ORIGINAL RESEARCH

Open Access

Fit of biokinetic data in molecular radiotherapy: a machine learning approach



Davide Ciucci¹, Bartolomeo Cassano^{3*}, Salvatore Donatiello¹, Federica Martire², Antonio Napolitano¹, Claudia Polito¹, Elena Solfaroli Camillocci¹, Gianluca Cervino⁴, Ludovica Pungitore⁴, Claudio Altini⁵, Maria Felicia Villani⁵, Milena Pizzoferro⁵, Maria Carmen Garganese⁵ and Vittorio Cannatà¹

*Correspondence: bartolomeo.cassano@gmail.com; bartolomeo.cassano@ifo.it

 ¹ Medical Physics Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy
² Tor Vergata Postgraduate School of Medical Physics, University of Rome, Rome, Italy
³ Medical Physics Department, IRCCS Regina Elena National Cancer Institute, Rome, Italy
⁴ Roma 3 University of Rome, Rome, Italy
⁵ Nuclear Medicine Unit/Imaging Department, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy

Abstract

Background: In literature are reported different analytical methods (AM) to choose the proper fit model and to fit data of the time-activity curve (TAC). On the other hand, Machine Learning algorithms (ML) are increasingly used for both classification and regression tasks. The aim of this work was to investigate the possibility of employing ML both to classify the most appropriate fit model and to predict the area under the curve (τ).

Methods: Two different ML systems have been developed for classifying the fit model and to predict the biokinetic parameters. The two systems were trained and tested with synthetic TACs simulating a whole-body Fraction Injected Activity for patients affected by metastatic Differentiated Thyroid Carcinoma, administered with [¹³¹I]I-Nal. Test performances, defined as classification accuracy (CA) and percentage difference between the actual and the estimated area under the curve ($\Delta \tau$), were compared with those obtained using AM varying the number of points (N) of the TACs. A comparison between AM and ML were performed using data of 20 real patients.

Results: As N varies, CA remains constant for ML (about 98%), while it improves for F-test (from 62 to 92%) and AICc (from 50 to 92%), as N increases. With AM, $\Delta \tau$ can reach down to - 67%, while using ML $\Delta \tau$ ranges within ± 25%. Using real TACs, there is a good agreement between τ obtained with ML system and AM.

Conclusions: The employing of ML systems may be feasible, having both a better classification and a better estimation of biokinetic parameters.

Keywords: Machine learning, Akaike information criterion, F-test, Fit function, Biokinetic curves

Background

Biokinetic parameters play a crucial role in molecular radiotherapy (MRT) in the assessment of the absorbed dose to lesions or organs at risk as they are closely related to treatment toxicity and efficacy [1].

Determining the time integrated activity (τ) accurately is challenging [2, 3]. The assessment of Time Activity Curve (TAC) depends on the data (TAC-Ps) collection and on the choice of the model that best fits data, in order to calculate the time integrated activity



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativecommons.org/licenses/by/4.0/.

coefficient (TIAc). Many parameters affect the fit model selection: the number of TAC-Ps, the time window, the time sampling and the error that affects each measurement.

In 2007 Glatting et al. [4] proposed the use of the corrected Akaike Information Criterion (AICc) and F-test [5–8] methods for the comparison of two different models in molecular radiotherapy. In 2013 Kletting et al. [9] showed the need of using a fitting method dedicated to MRT data whereby the physical and biological aspects of the biokinetics curve must be considered in order to compute meaningful parameters.

In recent years, several applications based on Machine Learning (ML) algorithms have been developed in nuclear medicine [10–12]. The first aim of this study was to implement ML systems to classify the proper curves model and predict, via regression, the TIAc. The secondary aim was to compare the ML performances with the ones obtained with fit algorithm and the analytical methods, AICc and F-test.

Methods

In order to have a better understanding of this manuscript we assume the following definitions: (i) Points of the Time-Activity Curve (TAC-Ps) are the set of points used to determine a TAC (ii) The training set (Tr_N) and the test set (Ts_N) consist of 10,000 and 2000 TAC-Ps, respectively, where each individual TAC-Ps is composed of a number of points equal to N.

The study was subdivided in 3 steps: (i) the generation of synthetic Tr_N and Ts_N to train and test the machine learning systems. (ii) The training of the ML systems, (iii) the test of the ML systems and performance evaluation in comparison with analytical methods. All these steps are fully described in the sections below and shown in Fig. 1.

