PAPER • OPEN ACCESS

A framework for prediction of personalized pediatric nuclear medical dosimetry based on machine learning and Monte Carlo techniques

IOP Publishing Phys. Med. Biol. 68 (2023) 084004 https://doi.org/10.1088/1361-6560/acc4a5

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
21 September 2022

REVISED
24 January 2023

ACCEPTED FOR PUBLICATION
15 March 2023

PUBLISHED
7 April 2023

PAPER

A framework for prediction of personalized pediatric nuclear medical dosimetry based on machine learning and Monte Carlo techniques

Vasileios Eleftheriadis[1], **Georgios Savvidis**[1], **Valentina Paneta**[1], **Konstantinos Chatzipapas**[2], **George C Kagadis**[2] and **Panagiotis Papadimitroulas**[1,2,*]

[1] BIOEMTECH, Mesogeion Av. 387, 15343, Athens, Greece
[2] Department of Medical Physics, University of Patras, 26504, Rion, Greece
[*] Author to whom any correspondence should be addressed.

E-mail: Papadimitroulaspanpap@bioemtech.com

Keywords: personalized dosimetry, GATE, MC simulations, pediatric dosimetry, machine learning, ensemble learning, prediction model

Supplementary material for this article is available online

Abstract

Objective: A methodology is introduced for the development of an internal dosimetry prediction toolkit for nuclear medical pediatric applications. The proposed study exploits Artificial Intelligence techniques using Monte Carlo simulations as ground truth for accurate prediction of absorbed doses per organ prior to the imaging acquisition considering only personalized anatomical characteristics of any new pediatric patient. *Approach:* GATE Monte Carlo simulations were performed using a population of computational pediatric models to calculate the specific absorbed dose rates (SADRs) in several organs. A simulated dosimetry database was developed for 28 pediatric phantoms (age range 2–17 years old, both genders) and 5 different radiopharmaceuticals. Machine Learning regression models were trained on the produced simulated dataset, with leave one out cross validation for the prediction model evaluation. Hyperparameter optimization and ensemble learning techniques for a variation of input features were applied for achieving the best predictive power, leading to the development of a SADR prediction toolkit for any new pediatric patient for the studied organs and radiopharmaceuticals. *Main results*. SADR values for 30 organs of interest were calculated via Monte Carlo simulations for 28 pediatric phantoms for the cases of five radiopharmaceuticals. The relative percentage uncertainty in the extracted dose values per organ was lower than 2.7%. An internal dosimetry prediction toolkit which can accurately predict SADRs in 30 organs for five different radiopharmaceuticals, with mean absolute percentage error on the level of 8% was developed, with specific focus on pediatric patients, by using Machine Learning regression algorithms, Single or Multiple organ training and Artificial Intelligence ensemble techniques. Significance: A large simulated dosimetry database was developed and utilized for the training of Machine Learning models. The developed predictive models provide very fast results (<2 s) with an accuracy >90% with respect to the ground truth of Monte Carlo, considering personalized anatomical characteristics and the biodistribution of each radiopharmaceutical. The proposed method is applicable to other medical dosimetry applications in different patients' populations.

List of abbreviations

AI	Artificial Intelligence
MC	Monte Carlo
SADR	Specific Absorbed Dose Rate
ML	Machine Learning
LOOCV	Leave One Out Cross Validation
MAE	Mean Absolute Error

MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
NM	Nuclear Medicine
DNN	Deep Neural Network
EMD	Empirical Mode Decomposition
SPECT	Single Photon Emission Computed Tomography
PET	Positron Emission Tomography
CT	Computed Tomography
HPC	High-performance computing

1. Introduction

Personalized internal dosimetry is of high interest in pediatric diagnostic and therapeutic applications involving ionizing radiation from radiopharmaceuticals (Khong *et al* 2013, Papadimitroulas *et al* 2019). Young patients provide a higher risk of stochastic effects under the radiation exposure from nuclear medicine (NM) procedures (Robbins 2008, Adelstein 2014, Treves *et al* 2014).

Modern medicine exploits advanced computational tools for assessing absorbed dose in organs of interest. To this basis, Monte Carlo (MC) simulations combined with detailed digital anthropomorphic models (Akhavanallaf *et al* 2022) are considered gold standard (Sarrut *et al* 2014). The well established MIRD dosimetry protocol considers patients' variability using interpolated *S*-values based on pre-defined calculations and mass correction (Bolch *et al* 2009).

The extensive development of Artificial Intelligence (AI) over the last decade, paired with the vast volume of data generated in healthcare systems has spiked the interest of both researchers and healthcare practitioners over its possible applications in medicine. This has led in an increase in AI applications in medical physics, including NM (Nensa *et al* 2019). The main applications of AI in molecular radiotherapy and internal radiation dosimetry are organ and tumour segmentation and classification, therapeutic dose calculation and internal dose prediction (Arabi and Zaidi 2020).

In NM therapy, internal dosimetry is the key to successful personalized treatment, since the risk of radiation-induced toxicity can be significantly reduced by patient-individualized dose calculations (Stabin et al 2019). Even though, MC simulations for voxel-based dosimetry are considered the gold standard for dosimetry in personalized therapy, they have not been applied in clinical use, due to the excessive computational cost and computing time that they require (Zaidi 1999). On the other hand, AI can quickly process and analyse large amounts of data. Once training is completed, AI can usually provide accurate results on specific tasks significantly faster than traditional methods like MC. In order to get the best out of these two techniques, several internal dose prediction studies have used MC simulations as ground truth in order to train ML, e.g. deep neural network (DNN), prediction models.

To overcome the limitations of the direct MC approach, Götz *et al* (Götz *et al* 2020) used a hybrid method based on a U-net DNN architecture in combination with empirical mode decomposition (EMD) techniques in conjunction with soft tissue kernel MC simulations to achieve a dose map of patients who had undergone ^{177}Lu-PSMA therapy. The system was trained using SPECT and CT from a patient cohort of 26 subjects as input and individual full MC simulation results as reference. The DNN-EMD hybrid method for internal dose prediction yielded superior results compared to the MIRD protocol with soft tissue DVK dose calculation method.

Lee *et al* (Lee *et al* 2019) proposed a voxel dose estimation method using dynamic PET/CT image patches of 10 patients as input and MC simulated dose rate maps as ground truth for the training of a 3D U-net CNN. The dose rate map obtained by this method agreed well with the ground truth with voxel dose rate errors of $2.54\% \pm 2.09\%$. The CNN-based method outperformed traditional personalized internal dosimetry approaches and showed results comparable with that of the direct MC simulation, but on notably less computing time since single dose rate maps were generated in less than 4 min using the trained CNN network, while the direct MC simulation took around 235 h to generate the single dose rate maps (Lee *et al* 2019).

Akhavanallaf *et al* (Akhavanallaf *et al* 2021) suggested a novel methodology for personalized organ-level, whole-body, voxel-based internal dosimetry using a ResNet composed of 20 convolutional layers. The DNN was trained using density maps generated by 24 CT images as input and considering the heterogeneity of activity distribution, non-uniformity of surrounding medium, and patient-specific anatomy. Voxelwise *S*-values generated using MC simulations were considered as ground truth. The DNN outperformed conventional

voxel-level and organ-level MIRD-based approaches, exhibiting performance comparable to the direct MC approach, having a mean relative absolute error of 4.5% ± 1.8%, while the computation time for building a whole-body voxel dose map was less than 0.1% of the time required for direct MC simulations.

In this context, it is essential to integrate modern AI models with the gold standard provided by MC to accurately assess the internal dosimetry (at organ level) for NM procedures performed on children. We propose a prediction framework for calculating the absorbed dose per organ that considers each pediatric patient's specific anatomy. More specifically, our aim is to train ML algorithms for predicting absorbed doses per organ based on the ground truth of dosimetry (pre-calculated through MC simulations). With this approach, we overcome the current procedure that the doses are calculated on predefined *S*-values and rescaling the organs. The idea is based on the prediction of absorbed doses per organ, considering different anatomical characteristics from the basis of the calculation that is done through MC. Our long-term goal is to extend the proposed method for other patient populations (i.e. adults, obese patients) and incorporate a large list of commonly used radiopharmaceuticals.

