

Original Paper

# Real-Time Respiratory Tumor Motion Prediction Based on a Temporal Convolutional Neural Network: Prediction Model Development Study

Panchun Chang<sup>1,2\*</sup>, MA; Jun Dang<sup>1\*</sup>, PhD; Jianrong Dai<sup>3</sup>, PhD; Wenzheng Sun<sup>4</sup>, PhD

<sup>1</sup>Department of Oncology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China

<sup>2</sup>School of Physics and Electronics, Shandong Normal University, Jinan, China

<sup>3</sup>Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

<sup>4</sup>Department of Radiation Oncology, School of Medicine, The Second Affiliated Hospital, Zhejiang University, Hangzhou, China

\*these authors contributed equally

**Corresponding Author:**

Wenzheng Sun, PhD

Department of Radiation Oncology, School of Medicine

The Second Affiliated Hospital

Zhejiang University

88 Jiefang Road

Hangzhou, 310009

China

Phone: 86 057187783538

Email: [sunwenzheng@zju.edu.cn](mailto:sunwenzheng@zju.edu.cn)

## Abstract

**Background:** The dynamic tracking of tumors with radiation beams in radiation therapy requires the prediction of real-time target locations prior to beam delivery, as treatment involving radiation beams and gating tracking results in time latency.

**Objective:** In this study, a deep learning model that was based on a temporal convolutional neural network was developed to predict internal target locations by using multiple external markers.

**Methods:** Respiratory signals from 69 treatment fractions of 21 patients with cancer who were treated with the CyberKnife Synchrony device (Accuray Incorporated) were used to train and test the model. The reported model's performance was evaluated by comparing the model to a long short-term memory model in terms of the root mean square errors (RMSEs) of real and predicted respiratory signals. The effect of the number of external markers was also investigated.

**Results:** The average RMSEs of predicted (ahead time=400 ms) respiratory motion in the superior-inferior, anterior-posterior, and left-right directions and in 3D space were 0.49 mm, 0.28 mm, 0.25 mm, and 0.67 mm, respectively.

**Conclusions:** The experiment results demonstrated that the temporal convolutional neural network-based respiratory prediction model could predict respiratory signals with submillimeter accuracy.

(*J Med Internet Res* 2021;23(8):e27235) doi: [10.2196/27235](https://doi.org/10.2196/27235)

**KEYWORDS**

radiation therapy; temporal convolutional neural network; respiratory signal prediction; neural network; deep learning model; dynamic tracking

## Introduction

The aim of radiation therapy is not only to deliver lethal doses of radiation to target tumors but also to minimize the dose of unnecessary radiation delivered to the surrounding healthy tissues and structures [1-5]. Modern technical advances, such

as intensity-modulated radiation therapy, have improved the accuracy of dose delivery. However, some targets, such as lung cancer and liver cancer tumors, may move substantially during the treatment delivery process due to respiratory motion [6-10]. Investigators have reported that lung and liver tumors can move up to 3 cm during a conventional radiation therapy treatment

session [11,12]. The motion of targets may substantially decrease the accuracy and efficiency of intensity-modulated radiation therapy or other advanced technologies.

Many methods have been investigated to reduce the effect of respiratory motion, which mainly include the following:

- Adding a margin around the target tumor: a 10- to 15-mm margin is always used as the radiation treatment field to avoid missing a tumor, which may result in unnecessary radiation exposure to healthy tissues and structures [13].
- Breath hold: patients need to hold their breath during the treatment to temporarily stop respiration, but this is not applicable for some patients, such as older patients and juvenile patients [14].
- Beam tracking: radiation beams track a moving tumor dynamically to ensure that the tumor target is constantly within the treatment field [15].

All beam tracking methods must compensate for the latency of various sources, such as latencies from beam adjustment and image capture times [5,16]. Hence, we must estimate the position of targets in advance to compensate for latency effects.

Recently, deep learning approaches based on long short-term memory (LSTM) have been successfully used to solve time series prediction problems in several fields. For example, Ma et al [17] used an LSTM model to capture traffic dynamics data for predicting short-term traffic speed. Bao et al [18] implemented an LSTM model to predict the one-step-ahead price (closing) of 6 stock indices for various financial markets. Lin et al [19] used an LSTM model to predict respiratory signals. Moreover, some recent studies have demonstrated that certain temporal convolutional neural network (TCN) architectures

could achieve state-of-the-art accuracy in time series prediction problems [20-23]. However, to our knowledge, there are no studies on using a TCN model to predict respiratory tumor motion. Hence, in this study, we developed a TCN-based respiratory prediction model by using external markers and compared the prediction performance of the TCN to that of an LSTM model. We also investigated the effect that the number of external markers had on prediction performance.