At the end, in order to show the applicability of the developed methods, the ML systems were used to calculate the TIAc of 20 patients' TAC-Ps.

Training and test dataset generation

To train and test the ML systems a series of Tr_N and Ts_N were synthetically generated. The aim of this first step was to generate TAC-Ps simulating whole-body fractions of injected activity (FIA_{WB}) of patient affected by metastatic Differentiated Thyroid Carcinoma (mDTC), defined as:

$$FIA_{WB}(t) = \frac{A_{WB}(t)}{A_{Adm}} \tag{1}$$

where $A_{WB}(t)$ is the whole-body activity at time t and A_{Adm} is the administered activity.

The synthetic TAC-Ps were generated with both a Mono-Exponential (MEf) and Bi-Exponential functions (BEf) into a band of possible curves, showed in Fig. 2, and their values are listed in Table 1. In the "Appendix" the generation method is reported.

To study how N affects the ML performances, 10 different Tr_N were created, with N ranging from 4 to 20, in a time window of 160 h. In order to simulate the patient's hospitalization, the first N-1 points were equally distributed in a range of [0; 54] h, while the last point was set at 160 h. Each Tr_N was used to train the ML systems and it was derived from 5000 MEf and 5000 BEf curves.

Similarly, 10 Ts_N were generated to test the ML systems; they were derived from 1000 MEf and 1000 BEf curves. The generation procedure is similar to the one of the



Fig. 1 Study workflow

 $Tr_{N_{s}}$ with the addition of Gaussian noise on the FIA coordinates, in order to simulate measuring errors. The amount of added noise is randomly extracted from a Gaussian distribution centred on the FIA coordinates, with a standard deviation of 5% on the FIA values [13].



Fig. 2 All the generated TACs are into the blue and red bands for Mono-Exponential and Bi-Exponential functions respectively

Table 1 TIAc Parameters for	the band of	possible	TACs
-----------------------------	-------------	----------	------

	MEf		BEf			
	A ₀	$\lambda_0[h^{-1}]$	A ₁	$\lambda_1 [h^{-1}]$	A ₂	$\lambda_2 [h^{-1}]$
Lower limit	0.78	0.0676	0.46	0.1078	0.34	0.0476
Upper limit	1.00	0.0165	0.70	0.0225	0.30	0.0044

Implementation and training of ML system

Two different supervised learning systems were implemented using Python scripts: one for a binary classification task (ML1) and one for a regression task (ML2). Scikit-Learn library provided different machine learning models.

Hardware was composed of a personal computer having 10th generation Intel I7 CPU with 16 GB RAM.

ML1 is an ensemble of two Logistic Regression models, which used the Soft Voting method to predict the proper class [14]. The TAC-Ps were implemented as features (2N features considering x and y coordinates), and it returns MEf or BEf class as output.

ML2 is an AdaBoost ensemble composed of 5 Gradient Boosting Regressors (GBRs), working sequentially. Each GBR was designed as a chain of 1000 Decision Tree Regressors [14, 16]. The ML2 task is to predict the TIAcs (A_i , λ_i), in order to calculate τ . In ML2, TAC-Ps and fit model (MEf or BEf) were used as input features (2N + 1 features). Each Tr_N with a specific number of points individually trains ML2. Therefore, even though the ML2 algorithms are always the same, ML2 consists of 10 ML2_N, depending on the number of points N. To simplify, the subscript N is omitted when referring to ML2 in the rest of this text.

The adopted models from Python Scikit-Learn library were customized with hyperparameters [17] reported in Tables 2 and 3, for the classifier and the regressor respectively. All the other ones that are not reported in tables were set with the default values proposed by the library.

Hyperparameter	1st Estimator	2nd Estimator	
solver	"liblinear"	"sag"	
max_iter	100 (default value)	10,000	

Table 2 Scikit-Learn hyperparameters used for the customization of ML1

Table 3 Scikit-Learn hyperparameters used for the customization of ML2

Hyperparameter	GBR	AdaBoost Regressor
loss	ls	Linear (default value)
n_estimators	1000	5
learning_rate	0.1 (default value)	0.1
max_depth	5	N.A

Analytical methods: AICc and F-test

Each TAC-Ps of Ts_N were fitted using the Trust-Region algorithm, implemented on the Matlab toolkit, while a stripping algorithm was used to evaluate the fit starting point [18]. Each TAC-Ps was fitted both with MEf and BEf and the two results were compared using the AICc and F-Test (with a confidence interval of 95%), as mentioned by Kletting et al. [4].

