18F-FDG PET in simultaneous PET/MR," *EJNMMI Phys.*, vol. 9, no. 1, p. 1, Dec. 2022, doi: 10.1186/s40658-021-00428-w.

[25]   E. Lindström, L. Lindsjö, A. Sundin, J. Sörensen, and M. Lubberink, "Evaluation of block-sequential regularized expectation maximization reconstruction of 68Ga-DOTATOC, 18F-fluoride, and 11C-acetate whole-body examinations acquired on a digital time-of-flight PET/CT scanner," *EJNMMI Phys.*, vol. 7, no. 1, p. 40, Dec. 2020, doi: 10.1186/s40658-020-00310-1.

[26]   W. Sureshbabu and O. Mawlawi, "PET/CT imaging artifacts," *J. Nucl. Med. Technol.*, vol. 33, no. 3, pp. 156–161; quiz 163–164, Sep. 2005.

[27]   T. M. Blodgett, A. S. Mehta, A. S. Mehta, C. M. Laymon, J. Carney, and D. W. Townsend, "PET/CT artifacts," *Clin. Imaging*, vol. 35, no. 1, pp. 49–63, Jan. 2011, doi: 10.1016/j.clinimag.2010.03.001.

[28]   B. L. Triche *et al.*, "Recognizing and Minimizing Artifacts at CT, MRI, US, and Molecular Imaging," *RadioGraphics*, vol. 39, no. 4, pp. 1017–1018, Jul. 2019, doi: 10.1148/rg.2019180022.

[29]   C. K. Hoh, "Clinical use of FDG PET," *Nucl. Med. Biol.*, vol. 34, no. 7, pp. 737–742, Oct. 2007, doi: 10.1016/j.nucmedbio.2007.07.001.

[30]   Y. Ichiya *et al.*, "FDG-PET in infectious lesions: The detection and assessment of lesion activity," *Ann. Nucl. Med.*, vol. 10, no. 2, pp. 185–191, Jun. 1996, doi: 10.1007/BF03165391.

[31]   K. A. O'Hanlan, "Resection of a 303.2-Pound Ovarian Tumor," *Gynecol. Oncol.*, vol. 54, no. 3, pp. 365–371, Sep. 1994, doi: 10.1006/gyno.1994.1225.

[32]   J. W. Wallis *et al.*, "Guidelines on Setting Up Stations for Remote Viewing of Nuclear Medicine and Molecular Imaging Studies During COVID-19," *J. Nucl. Med. Technol.*, vol. 49, no. 1, pp. 2–6, Mar. 2021, doi: 10.2967/jnmt.120.261890.

[33]   A. J. Degnan *et al.*, "Perceptual and Interpretive Error in Diagnostic Radiology—Causes and Potential Solutions," *Acad. Radiol.*, vol. 26, no. 6, pp. 833–845, Jun. 2019, doi: 10.1016/j.acra.2018.11.006.

[34]   N. Brucher, R. Mandegaran, T. Filleron, and T. Wagner, "Measurement of inter- and intra-observer variability in the routine clinical interpretation of brain 18-FDG PET-CT," *Ann. Nucl. Med.*, vol. 29, no. 3, pp. 233–239, Apr. 2015, doi: 10.1007/s12149-014-0932-8.

[35]   M. J. Campbell, *Statistics at square one*, 12th ed. Hoboken, N.J: Wiley, 2021.

[36]   F. A. Wichmann and N. J. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept. Psychophys.*, vol. 63, no. 8, pp. 1293–1313, Nov. 2001, doi: 10.3758/BF03194544.

[37]   F. A. Wichmann and N. J. Hill, "The psychometric function: II. Bootstrap-based confidence intervals and sampling," *Percept. Psychophys.*, vol. 63, no. 8, pp. 1314–1329, Nov. 2001, doi: 10.3758/BF03194545.

[38]   N. Prins, "The psychometric function: The lapse rate revisited," *J. Vis.*, vol. 12, no. 6, pp. 25–25, Jun. 2012, doi: 10.1167/12.6.25.

[39]   K. Wangerin, S. Ahn, S. G. Ross, P. E. Kinahan, and R. M. Manjeshwar, "Improving lesion detectability in PET imaging with a penalized likelihood reconstruction algorithm," presented at the SPIE Medical Imaging, C. R. Mello-Thoms and M. A. Kupinski, Eds., Orlando, Florida, United States, Mar. 2015, p. 94160W. doi: 10.1117/12.2082301.

[40]   W. D. Bidgood, S. C. Horii, F. W. Prior, and D. E. Van Syckle, "Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging," *J. Am. Med. Inform. Assoc.*, vol. 4, no. 3, pp. 199–212, May 1997, doi: 10.1136/jamia.1997.0040199.

[41]   H. Gabrani-Juma, Z. Al Bimani, L. S. Zuckier, and R. Klein, "Development and validation of the Lesion Synthesis Toolbox and the Perception Study Tool for quantifying observer limits of detection of lesions in positron emission tomography," *J. Med. Imaging*, vol. 7, no. 02, p. 1, Apr. 2020, doi: 10.1117/1.JMI.7.2.022412.

[42]   P. Blanc-Durand, A. Van Der Gucht, N. Schaefer, E. Itti, and J. O. Prior, "Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study," *PLOS ONE*, vol. 13, no. 4, p. e0195798, Apr. 2018, doi: 10.1371/journal.pone.0195798.

[43]   S. Ahamed and A. Rahmim, "Generalized Dice Focal Loss trained 3D Residual UNet for Automated Lesion Segmentation in Whole-Body FDG PET/CT Images," 2023, doi: 10.48550/ARXIV.2309.13553.

[44]   D. Dieckens, J. Lavalaye, L. Romijn, and J. Habraken, "Contrast-noise-ratio (CNR) analysis and optimisation of breast-specific gamma imaging (BSGI)

acquisition protocols," *EJNMMI Res.*, vol. 3, no. 1, p. 21, Mar. 2013, doi: 10.1186/2191-219X-3-21.

[45]   Y. Zhou *et al.*, "On the relationship of minimum detectable contrast to dose and lesion size in abdominal CT," *Phys. Med. Biol.*, vol. 60, no. 19, pp. 7671–7694, Oct. 2015, doi: 10.1088/0031-9155/60/19/7671.

[46]   Y. Zhou, J. Nute, A. Scott, and C. Lee, "Consistent low-contrast detectability for variable patient sizes and corresponding dose in abdominal CT," *Med. Phys.*, vol. 44, no. 3, pp. 861–872, Mar. 2017, doi: 10.1002/mp.12085.

[47]   D. J. Kadrmas, M. E. Casey, N. F. Black, J. J. Hamill, V. Y. Panin, and M. Conti, "Experimental Comparison of Lesion Detectability for Four Fully-3D PET Reconstruction Schemes," *IEEE Trans. Med. Imaging*, vol. 28, no. 4, pp. 523–534, Apr. 2009, doi: 10.1109/TMI.2008.2006520.

[48]   K. A. Wangerin, S. Ahn, S. Wollenweber, S. G. Ross, P. E. Kinahan, and R. M. Manjeshwar, "Evaluation of lesion detectability in positron emission tomography when using a convergent penalized likelihood image reconstruction method," *J. Med. Imaging*, vol. 4, no. 1, p. 011002, Nov. 2016, doi: 10.1117/1.JMI.4.1.011002.

[49]   P. L. Tilkemeier, J. Bourque, R. Doukky, R. Sanghani, and R. L. Weinberg, "ASNC imaging guidelines for nuclear cardiology procedures," *J. Nucl. Cardiol.*, vol. 24, no. 6, pp. 2064–2128, Dec. 2017, doi: 10.1007/s12350-017-1057-y.

[50]   C. L. Parks and K. L. Monson, "Automated Facial Recognition of Computed Tomography-Derived Facial Images: Patient Privacy Implications," *J. Digit. Imaging*, vol. 30, no. 2, pp. 204–214, Apr. 2017, doi: 10.1007/s10278-016-9932-7.

[51]   T. Uchida *et al.*, "De-Identification Technique with Facial Deformation in Head CT Images," *Neuroinformatics*, vol. 21, no. 3, pp. 575–587, Jul. 2023, doi: 10.1007/s12021-023-09631-9.

[52]   C. G. Schwarz *et al.*, "Face recognition from research brain PET: An unexpected PET problem," *NeuroImage*, vol. 258, p. 119357, Sep. 2022, doi: 10.1016/j.neuroimage.2022.119357.

[53]   H. Gabrani-Juma, Z. Al Bimani, L. S. Zuckier, and R. Klein, "Development and validation of the Lesion Synthesis Toolbox and the Perception Study Tool for quantifying observer limits of detection of lesions in positron emission tomography," *J. Med. Imaging*, vol. 7, no. 02, p. 1, Apr. 2020, doi: 10.1117/1.JMI.7.2.022412.