## Methods

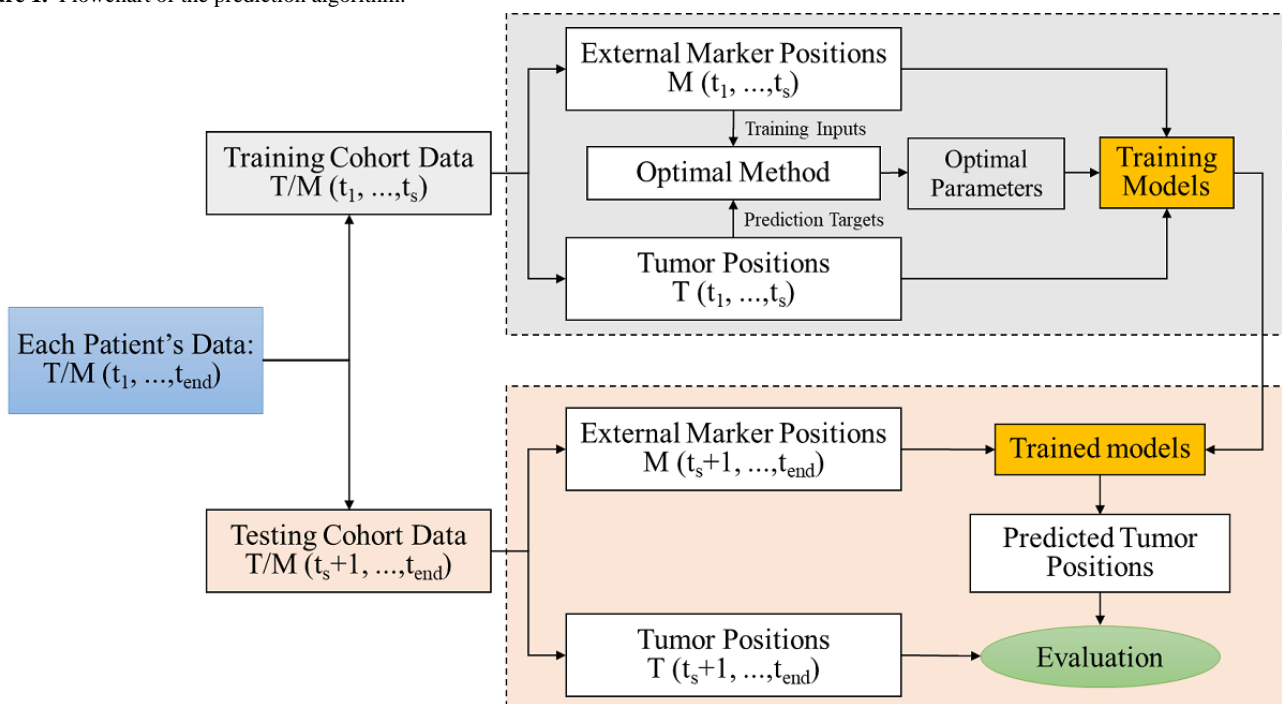
### Data Acquisition

The tumor motion data (69 treatment fractions of 21 patients) used in this study were obtained from an open data set, which was recorded by the CyberKnife Synchrony (Accuray Incorporated) tracking system with a recorded sampling rate of 25 Hz [24]. To analyze the external movements of patients, charge-coupled device cameras were used to monitor the luminous diodes located on a patient's abdomen and chest. To analyze internal fiducial positions, orthogonal diagnostic x-ray systems were used to observe implanted markers periodically.

### Prediction Process

The general scheme for the prediction process of 2 models is outlined in Figure 1, and the arrangement of the respiratory signals that were used for network training and validation is shown in Table 1. Each recorded position (internal tumor and external marker positions) was stratified into 2 cohorts based on time  $t_s$ . The positions prior to time  $t_s$  (the training signals) were used to train the TCN and LSTM models. The positions after  $t_s$  (the testing signals) were used to evaluate the developed model.

Figure 1. Flowchart of the prediction algorithm.



**Table 1.** The arrangement of respiratory signals used for network training and validation.

Position type	Data for training	Data for validation
<b>Inputs of the network</b>		
Position of marker 1	$M^a_{1_{SI, AP, LR}}(1, 2, \dots, t_s)$	$M_{1_{SI, AP, LR}}(t_s+1, 2, \dots, t_s+t_{end})$
Position of marker 2	$M_{2_{SI, AP, LR}}(1, 2, \dots, t_s)$	$M_{2_{SI, AP, LR}}(t_s+1, 2, \dots, t_s+t_{end})$
Position of marker 3	$M_{3_{SI, AP, LR}}(1, 2, \dots, t_s)$	$M_{3_{SI, AP, LR}}(t_s+1, 2, \dots, t_s+t_{end})$
<b>Targets of the network</b>		
Position of a tumor	$T^e_{SI, AP, LR}(1, 2, \dots, t_s)$	$T_{SI, AP, LR}(t_s+1, 2, \dots, t_s+t_{end})$

<sup>a</sup>M: external marker position.

<sup>b</sup>SI: superior-inferior.

<sup>c</sup>AP: anterior-posterior.

<sup>d</sup>LR: left-right.

<sup>e</sup>T: tumor position.

For the training process, the training input data and prediction target data were first used to tune the hyperparameters, which was done by using a cross-validation model. Afterward, they were used to train the model. The external markers' positions during the first input period of the training process (ie, the time between  $t=1$  and  $t=t_{delay}$ ) were used as the training input data for predicting the tumor positions (target positions) at a specific time frame ( $t=t_{delay}+t_{ahead}$ ). This training process was repeated and continued to predict the next tumor position until either the threshold of the cost function or the maximum iteration number, which was set in advance, was reached. Each pair of data points (ie, the input data,  $M[t+1, \dots, t+t_{delay}]$ , vs the output data,  $T[t+t_{delay}+t_{ahead}]$ ) consisted of a training data set. "M" denoted 3 external markers' positions (M1, M2, and M3), which were based on 3 directions (the superior-inferior, anterior-posterior, and left-right directions).  $t_{ahead}$  represented the ahead time we needed for making predictions.

For the evaluation process, the testing signals, which were represented as  $M(t_s+1, t_s+2, \dots, t_{end})$  and  $T(t_s+1, t_s+2, \dots, t_s+t_{end})$ , were used to evaluate the developed model. Similar to the process implemented in the training process, the external markers' positions during the first input period of the evaluation process (ie, the time between  $t=1$  and  $t=t_{delay}$ ) were used to predict a tumor's position ( $T'[t_s+t_{delay}+t_{ahead}]$ ) at a specific time ( $t=t_s+t_{delay}+t_{ahead}$ ). This process was also repeated to predict the next tumor position continuously. The external signals that were

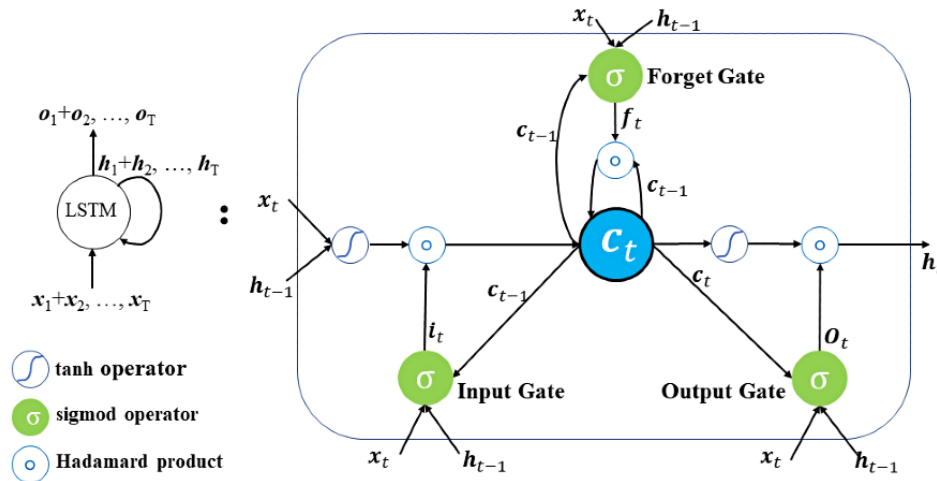
recorded during radiation therapy (ie, the time between  $t=t_{end}-t_{delay}-t_{ahead}+1$  and  $t=t_{end}-t_{ahead}$ ) were used to predict the final tumor position ( $T'[t_{end}]$ ). Finally, the predicted signals ( $T'[t_s+t_{delay}+t_{ahead}], \dots, T'[t_{end}]$ ) were compared to the real tumor positions ( $T[t_s+t_{delay}+t_{ahead}], \dots, T[t_{end}]$ ).

### LSTM Model

The recurrent neural network (RNN) is a particular type of neural network that allows for self-cycle connections and transmits parameters across different time stamps. An RNN model can store the information of former time stamps. However, it is difficult for the RNN to memorize long-term memory information due to vanishing and exploding gradients [25-27].

