Published in final edited form as:

IEEE Trans Radiat Plasma Med Sci. 2019 March ; 3(2): 242–249. doi:10.1109/TRPMS.2018.2884134.

Artificial Neural Network with Composite Architectures for Prediction of Local Control in Radiotherapy

Sunan Cui,

Applied Physics Program, University of Michigan, Ann Arbor, MI 48109, USA, sunan@umich.edu

Yi Luo,

Department of Radiation Oncology, University of Michigan, Medical School, MI 48103, USA, yiyiluo@med.umich.edu

Huan Hsin Tseng,

Department of Radiation Oncology, University of Michigan, Medical School, MI 48103, USA, thuanhsi@med.umich.edu

Randall K. Ten Haken, and

Department of Radiation Oncology, University of Michigan, Medical School, MI 48103, USA, rth@med.umich.edu

Issam El Naqa [Senior Member,IEEE]

Department of Radiation Oncology, University of Michigan, Medical School, MI 48103, USA, ielnaqa@med.umich.edu

Abstract

In this study, we investigated the application of artificial neural networks (ANNs) with composite architectures into the prediction of local control (LC) of lung cancer patients after radiotherapy. The motivation of this study was to take advantage of the temporal associations among longitudinal (sequential) data to improve the predictive performance of outcome models under the circumstance of limited sample sizes. Two composite architectures: (1) a one dimension (1D) convolutional + fully connected and (2) a locally-connected+ fully connected architectures were implemented for this purpose. Compared with the fully-connected architecture (multi-layer perceptron [MLP]), our composite architectures yielded better predictive performance of LC in lung cancer patients who received radiotherapy. Specifically, in a cohort of 98 patients (29 patients failed locally), the composite architecture of 1D convolutional layers and fully-connected layers achieved an AUC (area under receiver operating characteristic curve) of 0.83 (95% confidence interval (CI): 0.807~0.841) with 18 features (14 features are longitudinal data). Whereas, the composite architecture of locally- connected layers and fully-connected layers achieved an AUC of 0.80 (95%CI: 0.775~0.811). Both outperformed an MLP in the prediction performance with the same set of features, which achieved an AUC of 0.78 (95%CI: 0.751~0.790); (P-values for differences in AUC using the DeLong tests were 1.609×10^{-14} and 1.407×10^{-4} , respectively).

Keywords

Radiotherapy outcome modeling; local control; deep learning; multi-layer neural network; 1D convolutional layer

I. INTRODUCTION

OUTCOME modeling is an important process for personalization of dose prescription and optimization of response in radiotherapy of locally advanced lung cancer patients. In the early days of radiation treatment of cancer, analytical models based on simplified understanding of radiation biophysical effects on cells were the main guide for treatment in the field of radiotherapy outcome modeling. However, the assumptions of these models were too idealistic considering the underlying complexity involved in the interactions among biological, clinical and physical variables. As a result, data-driven models led by machine learning and deep learning techniques, which can incorporate a wide range of patients' specific information besides radiation doses have provided a more promising solution to outcome modeling [1]. In this study, we focused on modeling of local control (LC), which is an index indicating whether the tumor is locally eradicated or not. Former studies have applied various data-driven models, e.g., a kernel-based support vector machine (SVM), [2] or an artificial neural network (ANN) [3] into LC prediction.

