Voxel-based internal dosimetry using deep learning

PhD in Experimental Physics – Statistics seminar

Alberto Arzenton

Università degli Studi di Siena

May 11, 2023

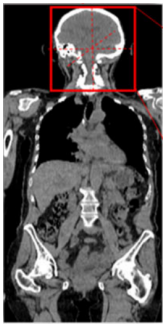
Overview

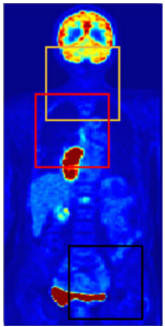
Overview

Monte Carlo voxel-based dosimetry

The **Monte Carlo (MC)** method, combined with **3D imaging**, is a powerful tool to perform **dosimetric analyses** in **nuclear medicine**.

- **Computed Tomography (CT)** can provide **anatomical** information to the MC software, which can create a **voxelized** volume resembling the patient's body.
- **Positron Emission Tomography (PET)** and **Single Photon Emission Computed Tomography (SPECT)** can provide **functional** information about the medical radioisotope distribution.





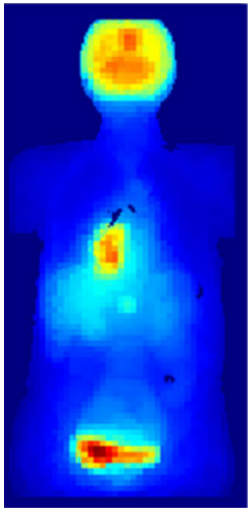
Monte Carlo voxel-based dosimetry

Radiation can be simulated sampling the activity distribution and the **absorbed dose** can be scored at voxel level, obtaining a **dose map**.

Monte Carlo codes

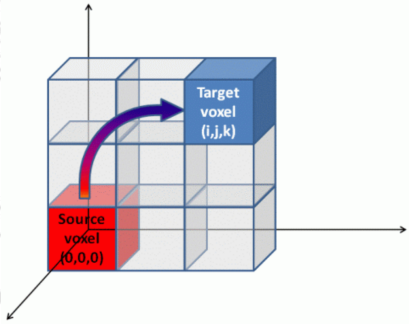
Common MC toolkits used for radiation tracking in medical applications are MCNP, GEANT4, PARTRAC and FLUKA.

The number of MC events required to have enough statistics is very high ($\geq 10^7$) and **time-consuming**.



The MIRD schema

An algorithm to **speed up** the process was proposed by the U.S. Committee on Medical Internal Radiation Dose (**MIRD**) [Bolch et al. (2009)].



The **dose-rate** in a target volume r_t from a series of sources r_s can be computed starting from the activity A (PET, SPECT) and a **convolution kernel** S (sized *ad lib*):

$$\dot{D}(r_t, t) = \sum_s A(r_s, t)\, S(r_t \leftarrow r_s)$$

The **S-values**, representing the mean absorbed dose in r_t per activity unit in r_s, can be simulated by MC *once for each radioisotope and tissue*.

Overview

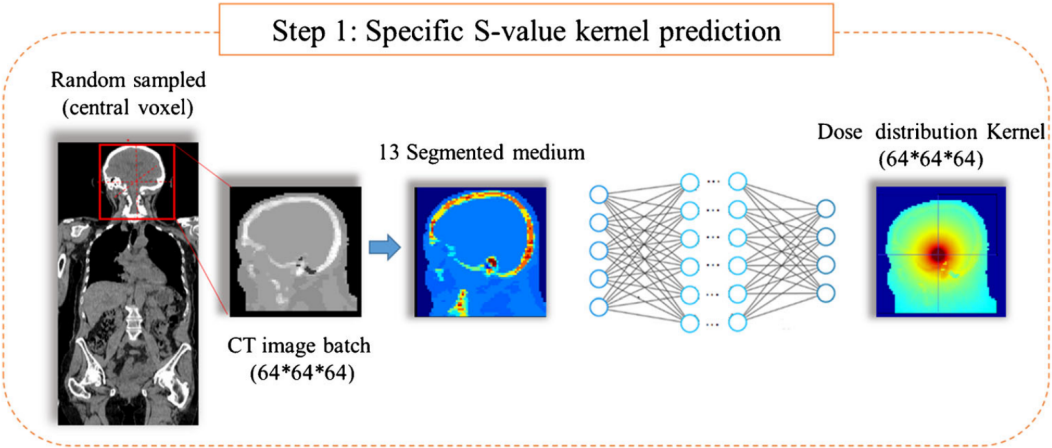
Why deep learning?

