

RESEARCH ARTICLE

Clinical applicability of deep learning-based respiratory signal prediction models for four-dimensional radiation therapy

Sangwoon Jeong¹ , Wonjoong Cheon² , Sungkoo Cho³, Youngyih Han^{1,4*} 

1 Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Korea, **2** Proton Therapy Center, National Cancer Center, Goyang, Korea, **3** Department of Radiation Oncology, Samsung Medical Center, Seoul, Korea, **4** Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

 These authors contributed equally to this work.

* youngyih@skku.edu



OPEN ACCESS

Citation: Jeong S, Cheon W, Cho S, Han Y (2022) Clinical applicability of deep learning-based respiratory signal prediction models for four-dimensional radiation therapy. PLoS ONE 17(10): e0275719. <https://doi.org/10.1371/journal.pone.0275719>

Editor: Ngie Min Ung, University of Malaya, MALAYSIA

Received: February 24, 2022

Accepted: September 21, 2022

Published: October 18, 2022

Copyright: © 2022 Jeong et al. 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 author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available from GitHub (<https://github.com/SangWoonJeong/Respiratory-prediction>).

Funding: This work was supported by National Research Foundation of Korea (NRF) and the Korean government (MSIT) (NRF-2019R1F1A1062775 and 2019M2A2B4096537) Receiver of the funds: YH URL: <https://www.nrf.re.kr/eng/index> The funders had no role in study

Abstract

For accurate respiration gated radiation therapy, compensation for the beam latency of the beam control system is necessary. Therefore, we evaluate deep learning models for predicting patient respiration signals and investigate their clinical feasibility. Herein, long short-term memory (LSTM), bidirectional LSTM (Bi-LSTM), and the Transformer are evaluated.

Among the 540 respiration signals, 60 signals are used as test data. Each of the remaining 480 signals was split into training and validation data in a 7:3 ratio. A total of 1000 ms of the signal sequence (T_s) is entered to the models, and the signal at 500 ms afterward (P_t) is predicted (standard training condition). The accuracy measures are: (1) root mean square error (RMSE) and Pearson correlation coefficient (CC), (2) accuracy dependency on T_s and P_t , (3) respiratory pattern dependency, and (4) error for 30% and 70% of the respiration gating for a 5 mm tumor motion for latencies of 300, 500, and 700 ms. Under standard conditions, the Transformer model exhibits the highest accuracy with an RMSE and CC of 0.1554 and 0.9768, respectively. An increase in T_s improves accuracy, whereas an increase in P_t decreases accuracy. An evaluation of the regularity of the respiratory signals reveals that the lowest predictive accuracy is achieved with irregular amplitude patterns. For 30% and 70% of the phases, the average error of the three models is <1.4 mm for a latency of 500 ms and >2.0 mm for a latency of 700 ms. The prediction accuracy of the Transformer is superior to LSTM and Bi-LSTM. Thus, the three models have clinically applicable accuracies for a latency <500 ms for 10 mm of regular tumor motion. The clinical acceptability of the deep learning models depends on the inherent latency and the strategy for reducing the irregularity of respiration.

1. Introduction

Radiation therapy can achieve high dose conformity with intensity-modulated radiation therapy and particle therapy, which can deliver a prescription dose to the target volume while minimizing undesirable doses near critical organs [1, 2]. However, respiratory movements of the

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

patient can result in the administration of undesired doses to the target volume and nearby organs at risk (OARs) [3–5]. Several studies have shown that patient respiration causes organ movements up to 40.0, 39.0, 23.0, and 10.0 mm for the liver, pancreas, kidney, and prostate, respectively [6, 7].

To reduce the dose delivery uncertainty associated with patient respiration during radiation treatment, various methods and technologies, such as deep inspiration breath-hold (DIBH), chest compression, real-time tracking, and respiratory gating, have been introduced.

The DIBH and chest compression methods can reduce dose uncertainty by physically minimizing the movement of the patient's chest caused by respiration during both computed tomography (CT) simulation and treatments [8–11]. However, for patients who find it difficult to maintain a breath-hold or sustain a chest compression, DIBH or chest compression is not clinically feasible [12, 13]. The real-time tracking method controls beam irradiation by following the tumor position, detected from X-ray fluoroscopy images taken simultaneously during the treatment. Respiratory phase or amplitude gating is a widely used four-dimensional radiation therapy (4DRT) strategy. It synchronizes treatment beam irradiation with patient respiration, represented by an external surrogate, and delivers a beam only at the planned respiration phases (amplitudes). A respiratory-gated treatment plan is developed, as follows. First, the beam delivery phases (amplitudes), such as 30–70% phases (amplitudes) of CT images, are selected from 10 time-resolved phases (amplitudes) of four-dimensional (4D) CT images. Second, the target and OARs are delineated in each of the selected phases (amplitudes) from the CT images, and a dose distribution is computed by average intensity projection (AIP) or maximum intensity projection. Finally, the radiation beam is delivered when the patient's respiration reaches and is within the selected phases (amplitudes). Although the respiratory gating technique may extend treatment time, it promises an improvement in treatment outcomes and reduces the probability of complications [14, 15].