Artificial neural networks (ANNs) [4, 5] have been previously applied in outcome modeling of radiotherapy. The applied architectures were based on one-hidden-layer multi-layer perceptrons (MLPs). However, with a recent resurgence in deep learning technologies, tremendous developments have been achieved in training and optimizing neural network architectures [6], which can be extended to modeling of radiotherapy outcomes. In this study, we revisited the application of ANNs for LC prediction. There are two main motivations. First, as mentioned earlier, with the prevalence of deep learning methods and their superior performance in many applications [7–9], many new techniques were developed to improve the training of neural networks. Regularization techniques, e.g., dropout [10] and batch normalization [11], were invented to prevent the common risk of overfitting. Optimization methods considering momentum and using adaptive learning rates, e.g., Adagrad [12], RMSprop [13] and Adam [14] have been shown to outperform their traditional counterparts such as stochastic gradients descent (SGD) in the optimization of the weight parameters of the network, which have been shown to improve the accuracy of the trained ANN models [12, 14]. Secondly, the traditional architecture of an MLP, the backbone of ANN, has its own limitations that degrade its performance. This is why modern deep learning technologies have expanded into new architectures. As a result, we were motivated to explore other possible architectures that are suitable for outcome modeling in radiotherapy. In an MLP, each neuron is connected to every neuron in its adjacent layers, which means the value of each neuron is calculated by taking a weighted sum of all the neurons in its previous layer. Such a fully-connected architecture takes each input as an independent unit and does not make use of inter-correlations among the inputs. In the field of image recognition, MLPs gained some early success. However, this required a huge amount of data to support training due to the large number of free parameters in such an

architecture. As a result, convolutional neural networks (CNNs) [15], which support shift invariance properties and exploit spatial local correlations, have emerged to overcome the limitations of traditional MLPs. CNNs largely reduce the number of weight parameters compared to MLPs, hence the required training size is dramatically reduced. In the field of radiation oncology, the number of patients is usually limited in a specific clinical scenario. However, with the advancement of biology and imaging technologies, more and more biological and imaging data are generated and collected together with traditional dosimetric and physical data. Under this circumstance, it is beneficial to take into account the associations among these patient specific variables. “Partially- connected” architectures such as CNNs, which are carefully designed to incorporate these associations, may potentially outperform a fully-connected counterpart (an MLP) in the prediction of radiotherapy response with fewer degrees of freedom when dealing with limited clinical sample sizes.

In this study, biological variables, e.g., cytokines and imaging data had been collected at several consecutive time points over the course of treatment. These longitudinal data may yield additional information about the treatment response and improve predictions of outcome. In this study, we investigated application of various neural network architectures to better consider the associations among longitudinal data and improve predictive performance. Specifically, two deep learning (partially-connected layers) architectures were proposed for LC prediction in lung cancer patient post- radiotherapy. Their performances were then compared with a fully-connected MLP, which is the most fundamental and prevalent ANN architectures, but which also disregards temporal correlations among its inputs.

II. MATERIALS

The study was conducted on a cohort of 98 non-small cell lung cancer (NSCLC) patients treated with radiotherapy. They were treated under four different protocols, among which two delivered constant dose per fraction and the rest adapted doses once during the treatment. Among those patients, 29 failed post treatment with a median follow-up of 61 months (range of 24– 84 months) in this NSCLC population.

Biological variables (including levels of cytokines, micro-RNAs and SNPs) and features extracted from PET imaging and dosimetric data were collected. Levels of cytokines such as eotaxin [16] and interleukin-1 α [17] were measured from blood samples collected before the treatment, and approximately after $\frac{1}{3}$ and $\frac{2}{3}$ of the prescribed doses were delivered. Micro-RNA-145 [18], 574 [19], 122 [20] and SNP SMAD9 [21] were also measured from blood samples collected before the treatment. PET images were collected twice, before the treatment and at the time of dose adaptation. The metabolic tumor volume (MTV) was extracted by applying a level set algorithm using the planned tumor volume (PTV) as a region of interest. Radiomics textural features were extracted after images were quantized into 32 levels using the Lloyd’s algorithm. They included the widely used global, gray level run-length matrix (GLRLM), and grey-level size-zone matrix (GLSZM) features. Radiation dose distributions were calculated in the Varian Eclipse treatment planning system by the AAA algorithm. The doses were converted to equivalent doses in 2Gy (EQD2s) [22] based

on physical fraction doses and the number of fractions, being using an $\frac{\alpha}{\beta} = 6 \text{ Gy}$.¹This served to account for overall dose effects among the different protocols. Derived dosimetric features included mean tumor doses before and after the adaption.

A summary of features applied in the predictive modeling is shown in Table 1. (18 features in total, 14 of which are from measurements of longitudinal variables at different time points.)