[54]   H. Juma, "Lesion Synthesis Toolbox: Development and Validation of Dedicated Software for Synthesis of Lesions in Raw PET and CT Patient Images," Carleton University, Ottawa, Ontario, Canada``, 2019. [Online]. Available: https://doi.org/10.22215/etd/2019-13866

Lesion Synthesis Toolbox
User Manual

Version 1.0

Revision Date: 2023/08/31

By: Quinn de Bourbon, BSc

Supervised by: Ran Klein, PhD

Credits: Hanif Juma, MASc

University of Ottawa, Department of Medicine, Division of Nuclear Medicine and

Molecular Imaging

The Ottawa Hospital, Department of Nuclear Medicine and Molecular Imaging

Ottawa Hospital Research Institute

A.1 Introduction

Lesion Synthesis Toolbox (LST) is a research tool developed by Ran Klein and his trainees at The Ottawa Hospital Research Institute in MATLAB®. LST is intended to generate well controlled fake lesions in raw positron emission tomography (PET) and reconstructed CT data and reconstruct the PET images in a realistic fashion.

Thus far, LST has only been implemented and tested for GE Healthcare PET/CT scanners, using their DUETTO image reconstruction toolbox, which requires a research collaboration agreement with GE Healthcare. For more information on DUETTO and research collaboration contracts, contact your GE Healthcare research representative and/or Elizbeth Philps (Elizabeth.Philps@med.ge.com). Nevertheless, LST was designed to be extendable to work with other vendors and products and we are looking forward to collaborating to achieve this.

LST can be used to reconstruct raw PET data. Lesions can be added with varying contrast, size, and shape, to both PET and CT datasets and can be previewed on the reconstructed images. In the case of PET, this is achieved by introducing the lesion signal into the projection data, prior to image reconstruction. In the case of CT, the lesions are painted into the reconstructed image. PET images can be reconstructed using any vendor supported algorithm (e.g., OSEM and Q.Clear^{TM}) and parameters.

LST can be used to create ground truth lesion reference images to train, validate and characterize human and AI observers. Furthermore, by changing reconstruction parameters, researchers can analyze how reconstruction changes the detectability of lesions and accurately models the noise and texture in a PET image.

Raw PET data and corresponding CTs can be imported from the PET scanner console using LST. Images with and without lesions can be displayed with various options for fusion, colormaps, and intensity levels directly within LST. Resulting image data (with and without lesions) can then be exported by DICOM transfer or DICOM files. Corresponding ground truth data on the lesions can also be exported.

LST is based of the Master's work of Hanif Gabrani-Juma[53], [54] and Quinn de Bourbon, under the supervision of Ran Klein at The Ottawa Hospital.

Figure A.1: The Lesion Synthesis Toolbox icon

A.2   Starting LST

LST can be run directly in MATLAB © by:

1) Copying the LST program files to a local directory.

2) Adding the local directory to the MATLAB path.

3) Adding the DUETTO to the path. Recommended to make the path in the *Recon Engines*.

4) Executing LesionSynthesisToolbox.mlapp in MATLAB

Alternatively, LST can be compiled to run as a standalone executable or webapp as per MATLAB documentations.

When LST runs for the first time, it will ask for a location to create the default data directories. The user can select either 'initialize new' or 'locate existing' to initialize the application data directory. Also, there is as an option to abort. If one of the first two options are selected, a file explorer window will open. From here the user can select a file folder in which a new folder is created, or an existing folder is selected, depending on if the user chose to initialize a new folder or locate an existing one.

- If these data are to be shared between multiple computer users, it is recommended to a shared directory such as "C:\Lesion Synthesis Toolbox".

- These folders will store the archive of raw data and reconstructed images, which can be very large files. It is recommended to select a drive with >1TB capacity.

The LST data subdirectories by default include:

- Application Data – where LST configuration settings are saved including general settings, project settings and reconstruction profiles.

- Raw Data – where raw PET associated CT data are archived when retrieved from the scanner console.

- Recon Queue – where reconstructions are queued for processing.

- Reconstructed Archive – where reconstructed image data are archived along with their CT and intermediate reconstruction files. Reconstructed PET are saved as DICOM directories and in a _fir3D.mat file format, both of which are named for the reconstruction type. The CTs are saved as DICOM in a CTAC_DICOM directory and as a CTAC.mat file. Each patient data is saved in a unique identifier directory.

- Simulation Archive – where simulations are queued for processing.

- Simulation Queue – where simulated studies are archived. Subdirectories identify the patient and sub-subdirectories identify the simulation name. In each simulation CT images are saved in a DICOM directory and _fir3D.mat file (same as for Reconstructed Archive). Likewise simulated PET are named according to the reconstruction profile. The simulation parameters are also saved in a _LesionParams.mat file.

When initializing a new installation, before the users can access the toolbox, the administrator must set up accounts for each user. The administrator is directed to the *Configure* tab to add users and make other general configurations such as setup/change the directories for the specific project. More details are available in section A.7.

A.3   The Login Tab

In this window (see Figure A.2), the user can enter their credentials to unlock the workspace. Users are configured in the Configuration tab as described in the General settings. Each user may have their own list of projects and their associated settings.
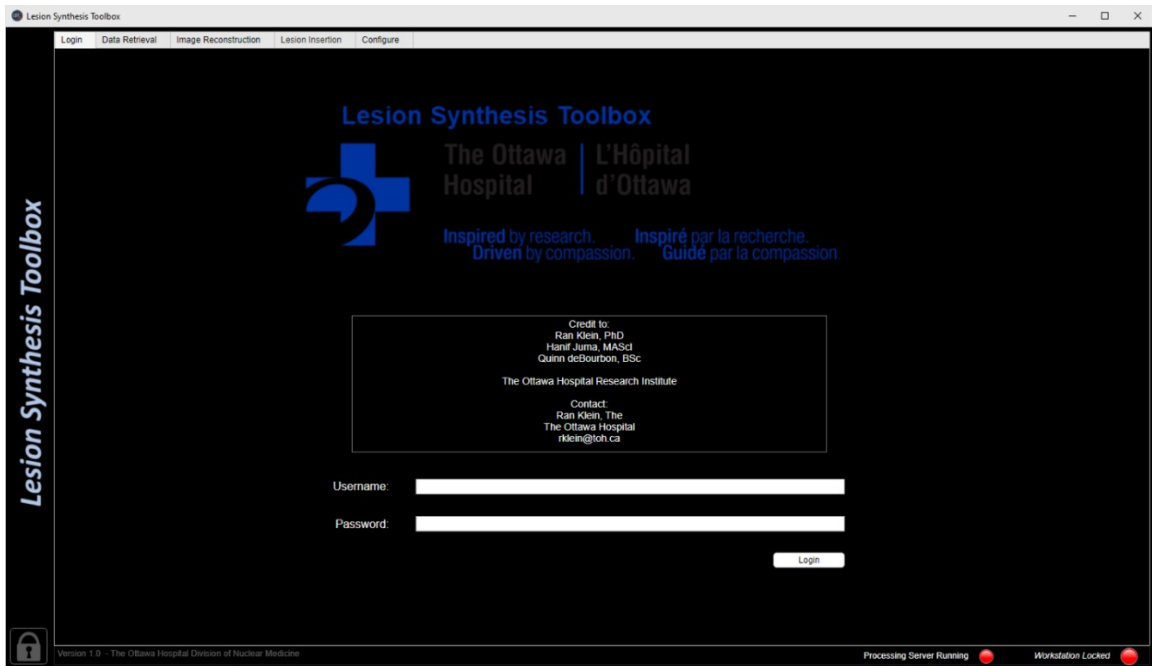

Figure A.2: The login tab

Table A.1: Elements from the login tab in Figure A.2

Element	Name	Description
Field	Username	The user enters their username.
Field	Password	The user enters their password.
Button	Login	If the username and password are correct, the workstation will be unlocked.
Lamp	Processing Server Running 🔴	Indicates whether the processing server is running in the background to process any reconstructions or simulations in the queues. A red circle 🔴 indicates that the server is not running, while a green circle 🟢 indicates that the server is running.
Lamp	Workstation Locked 🔴	If the username entered is incorrect, the "Workstation Locked" text will change to "username does not exist". If the username is correct, but the password is incorrect, the text will display "password does not match". If the workstation is unlocked, the red circle

		changes to a green circle 🟢 and the "Workstation Locked" text will change to the User's name.
Button	Logout 🔒	When clicked, the user is logged out and the workspace is locked.