As reported in [7] the AICc is described by the following equation:

$$AICc = N \cdot \ln\left(\frac{SS}{N}\right) + 2 \cdot K + \frac{2 \cdot K \cdot (K+1)}{N - K - 1}$$
(2)

where K is the number of estimated parameters included in the model, N is the number of points, and SS is the sum of squared deviations (between the measurement and the fitted curve). The model with the lower AICc score is the model that is more likely to be correct. Once AICc has been calculated for each fitting model, it is possible to compute the probability that the correct model (i) has been chosen, as follows:

$$w_i = \frac{e^{\left(-\frac{\Delta}{2}\right)}}{1 + e^{-\frac{\Delta}{2}}} \tag{3}$$

where Δ is the difference between the AICc score.

The model with the minimal AICc value between all candidate models indicates the best model.

The F-test is a hypothesis test in which only two models can be compared using the following equation [4, 7]:

$$F = \frac{(SS_{ME} - SS_{BE})/SS_{BE}}{(DF_{ME} - DF_{BE})/DF_{BE}}$$
(4)

where *SS* are the sums of squared deviations (for MEf and BEf) and DF are the degrees of freedom (DF = N-K). The decision to accept or discard the MEf model is based on the p-value calculated from the F ratio. For a P value below the chosen significance level

(0.05) the MEf model is rejected and therefore the more complex model is assumed to fit the data in a significantly better way.

Test and performance evaluation

A tenfold cross validation of ML1 system has been performed using 80% of the Tr_N in training and the remaining 20% for testing. The classification accuracy (CA) of the correct fit model was chosen as performance estimator. It is defined as:

$$CA = \frac{CC_{ME} + CC_{BE}}{2000} \tag{5}$$

where CC_{ME} and CC_{BE} are the number of correct classifications for mono- and bi-exponential functions respectively, and 2000 is the number of TACs.

The ability of ML1 system, AICc and F-test to correctly classify the model was evaluated, determining the CA using Ts_N as input.

The chosen parameter to evaluate the goodness of TIAc prediction was the area under each TAC (τ), assessed through the following equation [15]:

$$\tau = \sum_{i=1}^{n} \frac{A_i}{\lambda_i} \tag{6}$$

where *n* is equal to 1 for MEf and 2 for BEf and A_i and λ_i are the parameters obtained from the fit.

Two different tests were performed to evaluate the goodness of TIAc prediction: the first had the aim to evaluate the best possible performances of the fit algorithm and ML2, considering a classification with no errors. Meanwhile the second test evaluated the performances of the chains ML1 + ML2, fit + AICc and fit + F-Test, considering all the classification errors. In addition, a fourth chain was considered, named fit + AICc-W. In this method, an average model was considered, taking into account the probability of each model given by the parameter w_i . The area τ was calculated as follows:

$$\tau = \tau_{MEf} \cdot w_{MEf} + \tau_{BEf} \cdot w_{BEf} \tag{7}$$

The performances of the two tests were evaluated assessing the distributions of the percentage differences ($\Delta \tau$) between the calculated τ and true one, the interquartile range and the Maximum Error Range (MER). The MER was represented as the range between the minimum and maximum value of $\Delta \tau$.

Test on patient data

The whole-body TAC-Ps of 20 patients, affected by mDTC, were evaluated employing both methods: Fit + F-Test and ML1 + ML2. All TAC data were obtained by calculating the geometric mean of measurements taken with a plastic scintillator placed at a distance of 4 m from the patient and calibrated in terms of H*(10), providing a value of the dose rate in μ Sv/h.

Each TAC-Ps had 5 or 6 points and they were uniformly classified as 10 MEf and 10 BEf by the F-Test.



