

PAPER • OPEN ACCESS

## An artificial intelligence-based model for cell killing prediction: development, validation and explainability analysis of the ANAKIN model

To cite this article: Francesco G Cordonì *et al* 2023 *Phys. Med. Biol.* **68** 085017

View the [article online](#) for updates and enhancements.

You may also like

- [THE STRUCTURE OF THE LEONIS DEBRIS DISK](#)  
Nathan D. Stock, Kate Y. L. Su, Wilson Liu et al.
- [THE KECK INTERFEROMETER NULLER](#)  
E. Serabyn, B. Mennesson, M. M. Colavita et al.
- [Structural health monitoring system for bridges based on skin-like sensor](#)  
Konstantinos Loupos, Yannis Damigos, Angelos Amditis et al.



## PAPER

# An artificial intelligence-based model for cell killing prediction: development, validation and explainability analysis of the ANAKIN model

## OPEN ACCESS

## RECEIVED

7 November 2022

## REVISED

24 February 2023

## ACCEPTED FOR PUBLICATION

23 March 2023





## PUBLISHED

10 April 2023

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Francesco G Cordoni<sup>1,2,\*</sup> , Marta Missiaggia<sup>2,3</sup> , Emanuele Scifoni<sup>2</sup>  and Chiara La Tessa<sup>2,3,4</sup> 

<sup>1</sup> Department of Civil, Environmental and Mechanical Engineering, via Mesiano 77, I-38123, Trento, Italy

<sup>2</sup> Trento Institute for Fundamental Physics and Application (TIFPA), via Sommarive 15, I-38123, Trento, Italy

<sup>3</sup> Department of Radiation Oncology, University of Miami Miller School of Medicine, 33136, Miami FL, United States of America

<sup>4</sup> Department of Physics, via Sommarive 14, I-38123, Trento, Italy

\* Author to whom any correspondence should be addressed.

E-mail: [francesco.cordoni@unitn.it](mailto:francesco.cordoni@unitn.it), [marta.missiaggia@miami.edu](mailto:marta.missiaggia@miami.edu), [emanuele.scifoni@tifpa.infn.it](mailto:emanuele.scifoni@tifpa.infn.it) and [chiara.latessa@miami.edu](mailto:chiara.latessa@miami.edu)

**Keywords:** cell survival prediction, radiobiological modelling, RBE model, machine learning, deep learning

## Abstract

The present work develops ANAKIN: an *Artificial iNtelligence bAsed model for (radiation-induced) cell KilliNg prediction*. ANAKIN is trained and tested over 513 cell survival experiments with different types of radiation contained in the publicly available PIDE database. We show how ANAKIN accurately predicts several relevant biological endpoints over a wide broad range on ion beams and for a high number of cell-lines. We compare the prediction of ANAKIN to the only two radiobiological models for *Relative Biological Effectiveness* prediction used in clinics, that is the *Microdosimetric Kinetic Model* and the *Local Effect Model* (LEM version III), showing how ANAKIN has higher accuracy over the all considered cell survival fractions. At last, via modern techniques of *Explainable Artificial Intelligence* (XAI), we show how ANAKIN predictions can be understood and explained, highlighting how ANAKIN is in fact able to reproduce relevant well-known biological patterns, such as the overkilling effect.

## 1. Introduction

In the last decades, radiotherapy (RT) has increasingly proven to be an extremely effective cure against cancer. Within RT, particle therapy (PT), has been emerging (Durante and Flanz 2019), and at the end of 2021, about 325.000 patients have been treated worldwide with PT, of which close to 280.000 with protons and about 42.000 with carbon ions (PTCOG 2022). Furthermore, other ions have been recently gaining attention (Rovituso 2017): in 2021 the first patient was (re-)treated with helium (Mairani *et al* 2022) at the Heidelberg Ion Therapy Center (HIT) in Germany, while perspective studies are looking into the possible using of oxygen (Kurz *et al* 2012, Sokol *et al* 2017).

The physical rationale of using hadrons in cancer treatment is their characteristic energy loss mechanisms, which result in concrete biological advantages compared to photons, such as increased tumor control and a greater sparing of normal tissues, with a consequently lower risk of toxicity.

Despite the theoretically superior physical properties of hadrons compared to photons, further research is critical for increasing the PT application in the clinic. A correct and accurate estimation of radiation-induced biological damage remains one of the major limitations to the full exploitation of this treatment modality. The key quantity used to describe the radiation effectiveness in inducing specific damage is the *Relative Biological Effectiveness* (RBE), which is defined as the ratio between the dose delivered by a given radiation and the dose delivered by the reference radiation yielding the same biological effect:

$$\text{RBE} = \frac{D_{\text{reference}}}{D_{\text{radiation}} \Big|_{\text{isoeffect}}} .$$

RBE allows quantifying how much more lethal certain radiation is compared to the reference radiation, usually x-rays, and is used in *Treatment Planning Systems* (TPS) to calculate the biological dose, namely the physical dose multiplied by the RBE. For this reason, over the last decades a plethora of mathematical mechanistic models, (Kellerer and Rossi 1974, 1978, Tobias 1980, 1985, Hawkins 1994, Kase et al 2006, Elsässer et al 2010, Vassiliev 2012, Friedrich et al 2013a, 2013b, Manganaro et al 2017, Vassiliev et al 2017, Inaniwa and Kanematsu 2018, Bellinzona et al 2021, Cordonì et al 2021, McMahon and Prise 2021, Cordonì et al 2022a, 2022b), as well as data-driven phenomenological models, (Wilkins and Oelfke 2004, Tilly et al 2005, Carabe et al 2012, Chen and Ahmad 2012, McNamara et al 2015, Mairani et al 2017) have been developed to estimate RBE based on biological as well as physical quantities. At the base of most models is the linear-quadratic (LQ) behavior of the cell survival logarithm with respect to the imparted dose:

$$S(D) = e^{-\alpha D - \beta D^2},$$

where  $\alpha$  and  $\beta$  are some specific parameters that depend on both biological (e.g. tissue type) and physical (e.g. radiation quality) variables (McMahon 2018).

Currently, a constant RBE of 1.1 is conservatively used in proton therapy, although evidences show its variability, especially in the distal region (Paganetti et al 2002, Paganetti 2014, 2018, Missiaggia et al 2020, 2022a). For carbon and helium ions, the RBE variations across the irradiation field are significant enough that a constant value cannot be used. Currently, two radiobiological models are currently used to predict RBE in clinical practice: (i) the *Microdosimetric Kinetic Model* (MKM) (Inaniwa et al 2010, Inaniwa and Kanematsu 2018, Bellinzona et al 2021), and (ii) the *Local Effect Model* (LEM) (Elsässer et al 2010, Friedrich et al 2013a, Pfuhl et al 2022). Both models have been vastly tested against *in vitro* and *in vivo* data (Inaniwa and Kanematsu 2018, Pfuhl et al 2022), but the outcomes have not indicated a clear superiority of one model to the other. In addition, significant differences in the prediction of RBE across models are evident so that, at present days, the use in clinical practice of a variable RBE is highly subject to the model chosen, (Giovannini et al 2016, Missiaggia et al 2020, Bertolet et al 2021, Missiaggia et al 2022a).

The lack of a robust and generalized model for predicting RBE hinders the full exploitation of PT, including the use of ions heavier than carbon, such as oxygen, to successfully treat radio-resistant tumors, (Boulefour et al 2021), or multi-ion therapy, which is nowadays accessible from the technical point of view (Ebner et al 2021).

Furthermore, although some RBE models have a general mathematical formulation, their implementation in the TPS, especially for inverse planning, requires a heavy calculation effort. This issue is usually overcome both by using look-up tables and by making specific assumptions (Inaniwa and Kanematsu 2018), such as physical or biological approximations, which clearly limit the model generality and affect its RBE prediction accuracy.

Aiming at deriving a general model able to accurately predict RBE across a wide range of physical and biological variables, we developed ANAKIN (*an Artificial iNtelligence bAsed model for (radiation-induced) cell KilliNg prediction*), a new general AI-driven model for predicting cell survival and RBE. *Machine Learning* (ML) and *Deep Learning* (DL) algorithms have recently started to gain attention in the medical physics community with applications on imaging (Sahiner et al 2019), fast dose estimation (Götz et al 2020), Monte Carlo simulation (Sarrut and Krahe 2021), and particle tracking (Missiaggia et al 2022b) have been published. However, only Papakonstantinou et al (2021) apply ML for predicting radiation-induced biological quantities, conducting a study on the induction of DNA damage and its complexity, but no analysis on RBE is performed.

ANAKIN is composed by various ML and DL-based modules, each with a specific tool, and interconnected to each other. The model considers both physical variables such as the kinetic energy of the incident beam or also the *Linear Energy Transfer* (LET), that is the amount of energy that a particle transfers to the material traversed per unit distance, (Durante and Paganetti 2016), and biological variables, such as the  $\alpha$  and  $\beta$  values for the reference radiation response. To make the model as general as possible, we trained it on cell survival data for 20 cell lines widely used in radiobiology and 11 different ion types all available on the *Particle Irradiation Data Ensemble* (PIDE) (Friedrich et al 2013b, 2021). Together with particles of interest for clinical applications, we also included in the training process heavier ions, including iron. This choice extends the application of ANAKIN to other research fields, such as radiation protection in space. To verify ANAKIN predictions and assess their accuracy, we randomly divided the data available in PIDE into two sets, one for training and one for testing. Therefore all results reported in the present work refer to the test set, which consists entirely of experiments that have not been included in the training set.

*Artificial Intelligence* (AI) has had a disruptive impact both in the research field and in real-life applications. The potential of modern and advanced ML and DL algorithms have started to gain attention in the medical physics community, where several research papers on the application of DL to imaging, (Sahiner et al 2019), fast

dose estimation, (Götz *et al* 2020), Monte Carlo simulation, (Sarrut and Krah 2021), and particle tracking, (Missiaggia *et al* 2022b), have appeared. Quite surprisingly, to the best of our knowledge, the only results in the literature that use ML to predict radiation-induced biological quantities is (Papakonstantinou *et al* 2021), where the authors conduct a study on the induction of cellular damage, but no analysis on RBE is performed (Davidovic *et al* 2021).

ML and DL have been shown to be extremely powerful, accurate, and flexible tools to extract information and hidden relations as well as to predict the most likely outcome based on data of possibly different nature, (Khalid *et al* 2007, Ongsulee *et al* 2018, Shwartz-Ziv and Armon 2022). Moreover, an excellent, systematic and comprehensive data collection of cell survival experiments exists and is publicly available, the PIDE, (Friedrich *et al* 2013b, 2021).

ANAKIN is constituted by various ML and DL-based modules, each with a specific task, and interconnected to each other. Two different tree-based models, namely the *Random Forest* (RF) (Ho 1995, 1998), and the *Extreme Gradient Boosting* (XGBoost) (Chen and Guestrin 2016a, 2016b) algorithms are used to predict cell survival for a wide variety of radiation and cell-lines. It is worth stressing that the final goal of ANAKIN is to develop a robust and accurate model that is able to predict cell survival in the most general possible conditions. ANAKIN is trained to predict cell survival for 20 widely used cell lines and for 11 different ions type. Concerning this last point, despite the driving motivation being HT, many different ions, such as iron which is beyond the possible application in the clinic, are included in the model. This makes ANAKIN extremely general so that possible future application in space radioprotection is also envisaged.

ANAKIN is trained on the PIDE. It is worth stressing that, in order to be as realistic as possible, experiments contained in the PIDE dataset are divided into a training set and a testing set. Therefore all results reported in the present work refer to the test set, which consists entirely of experiments that have not been included in the training set. This means that ANAKIN is asked to predict the cell survival for experiments that have never been seen before. Besides the already mentioned variables, ANAKIN considers both physical variables such as the kinetic energy of the incident beam or also the LET, that is the amount of energy that a particle transfers to the material traversed per unit distance, (Durante and Paganetti 2016), and biological variables, such as the  $\alpha$  and  $\beta$  values for the reference radiation response.

ANAKIN is tested over several endpoints and metrics to establish the actual accuracy of its predictions. Further, ANAKIN predictions are compared with the MKM and the LEM, which are the only two radiobiological models currently used in the clinic. The two models are based on different funding assumptions, such as target size definition, the concept of locality, and parameters included, and therefore have significant differences in the predicted RBE. The analysis of the differences is beyond the scope of the present paper but has been deeply studied in literature (Friedrich *et al* 2013a, Stewart *et al* 2018, Scholz *et al* 2020, Bellinzona *et al* 2021). Regarding the LEM results, an extremely well-done and extensive analysis of the LEM has become available very recently (Pfuhl *et al* 2022). As a matter of fact, much of the analysis conducted in the current paper has been explicitly inspired by Pfuhl *et al* (2022). In this direction, it must be stressed that, in the current paper, version LEM III is used since LEM IV is not currently implemented in the *survival toolkit* and thus, the presented comparisons could not be translated to the state of the art version of the latter code. It is clear that the results reported in Pfuhl *et al* (2022) on the LEM IV are more accurate than the one reported in the current research using the LEM III, so this fact must be taken into account.

Finally, the current work further aims at demystifying the erroneous myth that ML and DL models are obscure black-box models whose predictions cannot be interpreted. If this argument can in fact be partially correct for extremely deep and sophisticated NN that has been built mostly in the field of the *Reinforcement Learning*, the same cannot be said for the vast majority of ML and DL developed in the last years. In fact, on one side, it must be said that some ML models, such as for instance tree-based models, are interpretable by nature and, on the other side, recently huge attention has been posed to the development of mathematical techniques aiming at explaining ML and DL models that are not of easy interpretation; such area of research is known as *Explainable AI* (XAI) (Gunning *et al* 2019).

The main focuses of the present research are to:

- (i) Develop for the first time a general AI-driven model to predict cell survival fraction over a wide range of biological cell lines and physical irradiation conditions.
- (ii) Compare ANAKIN with the two radiobiological models used in the clinic (MKM and LEM).
- (iii) Show that ML- and DL-based models are not only accurate but can also help in gaining new knowledge and understanding in radiobiology and medical physics.

## 2. Material and methods

### 2.1. The dataset

The development, training, and verification of ANAKIN are based on data from PIDE (Friedrich *et al* 2013b, 2021).

The MKM and LEM predictions are computed via the *survival toolkit*, (Attili and Manganaro 2018, Manganaro *et al* 2018). This toolkit is an open-source implementation that has been checked to be coherent with the published results of the models, but nonetheless, differences with the most advanced versions of the two formalisms may arise. Unfortunately, to date, no extensive and qualitative estimation of the MKM predictions over many cell lines exists so we could only rely on the *survival toolkit*. In particular, the MKM coupled with the amorphous track model, (Kase *et al* 2007), and the LEM III have been considered.

The PIDE database contains a series of cell survival experiments, conducted over a multitude of different irradiation conditions and cell lines. In addition to the original data, a set of LQ parameters are calculated for each experiment and is also reported. Following Pfuhl *et al* (2022), ANAKIN is thus trained over the exponential linear-quadratic fit on cell survival experiments. This is done as many experiments contained in the PIDE clearly show anomalous variability in the reported survival fraction. Experiments reporting less than 3 measurement points are removed from the dataset because at least 3 values are needed to fit an LQ curve. The dataset obtained from PIDE is then divided into two subclasses, one for training ANAKIN and one for testing its predictions. The selection is done so that each subset contains a sufficient amount of data for each cell line and ions to be statistically significant.