2. Methods

MC simulations were performed using the GATE MC toolkit for a population of computational pediatric models to calculate the specific absorbed dose rates (SADRs) in several organs and radiopharmaceuticals. The produced database will serve as training data for the development of a prediction toolkit based on SADRs for any new pediatric patient for the studied organs and radiopharmaceuticals.

2.1. Dosimetry-SADRs

In this work, we implement the method for calculating the SADRs which has been established by our group in a previous work (Papadimitroulas *et al* 2018). In this approach, the calculation of the absorbed dose per organ takes into account each patient's specific anatomy and estimates SADRs for each organ according to the specified clinical biodistribution of administered radiopharmaceutical throughout the whole body. SADRs (Gy/Mbq/sec) provide the instantaneous absorbed dose rate in a target organ from the activity of all organs of the patient, based on a specific biodistribution defined at time t_k:

$$\mathrm{SADR}(r_T \leftarrow r_{WB},\, t_k) = \frac{1}{m_{rT}} \sum_i E_{di}\, Y_{i,\,tk}, \tag{1}$$

where r_{WB} is whole-body source, E_{di} is the energy of the ith radiation per disintegration deposited in target organ r_T and m_{rT} is the mass of the target organ, while Y_i represents the yield per disintegration on the t_k biodistribution. The absorbed dose to a target organ through NM examination ($t_D = t_{\mathrm{final}} - t_0$) is given by the following equation (2):

$$D(r_T,\, t_D) = \int_{t_k=0}^{t_k=\mathrm{final}} A(r_{WB},\, t_D)\,\mathrm{SADR}(r_T \leftarrow r_{WB},\, t_D)\, dt, \tag{2}$$

where $A(r_{WB}, t_D)$ is the instantaneous whole-body activity at each post-administration time-point t_K. Based on the radiopharmaceutical t_k biodistribution and the duration (t_D) of the activity within the body, the integration of the SADRs for each target organ, on several times (t_k) of the radiopharmaceutical biodistributions, calculates the cumulative absorbed dose.

2.2. MC simulations

2.2.1. Pediatric population

For our purpose, a population of 28 pediatric computational phantoms was used for the development of the simulated dosimetry database. The pediatric phantom population consisted of male and female phantoms of varying ages and anatomical characteristics, such as mass and height. Indicatively, 22 of the phantoms were derived from the 4D pediatric XCAT (Segars *et al* 2015) reference models and 6 were based on the IT'IS Virtual Family models (Christ *et al* 2010). The characteristics of the pediatric phantoms are illustrated in table 1, while the voxel resolution of each phantom was set to $2 \times 2 \times 2\ \mathrm{mm}^3$.

The computational phantoms imported in GATE served both as radiation transport media and activity maps (identical voxel size of $2 \times 2 \times 2\ \mathrm{mm}^3$). In GateMaterials.db file all the materials used during the simulations were predefined, since the transport media in GATE take into consideration both the density and the elemental composition of each organ. Table 1 presents the characteristics of the pediatric population, while the density of the organs of interest is presented in table S1 of the supplementary material 'Supplementary data'.

Table 1. Characteristics of the pediatric phantoms used in the GATE simulations (voxel size of $2 \times 2 \times 2$ mm^3).

No. (#)	Age (year)	Gender	Type	Size, number of voxels	Weight (Kg)	Total height (m)
Phantom 1	15	Male	ITIS	242 × 156 × 863	50.4	1.7
Phantom 2	5	Female	ITIS	175 × 102 × 551	17.7	1.09
Phantom 3	6	Male	ITIS	202 × 113 × 591	18.6	1.16
Phantom 4	8	Female	ITIS	300 × 122 × 804	29.64	1.36
Phantom 5	8	Male	ITIS	218 × 126 × 1286	25.6	1.37
Phantom 6	11	Female	ITIS	250 × 140 × 780	34	1.49
Phantom 7	17.2	Male	XCAT	339 × 175 × 929	86.2	1.83
Phantom 8	15	Male	XCAT	295 × 155 × 845	58	1.66
Phantom 9	2.1	Female	XCAT	165 × 115 × 455	12.2	0.86
Phantom 10	2.8	Female	XCAT	205 × 125 × 475	14.9	0.92
Phantom 11	3.3	Female	XCAT	175 × 115 × 475	13.8	0.93
Phantom 12	5	Female	XCAT	207 × 145 × 585	19.9	1.13
Phantom 13	5.2	Female	XCAT	179 × 129 × 555	15.3	1.07
Phantom 14	13.8	Male	XCAT	325 × 165 × 915	67.4	1.79
Phantom 15	9.8	Female	XCAT	263 × 135 × 655	40.7	1.27
Phantom 16	10	Female	XCAT	255 × 125 × 715	33	1.39
Phantom 17	12.1	Female	XCAT	245 × 135 × 725	38.6	1.41
Phantom 18	14.3	Female	XCAT	305 × 165 × 855	73.2	1.68
Phantom 19	15	Female	XCAT	295 × 155 × 825	58	1.61
Phantom 20	16.8	Female	XCAT	285 × 155 × 805	50.5	1.57
Phantom 21	2.8	Male	XCAT	205 × 125 x 475	14.1	0.92
Phantom 22	3.7	Male	XCAT	185 × 115 × 505	16.2	0.97
Phantom 23	5	Male	XCAT	206 × 145 × 577	19.9	1.12
Phantom 24	5.3	Male	XCAT	215 × 115 × 575	22.7	1.1
Phantom 25	7.8	Male	XCAT	225 × 125 × 635	24.4	1.25
Phantom 26	9.6	Male	XCAT	245 × 135 × 709	33.9	1.39
Phantom 27	10	Male	XCAT	255 × 125 × 709	33	1.39
Phantom 28	12	Male	XCAT	255 × 145 × 765	43.5	1.5

2.2.2. GATE toolkit

The GATE MC toolkit (Jan *et al* 2004, Jan *et al* 2011, Sarrut *et al* 2022) was used for the development of the dosimetry database. GATE is based on the Geant4 code (Agostinelli *et al* 2003, Allison *et al* 2016) and is widely used and well validated for dosimetry applications (Papadimitroulas 2017, Sarrut *et al* 2014). Specifically, GATE v9.1 was used for the execution of the simulations. The 'standard model' (emstandard_opt3) which is appropriate for such electromagnetic processes is used in our GATE environment. As far as the method for calculating the absorbed dose per organ is concerned, the 'dose actor' tool was used, for scoring the energy deposition. The dose actor creates three-dimensional (3D) dose maps of the deposited energy and the absorbed dose at all organs of the phantoms with a specified voxel resolution. The dose actor takes into consideration the total energy and the interaction probability of the particles, as well as the density of each voxel.

The voxelized phantoms were imported in GATE using the 'ImageNestedParametrisedVolume' technique. This approach is based on a parameterized method which allows GATE to store a single voxel depiction in memory, changing its composition and location during the run of the simulation. Lastly, the 'ion' source type of Geant4 was used for the initialization of the primary particles. This is the most realistic and accurate way of simulating a radionuclide and incorporates both radioactive decay and atomic de-excitation. In our case the ^{131}I, ^{123}I and ^{153}Sm ion sources were used, while in the case of ^{99m}Tc we implemented the 'user spectrum', in which the user specifies the energy of the particles accompanied with their probability weight. Special reference for the used radioisotopes is presented in the following paragraphs.