From a technical point of view, an important prerequisite of precise respiratory-gated radiation therapy in the aforementioned methods is compensating for any system latency of the beam control or modulation system. The system latency is the time delay between the instructed and the actual beam on/off during respiratory-gated radiation therapy. Medical linear accelerators (linacs) have system latencies ranging from 300 ms to 800 ms; the Elekta (Stockholm, Sweden) linacs have latencies of 300–800 ms, and the Varian (Crawley, United Kingdom) linacs have latencies of 300–500 ms [16–18]. The system latency can cause position errors to the target and OARs of up to 7.6 mm [19]. Various mathematical models, such as the autoregressive moving average model (ARIMA) [20, 21], sinusoidal model [22, 23], and Kalman filter [24, 25], have been used to predict respiratory signals. Moreover, recent studies based on machine and deep-learning models demonstrated that the prediction accuracy of such models was superior to that of mathematical models in predicting time series data, and that they resulted in a 150% accuracy improvement [26].

Among various deep-learning models, the recurrent neural network (RNN) is designed to be suitable for time sequence data [27] and has been used with mathematical filters for respiration data prediction [28]. However, the RNN exhibited a deficiency in vanishing or exploding gradients problem [29, 30]. To resolve this deficiency, long short-term memory (LSTM) was introduced with the gradient clipping method by a forget-gate [31]. LSTM is characterized by a three-gate architecture that stores long-term memory and performs well on long time-series data [32]. Lin *et al.* [33] applied LSTM to respiration data acquired by a real-time patient monitoring (RPM) system and demonstrated good performance by optimizing the hyperparameters of the model. A bidirectional long short-term memory network (Bi-LSTM), another RNN variant, was developed to enhance prediction accuracy by using both directions of information on time sequence data [34]. Wang *et al.* [35] reported the superior performance of a

seven-layer Bi-LSTM over a mathematical model (autoregressive integrated moving average) and an artificial neural network model (adaptive boosting and multi-layer perceptron neural network) in predicting respiration with a 400 ms latency of the CyberKnife robotic radiosurgery system (Accuray, Sunnyvale, CA, USA). However, comparing the performance of Bi-LSTM with LSTM is somewhat controversial because a model based on Bi-LSTM achieves a superior performance with stock market prediction data [36], whereas a model based on LSTM has better accuracy in research on reservoir inflow forecasting [37]. Another deep-learning algorithm that is suitable for time sequence data is the Transformer. The Transformer is based on an attention mechanism approach, but without recurrent layers [38]. It has the potential to achieve a higher prediction accuracy for time-varying patient respiratory signals [39, 40].

These state-of-the-art deep learning algorithms can be used to realize accurate 4DRT technology. However, none of the studies thus far have evaluated these three models using the same set of clinical data. Moreover, only statistical measures, such as the mean absolute error and root mean square error (RMSE), have been provided as performance metrics in previous studies, thereby limiting clear understanding of the relevant error associated with each model in clinical practice.

Therefore, the performances of LSTM, Bi-LSTM, and the Transformer were evaluated for clinical respiratory signals acquired from patients undergoing proton therapy. In addition, the associated tumor targeting error in gated radiation therapy was measured for various machine latency models, and its clinical applicability was evaluated.

2. Materials and methods

2.A. Patient respiratory signal data

The data used in this study consisted of 540 respiration signals obtained from 442 patients, who received proton therapy for liver, lung, and breast cancer treatments. The patients took CT simulation with guided free-breathing and were trained in advance by a medical physicist to ensure regular respiratory signals could be obtained using an in-house developed respiration guiding system. Respiration signals were recorded during CT simulation using a respiration gating system (AZ-733VI, Anzai Medical Co. Ltd, Tokyo, Japan). The respiration gating system continuously recorded the respiratory pattern by measuring the distance from the self-emitting laser source to the body surface of the patient at 20 Hz [41]. The recorded time series of the patient's respiratory signal ranged from 84.35 to 272.50 s, with an average recording time of 145.26 s. This was a retrospective study of patients who received proton therapy. The patient's respiratory data were recorded from January 01, 2020 to December 31, 2021. The study protocol was approved by the Institutional Review Board of Samsung Medical Center (IRB number 2019-10-159). All respiratory data were fully anonymized before they were accessed.