Unlike Pfuhl *et al* (2022), ANAKIN is trained on both monoenergetic and Spread-out Bragg-peak (SOBP) ion beams, and for this reason, a specific variable is added to the data to specify the irradiation condition.

After applying all the selection criteria described above, the resulting dataset contains 513 experiments, including 20 cell lines and 11 ion types, of which 333 were randomly assigned for training and the remaining 180 for testing. Figure 1(a) gives an overall point of view on the number of considered experiments for each cell-lines as well as ion type.

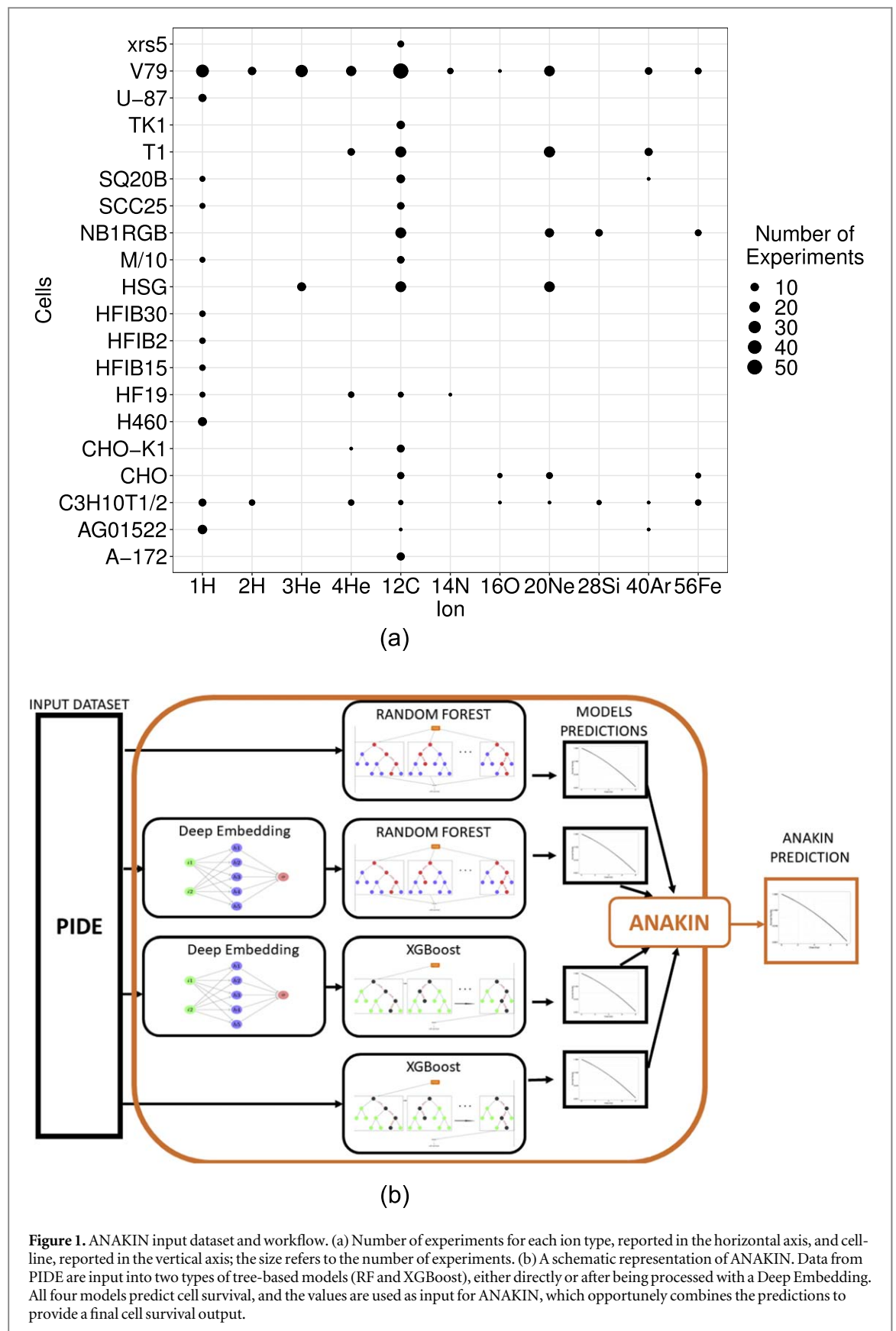
At the end of the cleaning of the data, we are left with 513 experiments, we randomly selected the 65% of the experiments, which correspond to 333 experiments, for training whereas the remaining 180 are used to test. It is worth stressing that the division between train and test has been performed over the experiments, meaning that ANAKIN is tested on experiments that have never been seen before by the model. All results that are shown in the present paper refer to the test set so that ANAKIN test reflects a realistic situation in which ANAKIN should predict the cell survival of an experiment or a real situation that has never been seen before.

ANAKIN takes as input 14 variables of both physical and biological parameters, either continuous, such as LET or energy, or discrete, such as ion type or cell-lines. The full list is reported in table 1. Variables names as reported in table 1 are taken from the PIDE and described in Friedrich *et al* (2013b, 2021). The only variable that has been added to the dataset is the square of the dose, named *Dose2*. The choice of considering the square of the dose is motivated by the well-known linear and quadratic form for the logarithm of survival. It is worth noticing that also  $\alpha_\gamma$  and  $\beta_\gamma$  values are passed as input to ANAKIN.

### 2.2. Machine and DL models

ANAKIN is an ensemble AI model composed of ML and DL modules, each with a different task, that together predicts the cell-survival probability. A schematic representation of ANAKIN is shown in figure 1(b). Four different tree-based models are trained on the PIDE: two RF (Ho 1995, 1998), and two XGBoost (Chen and Guestrin 2016a, 2016b). PIDE data are directly used as input for one RF and one XGBoost model, while for the other two they are first processed with the Deep Embedding (Micci-Barreca 2001, Guo and Berkahn 2016) *Neural Network* (NN), where categorical variables with high cardinality (in this case *Ion*, *Cells* and *CellCycle*) are pre-processed to learn a new meaningful data representation. Once the initial parameters are selected, (e.g. cell line, ion type, and kinetic energy), the survival is calculated with each of the four models, and these values are used as input to ANAKIN to predict the final cell survival.

Tree-based models have been chosen for the predictive modules rather than NN based algorithms because, to date, despite the groundbreaking impact that NN had on image detection, NN had a significantly less impact on tabular data; there are in fact several empirical pieces of evidence that standard ML approaches have comparable or even better results than NN, (Grinsztajn *et al* 2022). On the contrary, DL algorithms are used within ANAKIN in an innovative way to solve different tasks. As mentioned above, ANAKIN is trained to predict cell survival over many cell lines as well as ions. Such variables can assume only discrete values, typically referred to as *categorical variables* in the ML and DL community, and for this reason, they are not in principle easily handled by an ML or DL model. Even more problematic there is the fact that such categorical variables have a high number of possible values. This poses a serious issue in how these variables must be mapped to



**Figure 1.** ANAKIN input dataset and workflow. (a) Number of experiments for each ion type, reported in the horizontal axis, and cell-line, reported in the vertical axis; the size refers to the number of experiments. (b) A schematic representation of ANAKIN. Data from PIDE are input into two types of tree-based models (RF and XGBoost), either directly or after being processed with a Deep Embedding. All four models predict cell survival, and the values are used as input for ANAKIN, which opportunely combines the predictions to provide a final cell survival output.

numeric values to be efficiently treated by an ML model. Several possible solutions to the above problem exist, (Seger 2018), but recently, DL has gained huge attention not as solely a predictive tool but also as an extremely powerful data pre-processing tool, used for instance as a model to extract new information from data. For instance, DL has been recently proposed to specifically treat categorical variables with a high number of values. Such technique is called *Deep Embedding*, (Micci-Barreca 2001, Guo and Berkhahn 2016, Shreyas 2022), and



**Table 1.** List of all variables used as input to ANAKIN. The names are described according to PIDE documentation (Friedrich *et al* 2013b, 2021) with the exception of Dose2, which represents the square of the dose value and has been introduced in this work. Domain column refers to the origin of the variable considered, such as *physics* (p) or *biology* (b). Type refers to the type of the variable, that can be either *continuous* (c) or *discrete* (d). The other columns reports a description of the considered variable as well as a brief summary of statistics.

Variable	Domain	Type	Description	Statistics
Dose	p	c	Dose in Gy	Min value = 0 Gy Max value = 14 Gy
Dose2	p	c	Square of the dose in Gy	Min value = 0 Gy Max value = 196 Gy
LET	p	c	Linear Energy Transfer in keV $\mu\text{m}^{-1}$	Min value = 0.9 keV $\mu\text{m}^{-1}$ Max value = 2160 keV $\mu\text{m}^{-1}$
Energy	p	c	Specific energy of the ion in MeV $\text{u}^{-1}$	Min value = 0.275 MeV $\text{u}^{-1}$ max value = 680 MeV $\text{u}^{-1}$
Ion	p	d	Ion species	11 classes Most frequent: carbon ion (186 exp.) Least frequent: oxygen (4 exp.)
Charge	p	d	Charge of the ion	9 classes Most frequent: 6 (186 exp.) Least frequent: 8 (4 exp.)
IrradiationConditions	p	d	Irradiation modalities: monoenergetic (m) or SOBP (s)	2 classes Most frequent: m (433 exp.) Least frequent: s (80 exp.)
Cells	b	d	Cell line used	20 classes Most frequent: V79 (182 exp.) Least frequent: HGIB2 (4 exp.)
CellClass	b	d	Tumor cells (t) or normal cells (n)	2 classes Most frequent: n (383 exp.) Least frequent: t (130 exp.)
CellOrigin	b	d	Human cells (h) or rodent cells (r)	2 classes Most frequent: h (274 exp.) Least frequent: r (239 exp.)
CellCycle	b	d	Cycle of cells	5 classes Most frequent: a (491 exp.) Least frequent: G0/G1 2 exp.
DNAContent	b	d	Genomic length of diploid cells	2 classes Most frequent: 6 (274 exp.) Least frequent: 5.6 239 exp.
ax	b	c	Alpha parameter of reference radiation	Min value = 0.03 $\text{Gy}^{-1}$ Max value = 0.82 $^{-1}$
bx	b	c	Beta parameter of reference radiation	Min value = 0 $\text{Gy}^{-2}$ Max value = 0.11 $\text{Gy}^{-2}$

consists in training a NN that learns the most efficient way of encoding a categorical variable, such as in the present case the cell-line or also the ion type, into a low-dimensional numerical vector that can be efficiently used by another model to understand the most accurate relation between these variables and the target variable to predict. Therefore, ANAKIN has three specifically devoted modules to learn a new data representation for the cell lines, ion type, and also cell cycle. The DL-based Deep Encoding modules are connected to the previously mentioned tree-based predictive modules to create ANAKIN, the final ensemble model that takes each single module output and predicts the cell survival fraction.

Each input model has been validated using a 10-fold cross-validation, and their hyper-parameters have been obtained using a Bayesian optimization technique, as described in Missiaggia *et al* (2022b).

Consider dataset  $\mathcal{D} := \{\mathcal{X}_i, y_i\}_{i=1}^N$  composed by  $N$  samples, where

$$\mathcal{X}_i := (x_1^i, \dots, x_n^i), \quad n \in \mathbb{N}, \quad i = 1, \dots, N < \infty,$$

are the  $n$  input features on which a model is trained to predict the target variable  $y_i \in \mathbb{R}$ . In the current case,  $\mathcal{X}_i$  are the variables reported in table 1, whereas  $y_i$  is the cell survival.

Given a set of parameter  $W$ , that depends on the model, and a suitably chosen training set  $\mathcal{T} := \{\mathcal{X}_i, y_i\}_{i=1}^{N_{\mathcal{T}}}$ ,  $N_{\mathcal{T}} < N$ , the aim of the ML or DL models is to solve the following optimization problem

$$\min_W \sum_{i=1}^N \mathcal{L}(\mathcal{Y}_i, \hat{\phi}(\mathcal{X}_i; W)),$$

being  $\hat{\phi}$  the output function for the model and  $\mathcal{L}$  the loss function. To improve the accuracy and reduce the overfitting, a regularization is added to the loss function, as done in Bishop *et al* (1995), LeCun *et al* (2015).

The output function  $\hat{\phi}$  is the learned function approximating the ideal function  $\phi$ , that describes the link between the features  $\mathcal{X}_i$  and the target  $y_i$ .

### 2.2.1. Ensemble tree-based models: RF and XGBoost

*Random Forest* (RF) is an ensemble ML algorithm that combines weaker models, such as decision trees, to create a more robust final model (Ho 1995, 1998). Being a bagging algorithm, the ensemble model is created in parallel, and thus the output is the average of all the outcomes. Compared to decision trees, the RF reduces the overfitting on the train data, and thus it improves the prediction accuracy.

RF (Ho 1995, Friedman *et al* 2001) assumes that  $\hat{\phi}$  is the average of weaker learners decision-trees  $\psi_k$ , that is

$$\hat{\phi}(x_1^i, \dots, x_n^i) = \frac{1}{K} \sum_{k=1}^K \psi_k(x_1^i, \dots, x_n^i),$$

where  $\psi_k$  is the outcome of the  $k$ th decision tree.

Like RF, also XGBoost is an ensemble ML algorithm that combines weaker decision tree models to create a more robust final model (Chen and Guestrin 2016a, 2016b). XGBoost is a boosting algorithm so that the ensemble model is created in series, and thus the output of every single model is passed to another, with the aim of reducing the error of the previous one. Also bagging is mostly used to reduce the overfitting of the train data and to improve the accuracy of the predictions.

XGBoost starts with a potentially inaccurate model

$$\hat{\phi}_0(x; \bar{W}) = \arg \min_W \sum_{i=1}^N \mathcal{L}(\mathcal{Y}_i, \hat{\phi}(\mathcal{X}_i; W)), \quad (1)$$

and then it thus expanded in a greedy fashion as

$$\hat{\phi}_m(x; \bar{W}) = \hat{\phi}_{m-1}(x; \bar{W}) + \arg \min_W \left[ \sum_{i=1}^N \mathcal{L}(\mathcal{Y}_i, \hat{\phi}_{m-1}(x; W) + \hat{\phi}(\mathcal{X}_i; W)) \right]. \quad (2)$$

### 2.2.2. Deep embedding

Deep Embedding, (Micci-Barreca 2001, Guo and Berkhahn 2016, Shreyas 2022) is a NN- based technique for mapping a categorical variable into a vector. Being a supervised algorithm, the NN is trained to predict the cell survival fraction. Thus, the intermediate representation learned by the network is extracted and constitutes the new values used for the categorical variable.

In the context of NN, the  $W$  parameters defined in equations (1)–(2) are usually referred to as weights. In this work, we chose the *multilayer perceptron* (MLP) NN, which is the first and most classical type of network used.