All the dosimetry simulations were executed with 10^8 primaries. In order to accelerate the procedure, the ensemble of simulations was performed on a high-performance computing (HPC) center. Recently, the HPC advantages in the medical field and specifically in our pediatric internal dosimetry application have been reported (Koch *et al* 2023). This way, the simulations' execution time was reduced significantly, since 112 jobs were running in parallel, achieving low statistical uncertainty and demanding fewer memory consumption. The HPC consisted of nodes that each one included 28-Core Intel Broadwell CPUs and 512 GB of memory. These characteristics accelerate approximately ~100 times the simulations' total execution time in contrast to a typical 24 GB memory PC. Statistical uncertainty was calculated according to Chetty *et al* (Chetty *et al* 2006), with the following formula (3) that defines statistical uncertainty ε_k at voxel k, with N being the number of primary events and $d_{k,i}$ the deposited energy in voxel k for primary event i:

$$D_k = \sum_i^N d_{k,i},\ S_k = \sqrt{\frac{1}{N-1}\left(\frac{\sum_i^n d_{k,i}^2}{N} - \left(\frac{\sum_i^n d_{k,i}}{N}\right)^2\right)},\ \varepsilon_k = 100 \times \frac{S_k}{D_k}. \quad (3)$$

2.2.3. Radiopharmaceuticals used

The proposed methodology derived from the exploitation of the SADRs, considering the radioactivity distributed throughout the whole body (i.e. the organ's own radioactivity as well as the radioactive contribution from all the other organs), in order to calculate the total absorbed dose per organ. The biodistribution used as activity map for each one of the radiopharmaceuticals (^{99m}Tc-MDP, ^{123}I-MIBG, ^{131}I-INa, ^{131}I-MIBG and ^{153}Sm-EDTMP) derived from the study of Papadimitroulas *et al* (Papadimitroulas *et al* 2018). Time activity curves were used to simulate 4 different time points of the biodistribution in each radiopharmaceutical, except ^{131}I-INa, where 5 different time points were considered (due to the slow washout of the radiopharmaceutical). These biodistributions for each radiopharmaceutical are included in the 'Simulated_dosimetry_database' file in the supplementary material.

2.3. AI techniques

In this part of the study, we focus on the development of an internal dosimetry prediction toolkit, based on Machine Learning regression algorithms and AI ensemble techniques, which can accurately predict SADRs for pediatric patients per studied organ and radiopharmaceutical. The training and evaluation of the prediction models was performed using the simulated SADR database described in 2.1 and 2.2.

2.3.1. Machine learning regression algorithms

Eight supervised ML regression algorithms were evaluated: Least Squares Linear Regression (Lai *et al* 1979), Ridge Regression (Hoerl and Kennard 1970), AdaBoost (Schapire 2013) regressor, Gradient Boost (Friedman 2001) regressor, XGBoost (eXtreme Gradient Boosting) (Chen and Guestrin 2016) regressor, Random Forest (Breiman 2001) regressor, Decision Tree (Quinlan 1986) regressor and Support Vector Regressor (Awad and Khanna 2015). All the ML algorithm implementations used in this study can be found in the open-source software ML library Scikit-learn (Pedregosa *et al* 2011), except for XGBoost's implementation which can be found in the XGBoost[3] open-source software library.

2.3.2. Training procedure for a dosimetry prediction model

In order to train ML models to predict SADR values of a pediatric patient (target value) for each target organ, over time, for each of the 5 radiopharmaceuticals, we reshaped the simulated dataset as sets of input feature values (rows) that correspond to each target value. A row of input feature values will be referred as a snapshot. Our dataset consists of ∼3000 snapshots per radiopharmaceutical.

The set of input features consists of:

(a) the personalized anatomical characteristics of the phantoms,

(b) the specific organ of interest and

(c) the time point for each target value (SARD) for the specific radiopharmaceutical.

The input features, along with their assigned index, are listed in figure 1.

Since the tested radiopharmaceuticals display varying absorbed dose rate behaviour on the target organs over time, separate prediction models were trained for each radiopharmaceutical. Moreover, because anatomical characteristics measurements, such as Lung (total z-height of the lungs), Sitting height and Effective Diameter (as defined in Boone *et al* 2011) may not be as easily accessible to practitioners as the rest, we decided to also create different models according to the different combinations of available anatomical characteristics. In this regard, we include the first 7 features ('Organ', 'Time', 'Age (year)', 'Gender', 'Weight (Kg)', 'Total height (m)', 'BMI (kg/m^{-2})') in all feature combinations and added to these, all 7 possible combinations of the last 3 features ('Sitting height (cm)', 'Lung (cm)', 'Eff. diameter (cm)'), ending up with 8 feature combinations. Furthermore, we tested and evaluated the predictive accuracy of the ML algorithms, when a model was trained on all the available organs (multiple organs training) in the database versus when we trained separate models for each organ (single organ training). A schematic representation of all the combinations that were investigated with AI techniques among radiopharmaceuticals, features, algorithms, and model training procedure is seen in figure 2.

[3] https://xgboost.readthedocs.io/en/latest/index.html

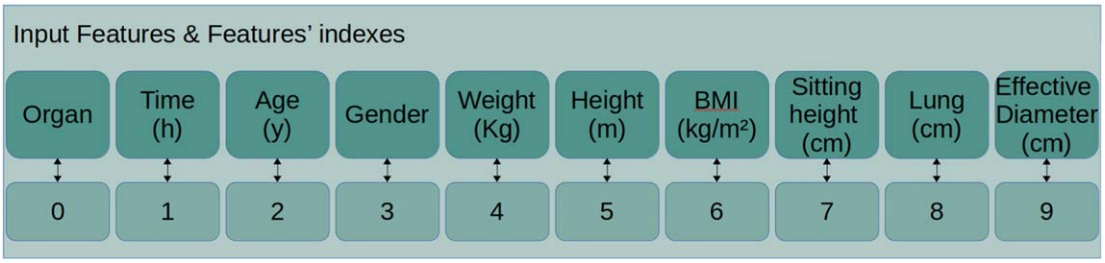

Figure 1. Feature list used as input to the prediction models. An index was assigned to each feature, for easier presentation of the input features' combinations.

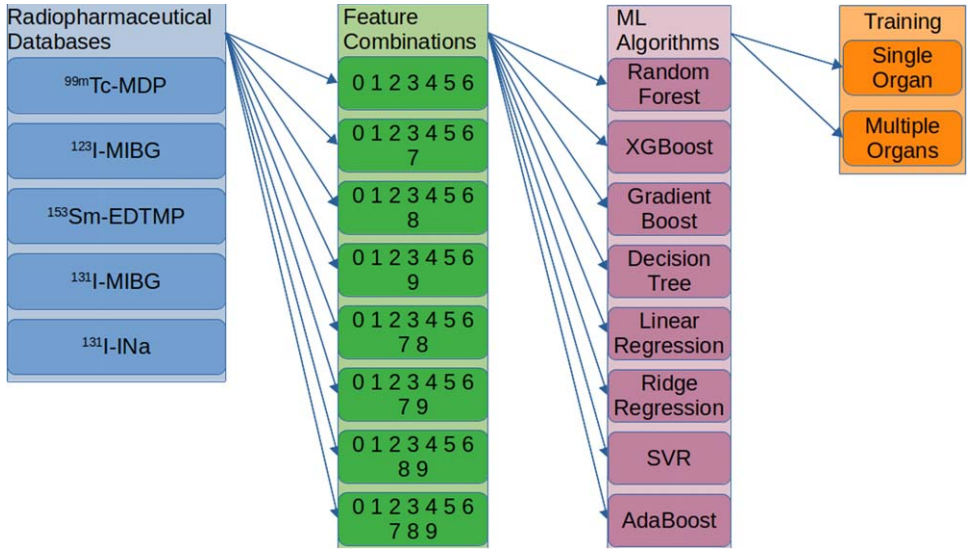

Figure 2. Schematic representation of all the ML training combinations.

The training method, on single or multiple organs, which yielded better performance, according to the metrics described in section 2.3.3, was chosen as the final predictive model for each feature combination, ML algorithm and radiopharmaceutical. By this point, it was clear that 4 algorithms (Random Forest, XGBoost, Gradient Boost and Decision Tree) were performing better than the rest, thus Hyper-parameter optimization was performed only on those.

