

Development and Validation of Radiation Sensitivity Prediction Model Using Gene Expression Profiling Data Based on US National Cancer Institute-60 Tumor Cell Line

Department of Nuclear Engineering, Hanyang University

Euidam Kim and Yoonsun Chung*

Purpose: In radiotherapy, the prediction of radiation sensitivity for various types of tumors is clinically relevant and important. Moreover, with the advance of the adaptive therapy for each patient, individual radiation sensitivity prediction has been required to determining the patient-specific treatment planning, doses, and fractionations. Since now, researchers have mainly used the conventional statistical methods to predict radiation sensitivity by measuring the correlation between radiation sensitivity represented by survival fraction at 2 Gy (SF2) and gene expression data. However, in this study, with the recent advance of the deep neural network (DNN) which is a novel modality of the decision making such as prediction of classification, we have developed the radiation sensitivity prediction model using DNN, which can predict the inherent radiation sensitivity of tumor cell lines represented by SF2 from the gene expression profiling data.

Materials and Methods: The microarray of gene expression profiling of the US National Cancer Institute-60 (NCI-60) tumor cell line was acquired from the gene expression omnibus (GEO) database. Since there were two or three samples for each tumor cell line of NCI-60, a total of 174 samples of 59 tumor cell lines of NCI-60, excluding MDA-N cell line. SF2 values used in this study which can be an indicator of radiation sensitivity, were obtained from previous publications. The radiation sensitivity prediction model was consisted of two main parts, convolutional neural network based feature vector extractor and fully connected prediction layer with residual blocks. Five rounds of 6-fold cross-validation approach, which divides the whole dataset into six subdataset and repeats by using five sets as the training set and one set as the test set, were applied to train and validate the prediction model. The absolute error and the relative error of the SF2 prediction was defined as the absolute deviation between the true and the predicted SF2, and the absolute error divided by the true SF2, respectively. The criteria for 'correct prediction' was defined as a case within 10% of relative error. All statistical analysis was performed with GraphPad Prism version 7.03. The model was evaluated and trained with the Nvidia TITAN RTX and the TensorFlow 1.14.0 framework based on Python version 3.6.8.

Results: Among the whole 174 samples, 94 (54.02%), 48 (27.59%), 25 (14.37%), 3 (1.72%), and 4 (2.30%) samples were included in the groups with relative errors of less than 1%, 1 to 2%, 2 to 5%, 5 to 10%, and 10% or more, respectively. Based on the criteria of correct prediction, the prediction model correctly predicted 170 among the 174 samples, marking 97.70% of prediction accuracy. The Pearson correlation coefficient (Pearson's r) was 0.9877 (95% CI: 0.9834 to 0.9909) between the true (measured) SF2 and predicted SF2. The average relative and absolute errors of 170 'correct prediction' cell lines were 1.226 ± 1.246 % and 0.00589 ± 0.00638 , respectively. On the other hand, the average relative and absolute errors of 4 incorrectly predicted cell lines were 187.241 ± 231.278 % and 0.184 ± 0.117 , respectively.

Conclusions: The result of this study suggests that the utilization of recently-developing DNN is reasonably feasible to develop the prediction model for radiation sensitivity using large amount of gene expression profiling data. Further study would be needed that reveals the 'black box' of the DNN which have been historically recognized as one of the biggest limitations of the deep learning.

Keywords: Radiation sensitivity; Deep neural network; Prediction model; Gene expression profiling