The LSTM layer is a special RNN layer that overcomes the weakness that the RNN has with memorizing long-term memory information [26,28]. Figure 2 shows an LSTM unit. Unlike the simple RNN unit, the LSTM unit has a memory cell state  $c_t$  at time  $t$ . The information that passes through state  $c_t$  is controlled by the following three gates: the input gate ( $i_t$ ), the forget gate ( $f_t$ ), and the output gate ( $o_t$ ). The input gate is used to control input data that flow into state  $c_t$ , the hidden state connection ( $h_t$ ) is used to control the forgetting of state  $c_t$ , and the output gate is used to moderate the output data that flow from state  $c_t$ . A plurality of LSTM layers can be stacked in a deeper neural network, which can fit the data of the complicated functions that are required to analyze the inputs and the targets.

**Figure 2.** The structure of an LSTM layer. LSTM: long short-term memory.



**TCN Model**

The TCN model was based on a transformation of a 1D fully convolutional network that was used for sequential prediction problems. The TCN model used a multilayer network to learn information over a long time span. Sequence information were transmitted layer by layer across the network until prediction results were obtained. The architecture of the TCN model is illustrated in Figure 3 [23], in which  $x_1, x_2, \dots, x_T$  are the original sequence signals (inputs), and  $\hat{o}_1, \hat{o}_2, \dots, \hat{o}_T$  are the prediction signals (outputs). The obvious characteristics of the TCN model, which were compared to those of the normal 1D fully convolutional network model, were as follows:

- The TCN model used causal convolutions, in which the output at time t was convolved only with elements from previous layers at time t and earlier, to ensure that no leakage occurred from the future into the past.
- The TCN model used dilated convolutions to ensure that each hidden layer had the same size as the input sequence

and to increase the receptive field (ie, learning longer lengths of information).

The input of the TCN model was interval sampled. The equation for the dilated convolution was as follows:

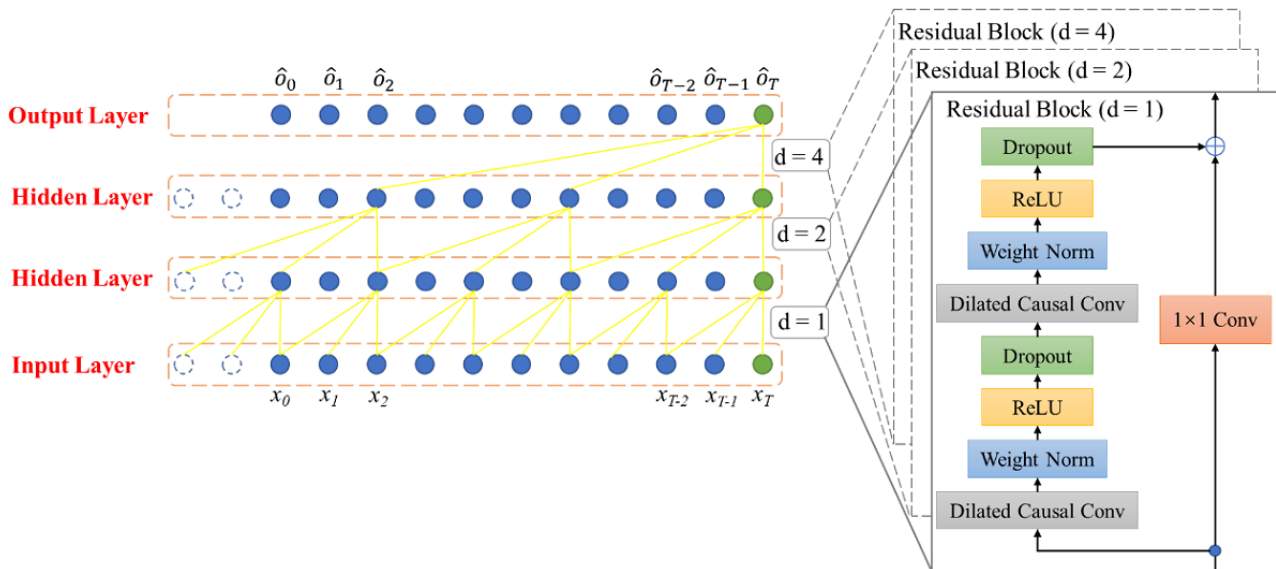
$$F(s) = (x *_d) * (s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (1)$$

In equation 1,  $d$  is the dilation factor (sampling rate). A  $d$  value of 1 in the lowest layer meant that every signal was sampled, whereas a  $d$  value of 2 in the middle layer meant that every 2 respiratory signals were sampled.

Residual networks [29], which are shown in Figure 3, were imported in this study to accelerate convergence and stabilize training. A residual block that included a branch was used to make a series of transformations ( $F$ ). Afterward, the outputs of the residual block (ie,  $F[X_{residual}]$ ) were added to the input (ie,  $X_{residual}$ ), as follows:

$$O_{residual} = \text{Activation}(X_{residual} + F[X_{residual}]) \quad (2)$$

**Figure 3.** The architecture of the temporal convolutional neural network model. "d" was the dilation factor. Conv: convolution; ReLU: rectified linear unit.



### Hyperparameter Tuning

With regard to the TCN model, previous TCN studies [20-23] reported (in the *Instruction* section) using the same TCN architecture and only sometimes varying the number of layers (*n*) and the filter size. Hence, we tested these two hyperparameters and used a dilation factor (*d*) of 2<sup>*n*</sup> for layer *n*. Moreover, the number of neurons in the input layer and the learning rate of the TCN model were also investigated in this study. For the LSTM model, the number of LSTM layers, learning rate, number of hidden units per layer, and number of neurons in the input layer were investigated. Furthermore, the Adam algorithm was used as the optimization algorithm for both the TCN model and LSTM model. The Kingma and Ba [30] study demonstrated that the hyperparameters in the Adam model required little tuning. Goodfellow et al [31] also approved of the robustness of the Adam model for their hyperparameter of choice and provided advice on how to tune the learning rate from the default value. Hence, we used the good default settings that were tested by Kingma and Ba [30] as the hyperparameters of the Adam optimizer and tuned the learning rate. The default settings were exponential decay rates of 0.9 and 0.999 and a decay exponent of 10<sup>-8</sup>. In this study, all hyperparameters were tuned synthetically by using a grid search model. It should be noted that we tested the hyperparameters in a 4D hyperparameter space instead of a subspace (ie, while a parameter was investigated, others were fixed) to maintain the accuracy of hyperparameter tuning.