A.4 Data Retrieval Tab

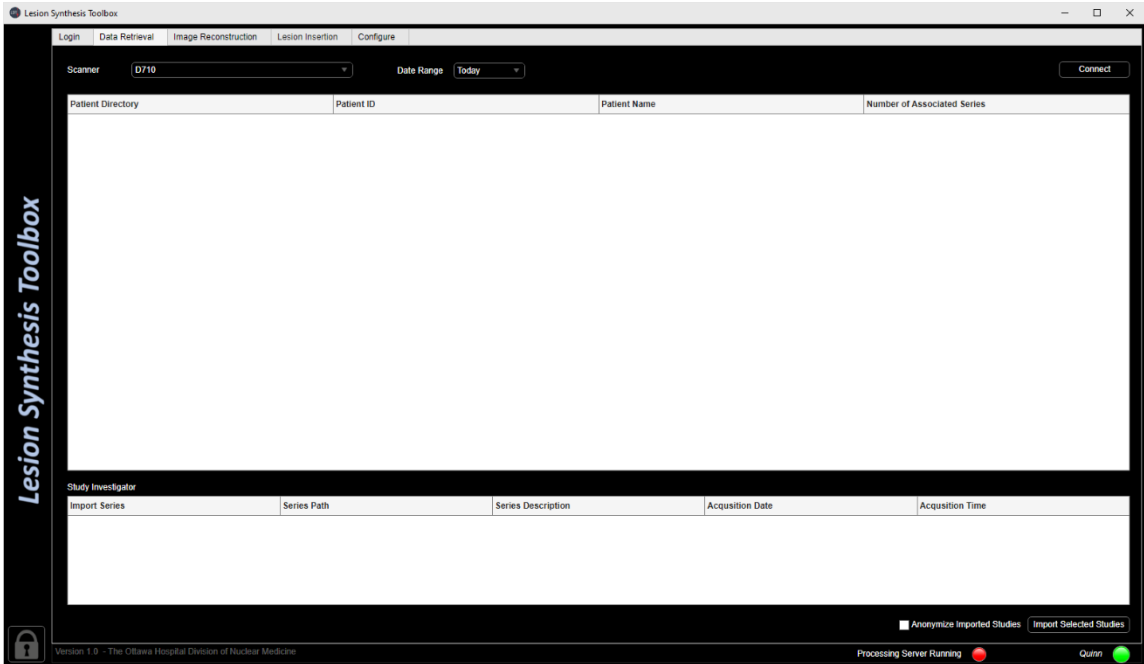The data retrieval window connects to the scanner to retrieve recent data files.


Figure A.3: Data Retrieval tab

Table A.2: Elements from the data retrieval tab in Figure A.3

Element	Name	Description
Drop Down List	Scanner	The user selects a scanner from a drop-down list. The scanner list can be changed in the 'Configure' tab.
Drop Down List	Date Range	The user selects the date range for the data from the scanner. The date options are 'today', 'past week', 'past 2 weeks', 'past 3 weeks', 'past month', and 'all'.
Button	Connect	If the connection to the selected scanner is unsuccessful, the button will turn red. If the connection is successful, the button will turn green and the data from the selected date range will load onto the patient data table.
Table	Patient Data	The patient data files are listed with the patient directory, patient ID, patient name, and number of associated series. The patient data files can be sorted in ascending or descending order based on the directory, ID, and name. The user can select a patient to get more information.

Table	Study Investigator	The selected patient information appears, including the series path, series description, acquisition date, and acquisition time. A check box in the import series column can be selected to be imported.
Checkbox	Anonymize Imported Studies	The selected patient data will be anonymized once the 'Import Selected Studies' button is pushed.
Button	Import Selected Studies	The selected patient data will be copied to the Image Reconstruction Tab. This process may take several minutes.
Pop-Up Window	Anonymization	The selected patient file is anonymized by replacing the patient family name, given name, ID, and the date of birth, however the year of birth is kept. The window is shown in Figure A.4.
Pop-Up Window	Confirm in place anonymization	The user can choose to copy over only the anonymized version of the data using the 'In place' button, or can create a copy and transfer both versions using the 'Duplicate' button to the Image Reconstruction Tab. This window is shown in Figure A.5.
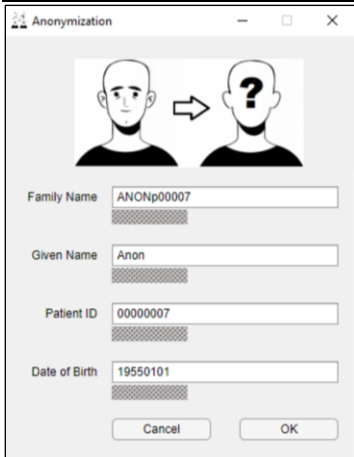
Figure A.4: Anonymization window
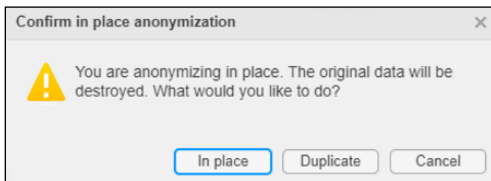
Figure A.5: In place anonymization

A.5   Image Reconstruction Tab

In the image reconstruction window, the patient data is set up to be reconstructed.
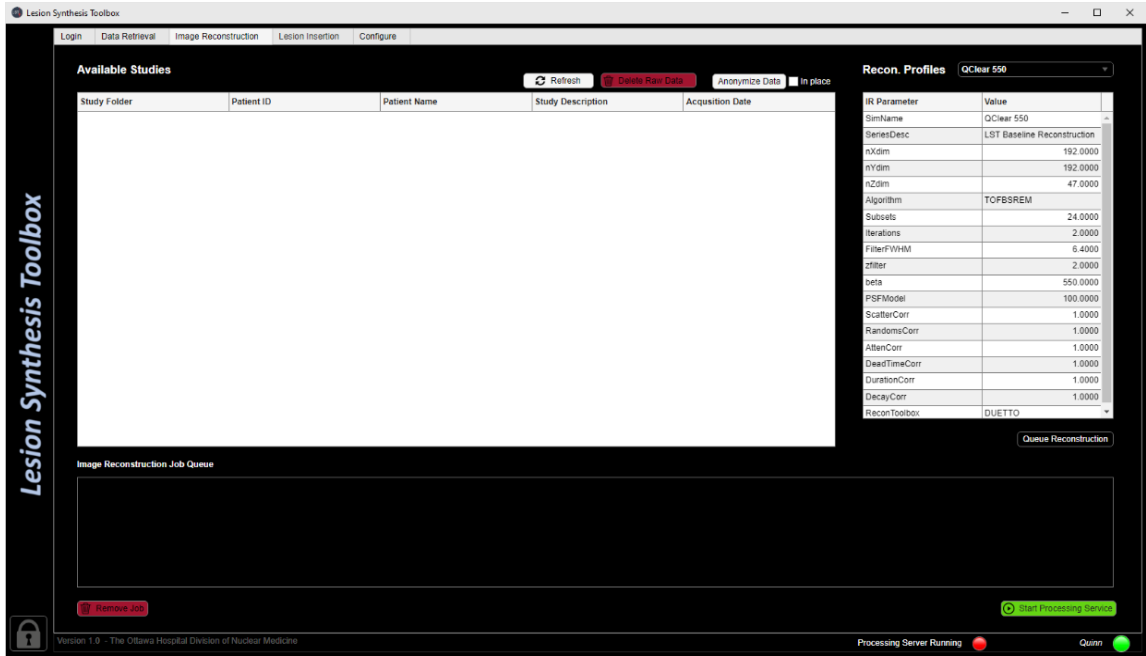

Figure A.6: Image Reconstruction tab

Table A.3: Elements from the image reconstruction tab in Figure A.6

Element	Name	Description
Table	Available studies	The studies available for reconstruction are listed by study folder, patient ID, patient name, study description, and acquisition date. The user can select a study to be reconstructed or deleted.
Button	Refresh	The list of available studies is refreshed.
Button	Delete Raw Data	The study selected by the user is deleted.
Button	Anonymize Data	The selected patient file is duplicated and anonymized. The pop-up window 'Anonymization' from Figure A.4 appears.
Checkbox	In Place	When the checkbox is selected and the button 'Anonymize Data' is clicked, the raw data are deleted after anonymization. The pop-up window 'Confirm in place anonymization' from Figure A.5 appears.
Drop Down List	Recon. Profiles	The user selects the type of reconstruction method from a drop-down list. These are defined in the *Recon settings* as described in section 0.
Table	IR Parameter	The reconstruction profile parameters are shown with their respective values. These values can be adjusted by the user.

Button	Queue Reconstruction	The patient file is added to the job queue.
List	Image Reconstruction Job Queue	Lists the patient file in the queue, listed as {patient name}\{profile method}_reconParams.mat. The user can select jobs from the queue to be removed.
Button	Remove Job	The selected patient file from the queue is removed.
Button	Start Processing Service	Starts the background image reconstruction and lesion simulation service.