A Multi-Layer Perceptron (MLP) is created by connecting several single-layer perceptrons, where several nodes are placed in a unique layer. The inputs  $(x_1, \dots, x_n)$  are fed to the network so that the final output  $z$  is produced. Typically the output is a nonlinear function of a weighted average of the input, i.e.

$$y = \hat{\phi}(x) = \sigma \left( \sum_{i=1}^n w_i x_i + b \right),$$

where  $w_i$  are the weights and  $b$  is the bias. Also,  $\sigma$  is a suitable (possibly) nonlinear function, like a sigmoid

$$\sigma(z) = \frac{1}{1 + e^{-z}}.$$

The connection between single-layer perceptrons is done in a preferred direction. This type of network is called feedforward because the inputs are fed to the first layer, then the output goes to the second layer, and so on until the data reaches the last output layer. By providing a series of correct results to the network, and thus making the problem supervised, the NN can learn the best weights and bias to reproduce any desired output (Bishop *et al* 1995, LeCun *et al* 2015).

## 2.3. Explainable AI

Several XAI techniques (Molnar 2020, Biecek and Burzykowski 2021) can be used to understand how an ML model work. In the present work, we focus on three specific very well-known and powerful techniques, namely (i) variable importance, (ii) Accumulated Local Effect (ALE) plot, and (iii) the Shapley value.



### 2.3.1. Variable importance

Variable importance (Breiman 2001, Fisher *et al* 2019) measures the global importance of each feature to the final output of the model. The main idea behind the calculation is that, if a variable is important for calculating the final output of the model, then after a permutation of the variable values, the model performance significantly decreases. Larger changes in the overall model performance are then associated with highly important features.

### 2.3.2. ALE plot

Accumulated Local Effect (ALE) plot, (Apley and Zhu 2020, Grömping 2020), is one of the most advanced and robust dependence plots for describing how variables influence on average predictions of an ML model. One of the most advanced aspects of this model is that it accounts for the correlation between variables. ALE plot thus calculates the average changes in the model prediction and sum (accumulate) them over the values assumed by a specific variable.

Ale plot is defined as Apley and Zhu (2020)

$$\hat{f}_{\text{ALE}}(x_1) := \int_{x_{\min}}^{x_1} \int \frac{\partial \hat{\phi}(z_1, x_2)}{\partial z_1} p(x_2|z_1) dx_2 dz_1 - \text{constant}. \quad (3)$$

Instead of considering the effect of the prediction  $\hat{\phi}$ , the ALE plot considers changes in the prediction  $\frac{\partial \hat{\phi}(z_1, x_2)}{\partial z_1}$ , which represents the local effect of the variable. This is averaged over all possible values of the other variable  $x_2$ , weighted by the actual probability of registering the value  $x_2$  given the considered value  $x_1$ . Then, the result is integrated, or accumulated, up to  $x_1$ . This value is centered around the average prediction, represented by the constant appearing in equation (3), so that the average effect over the data is 0. Therefore, ALE plots calculate the average difference in the prediction to be imputed to a local change in a variable.

### 2.3.3. The SHAP value

The Shapley Additive exPlanation (SHAP) value, (Lundberg and Lee 2017), is a local XAI technique extremely powerful that aims at explaining individual predictions and in particular what is the contribution of every single variable to the overall prediction. The SHAP method computes Shapley values (Hart 1989) as an additive feature attribution, alike a linear model, so that the prediction is decomposed as

$$\hat{\phi}(x) = \varphi_0 + \sum_{i=1}^n \varphi_i,$$

where  $\varphi_i$  represents the contribution of the  $i$ th feature and  $\varphi_0$  is an intercept.

## 2.4. Error assessment

To provide a comprehensive and accurate assessment of ANAKIN performances, many metrics are used throughout this paper. In order to compare cell survival fractions, for each experiment we computed the *logarithmic Root Mean Square Error* (logRMSE), defined as

$$\text{logRMSE}^i := \sqrt{\frac{1}{N_D} \sum_D (\log \hat{S}^i(D) - \log S^i(D))^2},$$

where  $D$  is the dose and  $N_D$  is the number of doses measured in the  $i$ th experiment.  $\hat{S}^i$  and  $S^i$  are the cell survivals predicted and measured, respectively. In the paper, the average and standard deviation of all the errors used are calculated by averaging the results for experiments included in the test set.

The RBE at the survival level  $\rho$  is defined as

$$\text{RBE}_\rho = \frac{\sqrt{\alpha_\gamma^2 - 4\beta_\gamma \log \rho} - \alpha_X}{2\beta_\gamma D},$$

where  $D$  is the dose giving  $\rho$  survival fraction. Also, we denote

$$\text{RBE}_\alpha = \frac{\alpha_{\text{ion}}}{\alpha_\gamma}, \quad \text{RBE}_\beta := \sqrt{\frac{\beta_{\text{ion}}}{\beta_\gamma}},$$

where  $\alpha_\gamma$  and  $\beta_\gamma$  represent the  $\alpha$  and  $\beta$  value for the reference radiation, respectively. We specifically consider three survival levels at  $\rho = 0.5, 0.1, 0.01$ . In the paper, we focus on  $\text{RBE}_{0.1}$  predictions, as this is the main value used in radiobiology for particle therapy.

The comparison of RBE measured or calculated with ANAKIN is investigated using the Mean Absolute Error (MAE) metric

$$\text{MAE} := \frac{1}{N} \sum_{i=1}^N |\overline{\text{RBE}}_{\rho}^i - \text{RBE}_{\rho}^i|,$$

and the Mean Absolute Percentage Error (MAPE) metric

$$\text{MAPE} := \frac{1}{N} \sum_{i=1}^N \frac{|\overline{\text{RBE}}_{\rho}^i - \text{RBE}_{\rho}^i|}{\text{RBE}_{\rho}^i}.$$

$\overline{\text{RBE}}_{\rho}^i$  and  $\text{RBE}_{\rho}^i$  represent ANAKIN values and measurements, respectively, for the endpoint  $\rho = 0.5, 0.1, 0.01$   $\alpha, \beta$  and the  $i$ th experiment. Since the range of RBE is extremely wide, the two metrics are often used together to provide a better evaluation of the performances of ANAKIN.

We also calculated the MAE values of  $\alpha$  and  $\beta$  as

$$\text{MAE}_{\alpha} := \frac{1}{N} \sum_{i=1}^N |\bar{\alpha}_{\text{ion}}^i - \alpha_{\text{ion}}^i|, \quad \text{MAE}_{\beta} := \frac{1}{N} \sum_{i=1}^N |\bar{\beta}_{\text{ion}}^i - \beta_{\text{ion}}^i|,$$

where  $\bar{\alpha}_{\text{ion}}^i$  and  $\bar{\beta}_{\text{ion}}^i$  are the predicted  $\alpha$  and  $\beta$  values for the  $i$ th experiment, whereas  $\alpha_{\text{ion}}^i$  and  $\beta_{\text{ion}}^i$  are the measured data. For those quantities, the MAPE values were not calculated, as the absolute value of both  $\alpha$  and  $\beta$  were close to 0.

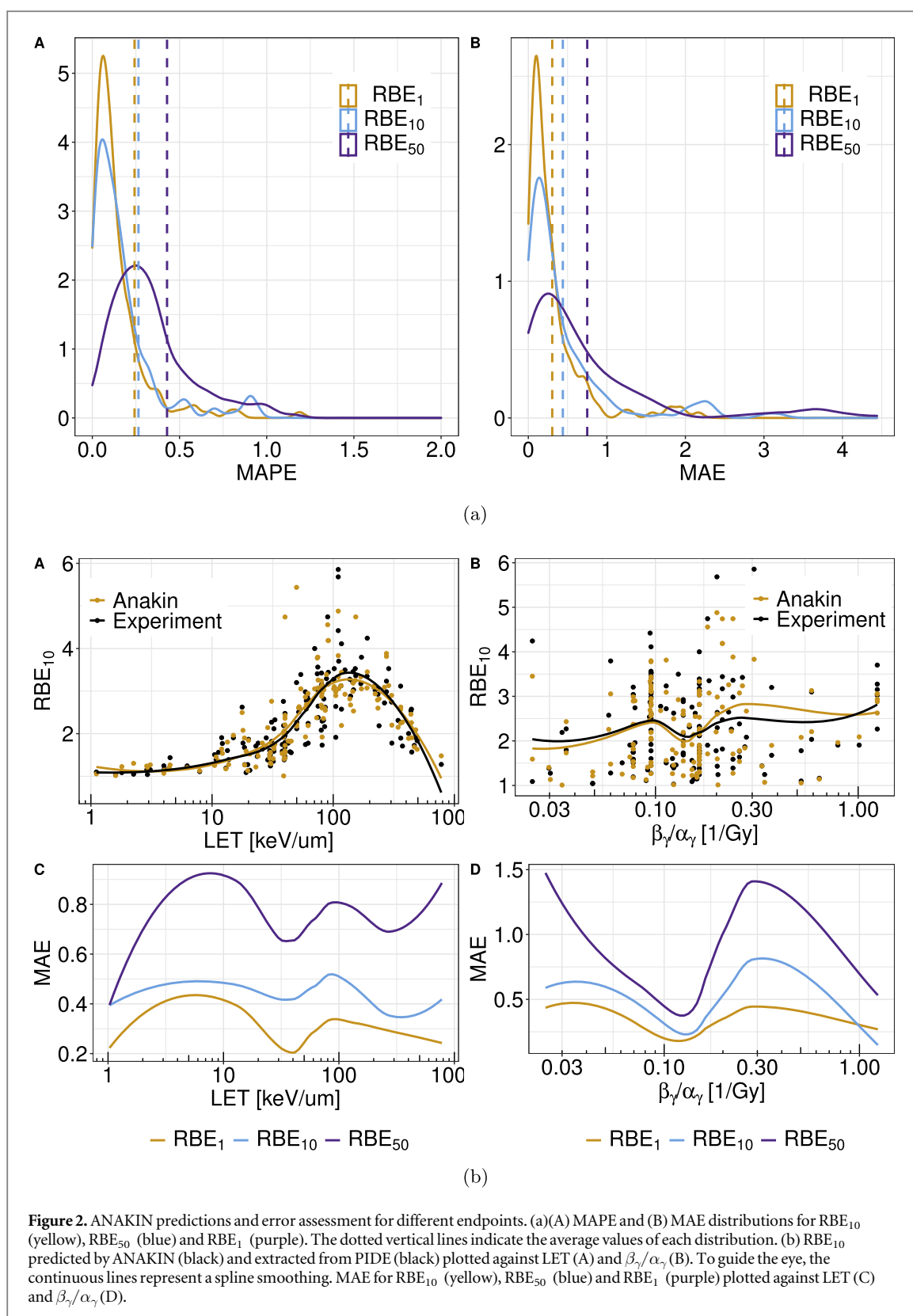
### 3. Results

Results of the current work include a quantitative and comprehensive analysis of the comparison between ANAKIN cell survival predictions and experimental measurements available in PIDE. A wide range of possible metrics, such as RBE at different cell survival probabilities,  $\alpha$  and  $\beta$  predictions as well as the cell survival at different doses are presented. A detailed description of the used metrics is reported in section 2.4.

Figure 2(a) shows (A) MAPE and (B) MAE values (section 2.4) for  $\text{RBE}_{10}$ ,  $\text{RBE}_{50}$  and  $\text{RBE}_1$ , while the numerical values, as well as the logRMSE, are reported in table 2. The results indicate that ANAKIN has similar errors for different endpoints, with  $\text{RBE}_{50}$  exhibiting a slightly higher MAE and MAPE than  $\text{RBE}_{10}$  and  $\text{RBE}_1$ .

Measured  $\text{RBE}_{10}$  values and ANAKIN predictions are reported in figure 2(b) as a function of the LET and  $\beta_{\gamma}/\alpha_{\gamma}$ . In addition, the MAE for  $\text{RBE}_{10}$ ,  $\text{RBE}_{50}$  and  $\text{RBE}_1$  are plotted against LET and  $\beta_{\gamma}/\alpha_{\gamma}$ . The results indicate an excellent agreement between  $\text{RBE}_{10}$  ANAKIN and the experimental data over the entire range of LET. The smoothing spline of the  $\text{RBE}_{10}$  predicted as a function of LET completely overlaps with the experimental curve. despite this is not necessarily a proof of a perfect agreement, it is nonetheless clear that the experimental trend is predicted by ANAKIN. A good agreement can be also seen by analyzing each experiment's results. This is also supported by the MAE for the other endpoints (figures 2(b) (C) and table 2), which remains mostly constant around for  $\text{LET} > 10 \text{ keV } \mu\text{m}^{-1}$ . Concerning errors as a function of  $\alpha_{\gamma}/\beta_{\gamma}$ , there is a higher variability than observed for LET. The discrepancy observed in the spline smoothing at high  $\beta_{\gamma}/\alpha_{\gamma}$  seems an artifact of the smoothing procedure, as it is not reflected in the MAE (panel (D)). On the contrary, at low  $\beta_{\gamma}/\alpha_{\gamma}$ , i.e. for high  $\alpha_{\gamma}/\beta_{\gamma}$  cell-lines, ANAKIN clearly underestimates the  $\text{RBE}_{10}$ , as it is also indicated by the high MAE in the low  $\beta_{\gamma}/\alpha_{\gamma}$  region. Figure 3 shows the experimental  $\text{RBE}_{10}$  against ANAKIN prediction. The results are sharply distributed around the bisector representing the ideal perfect prediction. The deviation between the bisector and the model prediction increases as the RBE grows. Figure 4 reports ANAKIN  $\text{RBE}_{10}$  predictions compared to the measurements, plotted against LET for 4 different ions (protons, helium, carbon and iron) in a very broad LET range. Overall, ANAKIN seems to reproduce well the experimental data. For protons, ANAKIN can reproduce the small RBE variability at low LET as well as the clear rise above  $20 \text{ keV } \mu\text{m}^{-1}$ . ANAKIN is accurate also for helium and carbon ions, and it is clearly able to reproduce the overkilling effect, which yields a decrease in the  $\text{RBE}_{10}$  around  $100 \text{ keV } \mu\text{m}^{-1}$ . ANAKIN values appear to be very close to the measurements also for iron.