S-values are tissue-specific, therefore a convolution using different kernels will not be fully reliable next to the **organ boundaries**. Thus, an innovative technique must be:

- faster than direct MC;
- more precise than S-values.

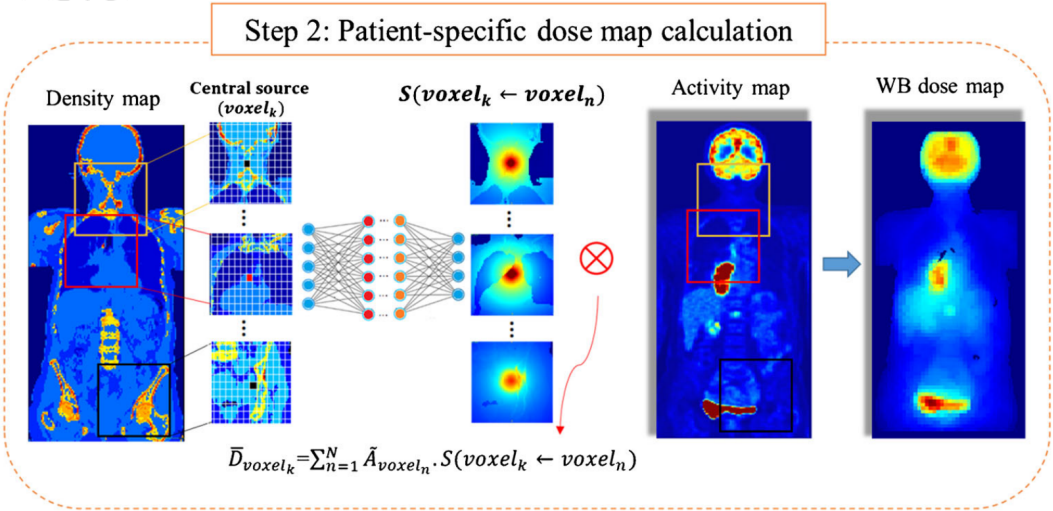
The AI state of the art suggested a possible improvement using **Deep Neural Network (DNN)** algorithms to build **mixed kernels** considering more than one tissue [Akhavanallaf et al. (2021)].

Workflow

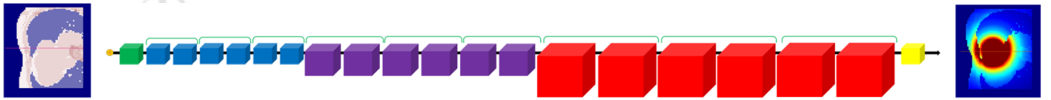


Each voxel has its specific S-value kernel, generated by the DNN using the CT.

Workflow

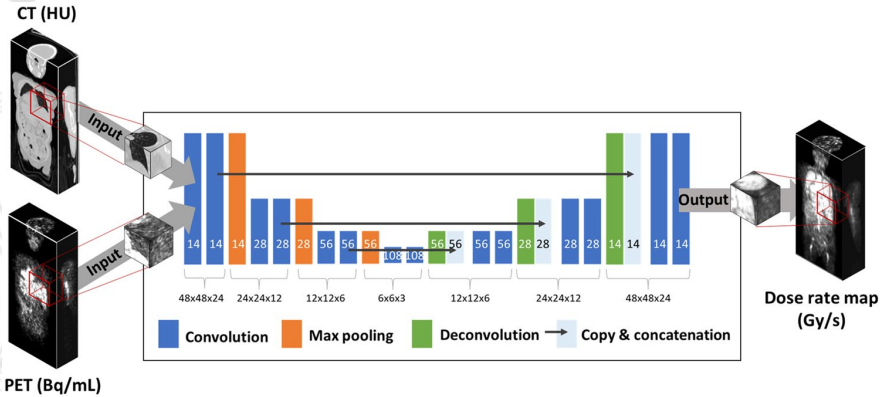

Step 2: Patient-specific dose map calculation
Density map
Central source ($voxel_k$)
$S(voxel_k \leftarrow voxel_n)$
Activity map
WB dose map
$\overline{D}_{voxel_k}=\sum_{n=1}^{N} \tilde{A}_{voxel_n}.S(voxel_k \leftarrow voxel_n)$

Workflow



The **ResNET** architecture, implemented on the TensorFlow platform, is suitable for internal dosimetry.