A.6   Lesion Insertion Tab

The Lesion Insertion tab is used to simulate define, simulate and validate lesions (e.g., emulated disease) within the patient data. It consists of a 4-step workflow, defined in sub-tabs:

A.6.1.   Select Patient Sub-Tab

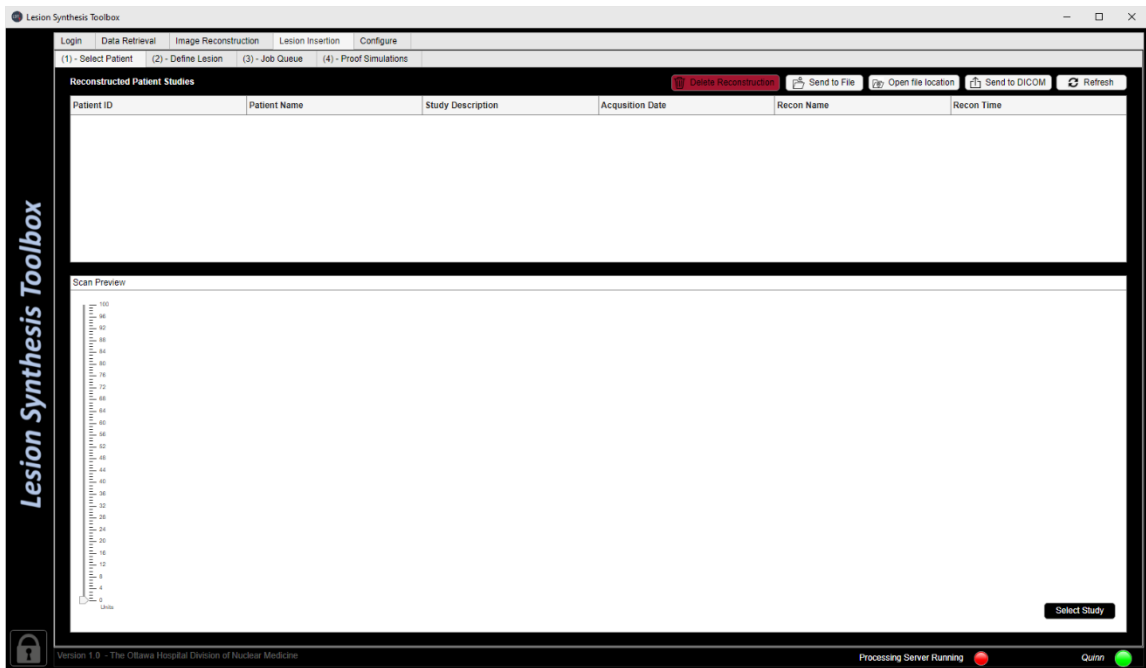The patient selection tab provides previews of reconstructed patient data.

Figure A.7: Lesion Insertion tab for Select Patient sub-tab.

Table A.4: Elements from the select patient sub-tab in Figure A.7

Element	Name	Description
List	Reconstructed Patient Studies	The list of patient studies that have been reconstructed will be displayed here. The user can select a study to be previewed or selected. Multiple studies can be selected to be sent to file.
Button	Delete Reconstruction	The selected studies are deleted from the list.
Button	Send to File	The selected studies CT and reconstructed PET images are sent to a folder selected by the user.
Button	Open File Location	The selected study file location is opened in the file explorer.

Button	Send to DICOM	The selected data is sent to a location specified by the DICOM send field in the 'Configure Tab'.
Button	Refresh	Refreshes the list of reconstructed studies.
Image	Scan Preview	A preview of the PET image of the selected study will be displayed with the sagittal, coronal, and transverse planes. The user can select the point of view by clicking on the image or can scroll through the image along a plane using a mouse scroll.
Slider	0-maximum intensity pixel	The user can adjust the maximum intensity of the PET image, between 0 and the maximum intensity pixel in the image. The image units depend on the corresponding project configuration setting.
Button	Select Study	The selected study will be opened in the 'Define Lesion Sub-Tab' for lesion insertion.

A.6.2.   Define Lesion Sub-Tab

In the define lesion tab, various lesion objects can be added to the patient's
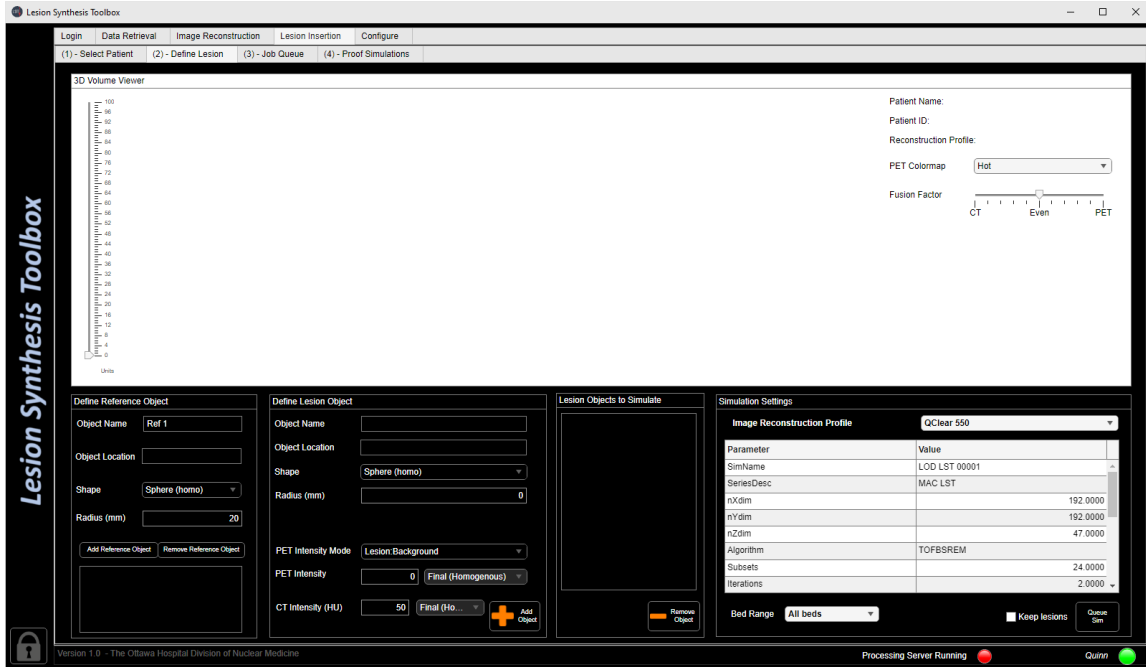
reconstruction image.

Figure A.8: Lesion Insertion tab for Define Lesion sub-tab.

Table A.5: Elements from the define lesion sub-tab in Figure A.8

Element	Name	Description
Panel	**3D Volume Viewer**	**Preview of the image volume with lesion insertion.**
Image	3D Volume Viewer	The 3D image of the selected patient study appears along the sagittal, coronal, and transverse planes. The user can scroll through the image using their mouse or can select to triangulate in three the three orthogonal slices by clicking on the image.
Slider	0-maximum intensity pixel	The user can adjust the maximum intensity of the PET image, between 0 and the maximum intensity pixel in the image. The image units depend on the corresponding project configuration setting.
List	PET Colormap	The user selects the colour map used for the PET images.
Slider	Fusion Factor	The user selects the ratio of image intensity between the CT and PET images.
Panel	**Define Reference object**	**The user selects a reference region for the intensity levels in the simulated lesion object.**

Field	Object Name	The user selects a name for the reference object.
Field	Object Location	The user selects a location for the reference object. This can be done by manually writing the location or selecting a location using the crosshairs on the 3D volume viewer.
List	Shape	The user selects the shape for the reference object. The shape selection is currently limited to a homogenous sphere (i.e., the average intensity of the pixels within the sphere).
Field	Radius (mm)	The user selects a radius for the reference object, in mm.
Button	Add Reference Object	The reference object is added.
Panel	**Define Lesion Object**	**The user selects parameters for the simulated lesion object.**
Field	Object Name	The user selects a name for the lesion object.
Field	Object Location	The user selects a location for the lesion object. This can be done by manually writing the location or selecting a location using the crosshairs on the 3D volume viewer.
List	Shape	The user selects a shape for the lesion object. The shape selection includes a sphere and a blobby sphere*.
Field	Radius (mm)	The user selects the radius of the lesion object, in mm.
List	PET Intensity Mode	The user selects the intensity mode for the lesion object on the PET image. The intensity modes include 'Bq/cc' (native PET intensity value in units of Bq/cc), 'SUV' (intensity value in units of standard uptake values by body weight), 'Lesion: Background' (relative to background), and for each reference object 'Lesion:{Reference object name}' (relative to the reference object).
Field/List	PET Intensity	The user selects the intensity for the lesion object for the PET image, based on the intensity mode. The intensity level can be written in the field, while the texture can be selected in the list. The texture options include 'Final: Homogenous' (equal intensity across lesion), 'Final: Maintain Texture' (average value changed but texture is the same), and 'Incremental'(additional over background).
Field/List	CT Intensity	The user selects the intensity for the lesion object on the CT image, in Hounsfield Units. The intensity level can be written in the field, while the texture can be selected in the list. The texture options include 'Final: Homogenous' (target uniform intensity across

		lesion), 'Final: Maintain Texture' (target average value is specified and existing image texture is preserved), and 'Incremental' (intensity is added to the existing background activity in the image).
Button	Add Object	The object is added to the list of lesion objects to simulate.
Panel	**Lesion Objects to Simulate**	**The list of lesion objects defined for simulation.**
List	Lesion Objects to Simulate	The lesion objects are listed here to be simulated. The objects can be selected by the user, which populates the Define Lesion Object Panel fields.
Button	Remove Object	The selected lesion object, from the lesion objects to simulate list, is deleted.
Panel	**Simulation Settings**	**The user selects the image simulation and reconstruction settings.**
List	Image Reconstruction Profile	The user selects the type of reconstruction method from a drop-down list. These are defined in the *Recon settings* as described in section 0.
Table	Parameter Values	The parameters of the selected image reconstruction profile are listed, and can be edited by the user.
List	Bed Range	The user selects the bed range for the reconstruction with the lesion object, the options are 'All beds' and 'Bed range with lesions'.*
Checkbox	Keep Lesions	When selected, the file and lesions stay in the 'Define Lesion' tab, which can be edited to be queued again.
Button	Queue Sim	The lesion objects will be added to queue to be simulated, using the parameters and bed positions that the user selected.