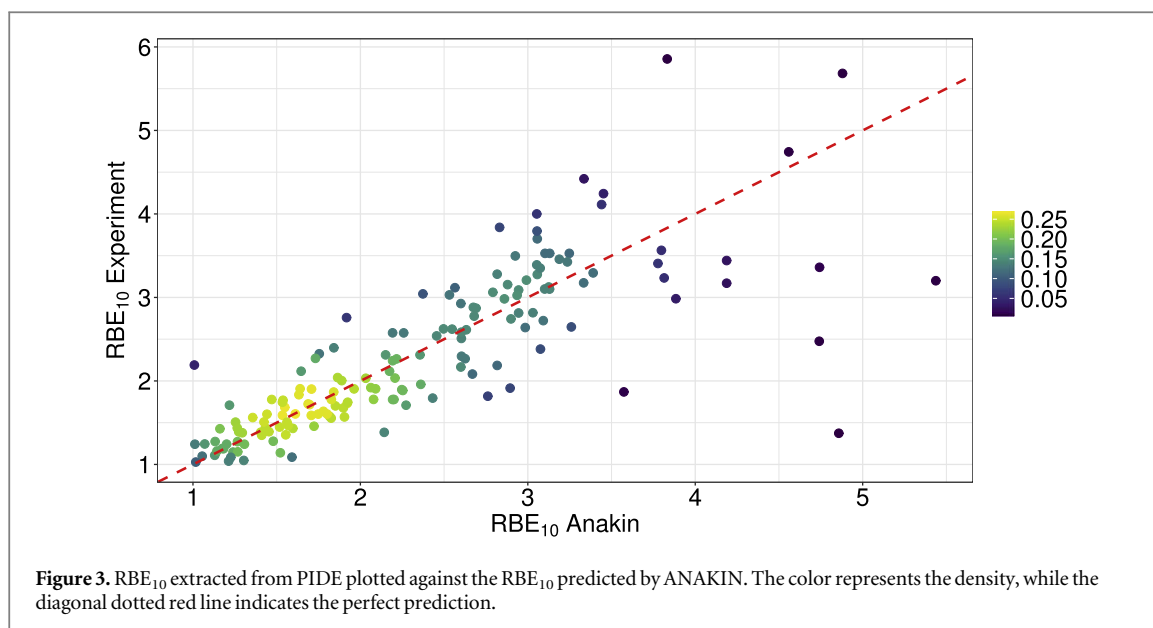
Similar conclusions can be drawn from figure 5(a). Iron shows a strongly peaked distribution because of the low number of available experiments; nonetheless, iron a low error in both metrics. Besides iron, the other ions show comparable results, with helium having a broader distribution in both MAE and MAPE reflecting a lower accuracy of ANAKIN. Protons exhibit an error distribution peaked around the average values, as well as some outliers with high errors, as clearly indicated by the spikes in the high errors region. However, these peaks are for MAPE, and we hypothesize that they might be mainly caused by low RBE values, which can result in high percentage errors. Figure 5(b) shows MAE and MAE error distributions evaluated for  $\text{RBE}_{10}$ , grouped by monoenergetic beam and SOBP. The peak of the distributions is similar for both cases, but the error distribution for the monoenergetic beams is clearly broader than for the SOBP. Figure 6 shows  $\text{RBE}_{\alpha}$  and  $\text{RBE}_{\beta}$  plotted against LET and  $\beta_{\gamma}/\alpha_{\gamma}$ .  $\text{RBE}_{\alpha}$  values are accurately predicted by ANAKIN independently of LET and  $\beta_{\gamma}/\alpha_{\gamma}$ . A higher inaccuracy is observed for  $\text{RBE}_{\beta}$  in the low  $\beta_{\gamma}/\alpha_{\gamma}$  region. The absolute errors in the  $\alpha$  and  $\beta$  predictions



show a steady behavior over the LET range (panel (E)), while the errors on the  $\alpha$  values clearly decrease as  $\beta_\gamma/\alpha_\gamma$  increases, coherently with previous analysis performed above.

### 3.1. Comparison with MKM and LEM

To further assess the accuracy of ANAKIN in predicting cell survival and RBE, we compared it with the only two RBE models that are currently used in clinical practice, namely the MKM (Hawkins 1994, Inaniwa *et al* 2010, Inaniwa and Kanematsu 2018, Bellinzona *et al* 2021) and LEM (Krämer *et al* 2000, Elsässer and Scholz 2007,



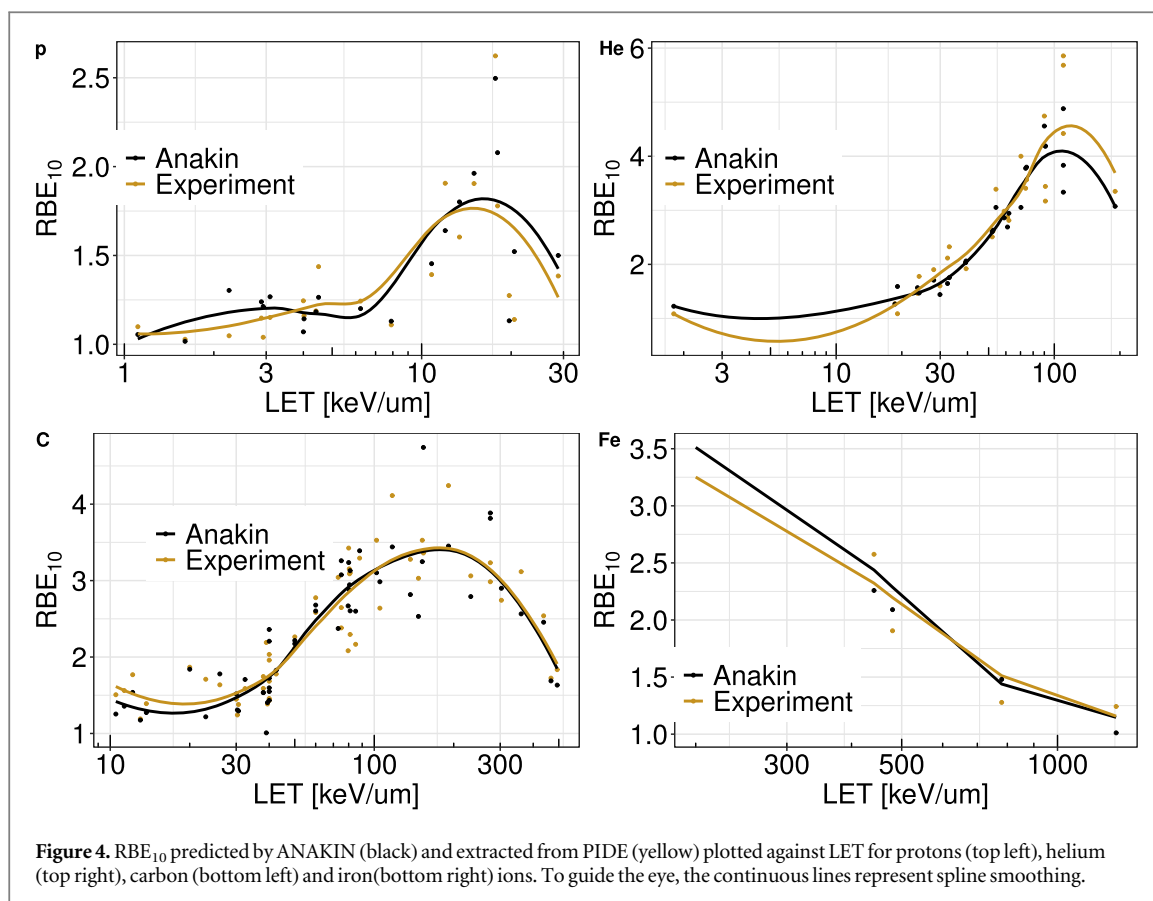
**Table 2.** Average errors and standard deviations of different error metrics and endpoints.

Endpoint	Error	Mean	Sd
$RBE_{10}$	logRMSE	1.06	1.26
	MAE	0.43	0.58
	MAPE	0.26	0.80
$RBE_1$	MAE	0.24	0.74
	MAPE	0.25	0.31
$RBE_{50}$	MAE	0.74	0.91
	MAPE	0.42	0.81
$\alpha$	MAE	0.24	0.24
$\beta$	MAE	0.03	0.05
$RBE_\alpha$	MAE	0.73	4
	MAPE	0.4	1.37
$RBE_\beta$	MAE	0.43	0.65
	MAPE	0.43	0.59

Elsässer *et al* 2008, Pfuhl *et al* 2022). To calculate the biological outcomes from the MKM and LEM III, we used the *survival toolkit* (Manganaro *et al* 2018). We performed the comparison for the HSG and V79 cell lines because they are among the most used in radiobiological experiments, and several datasets are available in the literature. For the V79 cell lines, we used 41 different experiments with proton, helium, and carbon ions, while for the HSG cell line, we included 15 experiments conducted with helium and carbon ions. To compare the models, the same metrics introduced in section 2.4 are used.

Predictions with MKM and LEM have been performed with the *survival toolkit* (Manganaro *et al* 2018, Attili and Manganaro 2018) including the implementation of a limited number of versions for the latter models. In particular, a newer version of the LEM, namely the LEM IV (Elsässer *et al* 2010), has been recently developed but it has not been used in the current study since a freely usable version is not available. For the LEM, we used version III, as the latest version (IV) is not available. However, an extensive quantitative study has been published (Pfuhl *et al* 2022), so a further quantitative comparison between the LEM IV accuracy with ANAKIN can be conducted.

The comparison between the models is shown in figures 7(a)–(b), while the numerical values are reported in table 3. The results indicate that overall ANAKIN is more accurate than both the MKM and the LEM in predicting all metrics and endpoints considered. For both cell lines, the LEM shows the largest deviations from the measurements, closely followed by the MKM with ANAKIN reporting lower errors. In particular, for the HSG cell-line, the LEM shows an MAE for the  $RBE_{10}$  of 1.55, whilst the MKM and ANAKIN have respectively a MAE of 1.18 and 0.43. For the V79 cell line, the MKM and the LEM predict comparable results with an MAE for  $RBE_{10}$  of 1.5 and 1.2. ANAKIN shows an MAE for  $RBE_{10}$  of 0.43. Other endpoints and metrics have comparable



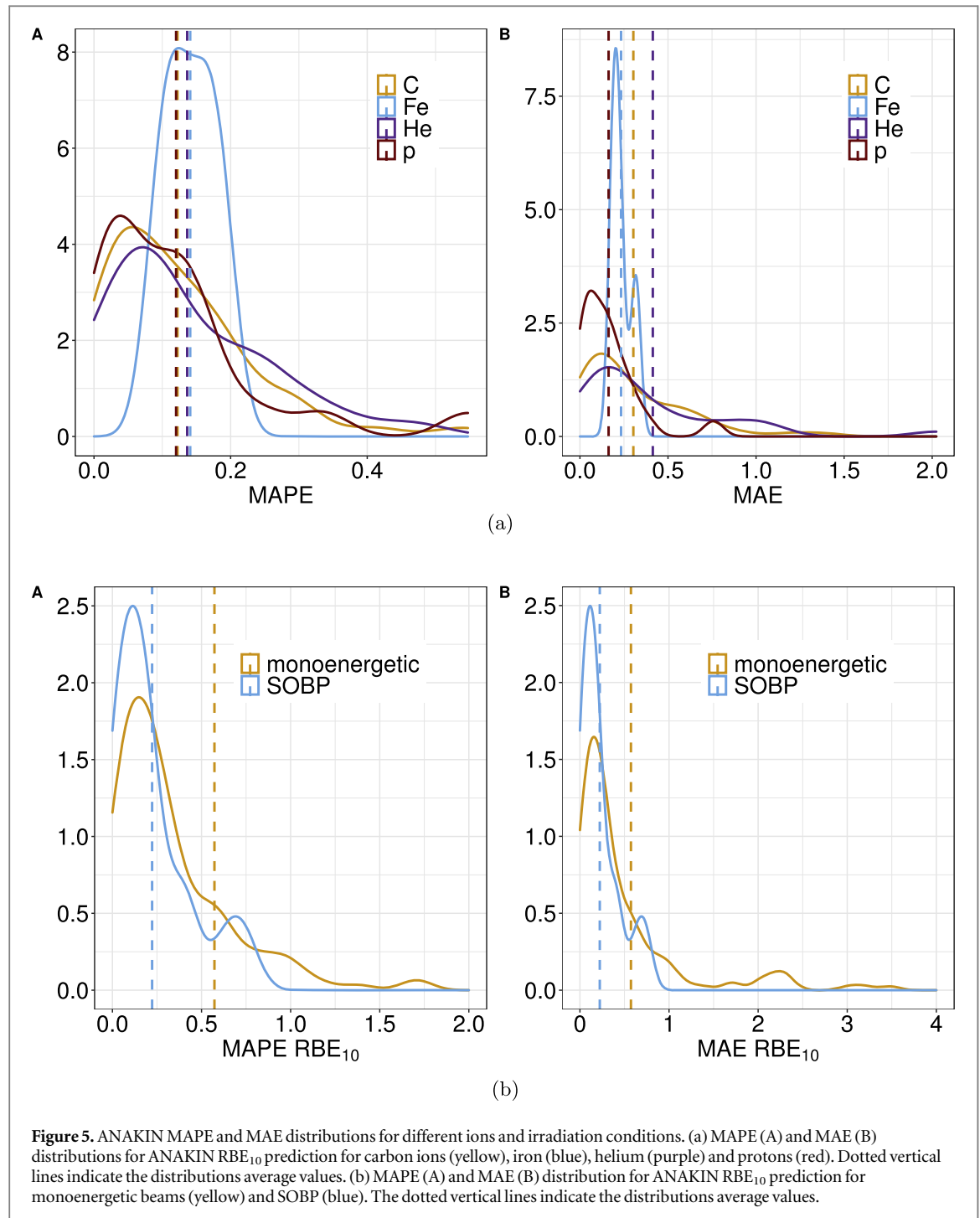
results. Together with having lower average errors, ANAKIN exhibits narrower error distributions and does never reach absolute errors as high as the MKM and LEM. Figures 8–9 illustrates the absolute difference of MAE between ANAKIN and MKM or LEM, calculated for the RBE<sub>10</sub> of both cell lines. All experimental datasets were obtained from PIDE.

This analysis confirms the results shown in figures 7(a)–(b) and table 3. ANAKIN is more accurate in predicting the selected biological outputs than both the MKM and LEM for the majority of experiments. The maximum discrepancy between MAE is significantly higher when ANAKIN is closer to the experimental data (yellow dots), reaching differences above 6 for V79 cell lines, but only 2 for the opposite case. This result indicates that for the cases where ANAKIN is less accurate than the other models, its error is smaller than when the predictions of the MKM and LEM III are off.

### 3.2. Explainable artificial intelligence

Figure 10(a) shows the global importance of all variables included in ANAKIN, calculated over the whole test dataset. The plot suggests that the dose is by far the most relevant parameter, followed by  $\beta_\gamma$ , LET, and Dose<sup>2</sup> (the dose squared), which are all close together. This finding indicates that ANAKIN uses both physical and biological variables to predict the biological outcome. The *Ions Cells* variables, on which a categorical embedding has been performed, denote the ion type and cell line, respectively, and have both a high impact on ANAKIN.