Alternative workflow



DNNs can also be used to directly predict the **whole dose map** [Lee et al. (2019)]; however, the **training time** grows dramatically ($\sim 10^3$), like the required **dataset**.

Overview

Statistical evaluation

How can we evaluate the goodness of a DNN approach? First of all, a proper **ground truth** has to be chosen.

Ground truth

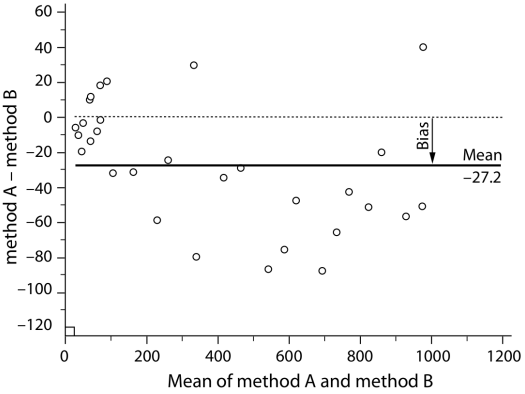
In internal dosimetry, the ground truth is usually chosen as the dose map produced by a direct MC simulation with a very high number of events.

N.B.

The ground truth is fundamental for both training and final evaluation.

Statistical evaluation

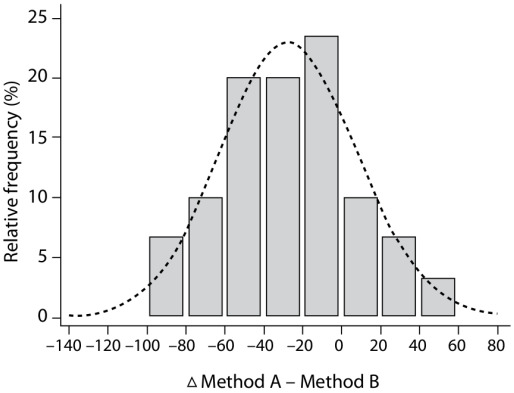
In medical statistics, a common method to compare the results of two different assays is the **Bland-Altman plot (B&A)**, which studies the **differences** between the measures against the **magnitude** of the observable [Giavarina (2015)].

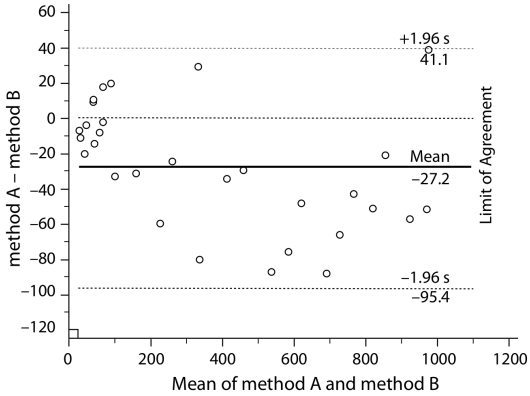


- Differences can be dimensional or **percentages**.
- The magnitude should be expressed as the **mean** of the two measures.
- An average **bias** can be identified with the mean difference $\bar{d}$, although it can actually change with the magnitude.