In addition, the actual target motion in the 4DCT of the patients was analyzed for clinical evaluation. In the anterior-posterior direction, the mean tumor motion was 6.53 mm and the standard deviation was 3.70 mm. In the superior-inferior direction, the mean tumor motion was 11.42 mm and the standard deviation was 8.28 mm. According to the recommendation of the American Association of Physicists in Medicine (AAPM) task group 76a (TG-76a), the maximum tumor motion caused by respiratory motion was limited to 5.0 mm (peak to peak) for external-beam radiation therapy [42].

2.B. Data preparation

Of the total 540 respiration data, 480 respiration data were used as training and validation data, and the remaining 60 respiration data were assigned to the test dataset. Each signal of the

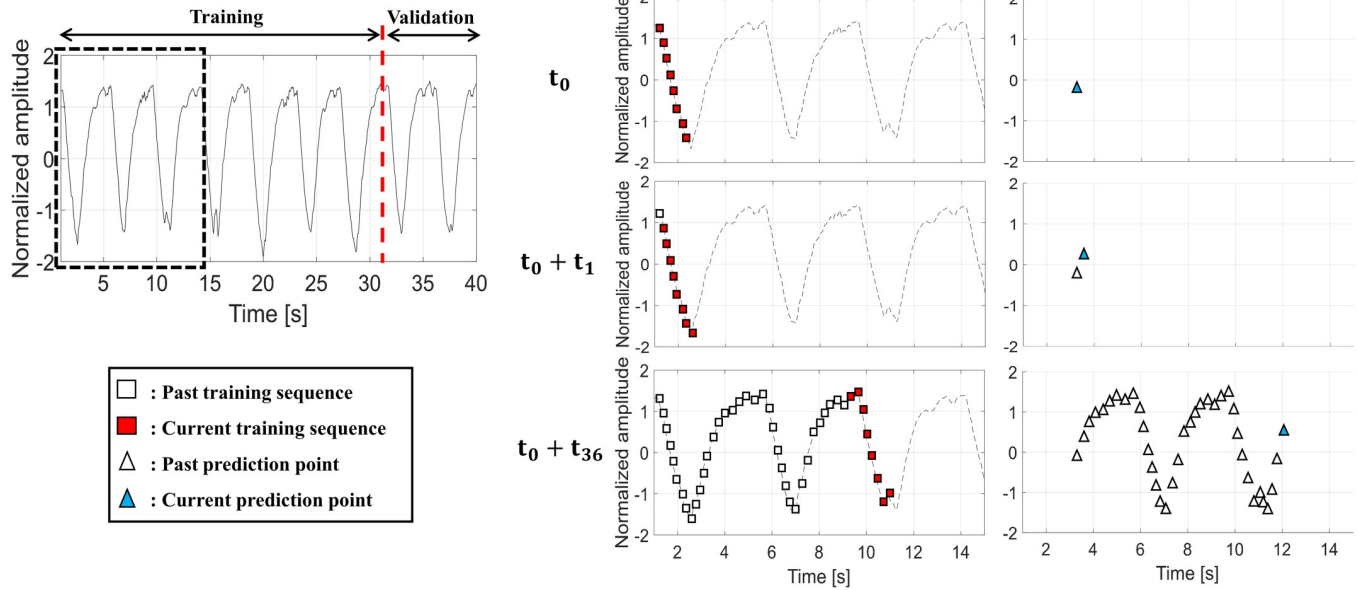


Fig 1. Process of learning a respiratory signal prediction model using deep learning.

<https://doi.org/10.1371/journal.pone.0275719.g001>

480 respiration data was divided into training and validation data for the deep learning models in a 7:3 ratio (Fig 1).

The input of a deep-learning model predicting the respiratory signal was a training sequence (T_s), and the output of the model was the prediction point (P_t) far away from the last point in the T_s by the system latency. An example of input and output for the prediction model is shown in Fig 1. In Fig 1, if $t = t_0$ and the T_s is eight time-points, the sequence consisting of eight points was used as an input and the point P_t was predicted by the network. When $t = t_0 + t_n$, the start point of T_s moved forward by t_n ; furthermore, P_t moved forward by t_n . In the numerical experiments, the standard condition was set to $T_s = 1000$ ms and $P_t = 500$ ms.