To have a better understanding of ANAKIN global behavior, and in particular of the correlation between LET and the predicted survival, we calculated the ALE plot as described in section 2.3 and reported in figure 10(b) as a function of the LET. To obtain an unbiased effect, the ALE has been evaluated at the same dose of 2 Gy for all the experiments. The typical trend of the overkilling effect is clearly visible: figure 10(b) implies that small positive variations in the LET yields a clear negative variation in the cell survival, with a consequent increase of the RBE, up to 100 keV  $\mu\text{m}^{-1}$ , after which the cell survival starts increasing again as the LET increases, with therefore a decrease in the RBE. Figure 11(a) reports the SHAP value for each experiment plotted against LET, considering a fixed dose of 2 Gy. Unlike the ALE plot, the SHAP value is a local technique, namely the SHAP is evaluated for each individual input variable, and thus figure 11(a) shows the importance of LET in the overall cell survival assessment, evaluating it for each experiment. As for the ALE plot, the typical behavior of the overkilling effect emerges. Protons exhibit a high positive SHAP value, which remains almost constant up to 15 keV  $\mu\text{m}^{-1}$ . As the LET rises above 15 keV  $\mu\text{m}^{-1}$ , protons shows a steep increase in the RBE and the SHAP value for LET drops. In this region, especially for low-energy protons, the kinetic energy becomes more



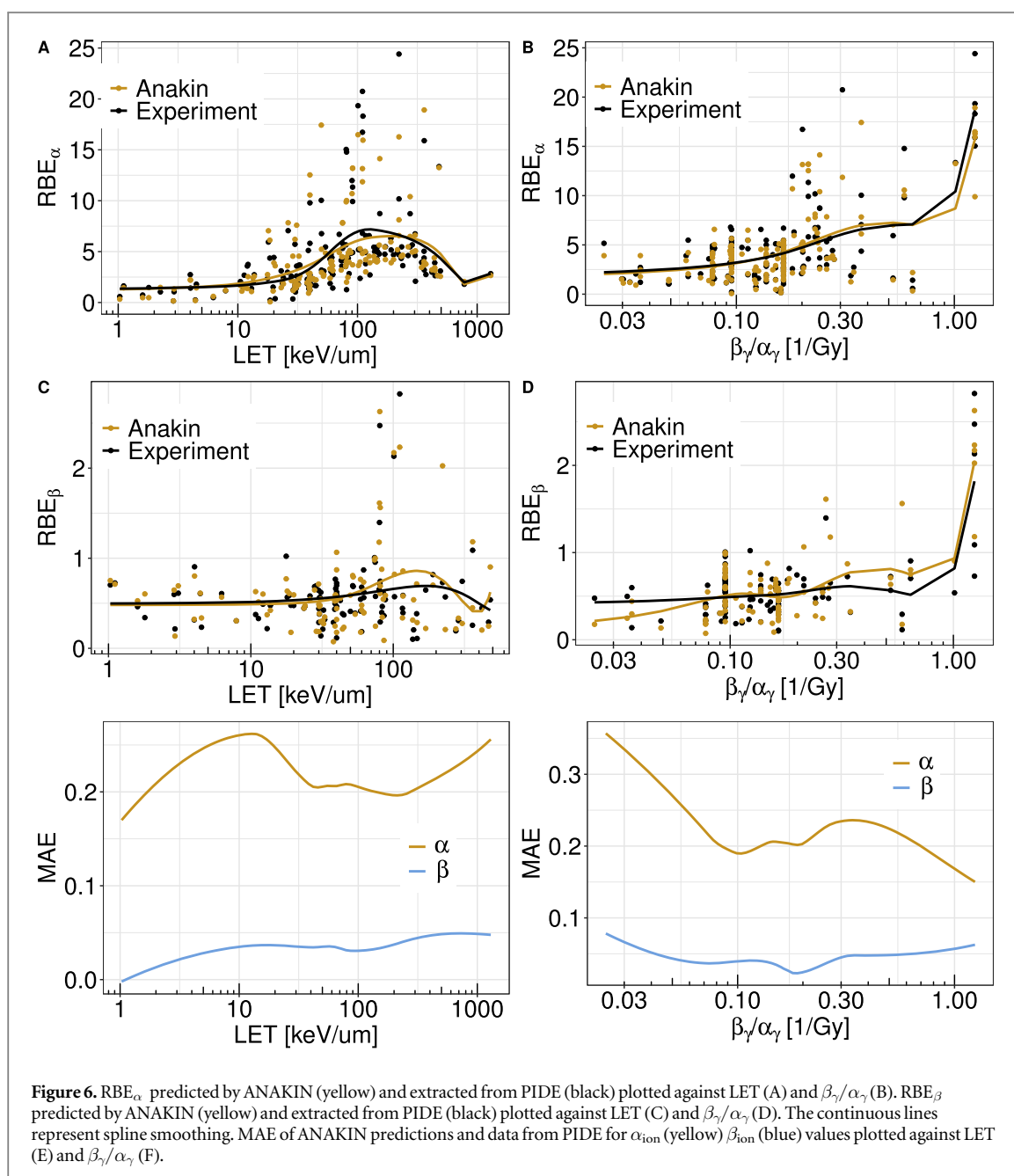
**Figure 5.** ANAKIN MAPE and MAE distributions for different ions and irradiation conditions. (a) MAPE (A) and MAE (B) distributions for ANAKIN RBE<sub>10</sub> prediction for carbon ions (yellow), iron (blue), helium (purple) and protons (red). Dotted vertical lines indicate the distributions average values. (b) MAPE (A) and MAE (B) distribution for ANAKIN RBE<sub>10</sub> prediction for monoenergetic beams (yellow) and SOBP (blue). The dotted vertical lines indicate the distributions average values.

important than the LET to predict cell survival (figure 12(b)). The SHAP value decreases steadily up to  $100 \text{ keV } \mu\text{m}^{-1}$ , after which it starts increasing again, reproducing the typical shape due to the overkill effect. Figure 11(b) shows the SHAP value for  $\alpha_\gamma$  and  $\beta_\gamma$ . The SHAP value for  $\alpha_\gamma$  shows that low  $\alpha_\gamma$  has a positive but almost equal importance to the model, but as  $\alpha_\gamma$  increases and  $\alpha_\gamma/\beta_\gamma$  goes over 5 Gy the SHAP values linearly decrease to have at last negative high values.

For  $\alpha_\gamma$  below a certain threshold, that coincides with cell lines with low  $\alpha_\gamma/\beta_\gamma < 5$ , the SHAP is positive, and then it starts decreasing linearly with  $\alpha_\gamma$ , reaching high negative values for high  $\alpha_\gamma$  and  $\alpha_\gamma/\beta_\gamma$ . A similar trend is shown by the  $\beta_\gamma$  SHAP values. For low  $\beta_\gamma$  and high  $\alpha_\gamma/\beta_\gamma$ , the SHAP value is positive, and then it begins to diminish. The data also indicate that the SHAP values for  $\beta_\gamma$  show both higher variability and higher absolute values than those for  $\alpha_\gamma$ .

Finally, we performed a comparison of experiments considering the SHAP values. Figure 12(a) shows the SHAP values for ANAKIN input features for two experiments performed with protons of similar LET of  $18 \text{ keV } \mu\text{m}^{-1}$  and  $19 \text{ keV } \mu\text{m}^{-1}$  for different cell lines. The corresponding RBE<sub>10</sub> values of the two experiments are

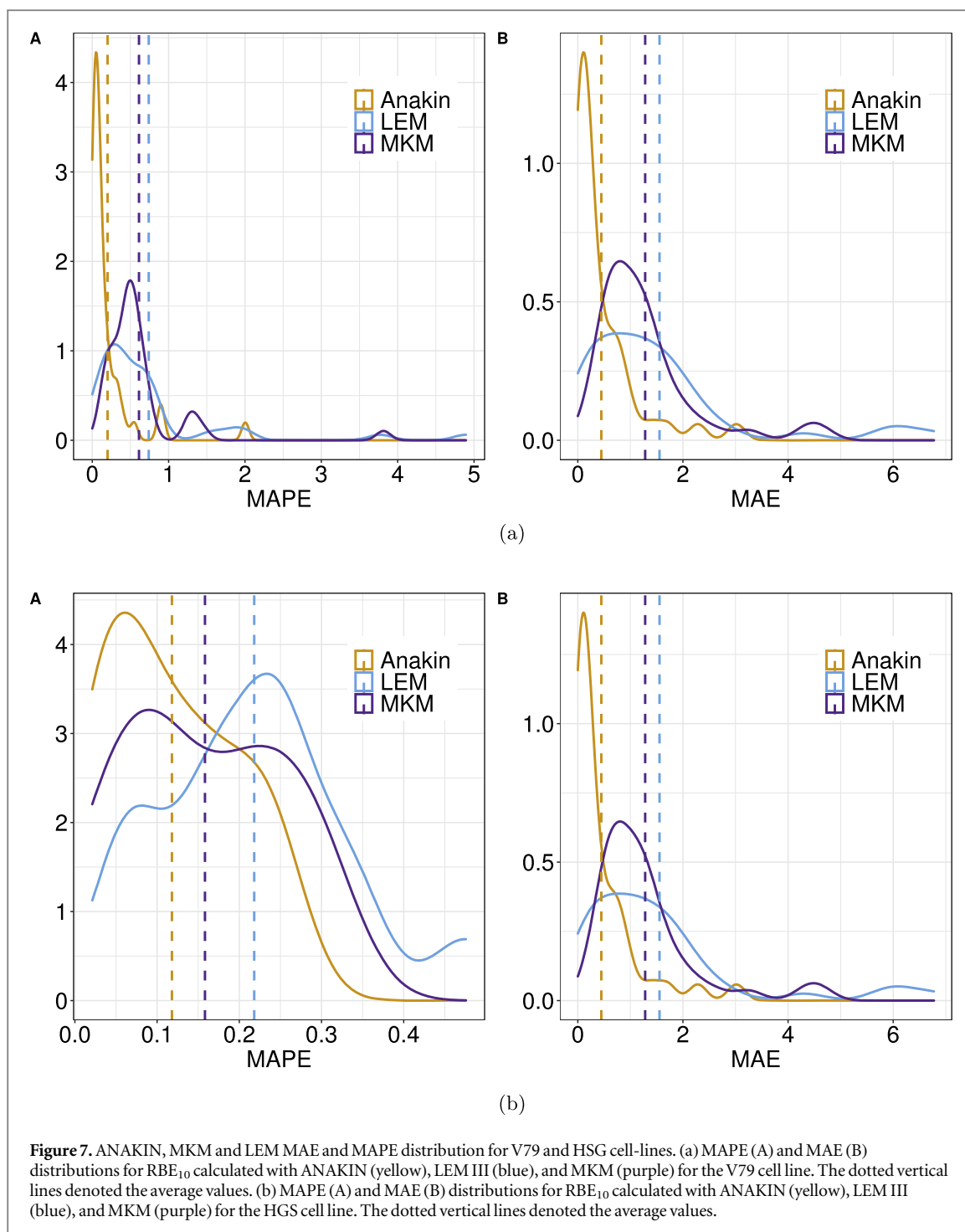




significantly different, being 1.2 and 2.7, respectively, as can be seen in figure 4 panel (A). ANAKIN outputs are extremely accurate for both experiments, being 1.13 and 2.5, with an MAE of 0.07 and 0.1 and a MAPE of 0.05 and 0.03, respectively. Figure 12(a) suggests that the only variables showing a significant difference between the two experiments are the  $\alpha_\gamma$  and  $\beta_\gamma$ , as it should be since the two experiments have been performed over different cell lines. Figure 12(b) compares the SHAP values for 4 different experiments, performed with either protons of helium of different LET (high or low). The rationale for the selection of the measurement is to test ANAKIN for different ions and LET. Besides differences in the cell-lines-specific parameters, focusing only on ion-specific variables, it can be seen how LET and energies are treated significantly differently. The SHAP value for LET is high and positive for both protons datasets and for low-LET helium, while is negative for high-LET helium. The SHAP related to the beam energy is positive and close to 0 for both particles when the LET is low, it is negative and close to 0 for high-LET helium, and negative with a high absolute value for high-LET protons.

#### 4. Discussion

The results reported in section 3 show that ANAKIN produces accurate results over a wide range of biological endpoints, beams of different particle species and energies, with a consistent behavior for different error metrics. Despite the fact that logRMSE is the most robust metric, being able to detect discrepancies between the predicted



cell survival and the experiments at different doses, to be easily comparable to existing radiobiological model, most of the analysis of ANAKIN results has been conducted for RBE. As often in predictive analysis, the choice of the metric is of fundamental importance and strongly depends on the specific variable that the model is built to predict. In this particular case, the cell survival curves considered in the study have an extremely wide range, with corresponding RBE up to 6, and for this reason just a single metric cannot give a robust evaluation of the model accuracy. Experiments conducted with high LET radiation, are characterized by high RBE values and might have high absolute errors. However, the relative percentage error might be lower in such cases, being perhaps a more appropriate metric for experiments with high RBE. On the contrary, for experiments where the RBE is close to 1, such as those conducted with high-energy protons, the MAPE might be misleading, and the MAE might be a better tool. Since the choice of the most relevant metric is not always trivial and depends on the chosen endpoint, our analysis was usually performed by studying MAPE and MAE together. Nonetheless, for the sake of readability and to avoid giving an excessive amount of information, when reasonable, only the MAE metric is reported as it is considered to be, for the present case, more informative as compared to the MAPE.

**Table 3.** Average errors and standard deviations of ANAKIN, MKM, and LEM III calculations for V79 and HSG cell lines considering different endpoints.

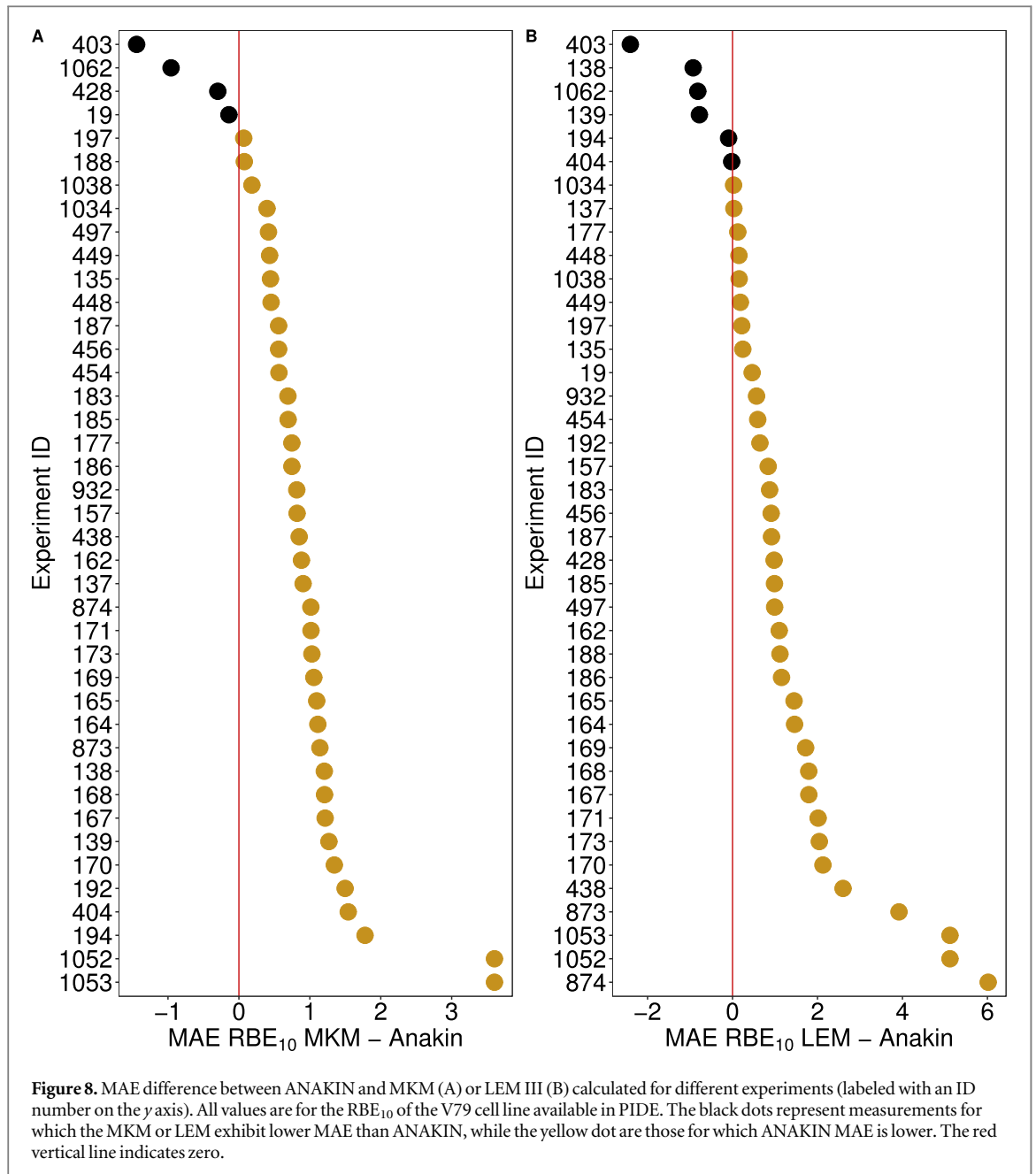
		V79					
		ANAKIN		MKM		LEM III	
Endpoint	Error	Mean	Sd	Mean	Sd	Mean	Sd
RBE <sub>10</sub>	logRMSE	0.70	0.71	1.9	1.43	1.66	1.4
	MAE	0.44	0.74	1.2	0.79	1.5	0.9
	MAPE	0.23	0.16	0.61	0.2	0.73	0.16
RBE <sub>1</sub>	MAE	0.57	0.85	1.4	1.62	1.71	1.7
	MAPE	0.2	0.14	0.42	0.43	0.46	0.47
RBE <sub>50</sub>	MAE	0.48	0.39	1.27	1.68	1.71	2.37
	MAPE	0.17	0.06	0.39	0.36	0.4	0.41
		HSG					
		ANAKIN		MKM		LEM III	
Endpoint	Error	Mean	Sd	Mean	Sd	Mean	Sd
RBE <sub>10</sub>	logRMSE	0.72	0.42	1.06	0.47	1.36	0.93
	MAE	0.43	0.3	1.28	0.38	1.55	0.26
	MAPE	0.11	0.07	0.16	0.09	0.21	0.1
RBE <sub>1</sub>	MAE	0.31	0.2	1.44	0.22	1.45	0.31
	MAPE	0.14	0.06	0.21	0.07	0.21	0.14
RBE <sub>50</sub>	MAE	0.51	0.49	1.63	0.56	1.69	0.34
	MAPE	0.15	0.1	0.19	0.11	0.28	0.08