2.3.2.1. Hyper-parameter optimization

Hyper-parameter optimization or tuning is the process of finding a set of hyper-parameter values which allows an ML algorithm to better fit the data, achieving the best possible performance according to a predefined metric (MAE in this case), on a cross validation set. Hyper-parameter optimization plays a vital role in the prediction accuracy of ML algorithms[4]. Bayesian optimization (Wu *et al* 2019) was selected due to its ability to achieve comparable improvement of the predictive performance of ML algorithms in significantly reduced computing time compared to other optimization methods, setting a prior distribution over the optimization function and updating its posterior gathering information from the previous sample.

2.3.2.2. Ensemble learning models

Ensemble learning (Dietterich 2000) refers to the process of developing a single 'strong' ML model that solves a computational problem by strategically combining multiple differently performing 'weaker' ML models, treating them as a 'committee' of solvers. The principle is that the prediction of the committee, when individual predictions are combined appropriately, should have better overall accuracy than any individual model (committee member).

[4] https://scikit-optimize.github.io/stable/

After the completion of the Hyper-parameter optimization process, we used the outputs of the 4 best performing models (Random Forest, XGBoost, Gradient Boost and Decision Tree) to create weighted average ensemble learning models.

Weighted average or weighted sum ensemble (Shahhosseini *et al* 2022) is an ensemble learning approach that combines the predictions from multiple models, where the contribution of each model is weighted proportionally to the model's predictive ability.

In weighted average ensembles, a weight is assigned to each contributing model. That weight is then multiplied by the model's prediction and is used for the calculation of the average prediction. In regression, the average prediction is calculated using the arithmetic mean, as shown in following equation:

$$P_e = \frac{\sum_{i=1}^{n} w_i \times P_i}{\sum_{i=1}^{n} w_i}, \tag{4}$$

where: P_e is the prediction of the ensemble

n is the total number of predictors contributing to the ensemble

P_i is the prediction of predictor i

w_i is the weight assigned to predictor i

To search for optimal model weights that result in improved performance comparing to any individual contributing model, we used a linear exhaustive approach. Integer weights ranging from 0 to 4 were assigned to each of the Random Forest, XGBoost, Gradient Boost models and from 0 to 2 for the Decision Tree models, producing 375 ensembles for each feature combination and each radiopharmaceutical.

2.3.2.3. Cross validation

The leave one out cross validation (LOOCV) (Sammut and Webb 2011) method was used to train and validate the models. The main reason the LOOCV method was selected for this study was due to the limited number ($n = 28$) of pediatric phantoms. The LOOCV method allows for the use of more data on the training of the models than any other training and validation method. According to the LOOCV method, the data is divided into two separate sets, a training and a validation set. The training set consists of snapshots of all the pediatric phantoms, apart from the snapshots of the one phantom which incorporates the validation set of each training iteration. So, the snapshots of one phantom are used for validation, and the rest of the dataset is used for the training of the model. This training and validation process will be repeated as many times as the total number of phantoms. The validation set's feature values of each snapshot are then entered as input to the trained model, which returns its prediction of the SARDs (target value) of the snapshots. This way we end up having a SARD prediction for each time point and organ for all 28 phantoms for validation purposes.

2.3.2.4. Metrics

To assess the predictive power of the ML models and ensembles, we computed the following performance measures using LOOCV:

1. Mean absolute error (MAE) is the average of the absolute errors of the model's predictions against the target values.
2. Root mean square error (RMSE) is the square root of the average of the squared errors of the model's predictions against the target values.
3. R-squared (R^2) or coefficient of determination represents the proportion of the variance of the target value that is explained by the input features in a regression model. R-squared values range from 0 to 1, with larger R^2 values indicating better fit of the data.
4. Mean absolute percentage error (MAPE) is the average of the absolute error percentage of the model's predictions against the target values and is a relative measure that essentially scales MAE to be in percentage units instead of the target value's units.

MAE and RMSE are scale dependent, so they can be used to compare the performance of different predictive regression models for a particular dataset but not between datasets (Hyndman and Koehler 2006). Smaller MAE and/or RMSE values indicate better predictive performance. Since, according to literature (Willmott and Matsuura 2005), MAE is the more natural measure of average error magnitude, and that, unlike RMSE, it is unambiguous, it was used as the primary model performance measure in this study for performance assessment and optimization purposes. For the presentation of the results, although, MAPE was preferred because it is straightforward and easier to interpret that other metrics, like MAE and RMSE, as it provides the error in terms of percentages.

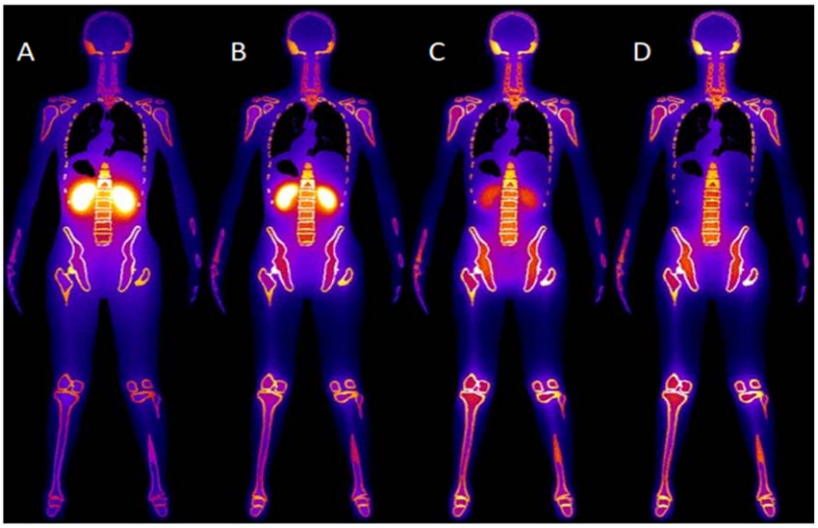

Figure 3. Dose maps of ^{99m}Tc-MDP radiopharmaceutical of a 15 year old female phantom at 4 different time points. (A) $T = 0$ h, (B) 1.42 h, (C) 4.11 h, (D) $T = 20.2$ h.

3. Results

3.1. Simulated dosimetry database

Dose rates for the 30 different organs of each 28 computational pediatric phantom were estimated through MC simulation. The output of the GATE toolkit is a 3D dose map of the anthropomorphic paediatric phantoms reflecting the amount of dose deposited at each organ. Figure 3 illustrates the dose deposition at 4 different time periods for a 15 year old female phantom for the case of ^{99m}Tc-MDP. At this example, the concept of bone scintigraphy is depicted, since ^{99m}Tc-MDP's main application concerns diagnostic purposes. The radiation was mainly stored at bones during the examination while much activity and consequently dose was collected at the bladder which presents an attenuation especially at the latest time point.

As a next step, we extract the dose maps and implement the SADR approach in the simulated outputs. The relative percentage statistical uncertainty, in the calculated dose values per organ, fluctuated between 0.05% and 2.7%, with a median value of 0.11%. The extracted absorbed dose rates presented large variation for the same organ on different phantoms up to ~70%.

Figure 4 presents indicative dose rate results for the case of ^{99m}Tc-MDP, illustrating the highest and lowest SADR values per organ that correspond to the youngest and oldest phantoms respectively. Each figure corresponds to different time point calculations while 8 of the most significant organs are presented. The same figure for 10 phantoms of various ages used in the present study is included in the supplementary material figure S1 for accessing SADR values across all phantoms too.

Dose rates distribute to the studied organs progressively (figure 4) and as is seen in figure S1 they present a similar pattern for phantoms with small age variation regardless the gender. At the supplementary material 'Simulated_dosimetry_database', the complete simulated dosimetry database of this study is presented concerning all the radiopharmaceuticals used in our study for each time point.

3.2. Prediction model performance

In this section we evaluate, using LOOCV, the predictive power of the ML and ensemble models that were developed during this study, for predicting SADR values of pediatric patients for 30 different organs of interest, over time, for each of the 5 radiopharmaceuticals, using as input features the personalised anatomical characteristics of the phantoms, the specific organ and time point.