In the case of preprocessing procedures, Z-score normalization was performed to match the baseline of the patient’s respiration signal and to quickly converge the deep learning model [43]. The Savitzky–Golay finite impulse response smoothing (S–G) filter can improve precision without distorting the signal trend [44]. The S–G filter was applied only to the output of the training and test data in the postprocessing step while raw respiratory signals were entered to network. AZ-733VI, a respiration measurement device, had no real scale value because respiration was measured using the phase gating method (there is no unit). Therefore, the quantity can be interpreted as a normalized amplitude.

To analyze the effect of the regularity of the respiratory pattern on the prediction accuracy, the respiration data were classified into four different groups: patterns with regular periods, patterns with irregular periods (type 1), patterns with irregular amplitude (type 2), and patterns with irregular periods and amplitudes (type 3). We defined a respiratory irregularity value using the mean of the standard deviation (STD) in the peaks and the valleys (Eq 1) [45].

$$\text{Irregularity value} = \frac{\text{STD}(\text{peaks}) + \text{STD}(\text{valleys})}{2}. \tag{1}$$

The amplitude irregularity was computed using the mean value of the STDs of the amplitudes at the peaks and valleys. For phase irregularity, the periods were computed from the peak and valley times in the signal. The phase irregularity values were computed using the

mean of the STDs of the peak-to-peak periods and valley-to-valley periods. After computing both types of irregularity values for all signals, the 48 signals with high amplitude irregularity values were assigned to the type 1 group, and 48 signals with high phase irregularity values were include in the type 2 group. After summation of both types of irregularity values, the 48 signals with high summation values were assigned into the type 3, and 48 signals with low values were assigned to the regular type.

2.C. Deep learning models for respiratory signal prediction

This section described the three deep learning model characteristics, and the structure of each model is presented in Fig 2.

2.C.1. LSTM

The LSTM structure is based on an RNN. LSTM is composed of a forget gate, input gate, and output gate. The forget gate (f_t) uses the previous LSTM output (h_{t-1}) and the current input (x_t) to determine how much of the previous cell state (C_{t-1}) information is to be maintained (Eq 2). The value of f_t is between 0 and 1. If the value of the f_t gate is 0, C_{t-1} is not used; if it is 1, all C_{t-1} information is used in Eq 5.

The input gate decides whether to add new information to the current cell state (C_t) by using h_{t-1} and x_t (Eq 4). The input gate consists of two layers: a layer (i_t) that selects the values to be updated using a σ (Eq 3) and a layer (\tilde{C}_t) that creates a new candidate value vector using the hyperbolic tangent function (\tanh) (Eq 4).

To update the current cell state (C_t), each element-wise product of the vectors at f_t and C_{t-1} and i_t and \tilde{C}_t are calculated, and the two resulting values are added (Eq 5).

The output gate (O_t) determines the part of C_t to be updated using h_{t-1} and x_t through σ (Eq 6). Output (h_t) is calculated by taking C_t to \tanh , mapping a value between -1 and 1 , and