* denotes feature identified as not fully validated.

A.6.3. Job Queue Sub-Tab

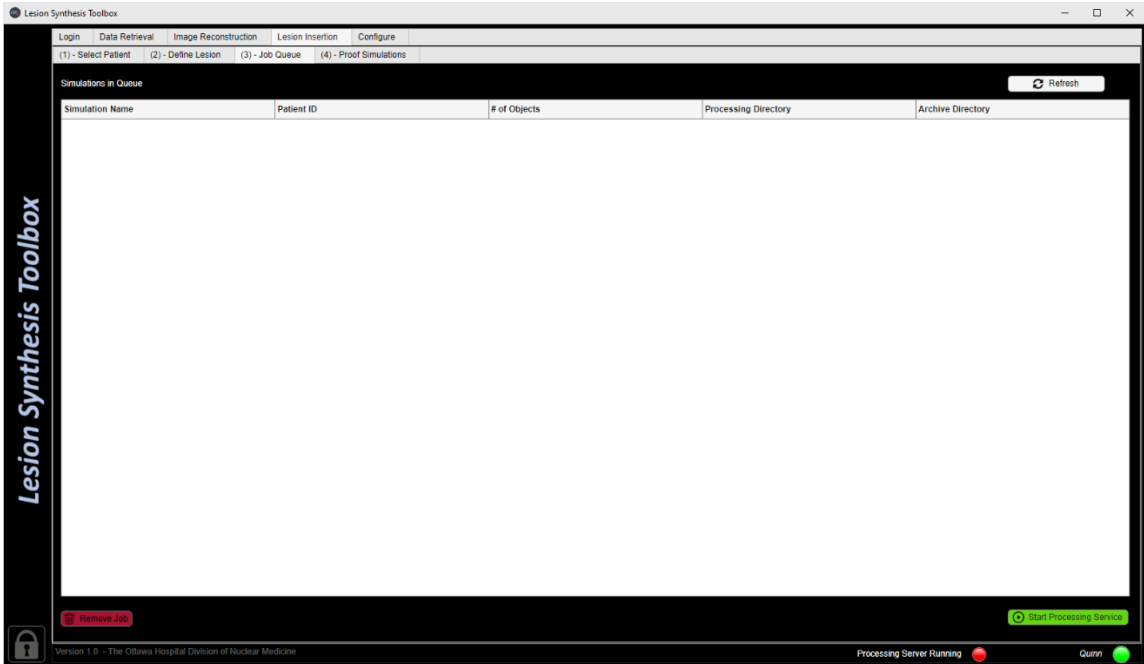In this tab, the job queue is displayed, and the lesion studies can be simulated.

Figure A.9: Lesion Insertion tab for Job Queue sub-tab.

Table A.6: Elements from the job queue sub-tab in Figure A.9

Element	Name	Description
Button	Refresh	The queue list is refreshed.
List	Simulation Queue	The job queue for lesion simulation is displayed. The user can select a job on the queue.
Button	Remove Job	The selected job on the simulation queue is deleted.
Button	Start Processing Service	Starts the background image reconstruction and lesion simulation service.

A.6.4. Proof Simulations Sub-tab

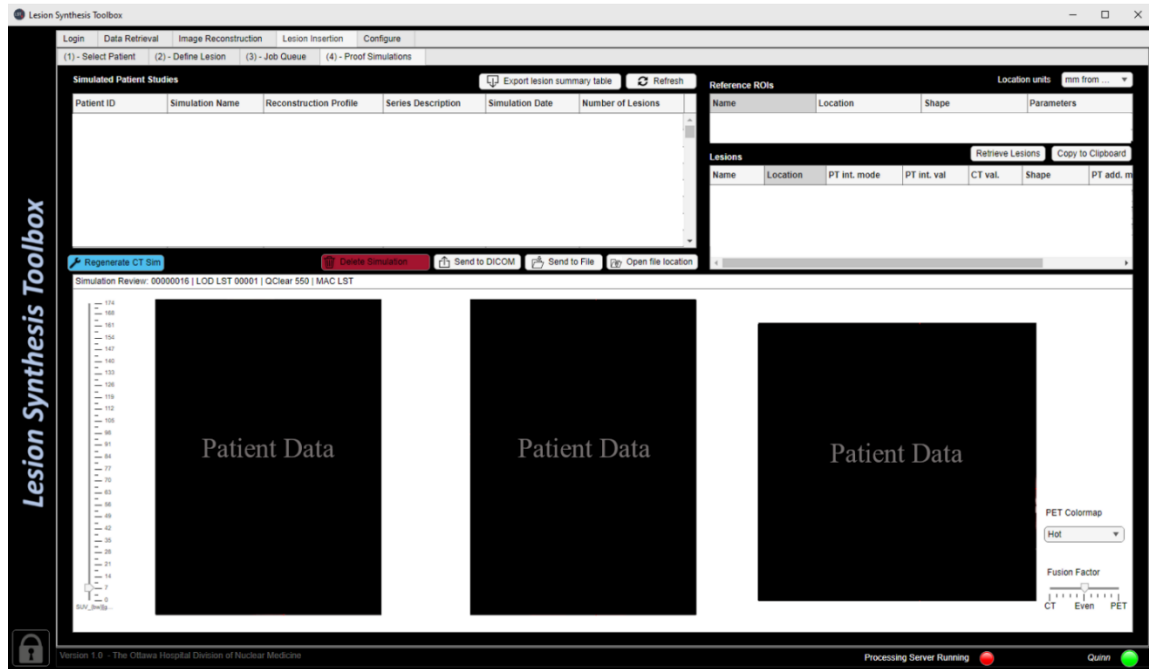The proof simulations tab allows for viewing the reconstructions after adding the lesion objects.

Figure A.10: Lesion Insertion tab for Proof Simulations sub-tab.

Table A.7: Elements from the proof simulations sub-tab in Figure A.10

Element	Name	Description
List	Simulated Patient Studies	A list of the patient studies that have been simulated appears here. The user can select a study to view.
Button	Export Lesion Summary Table	An excel spreadsheet is generated containing a list of the lesion objects and their parameters in the project simulations. See 0 for more details.
Button	Refresh	The simulated patient studies list is refreshed.
Button	Regenerate CT Sim	When clicked, the selected patient's CT is regenerated with the lesions included. This can be used if there is an issue creating the CT lesions.
Button	Delete Simulation	The selected patient simulations will be deleted.

Button	Send to DICOM	The selected patient study is sent to a DICOM file on another server, which can be specified in the 'Configure' tab.
Button	Send to File	The selected studies are sent to a folder specified by the user. The studies include the CT and PET images, but not the ground truth values of the lesions.
Button	Open File Location	The location of the patient simulation file is opened.
Drop Down List	Location Units	The units for location of points in the image are defined. The user can select either 'pixel', 'mm from centre', or 'mm from origin'. Note that the unit that is selected will appear in the Excel 'Lesion Summary Table'.
List	Reference ROIs	The location of the reference regions of interest are listed using the units selected by the user. To rename the reference ROI's, the user can right click on the ROI and a pop-up window will appear, as seen in Figure A.11.
Button	Retrieve Lesions	The selected patient's original reconstruction file is opened in the 'Define Lesion Sub-Tab' with lesions added to the 'Lesion Objects to Simulate' list. Note that this does not edit the simulated image, it recreates the 'Define Lesion Sub-Tab' before simulation.
Button	Copy to clipboard	The list of lesions for the selected patient study is copied to clipboard.
List	Lesions	The lesions are listed for the selected patient study with the parameters of the lesion, including name, size, shape, location, and intensity (mode, PET, and CT). To rename the lesions, the user can right click on the lesion and a pop-up window will appear as seen in Figure A.12.
Image	Simulation Review	The selected patient study containing lesion objects is displayed on the screen using three views, the sagittal, coronal, and transverse view.

Slider	0-maximum intensity pixel	The user can adjust the maximum intensity of the PET image, between 0 and the maximum intensity pixel in the image. The image units depend on the corresponding project configuration setting.
Drop Down List	PET Colormap	The user can select the colormap for the PET image.
Slider	Fusion Factor	The user can select the ratio of intensity for the PET and CT images.

A.6.5. Lesion Summary Table

When the *Export lesion summary table* button is pressed, the user is prompted to select a location and name of the Excel spreadsheet with the resulting ground-truth data for all the lesions in the project. By default, the file will be named *{Project Name}-ground truth.xlxs* and will be saved in the *simulation archive directory*.

