Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 22-23, 2025

Feature Importance Analysis of Ion Beam Exposure in Machine Learning for Cellular Damage Prediction

Dong-Seok Lee, Ali Abu Shqair , Eun-Hee Kim *
Department of Nuclear Engineering, Seoul National University
**Corresponding author: eunhee@snu.ac.kr*

***Keywords: radiation exposure, damage prediction, machine learning, feature importance**

1. Introduction

The linear-quadratic model (LQM) is widely used to relate the cellular response to radiation dose. The LQM is straightforward, only using two parameters α and β. Experimental data reported up to now are classified according to multiple features, including cell type and the mass, charge, and LET (or energy) of radiation. Those data cannot be modeled all together with a single set of α and β parametric values in identifying the relationship between radiation dose and cellular response.

Data-driven machine learning (ML) is useful for finding dose-to-response relationships. The neural network (NN) can recognize the non-linear and complicated relationships of each feature to the cellular response by training with large datasets. The Particle Irradiation Data Ensemble (PIDE) database [1, 2] is a collection of radiation experiments assessing the surviving fraction (SF) of cells from exposure, especially to ion beam radiation. This collection of data can be utilized for training neural networks to predict the SF of certain cells irradiated by specific types of ion beam. One can find out which parameters had a significant impact on SF or, simply put, which parameters are more important than others.

2. Backgrounds

2.1. Feed forward neural network

The Feed-Forward Neural Network (FFNN), which is one of the simplest types of artificial neural networks, can be used to train PIDE. It consists of layers of neurons where the information moves in a forward direction, from the input layer, through hidden layers, and to the output layer. Each neuron, if not in the input layer, receives a weighted sum of the previous layer and passes the activation function. Hidden layers and activation functions enable FFNN to understand non-linear relationships.

2.2. Permutation feature importance

After training, Permutation Feature Importance (PFI) [3] is calculated. PFI is one of the methods for estimating the impact of individual features. It measures the importance of features by observing the performance drop of a model at random shuffling of the values of specific features. It is model-agnostic and shows the direct impact of each feature on the model's prediction.

2.3 Z^{*2}/β^2 parameter

PIDE provides the linear energy transfer (LET) of the radiation source in each experiment. LET is considered to influence radiation biological effectiveness (RBE). However, some studies show that the Z^{*2}/β^2 value of radiation is better aligned than LET, regardless of the ion type [4], where the Z^* is the effective charge of the particle and β is the ratio of the particle speed to speed the of light. From PIDE, Z^{*2}/β^2 can be retrieved, and used as input feature for improved prediction of RBE.

3. Methods

3.1. Data preparation

The PIDE was preprocessed for training the neural network. PIDE provides the SFs from experiments at doses of a certain range and the values of α and β parameters for the corresponding LQM. The experiment data were excluded if R^2 of its LQM is lower than 0.9. Datasets with SF values exceeding 1 or lower than 0.001 and LET over 500 keV/μm were excluded as well. Categorical data, such as cell types or cell cycles, were label-encoded. For training NN with SFs, data augmentation was performed by using the LQM specific to each experiment. When making additional data, Gaussian noise, whose variance is mean squared error of LQM, was added.

For training with RBEs, doses that would lead to 10% of SF were calculated by the LQMs provided in PIDE. The Z^{*2}/β^2 values were calculated from the Bethe-Bloch formula, assuming the water medium. NN training with RBEs was performed by adopting a feature of either LET or Z^{*2}/β^2 to look for any difference in performance. The logarithmic conversion of Z^{*2}/β^2 values was suitable for training. Table 1 lists the input features considered in training the NN.

Table 1. Input features used for training neural network.