Fig. 3 Classification accuracy obtained by three methods varying the number of points of the TACs simulated



Fig. 4 Boxplots representing the $\Delta \tau$ distributions and the MER bands obtained using the fit algorithm (blue) and ML2 (red). For each set of measures the performances of the ML system and the fit algorithm were compared only for the assessment of the AUC, since it was known a priori if the points follow a mono-exponential or a bi-exponential trend. Through the estimation of the fit model parameters the AUC was calculated and the distributions were represented

Results

The computation time to train ML1 and ML2 sequentially is about eight minutes, while the computer takes few seconds to perform the test with 2000 curves. The same hardware takes about 10 min to perform fit + F-test and fit + AICc using the same datasets.

The CA in the cross validation, varying the number of points (N), is reasonably constant at a value of 99%. Figure 3 shows the CA of the three systems using the test dataset that is about constant at the value of 98% for ML1, while it increases from 50 to 92% for AICc and from 62 to 92% for F-test, when the number of points increases.

Figure 4 shows the $\Delta \tau$ distributions obtained using a classification without errors. The median values remain approximately constant around 0% for both methods, and also the

width of the distribution between the first and third interquartile is independent from the number of points with values equal to 4.5% and 3.6% for ML2 and the fit algorithm respectively. MER are represented by the bands in Fig. 4, which become thinner as the number of points increases, and their ranges vary from [-14.2; 14.5]% to [-8.1; 7.5]%, and from [-16.8; 22.1]% to [-6.7; 17.3]% for ML2 and fit algorithm respectively.

The $\Delta \tau$ distributions shown in Fig. 5 were obtained including any classification failure from AICc, AICc-W, F-test and ML1. Even in this case, varying N, the median values are approximately constant for all the considered methods, while interquartile range varies from 24.7 to 3.1% for the fit + AICc, from 24.6 to 3.1% for the fit + AICc-W and from 15.8% to 3.1% for the fit + F-test. The width remains constant around 4.5% for the ML1 + ML2 system. With the increase of the number of points, MER varies from [- 66.7; 11.9]% to [- 19.3; 20.9]% for fit + AICc, from [- 66.7; 12.0] to [- 20.7; 18.1] for fit + AICc-W, from [- 65.3; 32.1]% to [- 15.1; 20.9]% for fit + F-test and from [- 25.2; 23.6] to [- 10.6; 12.5]% for ML1 + ML2.

When N < 8, $\Delta \tau$ distributions obtained by AICc and F-Test are asymmetrical, while for N > 8 the distribution width, obtained with the three methods, are equally distributed. The MER band of ML2 is always thinner than the other two.

Considering the TAC-Ps of 20 patients, 9 out of the 10 MEfs were classified as BEfs, while all the 10 presented BEfs were confirmed to be BEfs by ML1. Figure 6 shows the correlation between the τ calculated with ML1 + ML2 and the one calculated with the fit Algorithm + F-Test on the 20 patients data, obtaining an R² equal to 0.93, and a difference percentage of the results in a [-26.9;12.8]% range, with a median value of 0.2%.

Discussion

The obtained results show the feasibility of using machine learning system both to classify the proper model and predict the TIAc with an improvement of the performances.



Fig. 5 Boxplots representing the $\Delta \tau$ distributions and the MER bands obtained using the fit + AlCc (blue), fit + AlCc-W (green), fit + F-test (grey) and ML1 + ML2 (red). For the analytical methods, points were fitted with both models and the AlCc and F-Test determine which AUC should be considered. For ML, first the best fit model was predicted and then the model parameters in order to calculate the AUC



Fig. 6 Comparison between the τ values obtained with ML1 + ML2 and Fit Algorithm + F-test. The red points represent the area under the curve of TACs classified as mono-exponential for the F-test while the blue points represent the area under the curve of TACs classified as bi-exponential for the F-test

The training phase plays a crucial role into obtaining results close to true values, and one of the major advantages shown by ML is the possibility for the model to be trained with error-free data and then tested with data simulating real curves, resulting in a high CA (about 98.5%) and a $\Delta \tau$ less than $\pm 25\%$. This aspect can be particularly advantageous in order to train the system without knowing the error of the measurement system "a priori". Training with error-free data is not only advantageous but also recommended, as reported by Geron et al. [14]. The training phase should be conducted with the cleanest possible data, to avoid confusing the system and to prevent overfitting. To confirm this, training was performed with noisy data similar to the ones used in Tr_N, and a decrease in classification accuracy of around 5% was recorded.