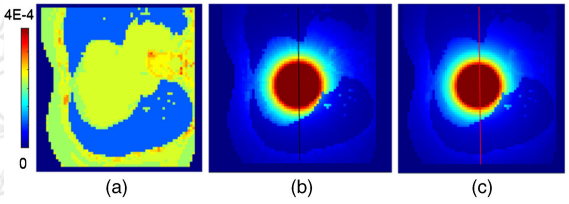
Statistical evaluation

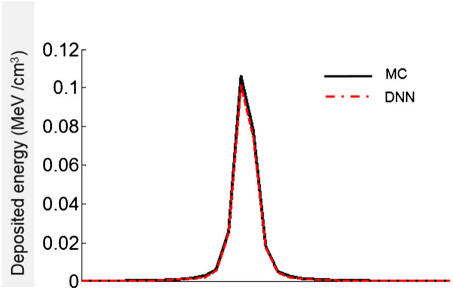
If the differences are normally distributed, their **95% Confidence Interval (CI)** will be defined by $\bar{d} \pm 1.96s$, where s is the **standard deviation**.





Comparison with Monte Carlo

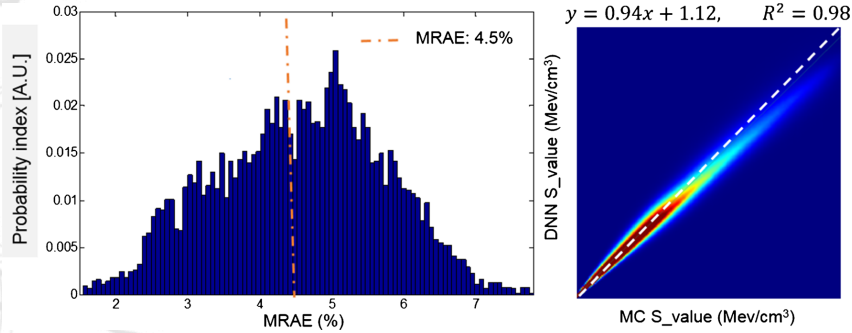




(a) Density image of the lung region taken from a CT. (b) Direct MC. (c) DNN.

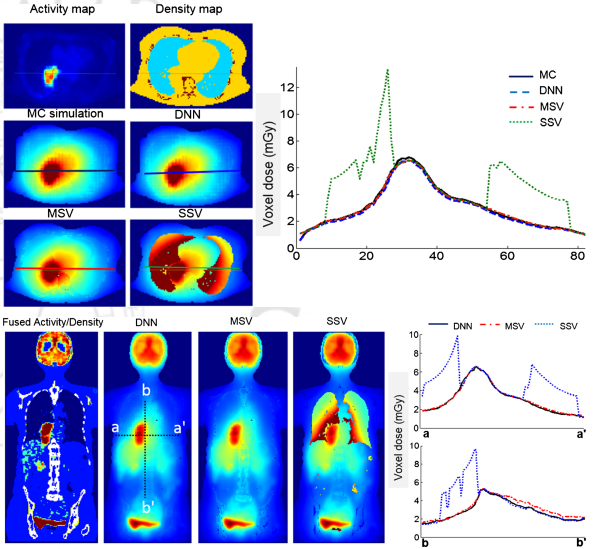
- The DNN produced $64 \times 64 \times 64$ **mixed kernels** for the β^+ decay of ^{18}F-FDG, which were compared to the ground truth, *i.e.* direct MC.
- The **kernel size**, 19.2 cm, was bigger than the range of the 511 keV annihilation γ-rays (~ 7 cm).
- The **training dataset** included 12100 frames taken from *just* 24 CT images.

Comparison with Monte Carlo



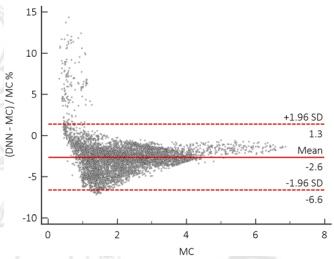
- The **distribution of the percentage differences** (here Mean Relative Absolute Error, MRAE), assessed before and after the convolution, looks coarsely normal.
- DNN shows a **good correlation** with MC kernels.

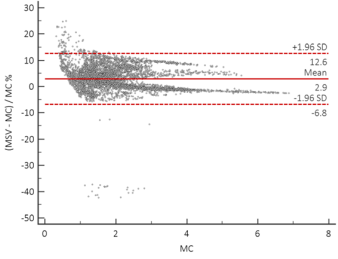
Comparison with Monte Carlo

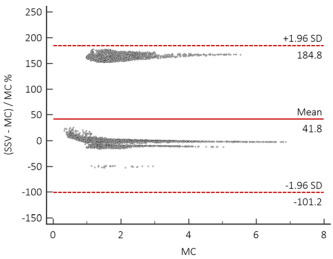


- DNN **dose maps** obtained by PET convolution were compared to MC, but also to single and multiple S-value approaches (SSV and MSV).
- SSV is very different and probably not worth the comparison.
- The DNN **execution time** was over 10^3 times shorter than MC.