### Model Evaluation

The respiratory signals from 69 treatment fractions of 21 patients with cancer who were treated with the CyberKnife Synchrony (Accuray Incorporated) device were used to evaluate the proposed model. Of the 69 treatment fractions, 5 were used to tune the hyperparameters. The rest of the patients were used to evaluate prediction performance. For each of the 69 treatment

fractions, signals that were acquired around the first 3 minutes (4500 data points) were used as the training signals for training the prediction model, and signals from the following 30 seconds were used as the test signals for assessing the effectiveness of the proposed model. The ahead time (*t*<sub>ahead</sub>) used in this study was 400 ms [1,5].

The root mean square errors (RMSEs) between real and predicted signals of respiratory motion in a 3D space were used for assessment [6,7]. The RMSEs for motion in each direction (*RMSE*<sub>*SI, LR, AP*</sub>) and motion in a 3D space (*RMSE*<sub>3D</sub>) were calculated by using equations 3 and 4, respectively, as follows:

$$RMSE_{SI,LR,AP} = \sqrt{\frac{1}{t_{end} - t_{start} + 1} \sum_{t=t_{start}}^{t_{end}} (\overline{T}_{SI,LR,AP}(t) - \overline{T}_{SI,LR,AP}(t))^2} \quad (3)$$

$$RMSE_{3D} = \sqrt{\frac{1}{t_{end} - t_{start} + 1} \sum_{t=t_{start}}^{t_{end}} (D(t))^2} \quad (4)$$

$$D(t) = \sqrt{(\overline{T}_{SI}(t) - \overline{T}_{SI}(t))^2 + (\overline{T}_{AP}(t) - \overline{T}_{AP}(t))^2 + (\overline{T}_{LR}(t) - \overline{T}_{LR}(t))^2} \quad (5)$$

In equation 5,  $\overline{T(t)}$  is the average of the true values, and  $\overline{T'(t)}$  is the average of predicted values. Time point *t* in equation 3 ranged from *t*<sub>start</sub> (*t*<sub>s</sub>+*t*<sub>delay</sub>+*t*<sub>ahead</sub>) to *t*<sub>end</sub>. The Wilcoxon signed-rank test was used as the statistical model for evaluating the differences between true values and predicted values.

### Results

Table 2 presents the RMSEs of the three models (ie, the LSTM, TCN, and no prediction models; ahead time=400 ms). Compared to the no prediction model, the RMSEs for motion in a 3D space were reduced by 46% in the LSTM model and 51% in the TCN model. For motion in all directions, the RMSEs of the TCN model were consistently lower than those of the LSTM model. The RMSE for motion in a 3D space decreased from 0.73 mm (LSTM model) to 0.67 mm (TCN model). The *P* value was <.001, indicating that the TCN method could significantly improve the prediction performance of the LSTM method.

**Table 2.** The root mean square errors (RMSEs) of the three prediction models.

Direction	RMSEs (mm) of the LSTM <sup>a</sup> model	RMSEs (mm) of the TCN <sup>b</sup> model	RMSEs (mm) of the no prediction model
Anterior-posterior direction	0.29	0.28	0.50
Left-right direction	0.27	0.25	0.45
Superior-inferior direction	0.55	0.49	1.04
3D space	0.73	0.67	1.36

<sup>a</sup>LSTM: long short-term memory.

<sup>b</sup>TCN: temporal convolutional neural network.

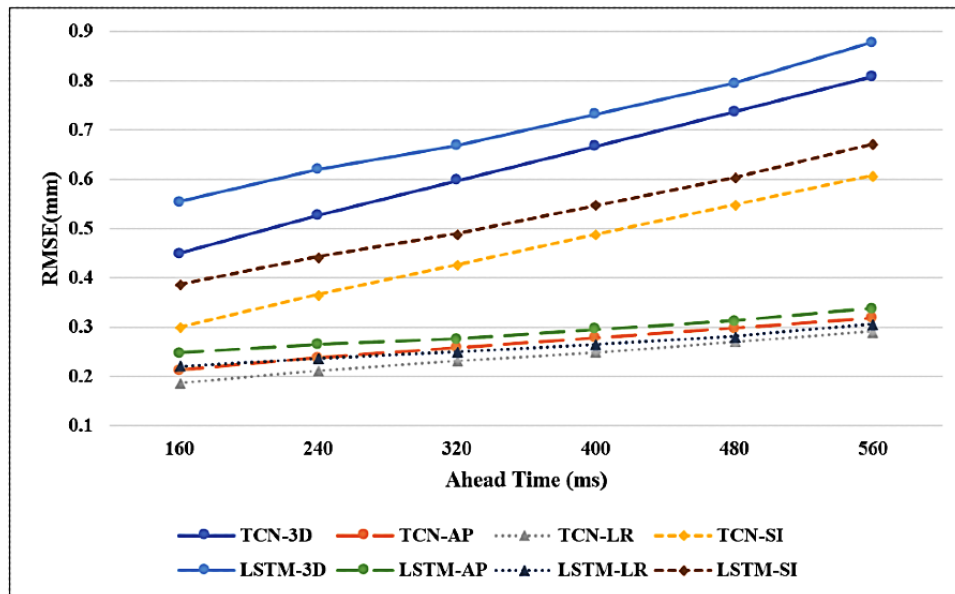
Figure 4 shows the RMSEs for motion in all directions with different ahead times. Obviously, the prediction performance of the TCN model was positive compared to that of the LSTM model for all ahead times. Further, the prediction performance of both models worsened as ahead times increased.

Figure 5 illustrates the performance comparison between the TCN and LSTM methods for predicting motion in the superior-inferior direction, anterior-posterior direction, and