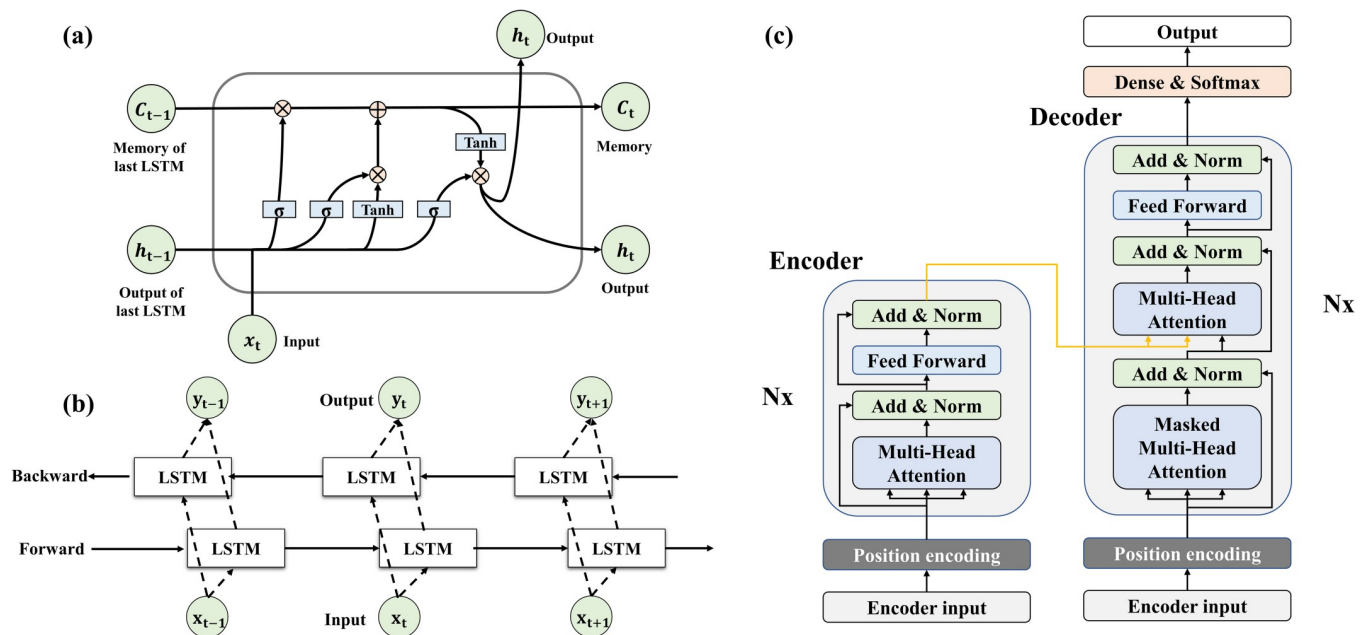


Fig 2. Network structures of the deep learning models used for respiration prediction. (a) Long-short term memory (LSTM): calculation using a forget gate based on RNN, (b) Bidirectional-LSTM: calculations using LSTMs in the backward and forward directions, (c) Transformer: iteratively computes N encoders and decoders.

<https://doi.org/10.1371/journal.pone.0275719.g002>

determining O_t and the element-wise product of the vectors (Eq 7).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{2}$$

$$i_t = \sigma(W_c \cdot [h_{t-1}, x_t] + b_c), \tag{3}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \tag{4}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \tag{5}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{6}$$

$$h_t = O_t \odot \tanh(C_t), \tag{7}$$

2.C.2. Bi-LSTM

Bi-LSTM has been widely used for natural language translation. Bi-LSTM is a variant of LSTM that uses bidirectional information. It adds information in a direction opposite that of LSTM.

The Bi-LSTM output (y_t) calculation multiplies and adds the \vec{h}_t and \overleftarrow{h}_t outputs of the forward and backward LSTMs, respectively (Eq 8).

$$y_t = W_{\vec{h}_y} \vec{h}_t + W_{\overleftarrow{h}_y} \overleftarrow{h}_t + b_y. \tag{8}$$

2.C.3. Transformer

The Transformer is a model that is implemented using the attention mechanism. To achieve computational efficiency, the Transformer uses only multi-head attention mechanisms without convolutional layers and recursive structures in encoders and decoders. The details of the Transformer are described in the original paper [38].

The Transformer encoders are composed of an input layer, four identical encoder layers, and a position-encoding layer with cosine functions. The identical encoder layer has a multi-head attention layer and a fully connected feed-forward network. The multi-head attention layer is a core structure of the Transformer that can be described as mapping a query (Q), key (K), and value (V) (Eq 9). Technically, these three entities were optimized during the training procedures.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{9}$$

where K^T is the transpose of K, and $\sqrt{d_k}$ is the dimension of Q and K. The role of the multi-head attention layer was to allow the model to jointly obtain information from different representation subspaces at different positions. The output of the multi-head attention was passed into the fully connected feed-forward network, which consists of two linear transformations with a rectified linear unit activation function. Subsequently, the output of the fully connected feed-forward network was followed by layer normalization and a residual connection with a decoder.

The decoder consists of an input layer, four identical decoder layers, and an output layer. In the case of the decoder, a third sublayer was inserted into the two sub-layers in each identical decoder layer. The third sublayer was a masked multi-head attention layer that could prevent self-attention. Based on the original paper, we used look-ahead masking and one-position

offset between the decoder input and target output in the decoder to ensure that the prediction of a time-series data point depended only on previous data points. Finally, the output layer mapped the output of the last decoder layer to the target time sequence.

2.D. Training of LSTM, Bi-LSTM, and the Transformer for respiratory signal prediction

To train the LSTM, Bi-LSTM, and Transformer to predict a respiratory signal for respiratory-gated radiation therapy, the hyperparameters of each model were determined as follows.