The location coordinate system and units in the spreadsheet depend on the selection in the *Location Units* drop-down list.

The parameters saved to the file are listed below.

- Patient ID
- Simulation Name
- Reconstruction Profile
- Series Description
- Simulation Date
- Lesion Name
- Location Coronal [location units]
- Location Sagittal [location units]
- Location Transaxial [location units]
- PET Intensity Mode
- PET Intensity Value
- CT Intensity Value
- Shape
- Shape Parameters
- Reference Lesion Activity (Bq/cc)
- Baseline Background activity (Bq/cc)
- Target Lesion Activity (Bq/cc)
- Surrounding Margin (mm)
- Surrounding Mean PET Intensity
- Surrounding SD PET Intensity
- Surrounding Min PET Intensity
- Surrounding Max PET Intensity
- Age (years)
- Sex
- Height (m)
- Weight (kg)
- Injected Activity (MBq)

A.6.6. Renaming simulation objects

Lesions and reference ROIs in the corresponding tables on the right side of the screen can be renamed (and the corresponding LesionParams.mat file updated), by selecting the object, right clicking on it, and selecting rename. A corresponding dialogue (see Figure A.11 and Figure A.12) is displayed where the new name can be typed manually or selected from a default list of names.
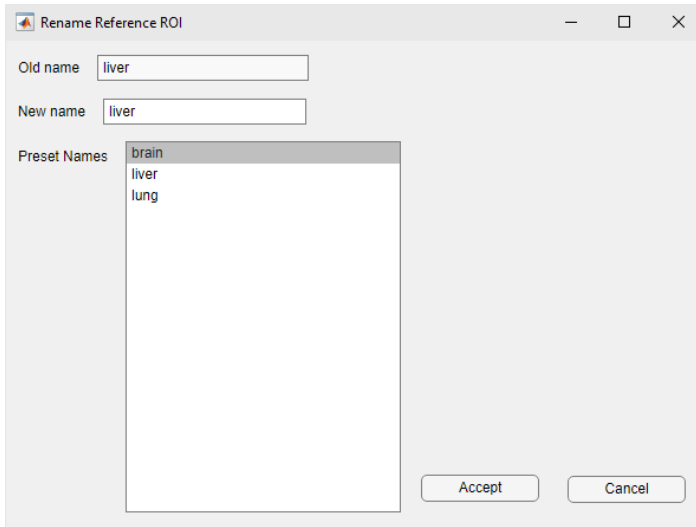

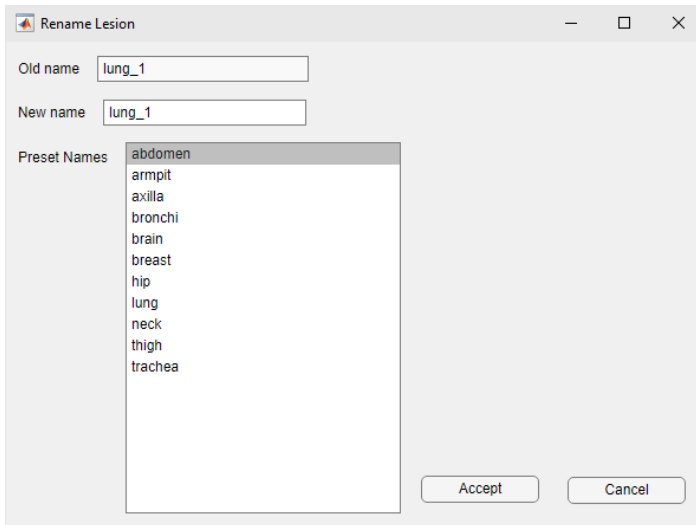**Figure A.11: Renaming a reference region of interest, with a list of preset names available.**

Figure A.12: Renaming a lesion, with a list of preset names available.

A.7 Configure Tab

In the configure window, the settings for the toolbox can be adjusted.

A.7.1. General settings

The general tab includes scanner and server settings, only accessible to admin users.
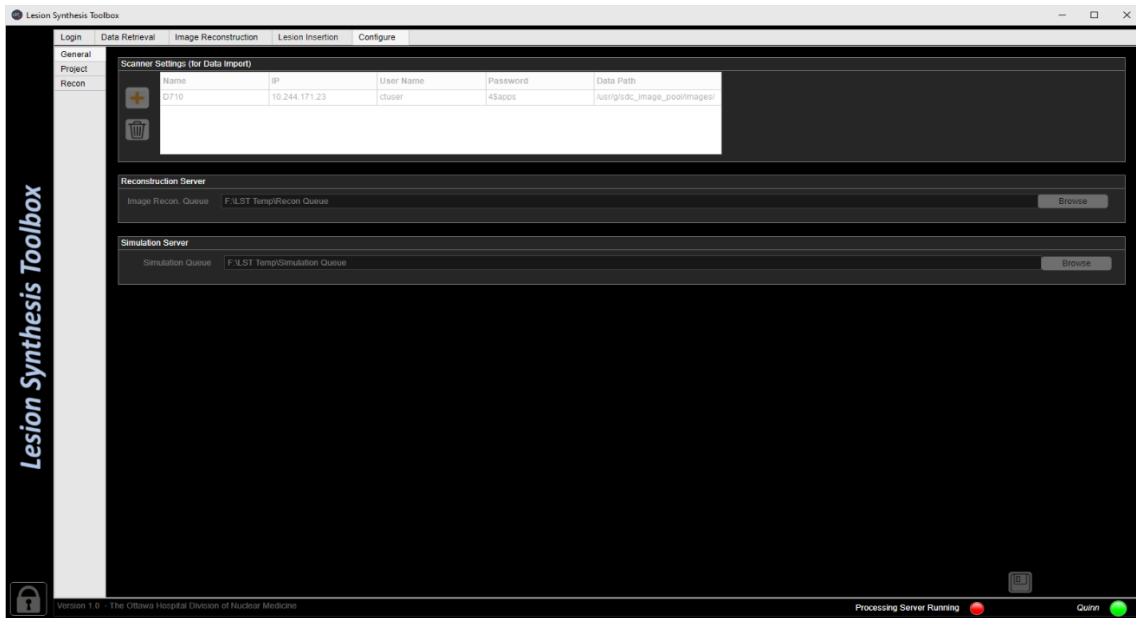

Figure A.13: Configure tab for general settings

Table A.8: Elements from the general settings sub-tab in Figure A.13

Element	Name	Description
List	Scanner Settings	The settings for the available scanners are listed, and only accessible to the admin user. The admin user can select a scanner from the list.
Button	Add ✚	The admin user can add a scanner.
Button	Delete 🗑	The selected scanner is removed from the list.
Button	Save 🖳	The scanner settings are saved.
Field	Image Recon. Queue	The admin user can add the directory for the reconstruction queue.
Field	Simulation Queue	The admin user can add the directory for the simulation queue.
Button	Browse	The user can use browse to select a directory from the file explorer instead of manual inserting it in the two fields above.

A.7.2. Project settings

The project tab contains the list of projects for the user, including the parameters for each project.
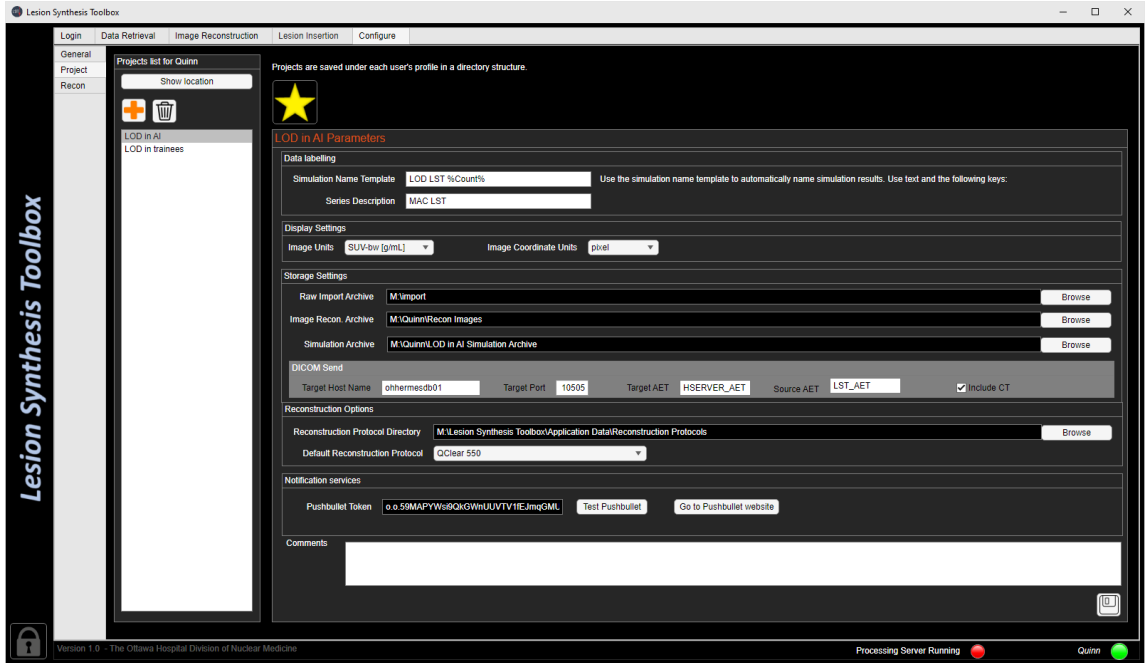

Figure A.14: Configure tab for project settings

Table A.9: Elements from the project setting sub-tab in Figure A.14

Element	Name	Description
Panel	**Projects List for {User}**	**The list of projects that the user has created are listed. The user can select a project and see the parameters on the screen.**
Button	Show Location	The location of the project file is opened in the File Explorer.
Button	Add ➕	A new project is created.
Button	Delete 🗑	The selected project file is deleted.
Button	Star ⭐	The user can select the default project that is opened when the user logs into the program.
Panel	**{Project Name} Parameters**	**The parameters for the selected project can be changed.**
Field	Simulation Name Template	The user can set a name to automatically name the simulation results. The user can still manually change the simulation name in the 'Define Lesion Sub-Tab'.