MAPE and MAE distributions (figure 2(a) and table 2) show that the errors for RBE<sub>10</sub>, RBE<sub>1</sub> and RBE<sub>50</sub> are all sharply peaked around the average values with low deviation, denoting an overall consistent cell-survival prediction despite the extremely large range of LET and cell-lines considered in the study. Further, it can be seen how ANAKIN is able to reproduce not only the average RBE, represented by the continuous splines but also the high variability of the RBE across many LET and cell-lines. The validation of ANAKIN against RBE<sub>10</sub> measurements (figure 2(b)) shows the model's accuracy. When the MAE is plotted against the LET, two key aspects emerge: (i) in the low-LET region, the MAE for RBE<sub>50</sub> is slightly lower compared to the high LET region; (ii) the MAE for RBE<sub>10</sub> and RBE<sub>1</sub> remains almost constant in the whole LET range, with a slight drop at around 30 keV  $\mu\text{m}^{-1}$ . Notable enough, in the range 80–120 keV  $\mu\text{m}^{-1}$ , the experiments exhibit a large variation in RBE, nonetheless ANAKIN error does not seem to be affected by this huge variability with no evident increase in ANAKIN inaccuracy. This could go in the same direction as noted above, meaning that ANAKIN is able to predict RBE fluctuations at high LET. Further, a feature that supports the potential of AI in modeling cell survival and RBE, is that ANAKIN predicts the overkill effect around 100 keV  $\mu\text{m}^{-1}$ , without any specific training.

ANAKIN RBE<sub>10</sub> predictions show a slowly higher discrepancy from the experimental values in the low  $\beta_\gamma/\alpha_\gamma$  region, again mostly for RBE<sub>50</sub>, which corresponds to cell lines with high  $\alpha_\gamma/\beta_\gamma$  (figure 2(b)). These cell lines are extremely radiosensitive and therefore are characterized by a larger experimental variability that is reflected in the low accuracy of ANAKIN prediction. Further, fewer experiments have been performed for cell-line with high  $\alpha_\gamma/\beta_\gamma$ , so a higher error might simply be a natural fluctuation due to lower statistics.

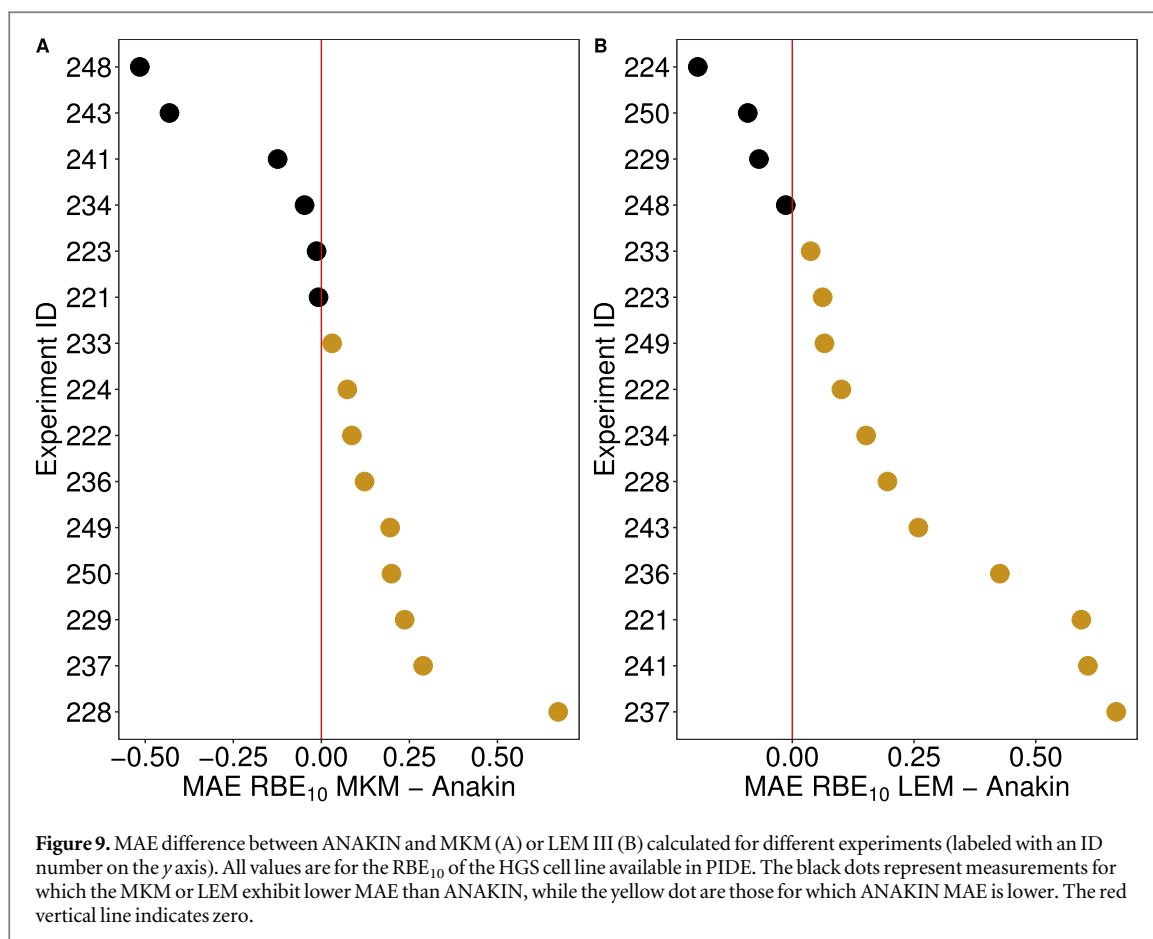
A specific analysis of single ions species prediction shows how ANAKIN accurately predicts cell survival over a wide range of ion species with very different LET, also guessing correctly the dependence of LET-RBE profiles on the ion type. For protons, ANAKIN is capable of reproducing the almost constant RBE at low-LET with a steep increase after 5 keV  $\mu\text{m}^{-1}$ , as well as the extremely high RBE at around 20 keV  $\mu\text{m}^{-1}$ . As shown, for example in (Missiaggia *et al* 2022a), currently used RBE models are often unable to accurately reproduce the RBE for very low energy protons. ANAKIN could then provide a robust and accurate tool to predict the RBE of clinical protons, thus allowing to develop TPS with a variable RBE instead of the fixed value of 1.1 currently used.

The comparison with experimental data acquired with helium and carbon ions show that ANAKIN prediction are accurate also for these species, even if exhibiting a larger variability on the errors. This is a direct consequence of the higher variability of RBE characterization connected to these two ions. These findings suggest that ANAKIN could provide an invaluable tool for predicting RBE for heavy ions, where it is commonly accepted that a constant value cannot be used in their clinical applications.



The error distribution for monoenergetic beams is lower than for SOBPs, as indicated by the main peak of the MAE distributions (figure 5(b)), but it is broader. We hypothesize that this behavior could be due to the fact that for a monoenergetic beam, an inaccurate LET estimation can result in a significantly different RBE prediction. Overall, ANAKIN is able to accurately predict RBE value for both monoenergetic and SOBPs, without the need of adding ad hoc adaptations.

Also,  $RBE_{\alpha}$  and  $RBE_{\beta}$  show a good accuracy between predicted and experimental values. Both  $\alpha$  and  $\beta$  errors seem to remain constants over the whole range of LET, whereas an analysis of the  $\alpha$  variability as a function of  $\beta_{\gamma}/\alpha_{\gamma}$  shows that for high  $\alpha_{\gamma}/\beta_{\gamma}$  cell lines the estimation of  $\alpha$  is subject to higher uncertainty. There is a clear underestimation of  $\alpha$  for high  $\alpha_{\gamma}/\beta_{\gamma}$  cell lines, which directly translated into the high RBE error in these cell lines, as already discussed above. Similar conclusions can be drawn for  $\beta$  with a slightly higher error in the high  $\beta_{\gamma}/\alpha_{\gamma}$  region. These results point out that ANAKIN is able to reproduce not only  $\alpha$ , but also  $\beta$ , which is typically subject to extremely high uncertainty, as shown for instance by low accuracy of many models in reproducing  $\beta$  variability, (Pfuhl *et al* 2022). In addition,  $\beta$  is predicted to be dependent on the radiation quality, as shown by the trend of the experimental data and in contrast to many other existing radiobiological models.



#### 4.1. Comparison with MKM and LEM

To further test ANAKIN capability and appreciate its accuracy, we compare its results with predictions from the two radiobiological models currently used in the clinics (MKM and LEM). In general, in fact, it is difficult to assess the accuracy of a model by simply considering its overall deviations from the experimental data. It is rather more effective to compare its predictions with other models assumed as benchmarks.

The finding of this comparison indicates that ANAKIN has an overall higher accuracy than both the MKM and the LEM in all the metrics, that is RMSE, MAPE and MAE. The MKM performs slightly better than the LEM, but this result could be related to the fact that we had to use LEM III, instead of the latest version LEM IV, which was not available. This hypothesis is supported by the results reported in Pfuhl *et al* (2022), which shows that LEM IV has significantly better accuracy than LEM III.

ANAKIN error distribution is significantly less broad than both the MKM and LEM distributions and its maximum error is lower. ANAKIN has not been specifically trained to predict the two cell lines selected for the comparison (i.e. V79 and HSG) but on a wide range of different cell lines available on PIDE.

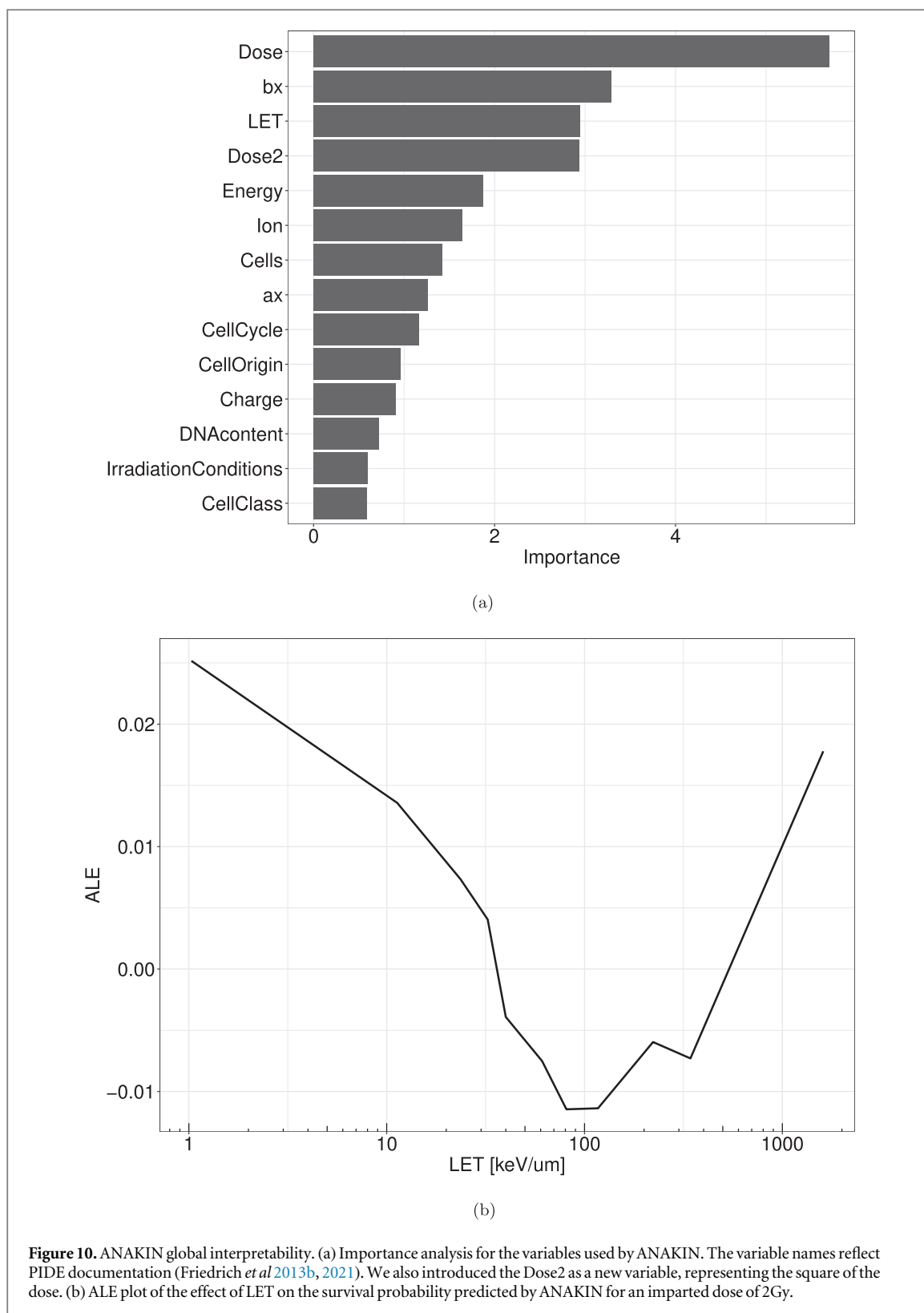
The analysis performed on several experiments suggests that ANAKIN is more accurate than both the MKM and LEM. Overall, ANAKIN shows a lower error than the MKM and LEM, and even when its prediction is less accurate than the other two models, the maximum error is lower than those obtained with the other two.