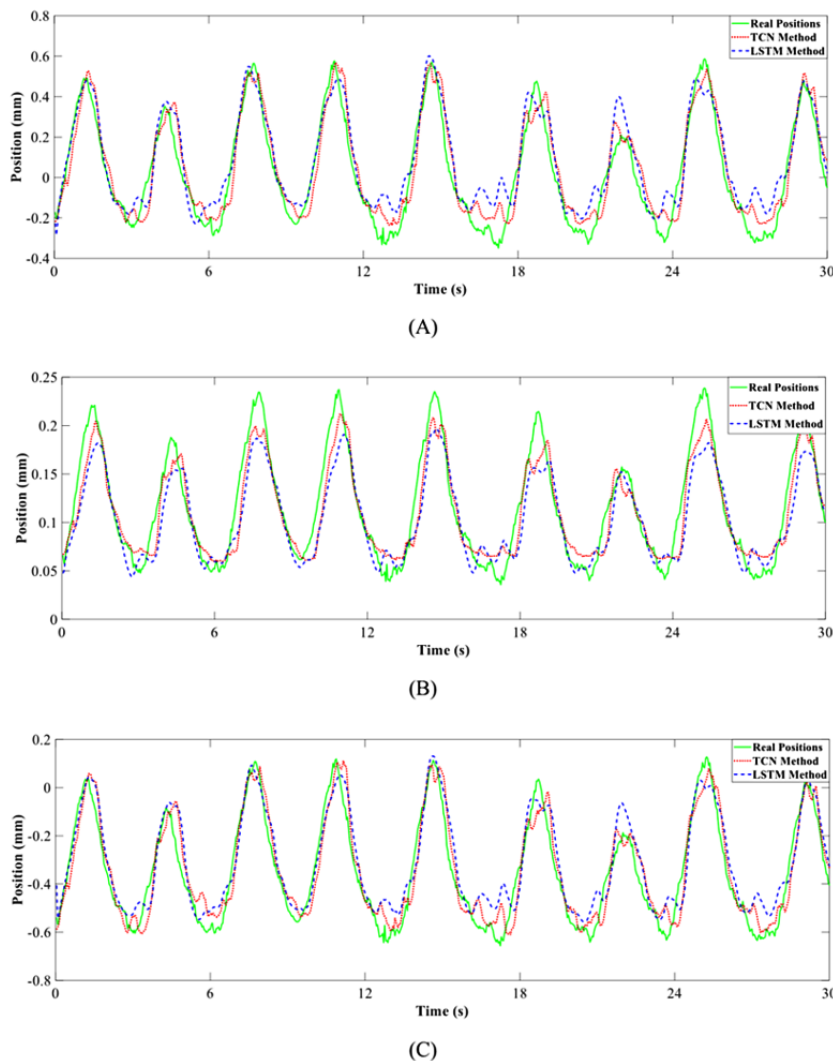
left-right direction. Obviously, the TCN method was more accurate and robust than the LSTM method.

We investigated the hyperparameters in the 4D hyperparameter space (625 experiments) for both the TCN and LSTM models by using the grid search method among 5 treatment fractions, which were selected randomly. The options and results of hyperparameter tuning are depicted in Table 3.

**Figure 4.** The RMSEs for respiratory motion in all directions. These were determined by using the LSTM and TCN models and different ahead times for each treatment fraction. AP: anterior-posterior; LR: left-right; LSTM: long short-term memory; RMSE: root mean square error; SI: superior-inferior; TCN: temporal convolutional neural network.



**Figure 5.** The performance comparison between the TCN and LSTM methods for predicting motion in the (A) superior-inferior direction, (B) left-right direction, and (C) anterior-posterior direction. LSTM: long short-term memory; TCN: temporal convolutional neural network.



**Table 3.** The options and results of hyperparameter tuning.

Models and hyperparameters	Hyperparameter options	Hyperparameter selected
<b>Temporal convolutional neural network model</b>		
Number of layers	4, 5, 6, 7, and 8	5
Filter size	1, 3, 5, 7, and 9	9
Number of neurons in the input layer	5, 10, 15, 20, and 25	15
Learning rate	0.0001, 0.001, 0.005, 0.01, and 0.1	0.001
<b>LSTM<sup>a</sup> model</b>		
Number of LSTM layers	1, 2, 3, 4, and 5	2
Learning rate	0.0001, 0.001, 0.005, 0.01, and 0.1	0.01
Number of hidden units per layer	10, 50, 100, 150, 200, and 250	200
Number of neurons in the input layer	5, 10, 15, 20, and 25	20

<sup>a</sup>LSTM: long short-term memory.

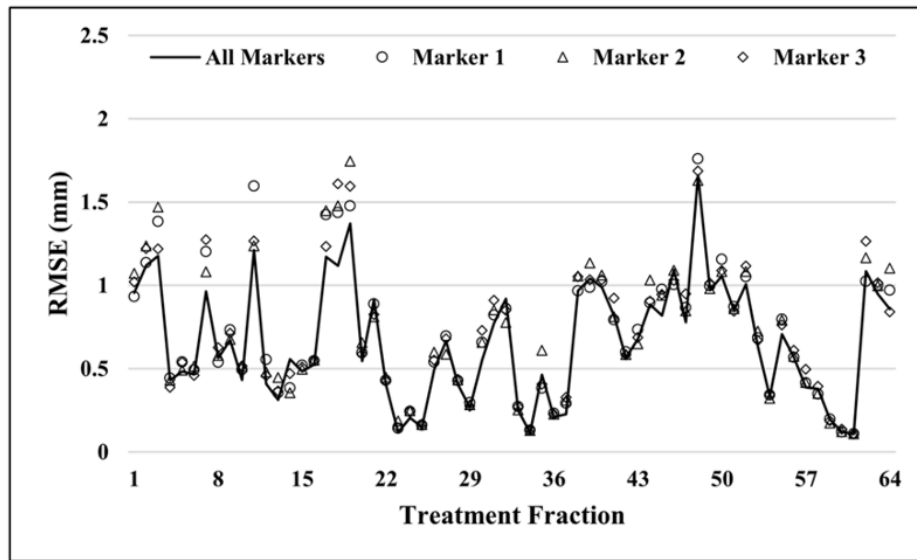
Table 4 presents the RMSEs of the TCN model for each external marker. Figure 6 shows the RMSEs for respiratory motion in a 3D space among each treatment fraction. The TCN model using 1 or 2 external markers was compared to the TCN model using all 3 external markers. The TCN model had the best performance in terms of predicting motion in all directions when all three external markers were used simultaneously. The average RMSEs for motion in a 3D space when using 1 marker and 2 markers were 0.72 mm and 0.68 mm, respectively. This decreased to 0.67 mm when using all three markers.

As illustrated in Figure 7, the ablative analysis of the TCN was also conducted. We focused on two components in this study—the filter size and the residual blocks. We found that the effect of the filter size was small when the filter size was larger than 3. The *P* values between 5 filter size pairs—filter sizes 1 and 3, 3 and 5, 5 and 7, and 7 and 9—were <.001, .11, .20, and .83, respectively. This indicated that prediction performance improved significantly before the filter size rose to 3. Further, we found that the residual blocks contributed significantly to prediction performance, as the *P* value was <.001.

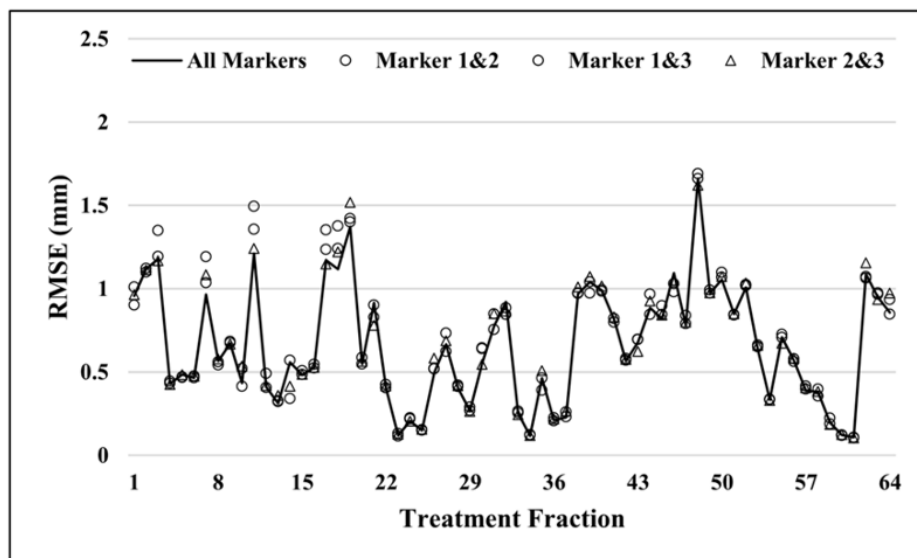
**Table 4.** The root mean square errors (RMSEs) of the temporal convolutional neural network model for each external marker (EM).