III. METHODS

In this work, we modeled the longitudinal data with 2 different partially-connected layers to take into account temporal associations. One layer was a 1D convolutional layer, where the convolution operation was applied along only one direction (time-axis). The other is a locally-connected layer, where only measurements of the same variable were connected into a single output neuron. These two layers, which perform data representation, are merged with the fully connected layers, which act as a classifier, into composite architectures for LC prediction, respectively.

A. Architecture A: 1D convolutional layer for modeling longitudinal data and the architecture of corresponding composite neural network.

1D convolutional layer computes the convolution along only one axis (e.g., time axis). It is worth noting that this is different from the classical 2D convolutional layer, which applies convolution operation along 2 spatial directions as in computer vision applications [23].

Figure 1 illustrates an example of applying a 1D convolutional layer to a 2D fictitious input whose size is 11×5 , where 11 and 5 are lengths in temporal (vertical) and non-temporal axis (horizontal) respectively. It is worth noting that “1D” means a convolution kernel is convolved with the input over 1D to produce outputs, it does not put any constraint to the input size dimension. For a 2D input, 2D convolutional layer is automatically reduced to 1D convolution when the width of a kernel equals to the width of inputs. Consequently, the width of a filter should be fixed to 5 in order to perform 1D convolution, while the length of filter (say 3) can be arbitrarily assigned. This layer has a filter of size 3×5 sliding along the time-axis by 1 pixel (stride) at a time. In each position, an element-wise multiplication of the input patch and the filter is carried out. The corresponding result would then be one of the output neurons.

For the longitudinal data in our dataset, PET images and dosimetric data were collected twice, and cytokines data were collected three times. We found that it was more advantageous to construct two separate convolutional layers for the inputs of time steps 3 and 2, respectively, rather than using a single convolutional layer by truncating the cytokine data. The rationale for this is further explained in the discussion section. In the next layer, the outputs from those two layers were concatenated with non-longitudinal data and fed into following fully-connected layers. One can refer to the flowchart of Figure 2, which embeds

¹ BEDs using $\alpha/\beta=10$ or 15 Gy have Pearson correlation coefficients of 0.999 and 0.997 with the 6 Gy values. Further, the single variable AUC for LC prediction for 6 Gy was slightly higher (0.636) than those for the other two (0.633, 0.624)

pseudo code based on the python deep learning library Keras [24] for clarification of the architecture.

B. Architecture B: Locally-connected layers for modeling longitudinal data and the architecture of corresponding composite neural network.

Besides 1D convolutional layers, a so-called locally-connected layer may also be considered for modeling longitudinal data. A locally-connected layer is of a partially connected structure, which only calculates the weighted sum of adjacent inputs. Assuming that the measurements of a variable at different time steps are highly correlated, one can then apply a locally- connected layer to the input vector where separate measurements of the same variables are arranged together.

Figure 3 illustrates a locally-connected layer with a stride of 3, whose input is composed of three measurements for three variables a, b and c. The layer is very similar to a 1D convolutional layer with a kernel size of 3×1 and stride 3, except there is no sharing weights scheme, which means weights associated with variables a, b and c are different from each other.

We applied a locally-connected layer with stride of 2 and 3 to PET image + dosimetric data and cytokines respectively. Then we followed a similar scheme as that in Figure 2, i.e., the outputs were then concatenated with the rest of features as a whole to be fed into an MLP. A clarification of the architecture can be found in the Figure 4.

IV. RESULTS

We implemented the proposed neural network architectures with the Python deep learning library Keras [24]. The regularization technique dropout [10] was adopted to reduce overfitting. The RMSprop optimization [14] method was implemented for the estimation of the network weights.

We performed five-fold cross-validation to evaluate the predictive models' performance and assess their generalizability. The oversampling technique synthetic minority over-sampling (SMOTE) [25] was applied to the training data to mitigate the class imbalance problem. The training of the ANN architecture was conducted on a GPU cluster of 2 nodes and takes an average of 70s (per training) while testing takes an average of 0.003s (per test).