To compare the accuracies of all three models when they achieved their best performances, the LSTM and Bi-LSTM models consisted of 15 hidden layers with a size of 3 nodes, respectively [33]. The Transformer model consisted of eight multi-head layers for the attention mechanism, six encoder layers, and six decoder layers [38]. The number of the hidden layers and multi-head layers were empirically determined.

Adaptive moment estimation [45] was used for the three deep-learning models, with a learning rate of 0.0001 and weight decay of 0.0002. The two beta parameters, which were default parameters relevant to the running average of the gradient in the adaptive moment optimizer, were set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The batch size for the training was set to 300. Model training was performed for 100 epochs, and the best validation performance model was used for evaluation. The training procedure was performed using an NVIDIA GeForce 2080Ti graphic processing unit on Pytorch 1.5.1. The best model was updated when the current validation loss was smaller than the previous validation loss obtained during the training procedure.

2.F. Evaluation

To evaluate the performance of the models, the difference between the actual and predicted respiratory signals was quantitatively assessed by computing the RMSE (Eq 10), and Pearson's correlation coefficient (CC, Eq 11).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}, \quad (10)$$

where y_i is the actual respiratory signal and \hat{y}_i is the predicted respiratory signal.

The CC was used to analyze the linear relationship between the two continuous variables. The CC takes values in the interval $[-1.0, 1.0]$. If the value of CC is closer to 1.0, on an absolute scale, the correlation between the ground truth and the predicted value is stronger.

$$\text{CC} = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}}, \quad (11)$$

where y_i is the actual respiratory signal; \hat{y}_i is the predicted respiratory signal; and \bar{y} and $\bar{\hat{y}}$ are the average values of y_i and \hat{y}_i , respectively. In addition, the statistical significance of the differences in the predictions achieved with the three models was analyzed. The P-values of the RMSE and CC of each prediction model were computed using a one-way analysis of variance (one-way ANOVA).

For the comparative study, the prediction accuracy was calculated on the i) standard condition of the T_s and P_t . In addition, we analyzed the ii) effect of the length of T_s and P_t on prediction accuracy and iii) the effect of the regularity of the respiratory pattern prediction accuracy. Finally, we analyzed the suitability of the three deep learning-based respiratory signal

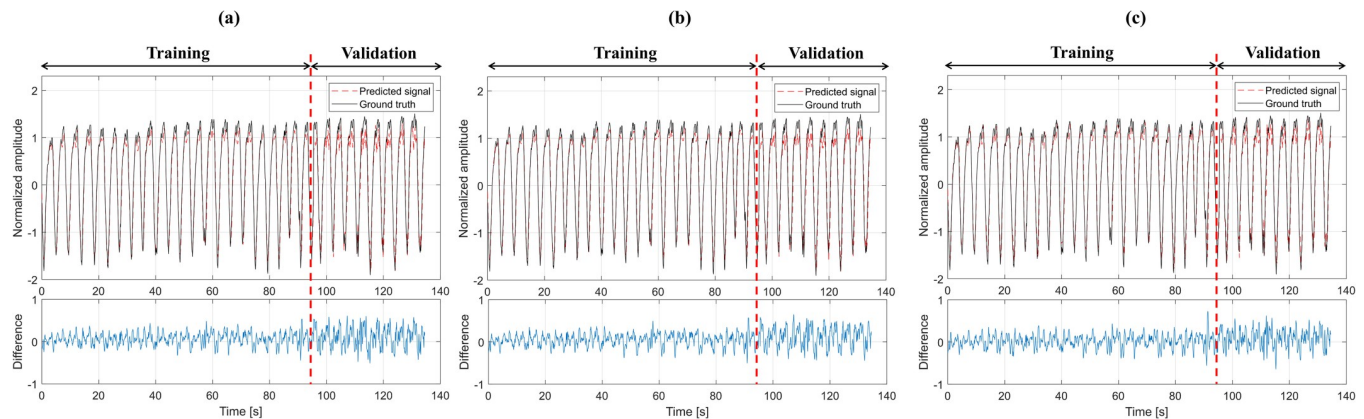


Fig 3. Plots of actual (solid line) and predicted respiratory signals (dotted line). (a) LSTM, (b) Bi-LSTM, and (c) Transformer. The difference is an absolute value (true signal amplitude–predicted signal amplitude).

<https://doi.org/10.1371/journal.pone.0275719.g003>

prediction models for respiration-gated radiation therapy to translate the statistical error into clinical measures. For this purpose, the respiration signal was assumed as identical to the tumor motion with an amplitude of 10 mm, and the model prediction error of 30% and 70% of the respiration phases was computed in mm, thereby allowing for evaluation of the clinical applicability of the models.