After the development of all individual ML and ensemble models we compared their performance based on the metrics described in 2.3.3 and selected the final predictive model for each combination of the 5 radiopharmaceuticals and the 8 input feature combinations. So, for each radiopharmaceutical, different predictive models will be applied according to the available features. The evaluation metrics of the best performing models for each radiopharmaceutical (among all feature combinations and studied organs) ranged to the values presented in table S2 of the supplementary material 'Supplementary data' and were found to be consistently good.

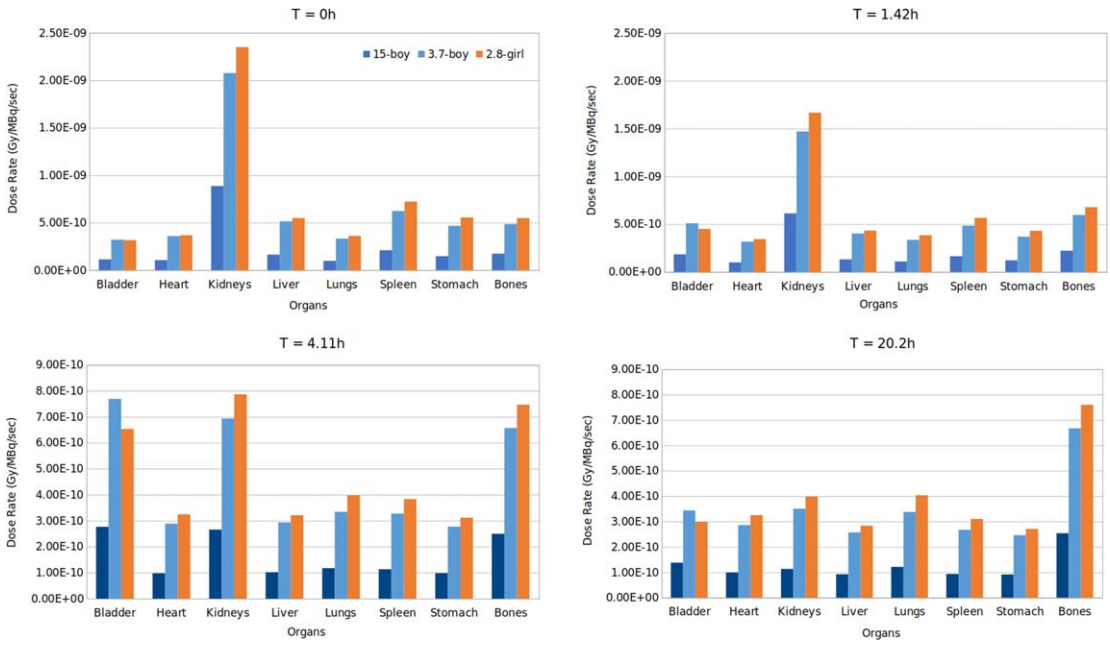

Figure 4. The lowest and highest dose rate (SADRs) results in several organs for the case of ^{99m}Tc-MDP, at 4 different times.

Indicative comparisons between the ML toolkit's predictions and the MC simulated SADR values in 8 significant organs for the cases of ^{99m}Tc-MDP and ^{123}I-MIBG during the 2nd and 3rd time points respectively, for 2 pediatric phantoms, a 14.3 year old female and a 5 year old male are illustrated in figure 5. The predicted values in the figure exhibit very good agreement with MC simulations ground truth, as seen by the indicated percentage differences. The distributions of mean absolute percentage errors (MAPE) of the best performing ML or ensemble models for each radiopharmaceutical and all organs are presented with a boxplot in figure 6, illustrating errors being around ~8% for all radiopharmaceuticals. The distribution across 3 age groups of MAPE values of the best performing models per organ in the case of ^{153}Sm-EDTMP are presented in boxplots in figure 7, showing slight variations.

Indicative metrics comparison between the best performing ensemble and individual ML models for the cases of ^{131}I-MIBG and ^{99m}Tc-MDP are presented in figure 8, with error values being normalised to the highest ones (worse performance) for MAE, MAPE and RMSE, depicting the resulted performance improvement via ensemble method.

Moreover, boxplots illustrating MAPE values of the best performing model of each radiopharmaceutical and organ, across all time points, are presented in figures S2–S6 of the supplementary material found in 'Supplementary data', for the model evaluation over time.

3.2.1. Computing time

The execution of the MC dosimetry simulation for one phantom and one radiopharmaceutical, took approximately 28.0 h on a system equipped with an AMD® Ryzen 9 5900x with 24 × 12-core processors and 32 GB of RAM. The development of the internal dosimetry prediction ML toolkit, for one radiopharmaceutical, including the training and evaluation process of all ML models, Hyperparameter optimization, and generating all ensembles, for all the combinations of input features, took similarly 23.3 h on the same system. However, this development procedure is performed once. Thereafter, ML predictions of SADR values for all organs using the developed ML toolkit, can be generated in under just 2 s for each pediatric patient on the same system. Table 2 summarises the computation time of each procedure required for the ML prediction and the MC calculation of the SADRs of a pediatric patient.

3.3. Evaluation of the prediction model

We evaluated the proposed methodology with the ground truth of dosimetry calculated by direct MC simulations as well as, with the well-validated and standardized MIRD schema in terms of absorbed doses per organ in mGy. More precisely we considered a pediatric computational model (Phantom 8: 15 year old boy, 58 kg) and performed a complete MC simulation in HPC for achieving low statistical uncertainty, for an acquisition of 20.2 h and an activity of 370 MBq. With such realistic simulations the absorbed doses per organ were extracted using the 'dose actors' provided by GATE (GATE Direct MC). In addition, we used the predicted SADRs in our

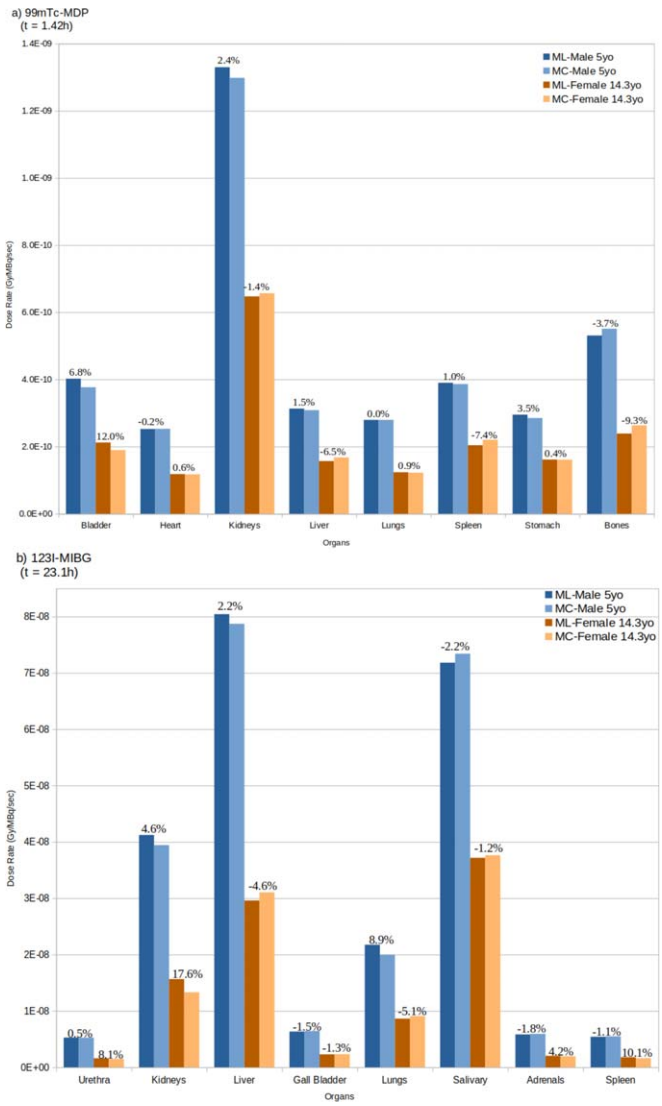

Figure 5. Comparison between ML dose rate predictions to MC dose rate calculations (difference in %) in significant organs for 2 indicative pediatric phantoms for the cases of (a) ^{99m}Tc-MDP during the 2nd time point ($t = 1.42$ h) and (b) ^{123}I-MIBG during the 3rd time point ($t = 23.1$ h).