A. Architecture A: composition of 1D convolutional layers and fully-connected layers

Architecture A is shown in Figure 2. In this architecture, the kernel size of 1D convolutional layer of 3×2 inputs (cytokines) was set to be 1×2 and 2×2 , respectively, as 3×2 would make it a fully-connected layer (Figure 5). Similarly, the kernel size of convolutional layers for the inputs 2×4 was fixed as 1×4 . The number of channels was set to be 2 for this layer to compensate for the reduction of size of output neurons compared to the size of input variables (4).

Activation functions “sigmoid”, “ReLU” [26] and “linear” were implemented for the two convolutional layers to conduct comparisons. The outputs of these two layers were then

concatenated with the rest of the features and fed into a fully-connected layer of size 3, with “sigmoid” activation and dropout rate 0.2. The output layer is fully-connected and of size 2 with “softmax” activation for the binary classification of LC.

B. Architecture B: composition of locally-connected layer and fully-connected layers

Architecture B is shown in Figure 4. In this architecture, the inputs of the two locally-connected layers are vectors of features where the measurements for the same variables were arranged together. The stride of the layer of the cytokines inputs was set to be 3, as the cytokines were measured for three times during the treatment. Similarly, the stride of the other locally-connected layers was set to be 2. The outputs of these two layers were then concatenated with other features and fed into exactly the same structure as described in A.

C. Architecture C: multi-layer perceptron

We also applied an MLP to conduct comparisons. The last two layers of the MLP were exactly the same as those of the architectures A and B. The size of the first fully-connected layer was set as 10 to be consistent with the size of the inputs of the second layer (the first fully-connected layer) in the architecture of A and B.

Table 2 is a summary of predictive performances of the three architectures A, B and C with their corresponding number of parameters (n). A single AUC in this table is an average of test AUCs from 25 times of 5-fold cross-validation (the random seed for partition of folds was changed each time). Brier scores (BS) [27], as a measure of the accuracy of prediction performance based on the test prediction and test labels in those cross-validations without SMOTE (for fair comparison), were also evaluated.

In general, architecture A which applied a 1D convolutional layer for longitudinal data yielded the best performance. Among the three implemented activations, “sigmoid” activation showed the best AUC of 0.83, “linear” activation also achieved a decent AUC of 0.81. However, “ReLU” activation did not work well in our case, achieving an AUC of only 0.79. The size of the kernel, specifically the length of the kernel which determined how many time steps would be considered in a single kernel almost did not affect the results. The AUCs under the two different settings of length of kernel were roughly the same (e.g., DeLong test of architecture A with “sigmoid” activation showed a p-value 0.814).

Architecture B, which applied locally-connected layers to the longitudinal data yielded an AUC of 0.80. This slightly outperformed the architecture C MLP in the prediction performance (AUC of 0.78).

Cross-validated ROC with confidence bands for architectures A, B and C are shown in Figure 6, Figure 7 and Figure 8. Architecture A (sigmoid activation) outperformed architecture B and C with p-values of 2.085×10^{-6} and 1.609×10^{-14} in the DeLong test. Architecture B outperformed C with a p-value of 1.407×10^{-4} in the DeLong test.

Brier score results also suggested similar performance to ROC with architecture A with “sigmoid” activation achieving the lowest score for predictive probability accuracy (BS = 0.157), as the best model.

In general, neural network architectures considering associations among longitudinal data yielded better predictive performance in LC prediction than the fully-connected architecture MLP. Specifically, the 1D convolutional layer with “sigmoid activation” performed the best in modeling longitudinal data.

V. DISCUSSION

This study aimed to investigate advanced neural network architectures that can take into account associations among biological, imaging and physical information in radiotherapy to improve outcome models. In radiation oncology, sample (e.g., patients) sizes for building predictive outcome models are usually limited. It is generally hard for such limited sample sizes to support the training of fully-connected architectures MLPs, which have many free parameters (weights). However, those biological, image and physical data are usually highly correlated with each other. Designing partially-connected architectures based new deep learning technologies can leverage these associations, reduce the number of free parameters, and potentially improve prediction performance.