Direction	RMSEs for all EMs	RMSEs for EMs 1 and 2	RMSEs for EMs 1 and 3	RMSEs for EMs 2 and 3	RMSEs for EM 1	RMSEs for EM 2	RMSEs for EM 3
Anterior-posterior direction	0.28	0.28	0.28	0.28	0.29	0.29	0.29
Left-right direction	0.25	0.26	0.26	0.25	0.27	0.26	0.26
Superior-inferior direction	0.49	0.51	0.50	0.50	0.52	0.53	0.53
3D space	0.67	0.69	0.68	0.68	0.71	0.72	0.72

**Figure 6.** A comparison of RMSEs for respiratory motion in a 3D space among each treatment fraction. A: Results of the TCN model using 1 external marker compared to those of the TCN model using all 3 external markers. B: Results of the TCN model using 2 external markers compared to those of the TCN model using all 3 external markers. RMSE: root mean square error; TCN: temporal convolutional neural network.



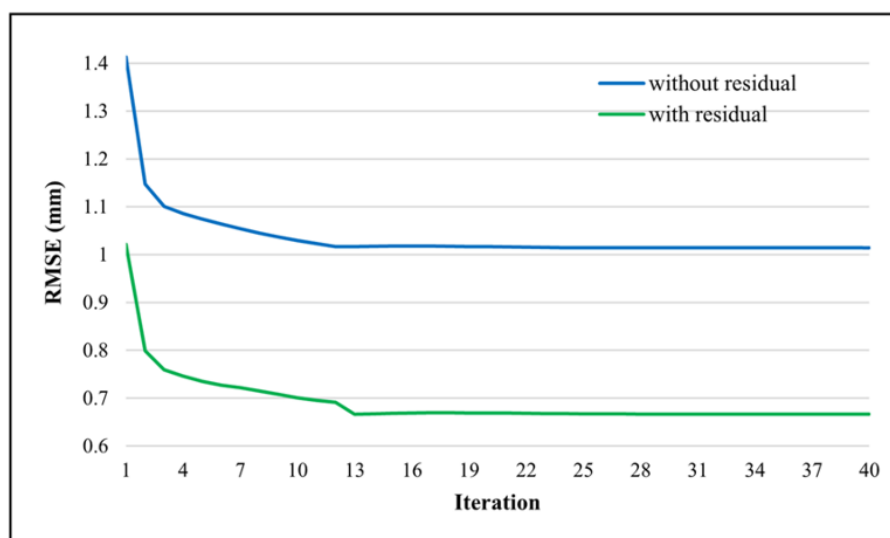
(A)



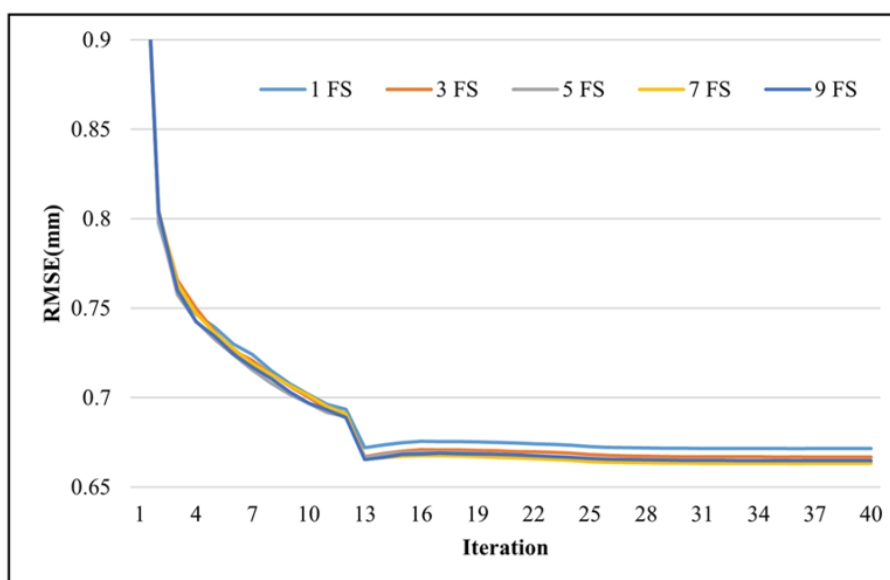
(B)



**Figure 7.** The effects of different components in the temporal convolutional neural network layer. A: Residual blocks. B: FS. FS: filter size; RMSE: root mean square error.



(A)



(B)

## Discussion

### Principal Findings

A TCN model for predicting respiratory motion by using external markers' prior signals was developed and tested in this study. The experiment demonstrated that the TCN model's performance in predicting future respiratory signals with a 400-ms ahead time was better than that of the LSTM model.

As is well known, hyperparameter settings have a large influence on the prediction performance of machine learning models. This also holds true for our TCN and LSTM models. We tuned 4 major hyperparameters for both of the TCN and LSTM models. Among these hyperparameters, the number of neurons in the input layer and the learning rate were tested for both models. Having a large number of neurons in the input layer allows for the inclusion of more features in models. Obviously, useful

features may increase prediction accuracy. However, redundancy features may also be brought in along with the useful features. Hence, if this hyperparameter is too large, prediction performance may degenerate. The best number of neurons in the input layer for the TCN and LSTM models in this study was 15 and 20, respectively. The learning rate was an important hyperparameter in the model optimization process. If the learning rate is too large, the model may oscillate around the global minimum value instead of achieving convergence. On the other hand, if this value is too small, the training time and the risk of overfitting increase. Learning rates of 0.001 and 0.01 were selected as the optimal hyperparameters of the TCN and LSTM models, respectively. In addition to the two abovementioned hyperparameters, the number of layers and filter sizes were also investigated for the TCN model, whereas the number of LSTM layers and number of hidden units per layer were tested for the LSTM model. With regard to the TCN

model, the size of the effective window (receptive field) increased as the number of layers and filter size increased. Hence, these two hyperparameters should guarantee that the receptive field of TCN model covers enough context for respiratory signal prediction. The optimal values for these two hyperparameters in our experiments were 5 and 9, respectively. With regard to the LSTM model, on one hand, a deeper LSTM model (a large number of LSTM layers) may be representative of a more complicated relationship among respiratory signals and improve prediction performance. On the other hand, a deeper LSTM model also has an increased risk of overfitting and increased convergence speed. In this study, the prediction performance results of the LSTM model were comparable when the number of LSTM layers was over 2. Hence, we selected 2 as the optimal number of LSTM layers. Further, the number of hidden units per layer determined the width of each LSTM layer. We also found that having a large number of hidden units per layer was helpful for establishing a more complicated prediction model, but at the same time, this increased the risk of overfitting and convergence speed.