#### 4.2. Explainable AI

The global variable importance study presented in figure 10(a) shows how both biological and physical variables are efficiently used by ANAKIN to predict cell survival. The dose is the most important variable, as expected. The analysis also identifies the square of the dose as a relevant variable, which is also reasonable as the quadratic relation between survival and the dose is widely used in many RBE models. Concerning the physical variables, LET is considered to be more important than ion kinetic energy.

For the biological variables,  $\alpha_\gamma$  and  $\beta_\gamma$  are among the most relevant input parameters together with the cell line. Our hypothesis is that ANAKIN uses these three variables to understand how a specific cell line responds to ionizing radiation.

The variables over which a Deep Embedding was performed, namely *Ion*, *Cells* and *CellCycle*, are also relevant to the model predictions, suggesting that such advanced DL based embedding has been able to uncover important information.

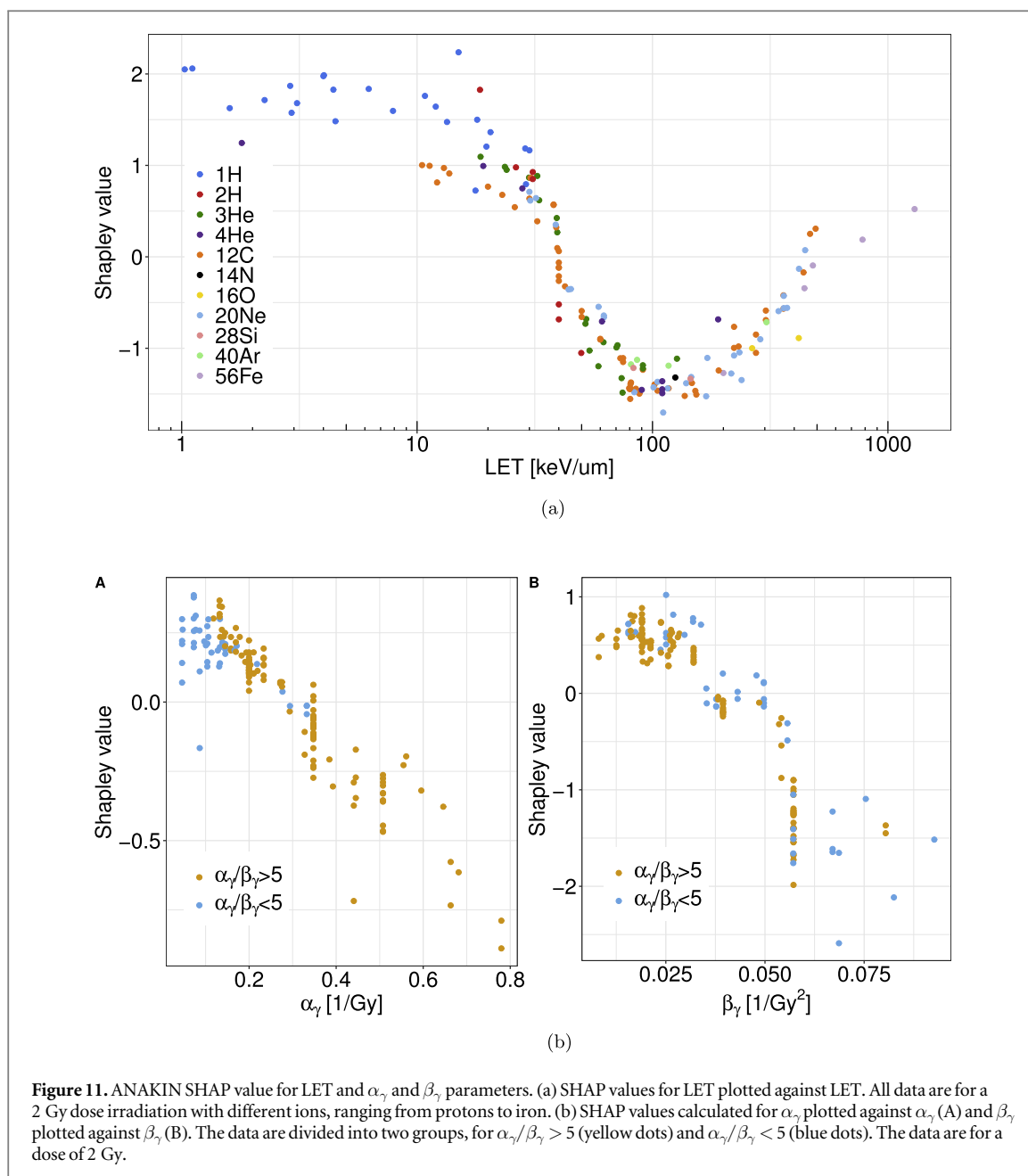


**Figure 10.** ANAKIN global interpretability. (a) Importance analysis for the variables used by ANAKIN. The variable names reflect PIDE documentation (Friedrich *et al* 2013b, 2021). We also introduced the Dose2 as a new variable, representing the square of the dose. (b) ALE plot of the effect of LET on the survival probability predicted by ANAKIN for an imparted dose of 2Gy.

The same effect emerges analyzing the dependence of the SHAP value from the LET. For protons, ANAKIN gives approximately the same positive and high importance to LET up to  $15 \text{ keV } \mu\text{m}^{-1}$ . In this region, protons show an almost constant RBE, and ANAKIN recognizes this behavior by giving the same importance to different values of LET.

The association between high  $\alpha_\gamma$  values and high negative importance in figure 11(b) reflects the fact that, for highly radiosensitive cell-lines, the  $\alpha_\gamma$  value, that describe the contribution of single track, should be more important than in cell-lines with lower  $\alpha_\gamma/\beta_\gamma$ .

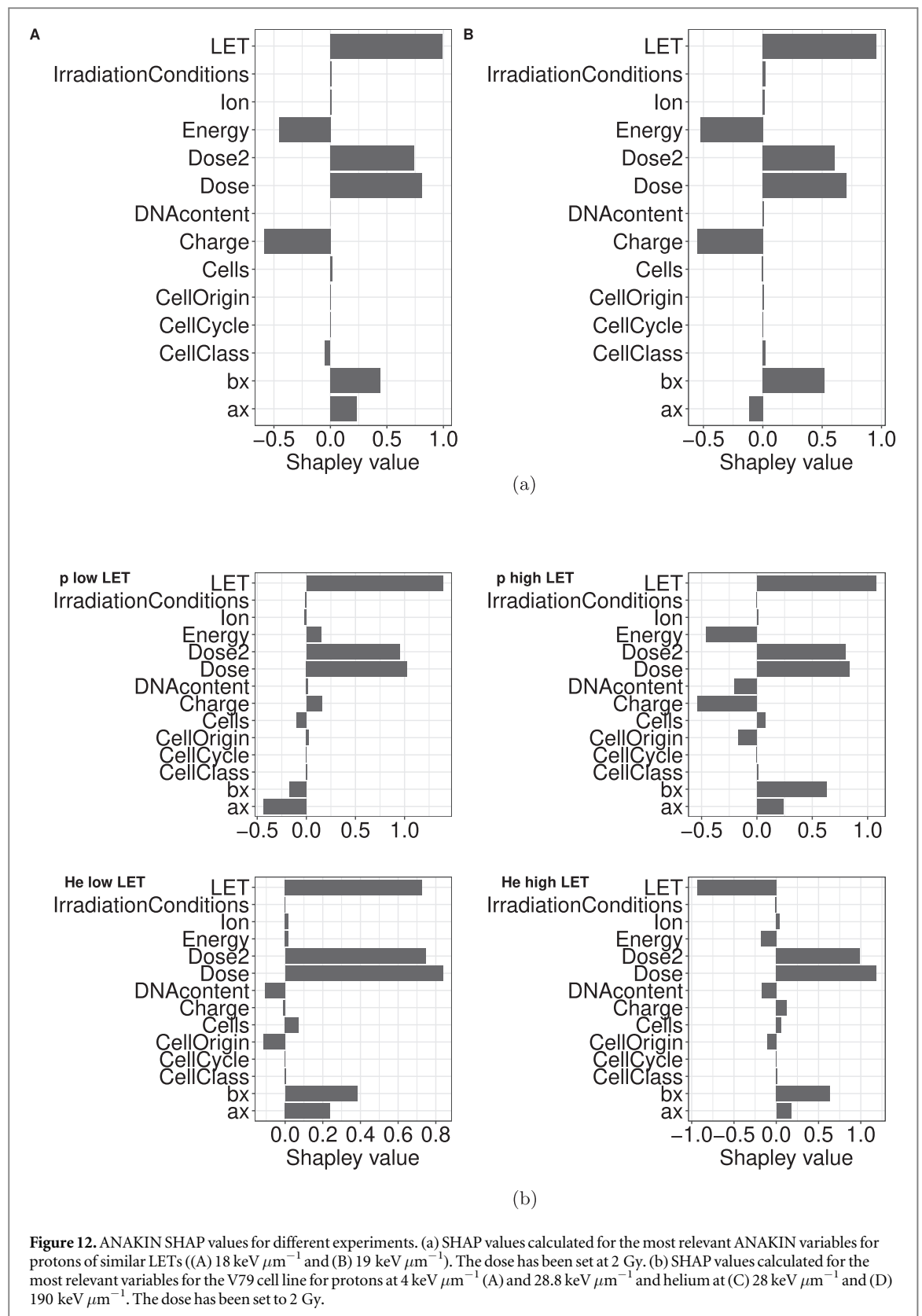




The SHAP value could be also extremely important in understanding which variables lead to a certain RBE. To support this hypothesis, we compared two experiments conducted with protons of comparable energies, namely  $18 \text{ keV } \mu\text{m}^{-1}$  and  $19 \text{ keV } \mu\text{m}^{-1}$ . The results indicate that ANAKIN can correctly predict a significant variability in the RBE, that in this case stemmed from the fact that different cell lines were considered, as pointed out by the SHAP values for the  $\alpha_\gamma$  and  $\beta_\gamma$  parameters.

To investigate how different physical variables affect ANAKIN outcomes, we considered proton and helium beams of different energies. We found that LET has always had a positive high impact only for high-LET helium, as this beam was the only one with a LET in the overkill region. Therefore it might be concluded that the LET is used by ANAKIN in a significantly different manner when an overkill effect is expected. The SHAP value for the beam kinetic energy is positive for protons and helium at low LET, whereas it is negative for the high LET beams, which have extremely low energies (below  $1 \text{ MeV } \text{u}^{-1}$ ), corresponding to depth downstream of the Bragg peak. The SHAP analysis indicates that the kinetic energy is much more important for protons than helium at low values. This behavior can be due to the fact that low-energy helium ions have a LET above the overkill threshold, and thus the main information to predict cell survival is carried by LET. Overall, it seems that ANAKIN is able to use jointly LET and kinetic energy to accurately predicts the cell survival fraction.

Advanced XAI techniques have been applied to understand what variables are relevant to ANAKIN predictions, as well as to show how specific biological features observed in experimental data, such as the



overkilling effect at high LET and the variable  $\beta$  coefficients, are reproduced by ANAKIN. The implementation of such behaviors is non-trivial in a purely mathematical model and it represents thus a strength of ANAKIN. Furthermore, these XAI techniques can play a major role in clinical applications since they allow the interpretation of ANAKIN prediction but also the understanding on how reliable the given prediction could be. It is worth stressing that, one of the major limitations in biophysical modeling of radiation effects, both for curative and radioprotective purposes, is exactly on the uncertainties estimation. Although an AI-based approach is not derived from mechanistic considerations, unlike existing radiobiological models, and thus

cannot provide validation of these mechanisms, on the other hand, its power in processing and filtering the data dependencies can help to reveal features hidden in the data, that on their turn can drive further comprehension of the phenomenon.

## 5. Conclusions

The present paper presents the AI-based model ANAKIN, for predicting the survival fraction of various cell lines exposed to different types of radiation. The findings contained in this paper prove that a single model is able to predict the behavior of different ion species, without the need to specifically train the model on data relative to a single ion. Although the main motivation for developing ANAKIN is to apply it in particle therapy, the model accuracy in predicting the biological effect of extremely high LET events could extend its application in other fields, such as space radioprotection.

The analysis described here indicates that ANAKIN is able to accurately predict cell survival and RBE over a wide range of different cell lines and ions type. Higher uncertainties and errors emerge for cell-lines characterized by low  $\alpha_\gamma/\beta_\gamma$  and LET in the range from 20 to 150 keV  $\mu\text{m}^{-1}$ . These uncertainties reflect the uncertainties in the experimental data, on which ANAKIN has been trained on. In fact, cell lines with high  $\alpha_\gamma/\beta_\gamma$  as well as experiments with high LET beams show a higher variability of RBE.

When compared with two of the most used radiobiological models, namely the MKM and LEM III, ANAKIN showed in average more accurate predictions. The gap between the models could be smaller if the latest versions of the MKM and LEM become available in the literature.

Although purely data-driven models are often considered to be less powerful than mechanistic models, ML and DL have the advantage of being extremely flexible. This is supported by the fact that ANAKIN predicts both the overkill effect and the variable  $\beta$  into the MKM. On the contrary, in mechanistic model ad hoc correction terms must typically be added to include the above effects.

The modular structure of ANAKIN makes it very easy to include advanced features. The most relevant example is the implementation of a radiation quality description different from the classical LET, such as microdosimetric or nanodosimetric quantities, as well as the coupling of ANAKIN with a mechanistic RBE model. In fact, while we notice the remarkable accuracy of the model based purely on LET, we are aware that the latter parameter is a suboptimal descriptor of a radiation field. Thus we plan to extend the model by including micro- or nanodosimetric information, which with the present dataset is difficult to retrieve, without introducing further inaccuracy. Further, whilst at the present stage, ANAKIN is trained and tested on *in vitro* data, in future extensions of ANAKIN, we plan to consider *in vivo* data and clinical values for  $\alpha$  and  $\beta$  parameters in order to get a more significant clinical descriptor in terms of TCP estimations. Also, when available sufficiently large datasets for other endpoints (mutation, transformation and chromosome aberration) ANAKIN can be easily extended to such predictions.

In conclusion, we showed that ANAKIN is an intuitive and understandable model, that demonstrates high accuracy in predicting cell survival and RBE. Any prediction given by ANAKIN can be analyzed in detail, so that the contribution of each input variable can be precisely assessed. Several advanced techniques of XAI can be used either to understand if a well-known biological or physical phenomena, such as the overkill effect or LET dependent  $\beta$ , has been included in ANAKIN, but also to gain further insight and unveil existing correlations between different variables. While AI has been broadly employed in radiotherapy treatment planning, either for physical dose optimization or image segmentation, ANAKIN is the first application on radiobiological calculations, which may open the possibility to use for the first time an AI-based model to *biological* treatment planning, i.e. the optimization of the dose delivery with explicit consideration of a voxel-dependent RBE. This problem is typically extremely heavy computationally and strongly linked to uncertainties, two features where a model like ANAKIN may be of the outermost advantage.