This study focused on investigating architectures for longitudinal data in outcome modeling. In the treatment course of radiotherapy, some biological, dosimetric and image information are collected over time. Thus, considering partially connected architectures based on temporal associations can be beneficial. The two architectures were considered for this objective. One architecture applied locally-connected layers, which took a weighted sum of only measurements from the same variable. The effect of this layer was to get an overall index of each variable for LC prediction by weighting the values measured at different times. It ignored interactions among measurements of different variables. The second architecture for longitudinal data is a 1D convolutional layer, where invariance along time-axis is considered. The temporal invariance is under the assumption that different variables would correlate with each other in the same way over time. Hence, a weight sharing scheme was implemented and reduced the number of free parameters in the architectures.

Although, as mentioned earlier, separate convolutional layers were considered for cytokine and other data because their time steps were not consistent. However, this separation did make sense since variables of the same categories were usually more correlated with each other than variables of different categories. To better visualize this, a heat map of the correlation matrix of longitudinal data is shown in Figure 9. From the heat map, one can easily find two distinguishable groups which would rationalize the separation assumption applied here.

Further analysis of weights gives one some insights about how the architecture extract information for prediction. The weights of locally-connected layers in architecture B for cytokines are shown in Table 3, as well as spearman correlation coefficients between LC and those variables. Clearly, the effect of the locally-connected layer is to weigh the cytokines at different time points, likely giving more weight to the more relevant (to LC) time point. The relative change of cytokines is not as considered as all the weights are with the same sign. This is consistent with the fact, the change of cytokines showed weak correlation with LC, as in Table 3 “Change_corr” row, where correlation between LC and *2_week-pre*,

4_week-2_week cytokines are shown. Similarly, weights corresponding to image features and dose inputs were also shown in Table 4. For variables GLSZM.ZP and GLRLM.GLN, clearly, weights corresponding to pre-measurement are far larger than those of mid-measurements, which is consistent with results of spearman correlation, where the mid-measurement of the two variables showed weaker correlation with LC (correlation coefficient -0.07 and 0.008 , respectively.) For dose and GLSZM.ZP whose pre- and mid-measurement both yields relatively higher correlations, weights seem to be roughly equally given to the two time points. In general, locally- connected layers were able to extract important and relevant information from the inputs to eventually aid the prediction task, which are consistent with individual correlations in this data.

One may be also interested in applying recurrent neural networks (RNNs) [28], which exhibit more dynamic temporary behavior and are prevalent in the field of natural languages processing to model longitudinal data in radiation oncology. However, according to other experiments (not fully described here), gated recurrent units (GRUs) [29], a common recurrent layers, did not work for modeling these longitudinal data. The corresponding composite architecture achieved an AUC of only 0.766 for LC prediction. One explanation for this was the complexity involving in a GRU was not well supported by the limited sample size and number of time points. GRUs largely increased the number of free parameters compared to MLP, while our proposed partially-connected architectures decreased the number of free parameters. To realize a RNN in prediction of LC, simplified but still functional recurrent units may need to be developed.

It is recognized that the traditional regression rule of thumb “one in ten” for reducing overfitting risk and guaranteeing generalization of a prediction model outside the training data was not fulfilled here, as seen in Table 2. In radiation oncology research the ratio of the number of variables to events is usually unbalanced. As a result, an elaborate cross-validation scheme was implemented, where the performance of models is tested outside of the training data for relatively fair evaluation of generalization [30]. Moreover, utilization of deep learning technologies to take advantage of temporal correlations, regularization methods such as dropouts, modern training algorithms aim to relax this rule of thumb while reducing the risk of overfitting. However, it is acknowledged that external validation would be still desirable to further consolidate the results. However, no such available external dataset exists so far to our knowledge; RTOG 0617 dataset would be of interest once it becomes available.