The effect that different numbers of external markers had on prediction performance was also investigated in this study. The TCN model had the best prediction performance when it used all three markers' positions. As shown in [Figure 6](#), the TCN model's prediction performance when using 3 markers was more robust than when using 1 marker or 2 markers. For most treatment fractions, the RMSEs of the TCN model using 3 markers was slightly smaller than those obtained by using 1 marker or 2 markers. However, for some treatment fractions, such as treatment fractions 7 and 11, the RMSEs of predictions

based on 1 or 2 external markers were quite larger than those of predictions based on 3 external markers. This was probably because having more external markers for different skin surface positions resulted in the inclusion of more useful features. Such useful features may alleviate the overfitting and underfitting problems.

Finally, we studied the influence of the different components (the filter size and residual blocks) in the TCN model. The size of the effective window (receptive field) increased with filter size. Hence, the model's prediction performance initially became better as the filter size increased. However, the model's prediction performance only slightly improved as the filter size increased continually. This may be because the receptive field that resulted from using a filter size of 3 provided enough context for the respiratory signal prediction task. On the other hand, we observed that the residual block architecture enhanced the model's prediction performance immensely. We believe that this was because the residual blocks effectively allowed the TCN model to be modified based on identity mapping instead of a full transformation, which was crucial for the deep neural network architecture.

## Conclusion

A deep learning approach based on the TCN architecture was developed to predict internal tumor positions with a 400-ms ahead time based on the external markers' positions in this study. The results demonstrated that this model could predict tumor positions accurately. Further, the prediction performance of the TCN model using multiple external markers was more robust and positive than that of the TCN model using 1 or 2 external markers.

## Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (62103366), the General Project of Chongqing Natural Science Foundation (grant cstc2020jcyj-msxm2928), Seed Grant of the First Affiliated Hospital of Chongqing Medical University (grant PYJJ2019-208), Chongqing Municipal Bureau of Human Resources and Social Security Fund (grant cx2018147), and Medical Research Key Project of Jiangsu Health Commission (grant ZDB 2020022).

## Conflicts of Interest

None declared.

## References

1. Ren Q, Nishioka S, Shirato H, Berbeco RI. Adaptive prediction of respiratory motion for motion compensation radiotherapy. *Phys Med Biol* 2007 Nov 21;52(22):6651-6661. [doi: [10.1088/0031-9155/52/22/007](https://doi.org/10.1088/0031-9155/52/22/007)] [Medline: [17975289](https://pubmed.ncbi.nlm.nih.gov/17975289/)]
2. McCall KC, Jeraj R. Dual-component model of respiratory motion based on the periodic autoregressive moving average (periodic ARMA) method. *Phys Med Biol* 2007 Jun 21;52(12):3455-3466. [doi: [10.1088/0031-9155/52/12/009](https://doi.org/10.1088/0031-9155/52/12/009)] [Medline: [17664554](https://pubmed.ncbi.nlm.nih.gov/17664554/)]
3. Bukhari W, Hong SM. Real-time prediction and gating of respiratory motion using an extended Kalman filter and Gaussian process regression. *Phys Med Biol* 2015 Jan 07;60(1):233-252. [doi: [10.1088/0031-9155/60/1/233](https://doi.org/10.1088/0031-9155/60/1/233)] [Medline: [25489980](https://pubmed.ncbi.nlm.nih.gov/25489980/)]
4. Verma P, Wu H, Langer M, Das I, Sandison G. Survey: Real-time tumor motion prediction for image-guided radiation treatment. *Comput Sci Eng* 2011 Sep;13(5):24-35. [doi: [10.1109/mcse.2010.99](https://doi.org/10.1109/mcse.2010.99)]
5. Riaz N, Shanker P, Wiersma R, Gudmundsson O, Mao W, Widrow B, et al. Predicting respiratory tumor motion with multi-dimensional adaptive filters and support vector regression. *Phys Med Biol* 2009 Oct 07;54(19):5735-5748. [doi: [10.1088/0031-9155/54/19/005](https://doi.org/10.1088/0031-9155/54/19/005)] [Medline: [19729711](https://pubmed.ncbi.nlm.nih.gov/19729711/)]
6. Yan H, Yin FF, Zhu GP, Ajlouni M, Kim JH. The correlation evaluation of a tumor tracking system using multiple external markers. *Med Phys* 2006 Nov;33(11):4073-4084. [doi: [10.1118/1.2358830](https://doi.org/10.1118/1.2358830)] [Medline: [17153387](https://pubmed.ncbi.nlm.nih.gov/17153387/)]