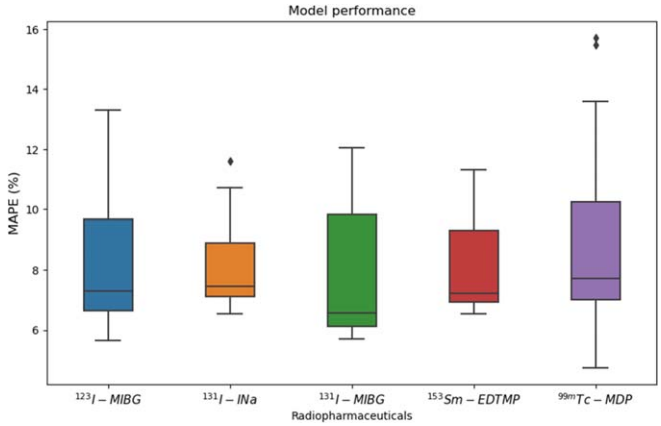

Figure 6. Boxplot of mean absolute percentage error (MAPE) of the best performing models for all studied radiopharmaceuticals and organs.

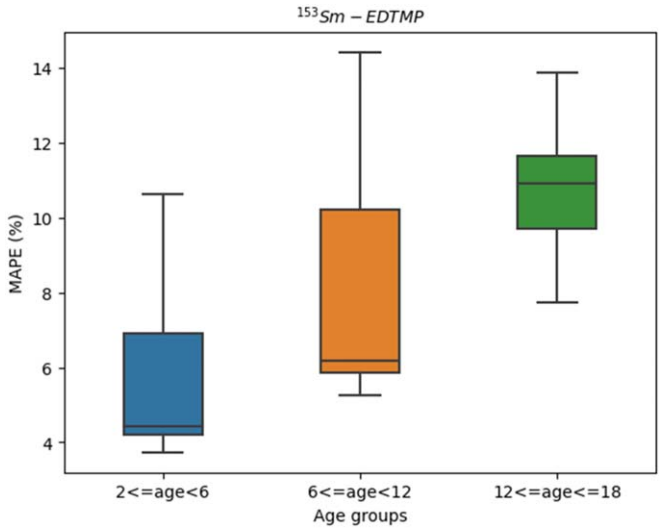

Figure 7. Indicative boxplot of mean absolute percentage error (MAPE) of best performing models regarding ^{153}Sm-EDTMP and all studied organs, for 3 different age groups.

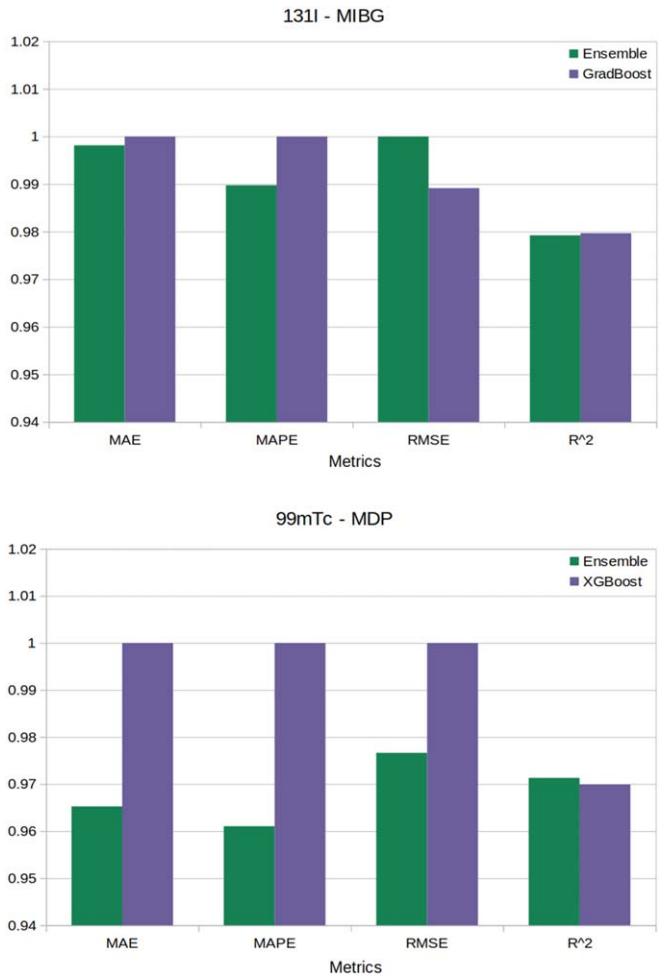

Figure 8. Metrics comparison between the best performing ensemble and individual ML models of ^{131}I-MIBG (top) and ^{99m}Tc-MDP radiopharmaceuticals (bottom).

final AI model, using the input features. Phantom 8 was considered as a totally new patient, meaning that we used the model, which was trained during LOOCV, with Phantom 8 being the validation set. The predicted SADRs were multiplied with the whole-body activity at each specific time point and the absorbed dose was

Table 2. Computing time of MC simulations for 1 phantom, ML algorithms training, optimization and ensembles generation, and prediction of SADR values of all the organs for the radiopharmaceutical ^{123}I-MIBG.

Procedure	Computing time
MC: SADR calculation	28.0 h
ML: training + optimizing	23.3 h
ML: SADR Prediction	<2 s

Table 3. Comparison of organ absorbed doses for radiopharmaceuticals ^{99m}Tc-MDP for 8 target organs calculated using MIRDcalc (Boone *et al* 2011), direct Monte Carlo method and AI-based method.

Organ of interest	Absorbed dose (mGy)			Percentage difference (%)	
	GATE Direct MC	MIRD *S*-values	AI SADRs	AI versus MC	AI versus MIRD
Brain	1.41	3.28	1.36	3.5	58.5
Kidneys	4.50	5.15	4.71	4.7	8.5
Liver	1.37	1.21	1.52	10.9	25.6
Spleen	1.58	1.61	1.82	15.2	13.0
Bladder	2.68	2.74	2.92	9.0	6.6
Stomach	1.30	1.26	1.49	14.6	18.3
Pancreas	1.85	2.03	1.99	7.6	2.0
Rest of Body	0.92	1.76	0.98	6.5	44.3

integrated in time (AI SADRs). Finally, the MIRDcalc program[5] was used in order to extract absorbed doses per organ, after correcting the masses of the organs according to Phantom 8. The absorbed doses are presented in the column of MIRD *S*-values in table 3. The percentage difference of our method with the other two methods is presented. Differences of up to 15% and up to 58% are reported in AI versus MC and AI versus MIRD respectively.

4. Discussion

GATE toolkit was used to execute the MC realistic simulations for a wide range of pediatric models, based on clinically derived biodistributions for each radiopharmaceutical and organ studied over time. SADR values were thus calculated for every combination of radiopharmaceutical and organ of interest, at four or five different time points after the injection. The produced extended simulated database now consists of SADRs for 28 computational models of pediatric patients with different anatomical characteristics of varying age (2–17 years old), gender, mass and height, regarding 30 organs and 5 radiopharmaceuticals, namely ^{99m}Tc-MDP, ^{123}I-MIBG, ^{131}I-MIBG, ^{131}I-INa and ^{153}Sm-EDTMP, at several time points.

The performed simulations provided a statistical uncertainty range between 0.05% and 2.7%. One of the most significant outcomes of this database concerns the fact that there are indeed fluctuations at the dose rate, for the same organ on different phantoms, namely up to ~71% difference at male phantoms and up to ~65% at female phantoms. This indication enhances the importance of taking into consideration the different anatomical and physiological characteristics of each patient before the definition of the injected activity.