It is worth noting that although we regarded the outcome prediction task here of LC as a binary classification problem, our proposed method is still valuable for survival analysis, where time-to-event is considered. In the medical field, survival analysis is a desirable tool to estimate the time of death, relapse or development of an adverse reaction for a group of patients. In this study, we adopted Kaplan-Meier estimator using the *R* package *prodlm* [31] to estimate local progression (LP) free (e.g., $LC=0$) survival probability for patients in the cross- validated test sets as predicted by the ANN outputs. The threshold between the two groups was set such that the number of high-risk patients equals to patients with LP in the original dataset to mitigate data imbalance issues. As shown in Figure 10, as expected, the patient group with the higher risks of LP has poorer performance over time. The differences

between the two groups is distinguishable across all time points and was confirmed by log-rank test (p -value < 0.0001). Despite the good survival prediction, the performance of our proposed architectures for survival analysis could be further improved by taking into consideration the event time explicitly as part of the ANN learning optimization problem, which is a subject of our future research.

This study investigated ways to take advantage of the temporal associations in longitudinal radiotherapy data (e.g., imaging and cytokines here). However, in radiotherapy, there are more associations among variables beyond temporal ones, e.g., variables along the same biophysical pathway may be highly- correlated. If an architecture which can take into account these correlations is developed, just as CNNs exploit spatial local correlation among adjacent pixels, it may even yield better prediction power and may reduce model's degrees of freedom under the circumstances of limited sample sizes. These associations can be obtained from better domain knowledge. Otherwise, an architecture which can automatically explore these associations and adjust the connections within the ANN would be desirable in the context of radiotherapy outcome modeling where training sizes are generally limited.

VI. CONCLUSIONS

This work investigated new architectures of artificial neural networks for prediction of LC in radiotherapy. In radiation oncology, because of the limited sample size, it is beneficial to take advantage of associations among predictive variables using new deep learning technologies to reduce model's complexity and increase model's generalization outside training data. According to this study, a 1D convolutional layer with "sigmoid" activation performed the best in modeling longitudinal data. Making use of temporal association by a 1D convolutional layer significantly improved LC prediction compared to a fully- connected MLP and a GRU-MLP composite neural networks.

Acknowledgments

This work was supported in part by grant from the NCI/NIH P01- CA059827.

VII. REFERENCES

- [1]. El Naqa I, "Outcome Modeling," in Informatics in radiation oncology, George Starkschall S; alfredo C R, Ed. Boca Raton, Florida: CRC press, 2014, pp. 257–275
- [2]. El Naqa I et al., "Datamining approaches for modeling tumor control probability," Acta oncologica (Stockholm, Sweden), vol. 49, no. 8, pp. 1363–1373, 3 2010.
- [3]. Coates J, Souhami L, and El Naqa I, "Big Data Analytics for Prostate Radiotherapy," Frontiers in Oncology, vol. 6, p. 149, 6 2016. [PubMed: 27379211]
- [4]. Su M, Miften M, Whiddon C, Sun X, Light K, and Marks L, "An artificial neural network for predicting the incidence of radiation pneumonitis," Med Phys, vol. 32, no. 2, pp. 318–325, 2 2005. [PubMed: 15789575]
- [5]. Chen S, Zhou S, Zhang J, Yin F-F, Marks LB, and Das SK, "A neural network model to predict lung radiation-induced pneumonitis," Med Phys, vol. 34, no. 9, pp. 3420–3427, 2007. [PubMed: 17926943]
- [6]. Ian G, "Optimization for Training Deep Models," in Deep Learning, Goodfellow I, Bengio Y, and Courville A, Eds. Cambridge, Massachusetts: MIT Press, 2016.