The time spent for training is very short, and it needs to be executed just once before running classifications and regressions.

The main difference between ML and the analytical method is the workflow. The former predicts the model and then assesses the biokinetic parameters. Instead, the latter calculates the biokinetic parameters for both models first and then chooses the optimal one. This aspect is crucial: the analytical method is not optimized because it necessarily performs both fit models and, in addition, the choice of the model is strongly dependent on the algorithm fit and on the goodness of the TIAc.

The cross-validation results are better than the ones obtained in the test phase, but this was to be expected: in fact, Ts_N simulates the measurements errors (thanks to the addition of Gaussian noise), and so the results get worse without being associated with an overfitting condition. Figure 3 shows that CA is not dependent on the number of points for ML1, while it is highly dependent for AICc and F-test. The analytical methods show a low accuracy when the number of points is near to the limit of applicability (from Eqs. 3 and 4 it is possible to notice that F-test and AICc are available

only if the number of points is at least 5 and 6 respectively). As reported by Kletting et al. in [19], F-test and AICc classify as mono-exponential most of the bi-exponential curves (CA = 61.0% and CA = 50.0%) if the number of points is 5 or 6 respectively. In addition, the CA obtained with ML1, equal to 98.5%, is higher than the results obtained with AICc and F-test.

Figure 4 shows the best results obtainable through the two systems, dedicated to the assessment of biokinetic parameters. Considering an error-free classification, the width of the $\Delta \tau$ distributions, obtained with ML2 and with the fit algorithm, are similar: hence, in most cases the two systems are equivalent. The MER obtained from ML2, represented by the red band in Fig. 4, does not exceed $\pm 15\%$ and it is smaller than the one obtained using the fit algorithm (about $\pm 20\%$ and represented by the blue band).

Glatting et al. in [4] use fitting algorithms that take into account uncertainties on individual points of the TAC. The fits used in the present study, on the other hand, do not consider such errors and, in future works, it will be necessary to perform a comparison. The use of these algorithms will likely lead to a convergence of performance between analytical methods and ML methods, but a priori knowledge of errors is not always possible. Therefore, "a priori"-trained ML systems remain a more general method for both model selection and obtaining kinetic parameters.

When the input to determine the biokinetic parameters are the classification results obtained through ML1, AICc, AICc-W and F-test, performances get worse for all the examined four systems (Fig. 5), but the ML1 + ML2 system keeps MER within \pm 25%. The analytical methods tend to underestimate the area under the curve when the number of points is lower than 8. This effect is due to the tendency of the two methods to classify bi-exponential curves as mono-exponential.

The use of ML in this field can also lead to the possibility of using it in a more radical way. For example, an attempt was made to train the ML2 system to obtain the area under the curve directly as the output, instead of the fit parameters. The obtained results were similar to those shown in Figs. 4 and 5, but, according to the authors, this is an incorrect way of using ML systems: in fact they are already considered black boxes, and using the system in the aforementioned way, without the possibility to assure the goodness of the fit, can be challenging to justify in the clinical use of dosimetry.

It is important to underline that the performance of ML systems has proven to be fairly independent from the number of points composing the TACs, whereas a more pronounced dependence has been observed in relation to errors in the data. Some tests were conducted by varying the error on the points, resulting in a decrease in performance. These analyses were not reported in the results section because they go beyond the scope of this work, which is to demonstrate the feasibility of using ML systems in dosimetry in molecular radiotherapy, for model selection and the calculation of the area under the TAC curve.

Figure 6 shows the applicability of the ML systems on real data: it confirms that, when ML1 and F-test classify the curve with the same model, the assessed biokinetic parameters are very similar. On the other hand, when classification does not coincide, the distance between the two results increases. An example of these results is shown



Fig. 7 TACs of Wholebody FIA of three different patients and fit obtained with fit + F-test and ML1 + ML2 systems. **a** The two model agree classifying and fitting the data with a MEf. **b** Also in this case ML1 and F-test classify with a BEf. **c** In the third case the F-test classify the curve as MEf while it is a BEf for ML2. In this case the two obtained fit curves are significantly different

in Fig. 7. Considering CA for 5 points (shown in Fig. 3) and the graph in Fig. 5, there is a high probability that the correct classification is given by ML1, and consequently the ML2 result could better represent the biokinetic curve of the patient.