3. Results

3.A. Evaluation of the prediction accuracy

The prediction accuracy of the respiratory signal was calculated for LSTM, Bi-LSTM, and the Transformer. The results were obtained under the standard conditions of $T_s = 1000$ ms and $P_t = 500$ ms. Examples of the prediction results of LSTM, Bi-LSTM, and the Transformer are shown in Fig 3. The averaged RMSE and CC of the validation and test sets for the three different deep learning models are summarized in Table 1. In the test set for the LSTM, Bi-LSTM, and the Transformer, the RMSEs were 0.1907, 0.1930, and 0.1554, and the CCs were 0.9689, 0.9661, and 0.9768, respectively. According to the data summarized in Table 1, the Transformer exhibited a relatively high performance in the two types of test sets compared with the other respiratory prediction methods. Based on a one-way ANOVA, the P-values were 0.0001 and 0.0002 for the RMSE and CC for the validation set, and 0.0051 and 0.0324 for the RMSE

Table 1. Respiratory signal prediction accuracies of LSTM, Bi-LSTM, and the Transformer.

		RMSE	CC
Validation set	LSTM	0.2501	0.9667
	Bi-LSTM	0.2554	0.9640
	Transformer	0.2309	0.9715
Test set	LSTM	0.1907	0.9689
	Bi-LSTM	0.1930	0.9661
	Transformer	0.1554	0.9768