- [7]. Silver D et al., "Mastering the game of Go without human knowledge," *Nature*, Article vol. 550, p. 354, 10/18/online 2017. [PubMed: 29052630]
- [8]. Alex Krizhevsky I. a. H. and Sutskever Geoffrey E, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. P. a. C. J. C. B. a. L. B. a. K. Q. Weinberger, Ed.: Curran Associates, Inc, 2012, pp. 1097–1105.
- [9]. Graves A, Mohamed A. r., and Hinton G, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 6645–6649.
- [10]. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, and Salakhutdinov R, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [11]. Ioffe S and Szegedy C, "Batch normalization: accelerating deep network training by reducing internal covariate shift," presented at the *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, Lille, France, 2015.
- [12]. Duchi J, Hazan E, and Singer Y, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *J. Mach. Learn. Res*, vol. 12, pp. 2121–2159, 2011.
- [13]. Hinton G, Srivastava N, and Swersky K. (2014). *Neural Networks for Machine Learning*, Lecture 6a, Overview of mini-batch gradient descent
- [14]. Kingma DP and Ba J, "Adam: A Method for Stochastic Optimization," *CoRR*, vol. abs/1412.6980, 2014.
- [15]. Krizhevsky A, Sutskever I, and Hinton GE, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [16]. Levina V et al., "Role of CCL11/eotaxin-1 signaling in ovarian cancer," *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 15, no. 8, pp. 2647–2656, 4 2009. [PubMed: 19351767]
- [17]. Lewis AM, Varghese S, Xu H, and Alexander HR, "Interleukin-1 and cancer progression: the emerging role of interleukin-1 receptor antagonist as a novel therapeutic agent in cancer treatment," *Journal of Translational Medicine*, vol. 4, p. 48, 11 2006. [PubMed: 17096856]
- [18]. Feng Y et al., "MicroRNA-145 inhibits tumour growth and metastasis in colorectal cancer by targeting fascin-1," *British Journal of Cancer*, vol. 110, no. 9, pp. 2300–2309, 3 2014. [PubMed: 24642628]
- [19]. Li Q et al., "MicroRNA-574–5p Was Pivotal for TLR9 Signaling Enhanced Tumor Progression via Down-Regulating Checkpoint Suppressor 1 in Human Lung Cancer," *PLOS ONE*, vol. 7, no. 11, p. e48278, 2012. [PubMed: 23133627]
- [20]. Hung C-H, Chiu Y-C, Chen C-H, and Hu T-H, "MicroRNAs in Hepatocellular Carcinoma: Carcinogenesis, Progression, and Therapeutic Target," *BioMed Research International*, vol. 2014, p. 11, 2014, Art. no. 486407.
- [21]. Tsukamoto S et al., "Smad9 is a new type of transcriptional regulator in bone morphogenetic protein signaling," *Scientific Reports*, Article vol. 4, p. 7596, 12 2014. [PubMed: 25534700]
- [22]. Hall EJ and Giaccia AJ, *Radiobiology for the Radiologist* Lippincott Williams & Wilkins, 2006.
- [23]. LeCun Y, Bengio Y, and Hinton G, "Deep learning," *Nature*, vol. 521, p. 436, 05/27/online 2015. [PubMed: 26017442]
- [24]. Chollet F. (2015). *Keras*
- [25]. Chawla NV, Bowyer KW, Hall LO, and Kegelmeyer WP, "SMOTE: synthetic minority over-sampling technique," *J. Artif. Int. Res*, vol. 16, no. 1, pp. 321–357, 2002.
- [26]. Nair V and Hinton GE, "Rectified linear units improve restricted boltzmann machines," presented at the *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Haifa, Israel, 2010.
- [27]. Steyerberg E, "Evaluation of Performance," in *Clinical Prediction Models A Practical Approach to Development, Validation, and Updating*, Steyerberg E, Ed.: Springer-Verlag New York, 2009.
- [28]. Zachary Chase L, "A Critical Review of Recurrent Neural Networks for Sequence Learning," *CoRR*, vol. abs/1506.00019, 2015.

- [29]. C. C. Junyoung Gulcehre; KyungHyun Cho; Yoshua Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," CoRR, vol. abs/1412.3555, 2014.
- [30]. Vittinghoff E and McCulloch CE, "Relaxing the rule of ten events per variable in logistic and Cox regression," Am J Epidemiol, vol. 165, no. 6, pp. 710–8, 3 15 2007. [PubMed: 17182981]
- [31]. Gerds Thomas A., "R:Fast and user friendly implementation of nonparametric estimators for censored event history (survival) analysis. Kaplan-Meier and Aalen-Johansen method," ed, 2018.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

An illustration of a 1D convolutional layer for a 2D input, to which kernels of size 3×5 were applied, where T represents the number of time points and f denotes the number of variables in one time point.

Figure 2.