Field	Series Description	The user can input the series description name to be identified in PACS using DICOM tag (0008,103E).
Drop-Down List	Image Units	The user can set the units for the reconstructed images. The user can select from 'Activity [Bq/cc]' and 'SUV-bw [g/ml]'.
Drop-Down List	Image Coordinate Units	The user can select the coordinate system for the reconstructed images. The user can select from 'pixel', 'mm from centre', and 'mm from origin'.
Field	Raw Import Archive	The user can enter the directory path for the raw imports, or use the browse button.
Field	Image Recon Archive	The user can enter the directory path for the reconstructed images, or use the browse button.
Field	Simulation Archive	The user can enter the directory path for the simulated images, or use the browse button.
Button	Browse	The user can use the browse buttons to select the directory path for the field in line with the respective buttons.
Field	Target Host Name	The user can enter the name of the target host that the DICOM files will be sent to.
Field	Target Port	The user can enter the target port number that the DICOM files will be sent to.
Field	Target AET	The user can enter the target DICOM node AE title.
Field	Source AET	The user can enter the source DICOM node AE title.
Checkbox	Include CT	When selected, the CT images are included in the DICOM file.
Field	Reconstruction Protocol Directory	The user can enter the directory path for the reconstruction protocol, or use the browse button.
Drop-Down List	Default Reconstruction Protocol	The user can select the default reconstruction protocol that will be used for this project, from the list of reconstruction protocols defined in the corresponding *Configure / Recon settings* (see A.7).
Field	Pushbullet Token	The user can enter their Pushbullet ID for receiving notification alerts on their mobile device.
Button	Test Pushbullet	The user can test the notification system to ensure it is active.
Button	Go to Pushbullet Website	The website for the pushbullet is opened.
Field	Comments	The user can add additional comments to the project file.
Button	Save 🖫	The project parameters are saved under the project name in the projects list.

A.7.3.   Recon settings

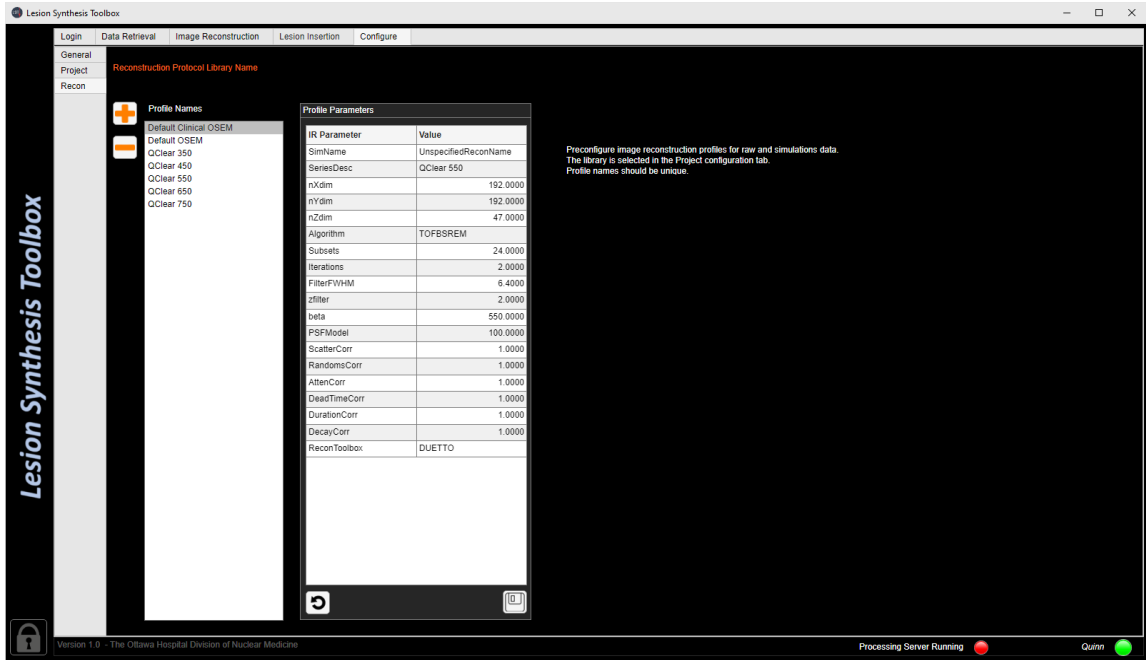The recon tab displays the reconstruction modes and parameters for each mode.

Figure A.15: Configure tab for reconstruction settings

Table A.10: Elements from the reconstruction sub-tab in Figure A.15

Element	Name	Description
Button	Add ➕	The user can add a new reconstruction profile, with a unique profile name.
Button	Delete ➖	The user can remove a reconstruction profile from the profile names list.
List	Profile Names	The reconstruction profiles are listed, and can be selected by the user to be viewed or deleted.
List	Profile Parameters	The parameters for the selected reconstruction profile are displayed.
Button	Refresh ↺	The parameters for the selected reconstruction profile is reverted to default.
Button	Save 🖫	The reconstruction profile is saved.

A.8   Tutorials

A.8.1.   Importing Data

Data can be imported on the 'Data Retrieval Tab', as seen in Figure A.3.

(1)   Select the scanner from the drop down 'Scanner' menu.

(2)   Next, select the date range using the following drop-down menu titled 'Date Range'.

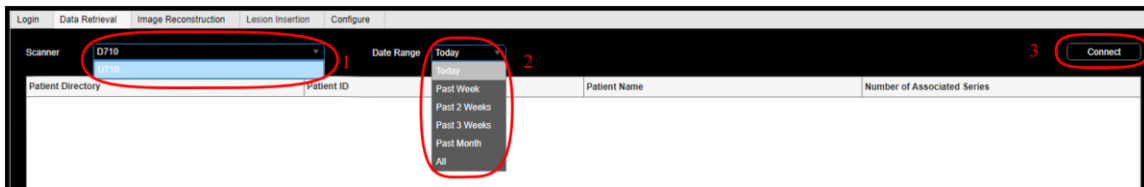(3)   Click the button 'Connect' to load the data.

Figure A.16: Tutorial - importing data from the scanner

A list of patient files will appear in the table below.

(4)   Select a file and it will appear in the study investigator at the bottom of the window. Click the check box next to the file in the study investigator to select the study.

(5)   The 'Anonymize Imported Studies' checkbox can selected to provide patient privacy.

(6)   Once selected, the study can be imported by clicking the 'Import Selected Studies' button on the bottom right corner of the window.
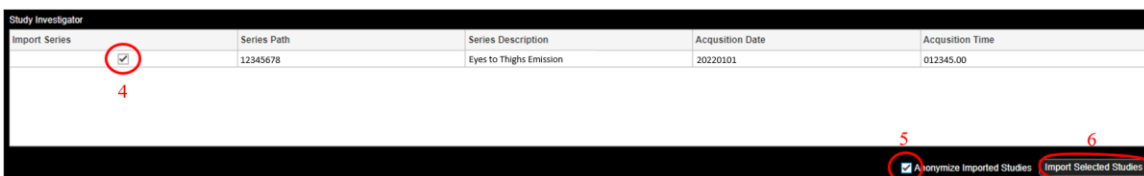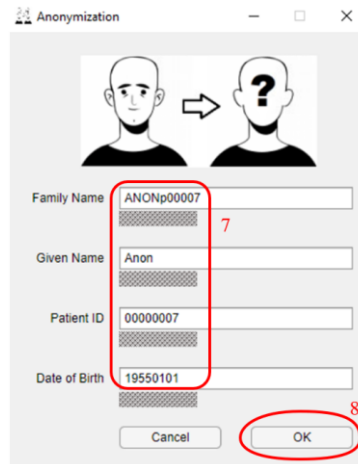

Figure A.17: Tutorial - selecting studies to import and whether to anonymize

(7)   If the file is to be anonymized, a window (Figure A.4) will appear which includes

fields to change the patient's name, surname, ID, and date of birth.