## Data availability statement

The data cannot be made publicly available upon publication because no suitable repository exists for hosting data in this field of study. The data that support the findings of this study are available upon reasonable request from the authors.

## ORCID iDs

Francesco G Cordonì  <https://orcid.org/0000-0002-1295-7884>

Emanuele Scifoni  <https://orcid.org/0000-0003-1851-5152>

Chiara La Tessa  <https://orcid.org/0000-0001-5742-6772>

## References

- Apley D W and Zhu J 2020 Visualizing the effects of predictor variables in black box supervised learning models *J. R. Stat. Soc. B* **82** 1059–86
- Attili A and Manganaro L 2018 Survival. (<https://github.com/batuff/Survival>)
- Bellinzona V E, Cordoni F, Missiaggia M, Tommasino F, Scifoni E, La Tessa C and Attili A 2021 Linking microdosimetric measurements to biological effectiveness in ion beam therapy: a review of theoretical aspects of mkm and other models *Front. Phys.* **8** 623–51
- Bertolet A, Cortés-Giraldo M and Carabe-Fernandez A 2021 Implementation of the microdosimetric kinetic model using analytical microdosimetry in a treatment planning system for proton therapy *Phys. Med.* **81** 69–76
- Biecek P and Burzykowski T 2021 *Explanatory Model Analysis: Explore, Explain and Examine Predictive Models* (San Francisco: Chapman and Hall/CRC)
- Bishop C M et al 1995 *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press)
- Boulevard W, Rowinski E, Louati S, Sotton S, Wozny A-S, Moreno-Acosta P, Mery B, Rodriguez-Lafrasse C and Magne N 2021 A review of the role of hypoxia in radioresistance in cancer therapy *Med. Sci. Monitor: Int. Med. J. Exp. Clin. Res.* **27** e934116
- Breiman L 2001 *Random Forests. Statistics Department* (Berkeley, CA: University of California) vol 94720, 5–32
- Carabe A, Moteabbed M, Depauw N, Schuemann J and Paganetti H 2012 Range uncertainty in proton therapy due to variable biological effectiveness *Phys. Med. Biol.* **57** 1159–72
- Chen T and Guestrin C 2016a XGBoost: a scalable tree boosting system *Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD '16, ACM (New York, NY, USA)* pp 785–94
- Chen T and Guestrin C 2016b Xgboost: a scalable tree boosting system *CoRR* abs/1603.02754 1 785–94
- Chen Y and Ahmad S 2012 Empirical model estimation of relative biological effectiveness for proton beam therapy *Radiat. Prot. Dosim.* **149** 116–23
- Cordoni F, Missiaggia M, Attili A, Welford S, Scifoni E and La Tessa C 2021 Generalized stochastic microdosimetric model: the main formulation *Phys. Rev. E* **103** 012412
- Cordoni F G, Missiaggia M, La Tessa C and Scifoni E 2022a Multiple levels of stochasticity accounted for in different radiation biophysical models: from physics to biology *Int. J. Radiat. Biol.* **0** 1–16
- Cordoni F G, Missiaggia M, Scifoni E and La Tessa C 2022b Cell survival computation via the generalized stochastic microdosimetric model (gsm2): I. The theoretical framework *Radiat. Res.* **197** 218–32
- Davidovic L M, Laketic D, Cumic J, Jordanova E and Pantic I 2021 Application of artificial intelligence for detection of chemico-biological interactions associated with oxidative stress and dna damage *Chem. Biol. Interact.* **345** 109533
- Durante M and Flanz J 2019 Charged particle beams to cure cancer: strengths and challenges *Semin. Oncol.* **46** 219–25
- Durante M and Paganetti H 2016 Nuclear physics in particle therapy: a review *Rep. Prog. Phys.* **79** 096702
- Ebner D K, Frank S J, Inaniwa T, Yamada S and Shirai T 2021 The emerging potential of multi-ion radiotherapy *Front. Oncol.* **11** 624786
- Elsässer T, Krämer M and Scholz M 2008 Accuracy of the local effect model for the prediction of biologic effects of carbon ion beams *in vitro* and *in vivo* *Int. J. Radiat. Oncol. Biol. Phys.* **71** 866–72
- Elsässer T and Scholz M 2007 Cluster effects within the local effect model *Radiat. Res.* **167** 319–29
- Elsässer T et al 2010 Quantification of the relative biological effectiveness for ion beam radiotherapy: direct experimental comparison of proton and carbon ion beams and a novel approach for treatment planning *Int. J. Radiat. Oncol. Biol. Phys.* **78** 1177–83
- Fisher A, Rudin C and Dominici F 2019 All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously *J. Mach. Learn. Res.* **20** 1–81
- Friedman J et al 2001 *The Elements of Statistical Learning* (New York: Springer series in statistics) vol 1
- Friedrich T, Durante M and Scholz M 2013a The local effect model—principles and applications *Health Risks Extraterr. Environ.* **2013** 1–14
- Friedrich T, Pfuhl T and Scholz M 2021 Update of the particle irradiation data ensemble (pide) for cell survival *J. Radiat. Res.* **62** 645–55
- Friedrich T, Scholz U, Elsässer T, Durante M and Scholz M 2013b Systematic analysis of rbe and related quantities using a database of cell survival experiments with ion beam irradiation *J. Radiat. Res.* **54** 494–514
- Giovannini G, Böhlen T, Cabal G, Bauer J, Tessonnier T, Frey K, Debus J, Mairani A and Parodi K 2016 Variable rbe in proton therapy: comparison of different model predictions and their influence on clinical-like scenarios *Radiat. Oncol.* **11** 1–16
- 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
- Grinsztajn L, Oyallon E and Varoquaux G 2022 Mon, 18 Jul 2022 arXiv:2207.08815 Why do tree-based models still outperform deep learning on tabular data?
- Grömping U 2020 Model-agnostic effects plots for interpreting machine learning models. Reports in mathematics *Phys. Chem. Rep.* **1** 1–2020
- Gunning D, Stefik M, Choi J, Miller T, Stumpf S and Yang G-Z 2019 Xai—explainable artificial intelligence *Sci. Robot.* **4** eaay7120
- Guo C and Berkhahn F 2016 Fri, 22 Apr 2016 arXiv:1604.06737 Entity embeddings of categorical variables
- Hart S 1989 Shapley value *Game Theory* (Berlin: Springer) pp 210–6
- Hawkins R B 1994 A statistical theory of cell killing by radiation of varying linear energy transfer *Radiat. Res.* **140** 366–74
- Ho T K 1995 Random decision forests *Proc. of 3rd Int. Conf. on Document Analysis and Recognition* vol 1 (IEEE) pp 278–82
- Ho T K 1998 The random subspace method for constructing decision forests *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 832–44
- Inaniwa T, Furukawa T, Kase Y, Matsufuji N, Toshito T, Matsumoto Y, Furusawa Y and Noda K 2010 Treatment planning for a scanned carbon beam with a modified microdosimetric kinetic model *Phys. Med. Biol.* **55** 6721–37
- Inaniwa T and Kanematsu N 2018 Adaptation of stochastic microdosimetric kinetic model for charged-particle therapy treatment planning *Phys. Med. Biol.* **63** 095011
- Kase Y, Kanai T, Matsufuji N, Furusawa Y, Elsässer T and Scholz M 2007 Biophysical calculation of cell survival probabilities using amorphous track structure models for heavy-ion irradiation *Phys. Med. Biol.* **53** 37–59
- Kase Y, Kanai T, Matsumoto Y, Furusawa Y, Okamoto H, Asaba T, Sakama M and Shinoda H 2006 Microdosimetric measurements and estimation of human cell survival for heavy-ion beams *Radiat. Res.* **166** 629–38
- Kellerer A M and Rossi H H 1974 The theory of dual radiation action *Curr. Top. Radiat. Res. Quarterly* **1** 85–158
- Kellerer A M and Rossi H H 1978 A generalized formulation of dual radiation action *Radiat. Res.* **75** 471–88
- Khalid M A, Jijkoun V and de Rijke M 2007 Machine learning for question answering from tabular data *18th Int. Workshop on Database and Expert Systems Applications (DEXA 2007)* (IEEE) pp 392–6
- Krämer M, Jäkel O, Haberer T, Kraft G, Scharadt D and Weber U 2000 Treatment planning for heavy-ion radiotherapy: physical beam model and dose optimization *Phys. Med. Biol.* **45** 3299–317

- Kurz C, Mairani A and Parodi K 2012 First experimental-based characterization of oxygen ion beam depth dose distributions at the heidelberg ion-beam therapy center *Phys. Med. Biol.* **57** 5017–34
- LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44
- Lundberg S M and Lee S-I 2017 A unified approach to interpreting model predictions *Advances in Neural Information Processing Systems* p 30
- Mairani A et al 2017 A phenomenological relative biological effectiveness approach for proton therapy based on an improved description of the mixed radiation field *Phys. Med. Biol.* **62** 1378–95
- Mairani A et al 2022 Roadmap: helium ion therapy *Phys. Med. Biol.* **67** 15TR02
- Manganaro L, Russo G, Bourhaleb F, Fausti F, Giordanengo S, Monaco V, Sacchi R, Vignati A, Cirio R and Attili A 2018 'Survival': a simulation toolkit introducing a modular approach for radiobiological evaluations in ion beam therapy *Phys. Med. Biol.* **63** 08NT01
- Manganaro L, Russo G, Cirio R, Dalmaso F, Giordanengo S, Monaco V, Muraro S, Sacchi R, Vignati A and Attili A 2017 A Monte Carlo approach to the microdosimetric kinetic model to account for dose rate time structure effects in ion beam therapy with application in treatment planning simulations *Med. Phys.* **44** 1577–89
- McMahon S J 2018 The linear quadratic model: usage, interpretation and challenges *Phys. Med. Biol.* **64** 01TR01
- McMahon S J and Prise K M 2021 A mechanistic dna repair and survival model (medras): applications to intrinsic radiosensitivity, relative biological effectiveness and dose-rate *Front. Oncol.* **11** 2319
- McNamara A L, Schuemann J and Paganetti H 2015 A phenomenological relative biological effectiveness (rbe) model for proton therapy based on all published in vitro cell survival data *Phys. Med. Biol.* **60** 8399–416
- Micci-Barreca D 2001 A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems *ACM SIGKDD Explorations Newsl.* **3** 27–32
- Missiaggia M, Cartechini G, Scifoni E, Rovituso M, Tommasino F, Verroi E, Durante M and La Tessa C 2020 Microdosimetric measurements as a tool to assess potential in-field and out-of-field toxicity regions in proton therapy *Phys. Med. Biol.* **65** 245024
- Missiaggia M, Cartechini G, Scifoni E, Tommasino F, Durante M and La Tessa C 2022a Investigation of in- and out-of-field radiation quality with microdosimetry and its impact on rbe in proton therapy *Int. J. Radiat. Oncol. Biol. Phys.* **115** 1269
- Missiaggia M, Pierobon E, La Tessa C and Cordonì F G 2022b An exploratory study of machine learning techniques applied to therapeutic energies particle tracking in microdosimetry using the novel hybrid detector for microdosimetry (hdm) *Phys. Med. Biol.* **67** 1–17
- Molnar C 2020 Survival. (<https://christophm.github.io/interpretable-ml-book/>)
- Ongsulee P, Chotchaung V, Bamrungsi E and Rodcheewit T 2018 Big data, predictive analytics and machine learning 2018 16th Int. Conf. on ICT and Knowledge Engineering (ICT&KE) (IEEE) pp 1–6
- Paganetti H 2014 Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer *Phys. Med. Biol.* **59** R419–72
- Paganetti H 2018 Proton relative biological effectiveness—uncertainties and opportunities *J. Part. Ther.* **5** 2–14
- Paganetti H, Niemierko A, Ancukiewicz M, Gerweck L E, Goitein M, Loeffler J S and Suit H D 2002 Relative biological effectiveness (rbe) values for proton beam therapy *Int. J. Radiat. Oncol. Biol. Phys.* **53** 407–21
- Papakostantinou D, Zanni V, Nikitaki Z, Vasileiou C, Kousouris K and Georgakilas A G 2021 Using machine learning techniques for asserting cellular damage induced by high-LET particle radiation *Radiation* **1** 45–64
- Pfuhl T, Friedrich T and Scholz M 2022 Comprehensive comparison of local effect model iv predictions with the particle irradiation data ensemble *Med. Phys.* **49** 714–26
- PTCOG 2022 Particle therapy co-operative group
- Rovituso M and La Tessa C 2017 Nuclear interactions of new ions in cancer therapy: impact on dosimetry *Transl. Cancer Res.* **6** 1310–26
- Sahiner B, Pezeshk A, Hadjiiski L M, Wang X, Drukker K, Cha K H, Summers R M and Giger M L 2019 Deep learning in medical imaging and radiation therapy *Med. Phys.* **46** e1–36
- Sarrut D and Krahn N 2021 Artificial intelligence and Monte Carlo simulation *Monte Carlo Techniques in Radiation Therapy* (Boca Raton, FL: CRC Press) pp 251–8
- Scholz M, Friedrich T, Magrin G, Colautti P, Ristić-Fira A and Petrović I 2020 Characterizing radiation effectiveness in ion beam therapy part I: Introduction and biophysical modeling of rbe using the lemviv *Front. Phys.* **8** 272–85
- Seger C 2018 An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing
- Shreyas P 2022 Deep embedding's for categorical variables (cat2vec)
- Shwartz-Ziv R and Armon A 2022 Tabular data: deep learning is not all you need *Inf. Fusion* **81** 84–90
- Sokol O et al 2017 Oxygen beams for therapy: advanced biological treatment planning and experimental verification *Phys. Med. Biol.* **62** 7798–813
- Stewart R D, Carlson D J, Butkus M P, Hawkins R, Friedrich T and Scholz M 2018 A comparison of mechanism-inspired models for particle relative biological effectiveness (rbe) *Med. Phys.* **45** e925–52
- Tilly N, Johansson J, Isacson U, Medin J, Blomquist E, Grusell E and Glimelius B 2005 The influence of rbe variations in a clinical proton treatment plan for a hypopharynx cancer *Phys. Med. Biol.* **50** 2765–77
- Tobias C A 1980 The repair-misrepair model of cell survival
- Tobias C A 1985 The repair-misrepair model in radiobiology: comparison to other models *Radiat. Res.* **104** S77–95
- Vassiliev O N 2012 Formulation of the multi-hit model with a non-poisson distribution of hits *Int. J. Radiat. Oncol. Biol. Phys.* **83** 1311–6
- Vassiliev O N, Grosshans D R and Mohan R 2017 A new formalism for modelling parameters  $\alpha$  and  $\beta$  of the linear-quadratic model of cell survival for hadron therapy *Phys. Med. Biol.* **62** 8041–59
- Wilkins J and Oelfke U 2004 A phenomenological model for the relative biological effectiveness in therapeutic proton beams *Phys. Med. Biol.* **49** 2811–25