The diagram of architecture A built base on Keras (built-in functions applied: Input, Con1D, Flatten, Concatenate, Dense). In the input bock, T and f represent number of time points and number of variables at each time point respectively.

Figure 3.

An illustration of a locally-connected layers with stride=3 for a 1D input.

Figure 4. Diagram of architecture B building based on Keras (built-in functions applied: Input, Localconnect1D, Flatten, concatenate, Dense)

Figure 5.
Kernel size options for cytokines (3×2) inputs

Figure 6.

Cross-validated ROC curve with confidence band of architecture A (sigmoid activation) for prediction of LC. (Based on the predictions and true labels for the testing data in 25 times of cross-validation)

Figure 7.
Cross-validated ROC curve with confidence band of architecture B for prediction of LC.
(Based on the predictions and true labels for the testing data in 25 times of cross-validation)

Figure 8.

Cross-validated ROC curve with confidence band of architecture C for prediction of LC.
(Based on the predictions and true labels for the testing data in 25 times of cross-validation)

Figure 9.
Absolute value of Spearman correlations among longitudinal data.

Risk of LP (high) :	29	26	24	21	19	16	15	15	13	13	12
Risk of LP (low) :	69	69	64	64	62	61	60	59	59	59	58

Figure 10.

Survival curves with 95% CI for patient groups defined by ANN outputs (threshold=0.397).
(Based on survival time, true labels and predictions by architecture C for test patients in the cross validation).

Table 1.

Features that were applied for LC prediction.

Categories	Variables
Biological data: cytokines, (3 [*]) SNPs,(1) Micro-RNA(1)	Eotaxin, interleukin-1- α
	SMAD9
	145–5p, 574–5p, 122–5p
Dosimetric data (2)	Mean tumor doses
PET image data (2)	GLSZM.ZP, GLSZM.LGZ E, GLRLM.GLN

* The number in the brackets following the categories of variables denotes how many times the same variable was measured over in the treatment.
3* denotes longitudinal data.

Table 2.

Cross-validated AUC predictions of LC in architectures A, B and C. The activation applied for the convolutional layers in architecture A and the size of kernel of the convolutional layer for cytokines were shown in the table. For reference, Brier score of null models where all the patients were given an LC probability as population LC rates is 0.209.

Architecture	Number of free parameters	AUC with 95%CI	Brier* score
A: "ReLU" 1×2	54	0.785 (0.752–0.792)	0.189
A: "ReLU" 2×2	56	0.786 (0.757–0.796)	0.189
A: "linear" 1×2	54	0.814 (0.787–0.823)	0.182
A: "linear" 2×2	56	0.812 (0.779–0.820)	0.182
A: "sigmoid" 1×2	54	0.832 (0.807–0.841)	0.157
A: "sigmoid" 2×2	56	0.829 (0.804–0.838)	0.160
B	60	0.802 (0.775–0.811)	0.161
C	231	0.778 (0.751–0.790)	0.190

*
A lower Brier score indicate a better performance.

Table 3.

Weights corresponding to cytokines in architecture B as well as spearman rank correlation between LC and raw values of cytokines (Corr); and the change of cytokines (Change_corr)

		pre	2_week	4_week
Eotaxin	Weights	-0.25	-0.05	-0.21
	Corr	-0.176	-0.136	-0.175
	Change_corr	-0.013 -0.035		
Interlukin-1- α	Weights	-0.10	-0.07	-0.18
	Corr	-0.210	-0.175	-0.216
	Change_corr	0.05 -0.076		

Table 4.

Weights corresponding to dose and image features in architecture B as well as spearman rank correlation between LC and these variables (Corr); and the change of these variables (Change_corr)

		pre	mid
Mean tumor doses	Weights	0.28	0.22
	Corr	0.273	0.231
	Change_corr	-0.016	
GLSZM.ZP	Weights	0.42	0.52
	Corr	0.148	0.123
	Change_corr	-0.122	
GLSZM.LGZE	W	-0.44	-0.01
	Corr	-0.18	-0.07
	Change_corr	-0.070	
GLRLM.GLN	Weights	0.84	0.11
	Corr	0.10	0.008
	Change_corr	0.072	