(8)   Click 'Ok' when finished anonymizing the patient information.

Figure A.18: Tutorial - anonymization information

(9)   A new window will appear (Figure A.5) with the option to anonymize in place or

create a copy. If the anonymization is done in place, the file downloaded to the

LST will be overwritten with an anonymized version but will not affect the

original file on the scanner.
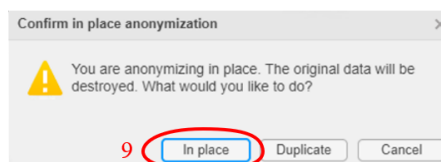
Figure A.19: Tutorial - anonymization data deletion

The importing process may take a couple of minutes to complete. The imported

studies will appear in the next tab (Image Reconstruction Tab).

A.8.2.   Reconstructing Data

On the Image Reconstruction Tab, the imported patient files are listed in the

'Available Studies' table.

(1) Select a study to be reconstructed from the table of available studies.

(2) If the file that was not anonymized when imported, it can be anonymized by selecting the 'Anonymize' button and choosing whether or not to anonymize in place. To see how to do anonymization refer to the tutorial 'A.8.1. Importing Data'.

(3) Choose a reconstruction method from the drop-down list 'Recon. Profiles' on the top right of the screen. The reconstruction parameters for the selected method will appear in the table below.

Figure A.20: Tutorial – available studies and reconstruction

(4) Click 'Queue Reconstruction' to add the reconstruction to the job queue.

(5) The job queue can be seen at the bottom of the tab.

(6) To remove a job from the queue, select the job from the queue and click 'Remove Job'.

(7) When the queue is ready, click 'Start Processing Service' to start processing the job queue. The reconstructed files can be found on the 'Lesion Insertion: Select Patient' tab.

Figure A.21: Tutorial – initial reconstruction

A.8.3. Define lesions

On the Lesion Insertion tab, the first sub-tab (Select Patient) should automatically

open.

(1)   Select the patient file that will be used for lesion synthesis.

(2)   If the study was newly added and does not appear, the list can be refreshed by
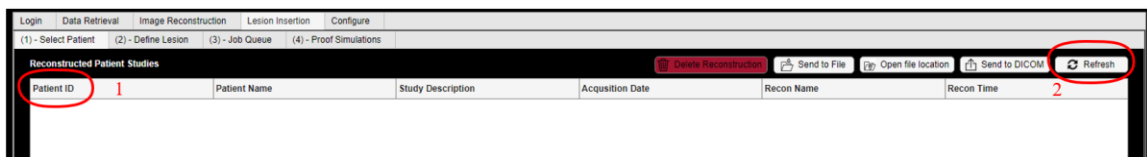
clicking 'Refresh'.

Figure A.22: Tutorial – selecting a patient for lesion insertion

(3)   The preview of the selected study will be shown at the bottom of the page, with a

3D view. The intensity of the image can be adjusted using the slider on the left of

the 3D view.

(4)   Click the 'Select Study' button and it will open the study in the next sub-tab

(Define Lesion).

Figure A.23: Tutorial – viewing patient before selection

(5)   The 3D viewer shows the fused PET and CT data on the top of the screen, with

information about the patient on the top bar.

(6)   Adjust the PET colormap and fusion factor on the right,

(7)   And intensity level on the left, until it is best for viewing.

(8)   The focal point on the image can be selected by clicking on one or more views. Adjusting the focal point will populate the 'Object Location' fields, so that they do not need to be entered manually.

Figure A.24: Tutorial – adjusting view during lesion insertion

Before adding any objects, a reference object should be set in the 'Define Reference Object' panel. This allows the lesion intensity to be relative to another organ, typically the liver.

(9)   The reference object name is automatically populated with Ref_{reference object number}, but can be manually changed to the region name. Set the object location, either by manually typing in the coordinates or by selecting the focal point on the image. Enter the shape and radius values. ROIs can be removed by selecting from the list and clicking 'Remove Reference Object'.

(10)   The object name is automatically populated with Object_{object number}, but can be manually changed to reflect the type of lesion. The location can be set manually or by selecting the focal point on the image. Select the shape and radius of the object.

(11) Adjust the PET intensity mode, level, and overlay method, as well as the CT
      intensity in Hounsfield Units and the overlay method. Add the lesion by clicking
      'Add Object' and repeat for as many lesions as needed.

(12) A list of lesions to simulate will be populated as lesions are created. Select a
      lesion to remove, or create another lesion with the same parameters, but ensure to
      change location coordinates to prevent possible issues.

(13) Select a reconstruction profile for the image reconstruction from the drop-down
      menu,

(14) And adjust the parameters in the table or bed range if needed. If the same
      reconstruction is to be repeated with different parameters, select 'Keep lesions' to
      keep the patient and lesions in the 'Define Lesion Sub-Tab'. Queue the simulation
      once all lesions are added, and the 'Job Queue Sub-Tab will open.

Figure A.25: Tutorial – defining lesions and reconstruction parameters

(15) Jobs can be selected from the list to be removed from the job queue.

(16) If the processing service is not already running, start it by selecting the green
      button at the bottom of the screen. A progress window will open, as is described
      in Section A.9 below.

Figure A.26: Tutorial – job queue and begin processing

A.8.4.   Check lesions

On the last sub-tab of the Lesion Insertion tab (Proof Simulation), the simulated lesions can be checked.

(1)   The list of files with simulated lesions can be seen at the top of the tab. Select a patient file to be viewed.

(2)   The simulated lesions and their properties are shown on the top right of the tab, as well as the reference ROI. Reference ROIs and lesions can be selected from the lists.

(3)   The selected lesion or ROI will be shown on the images, with the cross-hairs pinpointing the location. The view can be adjusted using the slider and menu in the bottom right.

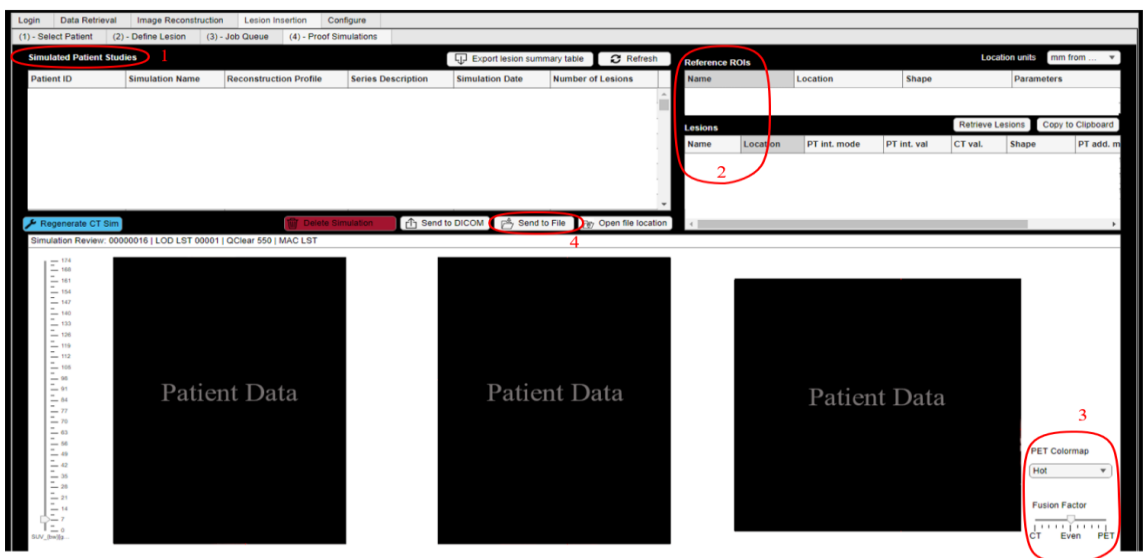(4)   Once the data are reviewed, the data can be exported to a file folder.

Figure A.27: Tutorial – reviewing lesions

A.9 Image Reconstruction and Lesion Synthesis Service

Image reconstruction and lesion simulation are computationally demanding. Depending on computer resources, these may require minutes to hours per image series. These processes are implemented as a computer service, for serial processing, which is especially advantageous in settings where the computer is acting as a web app service for multiple users. The lesion synthesis toolbox processing service (LSTProcessingService.m) is a MATLAB script that monitors directories for queued reconstruction and simulations. These are run one at a time using parallel processing for multiple bed positions.

The LST Processing Service can be launched from within the LST Toolbox, where a separate MATLAB instance is started in the background, enabling to continue using the LST Toolbox. Likewise, the LST Processing Service can be launched in MATLAB command line using the following parameters:

LSTProcessingService(command, dataDirs) where command can be any of the following:

- 'start' - start the server as a background service.
- 'stop' or 'kill' - stop the server.
- 'one time' - run the processing routine once for the existing files and then stop (i.e., not as a service)

Example uses are:

- LSTProcessingService('start',{'F:\LST Temp\Simulation Queue','F:\LST Temp\Recon Queue'})

- LSTProcessingService('stop')

- LSTProcessingService('one time',{'F:\LST Temp\Simulation Queue','F:\LST Temp\Recon Queue'})