Cells	cell line
Cell Class	tumor or normal cells
Cell Origin	human or rodent cells
Cell Cycle	cell cycles, only used for SF prediction
Irradiation Conditions	monoenergetic or spread-out Bragg's peak

Ion	mass of ions in amu
Charge	charge of ion
LET	linear energy transfer in keV/μm
Z^{*2}/β^2	only used in RBE estimation
Photon experiments	photon experiment condition, only used in RBE estimation
Dose	dose delivered

3.2 Neural network modeling

From repetitive testing, an NN structure consisting of 6 hidden layers was chosen, with 6 layers of 512, 256, 256, 128, 128, and 64 nodes for training with SFs and those of 64, 32, 32, 16, 16, and 16 nodes for training with RBEs. Each node in the hidden layers was activated with the GELU function: $GELU(x) = 0.5x\left(1 + \tanh\left(\sqrt{2/\pi}\,(x + 0.044715x^3)\right)\right)$. Output nodes were activated with a sigmoid function for training with SF, whereas linear activation was employed for training with RBE.

In the optimization process, not only the mean-squared error (MSE) but also additional loss due to SF(0) being 1 was considered. Ten neural networks were created by 10 separate trainings. The final prediction was an average of the individual models' predictions.

3.3 Estimation of feature importance

After training, permutation feature importance was calculated for the component features including LET and Z^{*2}/β^2. Each feature's importance was quantitated by the drop of R^2 score for test sets with randomly shuffled values of the corresponding feature. The shuffling process was conducted 100 times.

4. Results

FFNN predicted the SF of test sets with R^2 over 0.9 (Fig. 1). The permutation feature importance for SF is presented in Fig. 2. Apart from dose, LET was the most significant feature, followed by ion mass. In predicting RBE10 (RBE at SF=0.1) with the LET feature adopted, the R^2 score was not significant. However, the adoption of Z^{*2}/β^2 instead of LET provided better predictions of RBE (Figs. 3 and 4). Although performance was limited, permutation feature importance for RBE prediction was calculated (Figs. 5 and 6). Z^{*2}/β^2 and LET were the most important features for RBE prediction. Mass and charge of ion followed. Cell origin, cell class, and irradiation condition insignificantly affected SF and RBE prediction.

5. Conclusion

LET was the most important feature in the machine learning by FFNN model for SF prediction with cells exposed to ion beam, with other features insignificant compared to LET. FFNN better performed in RBE10 prediction when Z^{*2}/β^2 instead of LET was featured in training.

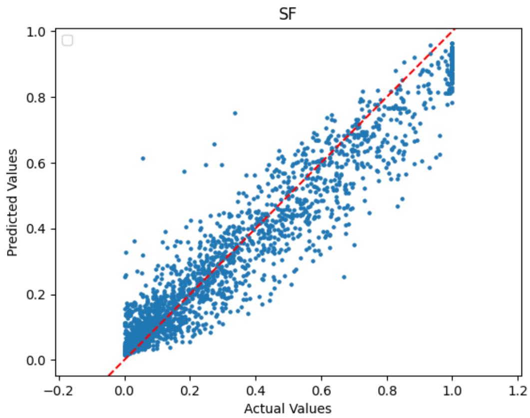

Fig. 1. The correlation of NN-predicted SFs and the test SFs. (R^2=0.92)

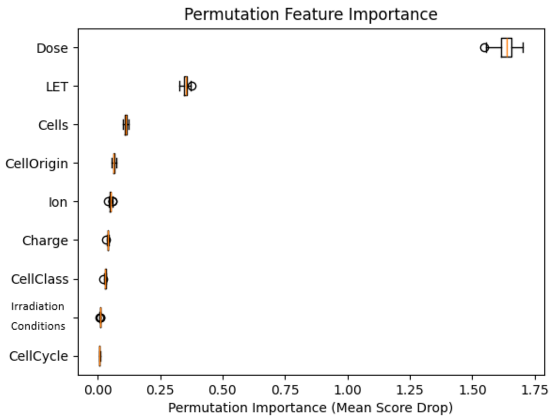

Fig. 2. Permutation feature importances in SF prediction.