ML1 and ML2 have been trained with curves simulating the whole-body FIA of patients administered with [¹³¹I]-NaI, but can be trained to perform cumulated activity calculations also for different radionuclides in metabolic radiotherapy, and with different dosimetry calculation methods (i.e. voxel dosimetry).

In 2017 Sarrut et al. [20] showed that the selection of the proper fit model reduces the number of fit failures in voxel dosimetry (defined as the case in which the optimizer does not converge and reaches the maximum number of iterations, or if the R^2 is lower than a certain threshold). The brief time to perform the 2000 test fits and the independence of the performance from the number of points make these systems suitable for this application.

It is necessary to emphasize that each new application (such as the study of new curve models) needs a new training, and that the performance is strongly dependent on the similarity between the training samples and the clinical case. In addition, in this study, Gaussian-type errors were used on the data because the purpose of the work was to demonstrate the feasibility of using ML systems. However, in future studies, it will be necessary to investigate how data errors affect performance using different thresholds and various types of distributions, such as the Poisson distribution.

This is a study on the feasibility of using the ML system for classifying the correct fit model and predicting the TIAc, and it doesn't have the purpose to study which ML algorithm is the optimal one in performing these two tasks. The choice of the logistic regression and the AdaBoost algorithm is arbitrary, and a subsequent study is necessary to identify which algorithms could increase the performance results.

In conclusion, to the knowledge of the authors, this study is the first to propose the use of ML systems for TIAc calculations for dosimetry in MRT. As demonstrated, the use is feasible and promising, but it requires further investigations in different fields.

Therefore, investigations will be performed in order to train such systems with different algorithms, treatments, radionuclides, curve models and for voxel dosimetry.

Appendix

Method for synthetic TAC-Ps generation

The datasets (TR_N and TS_N) used to train and test the systems consist of TAC-Ps, where each point composing them simulates the FIA_{WB} of patients with metastatic differentiated thyroid carcinoma (mDTC), administered with [¹³¹I]I-NaI. Each point in the TAC-Ps simulates the result of the geometric mean between measurements acquired from the anterior and posterior positions, using an external probe located 4 m from the patient. The TIAc values for the two bands are reported in Table 1 and are obtained using the upper and lower bounds of the TACs calculated for 50 patients administered in our institution and subjected to dosimetric study.

To generate each of the TAC-Ps, the following method was used for both mono-exponential and bi-exponential trends. For BEf, four time points were selected at 0, 24, 90, and 160 h and, for each of these times, a range of possible values was determined by the band showed in Fig. 2. From each of these ranges, a random FIA_{WB} value was extracted, resulting in 4 values corresponding to the 4 different times. These values were fitted with a bi-exponential curve, obtaining the parameters of the generated curve. Finally, a check is performed to ensure that the generated curve does not intersect the bands up to 500 h. If the check is successful, the curve parameters are recorded, and the TAC-Ps are created with the chosen number of points for either Tr_N or Ts_N . The procedure for TAC-Ps with a mono-exponential trend is identical, with the only difference being that two points are extracted at times 0 and 90 h.

Acknowledgements

None.

Author contributions

(I) Conception and design: DC, BC and VC; (II) Patients data acquisition: CA, MFV, MP and MCG; (III) Generation Curve algorithm: DC, BC, AN, GC and LP; (IV) Programming in Python and MATLAB: DC, BC, AN, GC, LP and CP; (V) Synthetic Data Analysis: DC, BC, DS, FM, AN, CP, ESC, GC and LP; (VI) Patient Data Analysis: DC, BC, SD, CP, ESC, CA, MFV, MP, MCG and VC; (VII) Interpretation of data: DC, BC, SD, FM, AN, CP, ESC, GC and LP; (VII) Manuscript writing: DC, BC and VC (IX) Final approval of manuscript: All authors. Large Language Models: No Large Language Models (LLMs), such as ChatGPT, has been used.

Funding

This work was supported also by the Italian Ministry of Health with Current Research funds.

Availability of data and material

The data that support the findings of this study are available on request from the corresponding author.

Declarations

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Informed consents for the publication of the study, the diagnostic and therapeutic procedure were obtained from the legally authorized representative.

Competing interests

All authors declare that they have no conflict of interest.