7. Sun WZ, Jiang MY, Ren L, Dang J, You T, Yin FF. Respiratory signal prediction based on adaptive boosting and multi-layer perceptron neural network. *Phys Med Biol* 2017 Aug 03;62(17):6822-6835 [FREE Full text] [doi: [10.1088/1361-6560/aa7cd4](https://doi.org/10.1088/1361-6560/aa7cd4)] [Medline: [28665297](https://pubmed.ncbi.nlm.nih.gov/28665297/)]
8. Ernst F, Schlaefer A, Schweikard A. Predicting the outcome of respiratory motion prediction. *Med Phys* 2011 Oct;38(10):5569-5581. [doi: [10.1118/1.3633907](https://doi.org/10.1118/1.3633907)] [Medline: [21992375](https://pubmed.ncbi.nlm.nih.gov/21992375/)]
9. Torshabi AE, Riboldi M, Fooladi AAI, Mosalla SMM, Baroni G. An adaptive fuzzy prediction model for real time tumor tracking in radiotherapy via external surrogates. *J Appl Clin Med Phys* 2013 Jan 07;14(1):4008 [FREE Full text] [doi: [10.1120/jacmp.v14i1.4008](https://doi.org/10.1120/jacmp.v14i1.4008)] [Medline: [23318386](https://pubmed.ncbi.nlm.nih.gov/23318386/)]
10. Vergalasova I, Cai J, Yin FF. A novel technique for markerless, self-sorted 4D-CBCT: feasibility study. *Med Phys* 2012 Mar;39(3):1442-1451 [FREE Full text] [doi: [10.1118/1.3685443](https://doi.org/10.1118/1.3685443)] [Medline: [22380377](https://pubmed.ncbi.nlm.nih.gov/22380377/)]
11. Berbeco RI, Nishioka S, Shirato H, Jiang SB. Residual motion of lung tumors in end-of-inhale respiratory gated radiotherapy based on external surrogates. *Med Phys* 2006 Nov;33(11):4149-4156. [doi: [10.1118/1.2358197](https://doi.org/10.1118/1.2358197)] [Medline: [17153393](https://pubmed.ncbi.nlm.nih.gov/17153393/)]
12. Shirato H, Suzuki K, Sharp GC, Fujita K, Onimaru R, Fujino M, et al. Speed and amplitude of lung tumor motion precisely detected in four-dimensional setup and in real-time tumor-tracking radiotherapy. *Int J Radiat Oncol Biol Phys* 2006 Mar 15;64(4):1229-1236. [doi: [10.1016/j.ijrobp.2005.11.016](https://doi.org/10.1016/j.ijrobp.2005.11.016)] [Medline: [16504762](https://pubmed.ncbi.nlm.nih.gov/16504762/)]
13. Goodband JH, Haas OCL, Mills JA. A comparison of neural network approaches for on-line prediction in IGRT. *Med Phys* 2008 Mar;35(3):1113-1122. [doi: [10.1118/1.2836416](https://doi.org/10.1118/1.2836416)] [Medline: [18404946](https://pubmed.ncbi.nlm.nih.gov/18404946/)]
14. Nelson C, Starkschall G, Balter P, Fitzpatrick MJ, Antolak JA, Tolani N, et al. Respiration-correlated treatment delivery using feedback-guided breath hold: a technical study. *Med Phys* 2005 Jan;32(1):175-181. [doi: [10.1118/1.1836332](https://doi.org/10.1118/1.1836332)] [Medline: [15719968](https://pubmed.ncbi.nlm.nih.gov/15719968/)]
15. Hansen R, Ravkilde T, Worm ES, Toftegaard J, Grau C, Macek K, et al. Electromagnetic guided couch and multileaf collimator tracking on a TrueBeam accelerator. *Med Phys* 2016 May;43(5):2387. [doi: [10.1118/1.4946815](https://doi.org/10.1118/1.4946815)] [Medline: [27147350](https://pubmed.ncbi.nlm.nih.gov/27147350/)]
16. Sharp GC, Jiang SB, Shimizu S, Shirato H. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Phys Med Biol* 2004 Feb 07;49(3):425-440. [doi: [10.1088/0031-9155/49/3/006](https://doi.org/10.1088/0031-9155/49/3/006)] [Medline: [15012011](https://pubmed.ncbi.nlm.nih.gov/15012011/)]
17. Ma X, Tao Z, Wang Y, Yu H, Wang Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp Res Part C Emerg Technol* 2015 May;54:187-197. [doi: [10.1016/j.trc.2015.03.014](https://doi.org/10.1016/j.trc.2015.03.014)]
18. Bao W, Yue J, Rao Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS One* 2017 Jul 14;12(7):e0180944. [doi: [10.1371/journal.pone.0180944](https://doi.org/10.1371/journal.pone.0180944)] [Medline: [28708865](https://pubmed.ncbi.nlm.nih.gov/28708865/)]
19. Lin H, Shi C, Wang B, Chan MF, Tang X, Ji W. Towards real-time respiratory motion prediction based on long short-term memory neural networks. *Phys Med Biol* 2019 Apr 10;64(8):085010 [FREE Full text] [doi: [10.1088/1361-6560/ab13fa](https://doi.org/10.1088/1361-6560/ab13fa)] [Medline: [30917344](https://pubmed.ncbi.nlm.nih.gov/30917344/)]
20. Dauphin YN, Fan A, Auli M, Grangier D. Language modeling with gated convolutional networks. 2017 Aug Presented at: The 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia.
21. Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN. Convolutional sequence to sequence learning. 2017 Aug Presented at: The 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia.
22. Kalchbrenner N, Espeholt L, Simonyan K, van den Oord A, Graves A, Kavukcuoglu K. Neural machine translation in linear time. arXiv. Preprint posted online on October 31, 2016 [FREE Full text]
23. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv. Preprint posted online on April 19, 2018 [FREE Full text]
24. Suh Y, Dieterich S, Cho B, Keall PJ. An analysis of thoracic and abdominal tumour motion for stereotactic body radiotherapy patients. *Phys Med Biol* 2008 Jul 07;53(13):3623-3640. [doi: [10.1088/0031-9155/53/13/016](https://doi.org/10.1088/0031-9155/53/13/016)] [Medline: [18560046](https://pubmed.ncbi.nlm.nih.gov/18560046/)]
25. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Netw* 1994;5(2):157-166. [doi: [10.1109/72.279181](https://doi.org/10.1109/72.279181)] [Medline: [18267787](https://pubmed.ncbi.nlm.nih.gov/18267787/)]
26. Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: A deep learning approach. *J Biomed Inform* 2017 May;69:218-229 [FREE Full text] [doi: [10.1016/j.jbi.2017.04.001](https://doi.org/10.1016/j.jbi.2017.04.001)] [Medline: [28410981](https://pubmed.ncbi.nlm.nih.gov/28410981/)]
27. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. 2013 Jun 16 Presented at: The 30th International Conference on International Conference on Machine Learning (ICML); June 16-21, 2013; Atlanta, USA.
28. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997 Nov 15;9(8):1735-1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)] [Medline: [9377276](https://pubmed.ncbi.nlm.nih.gov/9377276/)]
29. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, Nevada, USA. [doi: [10.1109/cvpr.2016.90](https://doi.org/10.1109/cvpr.2016.90)]
30. Kingma DP, Ba JL. Adam: A method for stochastic optimization. arXiv. Preprint posted online on December 22, 2014 [FREE Full text]
31. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, Massachusetts: MIT Press; Nov 2016.

## Abbreviations

**LSTM:** long short-term memory  
**RMSE:** root mean square error  
**RNN:** recurrent neural network  
**TCN:** temporal convolutional neural network

*Edited by R Kukafka; submitted 18.01.21; peer-reviewed by B Li, N Kim; comments to author 17.03.21; revised version received 04.06.21; accepted 05.07.21; published 27.08.21*

*Please cite as:*

*Chang P, Dang J, Dai J, Sun W*

*Real-Time Respiratory Tumor Motion Prediction Based on a Temporal Convolutional Neural Network: Prediction Model Development Study*

*J Med Internet Res 2021;23(8):e27235*

*URL: <https://www.jmir.org/2021/8/e27235>*

*doi: [10.2196/27235](https://doi.org/10.2196/27235)*

*PMID: [34236336](https://pubmed.ncbi.nlm.nih.gov/34236336/)*

©Panchun Chang, Jun Dang, Jianrong Dai, Wenzheng Sun. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.