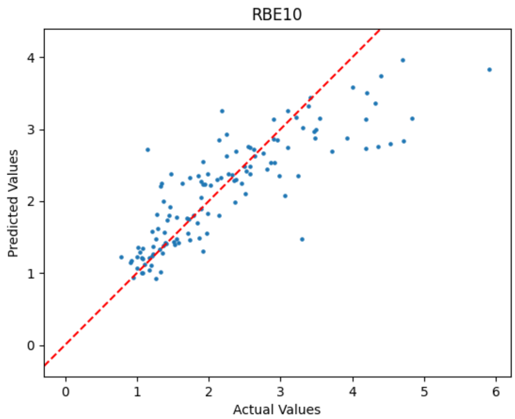

Fig. 3. The correlation of NN-predicted RBE10 and test RBE10 with the feature Z^{*2}/β^2 adopted. (R^2=0.68)

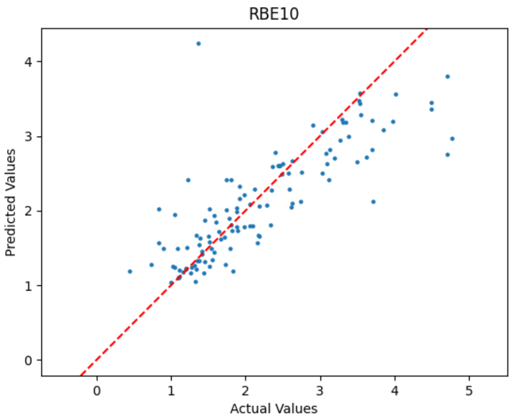

Fig. 4. The correlation of NN-predicted RBE10 and test RBE10 with the feature LET adopted. (R^2=0.64)

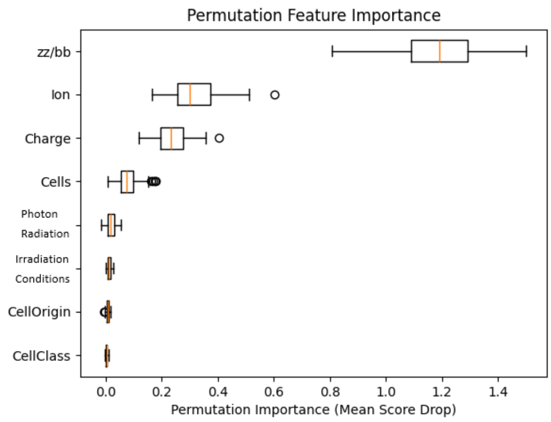

Fig. 5. Permutation feature importance in RBE10 prediction with Z^{*2}/β^2 considered.

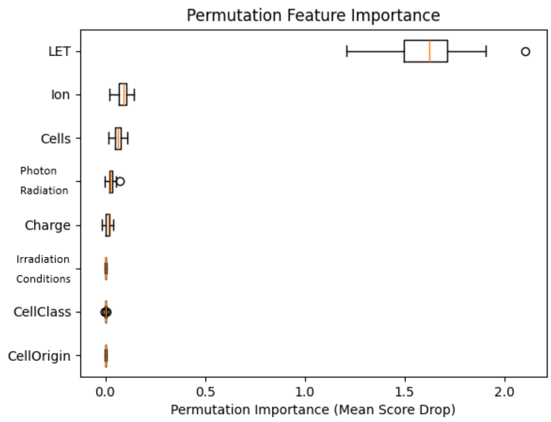

Fig. 6. Permutation feature importance in RBE10 prediction with LET considered.

REFERENCES

[1] T. Friedrich, U. Scholz, T. ElsäSser, M. Durante, M. Scholz, Systematic analysis of RBE and related quantities using a database of cell survival experiments with ion beam irradiation, Journal of Radiation Research, vol. 54, no. 3, pp. 494-514, 2013.

[2] T. Friedrich, T. Pfuhl, M. Scholz, Update of the particle irradiation data ensemble (PIDE) for cell survival, Journal of Radiation Research, vol. 62, no. 4, pp. 645-655, 2021.

[3] André Altmann, Laura Toloşi, Oliver Sander Thomas Lengauer, Permutation importance: a corrected feature importance measure, Bioinfometics, vol. 26, issue 10, pp. 1340-1347, 2010.

[4] Cucinotta F.A., Kim Myung-Hee.Y., Chappell, Lori J., Space Radiation Cancer Risk Projections and Uncertainties – 2010, NASA Technical Publication NASA/TP-2011-216155, National Aeronautics and Space Administration, 2011