Received: 13 June 2023 Accepted: 15 February 2024 Published online: 22 February 2024

References

- Dewaraja YK, Schipper MJ, Shen J, et al. Tumor-absorbed dose predicts progression-free survival following (131) I-tositumomab radioimmunotherapy. J Nucl Med. 2014;55(7):1047–53. https://doi.org/10.2967/jnumed.113.136044.
- Bolch WE, Eckerman KF, Sgouros G, et al. MIRD pamphlet No. 21: a generalized schema for radiopharmaceutical dosimetry-standardization of nomenclature. J Nucl Med. 2009;50(3):477–84. https://doi.org/10.2967/jnumed.108. 056036.
- Siegel JA, Thomas SR, Stubbs JB, et al. MIRD pamphlet no. 16: Techniques for quantitative radiopharmaceutical biodistribution data acquisition and analysis for use in human radiation dose estimates. J Nucl Med. 1999;40(2):37S-61S.
- 4. Glatting G, Kletting P, Reske SN, et al. Choosing the optimal fit function: comparison of the Akaike information criterion and the F-test. Med Phys. 2007;34(11):4285–92. https://doi.org/10.1118/1.2794176.
- Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Second international symposium on inference theory. Budapest: Akademiai Kiado; 1973. p. 267–81.
- Burnham KP, Anderson DR. Model selection and multimodel interference-a practical information-theoretic approach. New York: Springer; 2002.
- Motulsky HJ, Christopoulos A. Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting. Boston: GraphPad Software Inc; 2004.
- Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika. 1989;76(2):297–307. https://doi.org/10.1093/biomet/76.2.297.
- Kletting P, Schimmel S, Kestler HA, et al. Molecular radiotherapy: the NUKFIT software for calculating the time-integrated activity coefficient. Med Phys. 2013;40(10):102504. https://doi.org/10.1118/1.4820367.
- Currie G, Hawk KE, Rohren E, et al. Machine learning and deep learning in medical imaging: intelligent imaging. J Med Imaging Radiat Sci. 2019;50(4):477–87. https://doi.org/10.1016/j.jmir.2019.09.005.
- Uribe CF, Mathotaarachchi S, Gaudet V, et al. Machine learning in nuclear medicine: part 1-introduction. J Nucl Med. 2019;60(4):451–8. https://doi.org/10.2967/jnumed.118.223495.
- Zukotynski K, Gaudet V, Uribe CF, et al. Machine learning in nuclear medicine: part 2-neural networks and clinical aspects. J Nucl Med. 2021;62(1):22–9. https://doi.org/10.2967/jnumed.119.231837.
- Flux GD, Guy MJ, Beddows R, Pryor M, et al. Estimation and implications of random errors in whole-body dosimetry for targeted radionuclide therapy. Phys Med Biol. 2002;47(17):3211–23. https://doi.org/10.1088/0031-9155/47/17/ 311.
- 14. Géron A. Hands-on machine learning with scikit-learn, keras, and TensorFlow. Sebastopol: O'Reilly Media; 2019.
- Lassmann M, Hänscheid H, Chiesa C, et al. EANM Dosimetry Committee. EANM Dosimetry Committee series on standard operational procedures for pre-therapeutic dosimetry I: blood and bone marrow dosimetry in differentiated thyroid cancer therapy. Eur J Nucl Med Mol Imaging. 2008;35(7):1405–12. https://doi.org/10.1007/ s00259-008-0761-x.
- 16. Schapire RE. The strength of weak learnability. Mach Learn. 1990;5:197–227. https://doi.org/10.1007/BF00116037.
- 17. https://scikit-learn.org/
- Kirkup L, Sutherland J. Curve stripping and nonlinear fitting of polyexponential functions to data using a microcomputer. Comput Phys. 1988;2:64. https://doi.org/10.1063/1.168313.
- 19. Kletting P, Glatting G. Model selection for time-activity curves: the corrected Akaike information criterion and the F-test. Z Med Phys. 2009;19(3):200–6. https://doi.org/10.1016/j.zemedi.2009.05.003.
- Sarrut D, Halty A, Badel JN, et al. Voxel-based multimodel fitting method for modeling time activity curves in SPECT images. Med Phys. 2017;44(12):6280–8. https://doi.org/10.1002/mp.12586.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.