A significant point to mention concerns the pattern of dose rates values in relation to the age of phantoms. As expected, dose rates are in all cases higher for the youngest children, due to the overall smaller size of their body and the greater contribution of the cross-irradiating organs. Respectively, older ages illustrate lower values at dose deposition at all tested organs. In addition to the latter observation, it is useful to mention that dose rates present a similarity in pattern at models with small age variations between them, which coincides to similar weight and anatomical characteristics, as was also observed in our previous work (Papadimitroulas *et al* 2018) and confirmed in the present study with the extension of the database. Indicatively, as seen in the figure S1, the

[5] https://mirdsoft.org/

15 years old boy (58 kg) illustrates similar dose rate distribution with the 13.8 male phantom (67.4 kg) while the 6 year old male phantom (18.6 kg) coincides also with the 5 year old female phantom (17.7 kg) in dose rates, as expected, although they differ in gender.

The produced extended database of simulated SADR values enabled the development of ML regression techniques for fast predicting personalised internal absorbed dose rates for the organs and radiopharmaceuticals included in the database, for any pediatric patient. Hyperparameter tuning and ensemble AI techniques were applied, while the best performing models were selected. It is notable that model performance got indeed highly boosted by ensemble technique in several cases (up to 4% in all metrics besides R^2 that was found at the same high-level value of 0.97), while in other cases the ensemble model was equally or slightly worse performing. Computing time of SADR determination with the developed predictive models is tremendously reduced compared to the values extracted via MC realistic simulations. Indicative predictions, seen in figure 5 for the case of ^{99m}Tc-MDP for a very young boy (5 years old) and an older girl (14.3 years old) in selected organs, agree very well with the corresponding actual values from the simulated database.

SADR predictions are produced with a MAE below 10% for most of the models that were developed in the present work (for each radiopharmaceutical and organ), as reported in figure 6 in boxplots. 25% of the developed models present mean absolute error (MAPE) below 5% with the median value being at 8%, whereas an uncertainty of 10% is considered more than acceptable in the field. Such differences are common and acceptable in internal dosimetry. In Divoli *et al* (Divoli *et al* 2009) a comparison was implemented to investigate differences (due to anatomical variations) of the well-established MIRD protocol using *S*-values with direct MC dosimetry. Differences up to 140% were reported when realistic cumulative activity was used but decreased to up to 26% after mass correction. Error levels vary slightly with age, while still lying on low values around 8%, as seen in figure 7, also depicting a wider distribution of error for intermediate ages (6–12 years old). The highest age group (12–17 years old) exhibits higher error values on a narrower distribution (figure 7). Finally, figures S2—S6 in the supplementary material ('Supplementary data') show MAPE values of our models over time for every pharmaceutical in boxplots across all organs and illustrates that predictivity performance remains at the same low level over time, as desirable.

Several studies in the literature reported differences in internal dosimetry due to anatomical variations for a variety of applications incorporating radioimmunotherapy. Differences up to 36% in red marrow were reported in a study that investigated the influence of the total body mass on the scaling of the *S*-values, for therapeutic radiopharmaceuticals (Traino *et al* 2007). In another study, comparison was applied on the calculation of effective doses for internal photon dosimetry in voxelized and stylized anthropomorphic phantoms, where differences of 15%, 25%, 37% and 60% were reported for thyroid, lungs, bones and liver respectively (Kramer *et al* 2005). Marine *et al* also mentioned differences in specific absorbed fractions in the range of 10%–33% between adult men with normal body mass indices (Marine *et al* 2010).

MIRD schema is a well-established and well-validated dosimetry protocol, where interpolated *S*-values are considered for internal dosimetry assessment, considering the patients' variability (rescaled organ masses). A comparison of our proposed approach (AI), using state-of-the-art ML techniques, has been performed with the ground truth of direct MC dosimetry and with the standardised MIRD schema using the MIRDcalc program. Such a comparison is presented in table 3, where the differences of the final absorbed doses of 8 different organs of interest is presented for the ^{99m}Tc case. The maximum differences reported between AI and MC is almost ∼15% for spleen and stomach, while the minimum differences are reported in kidneys and brain in the range of 3%–4%. A largest variation is reported in the comparison of absorbed doses/organ when comparing AI versus MIRD reaching up to 58%.

The novelty of the proposed approach lies on the prediction of SADRs for each new patient based on the personalized anatomical characteristics (such as age, gender, weight, height, effective diameter). However, it should be noted that although the high accuracy on the predictive absorbed doses per organ, there are specific limitations needed to be considered. SADRs are dependent on the specific biodistribution of each radiopharmaceutical which is used in the simulation procedure to calculate the simulated SADRs. Such a limitation is an obstacle in the current form of the model to be generalized for clinical use. However, our methodology can be also extended towards different biodistributions (which was not the scope of the current study), providing a ground truth dataset with varying biodistributions in a similar manner with the anatomical characteristics of this study. Then, ML models can learn the biodistribution variation (of the same radiopharmaceutical) among different patients, coupled with the varying anatomical characteristics. Another limitation of the proposed study is the limited representation of the pediatric population. Based on the Society of Nuclear Medicine it is a standard procedure to use anthropomorphic models for such dosimetry applications. However, considering the need of high accuracy, increasing the number of the pediatric models and their variability (different types of models - highly heterogeneous population), could extensively make the prediction model more accurate and more robust, providing personalized dosimetry assessment. This could be a future work for optimizing the models, as the purpose of this study was to develop, introduce and evaluate a novel predictive framework for internal dosimetry pediatric applications. The size of the training dataset is an inherent

issue of all AI procedures to aim for increased model generalization and predictive power. Finally, the proposed AI approach and methodology on internal dosimetry prediction for a targeted patient group, can be further extended to other applications or other patient groups (e.g. obese patients), as well as other organs and radiopharmaceuticals than the ones studied in the present work. Recently an application of the proposed approach has been presented showing a Graphical User Interface for clinical use (section 4) (Koch *et al* 2023).

5. Conclusion

The present study implemented the methodology of the previous work by Papadimitroulas *et al* (Papadimitroulas *et al* 2018) on the SADRs and extended its simulated dosimetry database for the purpose of exploiting it, towards the development of a prediction dosimetry model. The varying absorbed dose rates of this wider database, related to anatomical characteristics, age and gender, have been modelled in the present work using ML techniques, thus facilitating the individualized determination of SADRs for any pediatric patient, for a list of 5 commonly used radiopharmaceuticals, very fast and accurately. The produced predictive models are therefore expected to have a significant contribution in nuclear medical pediatric applications towards the optimization and personalization of dosimetry protocols. The produced enriched and broad database of simulated SADRs on anatomical characteristics, age and gender enabled the training and development of ML regression models, resulting to an internal dosimetry prediction toolkit, which predicts very fast the corresponding SADR values for each new pediatric patient, considering her/his personalised anatomical characteristics. The proposed methodology of combining the predictive power of AI utilizing MC ground truth for dosimetry assessment, can be further extended to other populations (adult, obese, pregnant) and medical applications (radioimmunotherapy), where fast and personalized absorbed dose determination is critical, which is the case in modern medicine in both diagnostic and therapeutic applications.

A challenging investigation for our future work is to extend the proposed methodology, with the ML developed prediction models, on *S*-values calculations (instead of SADR values) aiming to a prediction of the absorbed doses per organ based on the overall anatomical characteristics of the patients, and not by rescaling pre-calculated *S*-values. Thus, new predicted personalized *S*-values could be generated per patient enhancing the MIRD schema to more personalized approaches.