Comparison with Monte Carlo

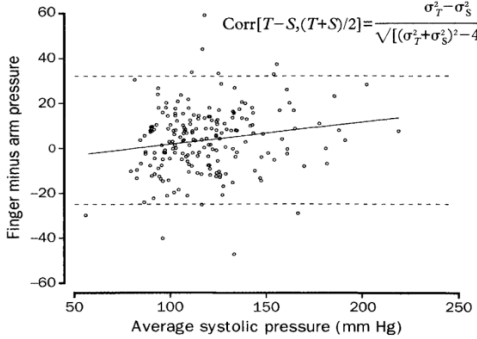


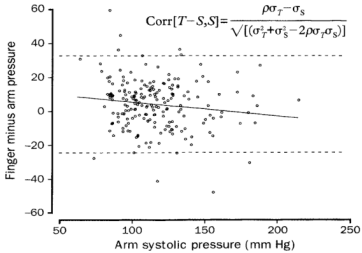


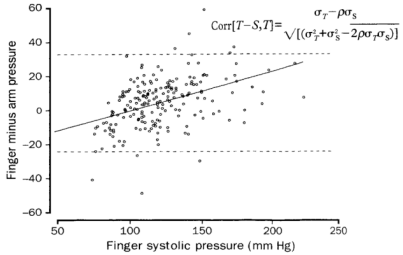


- In B&A analysis, the **smallest bias and CI** were given by DNN.
- The points out of the CI were identified as negligible boundary effects.
- Instead of the mean, MC was used in the abscissa: in literature this practice is marked as *misleading*, as it may create itself a dependence from the magnitude.

Comparison with Monte Carlo







In fact, the **correlation between differences and magnitude** can completely change depending on which method is chosen as the standard [Bland and Altman (1995)].

Overview

Research goals

In general, we can say that the DNN mixed kernels:

- bring an **improvement** towards the MC gold standard, with respect to MSV;
- require $< 0.1\%$ of MC **execution time**;
- need a **much smaller training dataset** from medical imaging than the DNN approach without kernels, as well as a **shorter training time**;
- imply a training which is only **isotope-specific** and not prone to any bias caused by the radiopharmaceutical biodistribution.

Statistical review

In general, the authors made **quite good use of B&A analysis**.

- ✓ DNN was correctly proved to be better than MSV.
- ✓ The data beyond the CI were verified to be negligible.
- ✗ Not putting the mean in the abscissa prevented from understanding a possible difference/magnitude relation.

Finally, one may expect the DNN approach without kernels to be more precise for long-range radiation, but:

- such a comparison is still missing in literature;
- the authors who used that method did not perform B&A analysis.

Thank you for your kind attention!

References I

Azadeh Akhavanallaf, Iscaac Shiri, Hossein Arabi, and Habib Zaidi. Whole-body voxel-based dosimetry using deep learning. *European Journal of Nuclear Medicine and Molecular Imaging*, 48:670–682, 2021.

J. Martin Bland and Douglas G. Altman. Comparing methods of measurement: why plotting difference against standard method is misleading. *The Lancet*, 346:1085–1087, 1995.

Wesley E. Bolch, Keith F. Eckerman, George Sgouros, and Stephen R. Thomas. Mird pamphlet no. 21: A generalized schema for radiopharmaceutical dosimetry—standardization of nomenclature. *The Journal of Nuclear Medicine*, 50(3):477–484, 2009.

Davide Giavarina. Understanding Bland Altman analysis. *Biochemia Medica*, 25(2): 141–151, 2015.

References II

Min Sun Lee, Donghwi Hwang, Joong Hyun Kim, and Jae Sung Lee. Deep-dose: a voxel dose estimation method using deep convolutional neural network for personalized internal dosimetry. *Scientific Reports*, 9:10308, 2019.