Acknowledgments

This study was co-financed by: a. The experiment 'PediDose' from the European High-Performance Computing Joint Undertaking (JU) through the FF4EuroHPC project under Grant agreement No 951745. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, Italy, Slovenia, France, Spain. b. The CHIST-ERA Grant [CHIST-ERA-19-XAI-007] with project acronym INFORM, by General Secretariat for Research and Innovation (GSRI) of Greece [T12EPA5–00053], National Science Centre (NCN) of Poland [2020/02/ Y/ST6/00071] and Agence Nationale de la Recherche (ANR) of France [ANR-21-CHR4–0006]. c. The Hellenic Foundation for Research and Innovation (H.F.R.I.) under the '2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers' (Project Number: 2692).

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary information files).

ORCID iDs

Vasileios Eleftheriadis https://orcid.org/0000-0002-2103-1047
Georgios Savvidis https://orcid.org/0000-0001-8308-0090
Valentina Paneta https://orcid.org/0000-0002-9485-8993
Konstantinos Chatzipapas https://orcid.org/0000-0003-4006-9304
George C Kagadis https://orcid.org/0000-0002-6983-2863
Panagiotis Papadimitroulas https://orcid.org/0000-0002-5981-6149

References

Adelstein S J 2014 Radiation risk in nuclear medicine *Semin. Nucl. Med.* **44** 187–92
Agostinelli S *et al* 2003 Geant4 a simulation toolkit *Nucl. Instrum. Meth. A* **506** 250–303

Akhavanallaf A, Fayad H, Salimi Y, Aly A, Kharita H, Naemi H and Zaidi H 2022 An update on computational anthropomorphic anatomical models *Digit Health.* **8** 20552076221111941
Akhavanallaf A, Shiri I, Arabi H and Zaidi H 2021 Whole-body voxel-based internal dosimetry using deep learning *Eur. J. Nucl. Med. Mol. Imaging* **48** 670–82
Allison J *et al* 2016 Recent developments in Geant4 *Nucl. Instrum. Meth. A* **835** 186–225
Arabi H and Zaidi H 2020 Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy *Eur. J. Hybrid Imaging* **4** 17
Awad M and Khanna R 2015 Support vector regression *Efficient Learning Machines* (Berkeley, CA.: Apress) pp 67–80
Boone J *et al* 2011 *Size-Specific Dose Estimates (SSDE) in Pediatric and Adult Body CT ExaminationsAAPM* Report No. 204 American Association of Physicists in Medicine
Bolch Wesley E, Eckerman Keith F, Sgouros George and Thomas Stephen R 2009 MIRD Pamphlet No. 21: A generalized schema for radiopharmaceutical dosimetry—standardization of nomenclature *Journal of Nuclear Medicine* **50** 477–484
Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system *Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining: ACM New York, NY, USA* pp 785–94
Chetty I J *et al* 2006 Reporting and analyzing statistical uncertainties in Monte Carlo-based treatment planning *Int. J. Radiat. Oncol. Biol. Phys.* **65** 1249–59
Christ A *et al* 2010 The Virtual Family–development of surface-based anatomical models of two adults and two children for dosimetric simulations *Phys. Med. Biol.* **55** 15–23
Dietterich T G 2000 Multiple classifier systems *Lect. Notes Comput. Sci.* (Berlin: Springer) vol 1857
Divoli A *et al* 2009 Effect of patient morphology on dosimetric calculations for internal irradiation as assessed by comparisons of Monte Carlo versus conventional methodologies *J. Nucl. Med.* **50** 316–23
Friedman J H 2001 Greedy function approximation: a gradient boosting machine *Ann. Stat.* **29** 1189–232
Götz T I, Schmidkonz C, Chen S, Al-Baddai S, Kuwert T and Lang E W 2020 A deep learning approach to radiation dose estimation *Phys. Med. Biol.* **65** 035007
Hoerl A E and Kennard R W 1970 Ridge regression: Biased estimation for nonorthogonal problems *Technometrics* **12** 55–67
Hyndman R J and Koehler A B 2006 Another look at measures of forecast accuracy *Int. J. Forecast.* **22** 679–88
Jan S *et al* 2004 GATE: a simulation toolkit for PET and SPECT *Phys. Med. Biol.* **49** 4543–61
Jan S *et al* 2011 GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy *Phys. Med. Biol.* **56** 881–901
Khong P L *et al* 2013 ICRP publication 121: Radiological protection in paediatric diagnostic and interventional radiology *Ann. ICRP* **42** 1–63
Koch M *et al* 2023 HPC + in the medical field: Overview *Technol. Health Care* Pre-press 1–15 HPC+in the medical field: Overview and current examples
Kramer R, Khoury H J and Vieira J W 2005 Comparison between effective doses for voxel-based and stylized exposure models from photon and electron irradiation *Phys. Med. Biol.* **50** 5105–26
Lai T L, Robbins H and Wei C Z 1979 Strong consistency of least squares estimates in multiple regression II *J. Multivar. Anal.* **9** 343–61
Lee M S, Hwang D, Kim J H and Lee J S 2019 Deep-dose: a voxel dose estimation method using deep convolutional neural network for personalized internal dosimetry *Sci. Rep.* **9** 10308
Marine P M, Stabin M G, Fernald M J and Brill A B 2010 Changes in radiation dose with variations in human anatomy: larger and smaller normal-stature adults *J. Nucl. Med.* **51** 806–11
Nensa F, Demircioglu A and Rischpler C 2019 Artificial intelligence in nuclear medicine *J. Nucl. Med.* **60** 29S–7S
Papadimitroulas P 2017 Dosimetry applications in GATE Monte Carlo toolkit *Phys. Med.* **41** 136–40
Papadimitroulas P *et al* 2019 A Review on personalized pediatric dosimetry applications using advanced computational tools *IEEE Trans. Radiat. Plasma Med. Sci.* **3** 607–20
Papadimitroulas P, Erwin W D, Iliadou V, Kostou T, Loudos G and Kagadis G C 2018 A personalized, Monte Carlo-based method for internal dosimetric evaluation of radiopharmaceuticals in children *Med. Phys.* **45** 3939–49
Pedregosa F *et al* 2011 Scikit-learn: machine learning in python *J. Mach. Learn. Res.* **12** 2825–2830
Quinlan J R 1986 Induction of decision trees *Mach. Learn.* **1** 81–106
Robbins E 2008 Radiation risks from imaging studies in children with cancer *Pediatr Blood Cancer* **51** 453–7
Sammut C and Webb G I (ed) 2011 *Encyclopedia of Machine Learning* (Berlin: Springer)
Sarrut D *et al* 2014 A review of the use and potential of the GATE Monte Carlo simulation code for radiation therapy and dosimetry applications *Med. Phys.* **41** 064301
Sarrut D *et al* 2022 The OpenGATE ecosystem for Monte Carlo simulation in medical physics *Phys. Med. Biol.* **67** 184001
Schapire R E 2013 Explaining adaboost *Empirical Inference* (Berlin,Heidelberg: Springer) pp 37–52
Segars W P *et al* 2015 The development of a population of 4D pediatric XCAT phantoms for imaging research and optimization *Med. Phys.* **42** 4719–26
Shahhosseini M, Hu G and Pham H 2022 Optimizing ensemble weights and hyperparameters of machine learning models for regression problems *MLWA* **7** 100251
Stabin M G, Madsen M T and Zaidi H 2019 Personalized dosimetry is a must for appropriate molecular radiotherapy *Med. Phys.* **46** 4713–6
Traino A C, Ferrari M, Cremonesi M and Stabin M G 2007 Influence of total-body mass on the scaling of S-factors for patient-specific, blood-based red-marrow dosimetry *Phys. Med. Biol.* **52** 5231–48
Treves S T, Falone A E and Fahey F H 2014 Pediatric nuclear medicine and radiation dose *Semin. Nucl. Med.* **44** 202–9
Willmott C J and Matsuura K 2005 Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance *Clim. Res.* **30** 79–82
Wu J, Chen X Y, Zhang H, Xiong L D, Lei H and Deng S H 2019 Hyperparameter optimization for machine learning models based on Bayesian optimization *J. Electron. Sci. Technol.* **17** 26–40
Zaidi H 1999 Relevance of accurate Monte Carlo modeling in nuclear medical imaging *Med. Phys.* **26** 574–608