*The values in bold are the high scores in this comparison study

<https://doi.org/10.1371/journal.pone.0275719.t001>

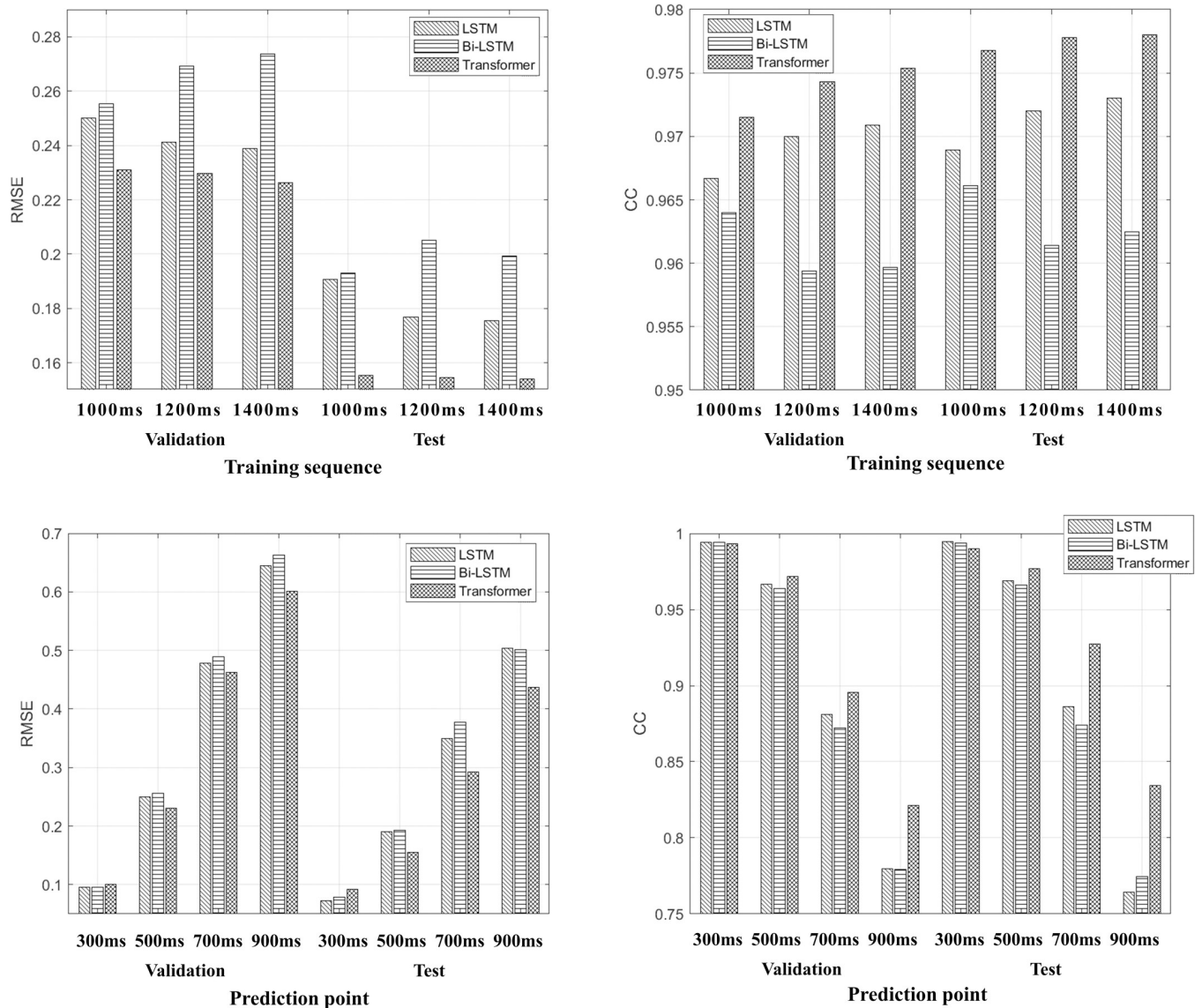


Fig 4. Comparison of the prediction accuracies of LSTM, Bi-LSTM, and the Transformer for different training sequences (T_s) and prediction points (P_t).

<https://doi.org/10.1371/journal.pone.0275719.g004>

and CC for the test set. All of the P-values were less than 0.05, thereby indicating a significant difference.

3.B. Effect of the training sequence (T_s) and prediction point (P_t) on prediction accuracy

To analyze the effect of the training data length of T_s on prediction accuracy, the prediction accuracy of respiratory signals among the three different models was assessed using three different T_s : 1000, 1200, and 1400 ms. The P_t was fixed at 500 ms. The RMSE and CC were calculated for three different T_s values. The results are summarized in Fig 4.

In the validation set with 1200 ms of T_s , the RMSEs were 0.2412, 0.2692, and 0.2398 for LSTM, Bi-LSTM, and Transformer, respectively. In the test set with 1200 ms of T_s , the RMSEs

Table 2. Accuracy evaluation of the LSTM, Bi-LSTM, and Transformer methods using the RMSE and CC.

	irregular						regular	
	Type 1 (Period)		Type 2 (Amplitude)		Type 3 (Period and amplitude)		RMSE	CC
	RMSE	CC	RMSE	CC	RMSE	CC		
LSTM	0.2601	0.9636	0.2733	0.9547	0.2882	0.9555	0.2501	0.9667
Bi-LSTM	0.2602	0.9624	0.2802	0.9517	0.2937	0.9555	0.2554	0.964
Transformer	0.2436	0.9696	0.261	0.9613	0.2773	0.9613	0.2367	0.9717

<https://doi.org/10.1371/journal.pone.0275719.t002>

were 0.1769, 0.2053, and 0.1559 for LSTM, Bi-LSTM, and the Transformer, respectively. As shown in Fig 4, the Transformer exhibited a relatively high prediction accuracy for the three T_s values.

To analyze the effect of the length of P_t on prediction accuracy, T_s was fixed at 1000 ms, and P_t was set as 300, 500, 700, and 900 ms. The RMSE and CC values were calculated and are summarized in Fig 4. As the length of P_t increased, the performance of all models reduced. At $P_t = 300$ ms, the prediction errors of the three models were equivalent. However, for P_t longer than 500 ms, the performance of the Transformer was higher than that of the other models.

3.C. Evaluation of the prediction accuracy for different respiration patterns

The RMSE and CC were calculated and summarized for each irregular respiratory pattern (Table 2). Herein, the accuracy for regular respiration is presented as a reference. For the irregular pattern of type 3, the RMSEs were 0.2882, 0.2837, and 0.2773 for LSTM, Bi-LSTM, and the Transformer, respectively. In these quantitative assessments, the Transformer exhibited the highest prediction accuracy for the three groups with irregular respiratory patterns. When compared to the accuracy for regular signals, the highest increment of RMSE and the highest decrement of CC were observed for signals with an irregular amplitude pattern. The prediction deviated in regions where respiration signals were not smooth, and in particular, respiration amplitude irregularity was observed (Fig 5).

3.D. Analysis of clinical feasibility

To analyze the clinical feasibility of the deep-learning methods for respiration-gated radiation therapy, we assumed that the maximum tumor motion caused by respiration motion was limited to 10.0 mm.

The treatment plan configured with respiratory-gated radiotherapy is developed typically based on the AIP. CT images were calculated from 4D CT images by averaging or accumulating the HU number at each voxel from the CT images of selected respiration phases, such as 40–60% or 30–70% of the phases among 10 phases of respiration. Thus, each full respiration cycle of the obtained signals was divided into 10 phases, and 30% and 70% of the phases were selected as beam on/off phases, respectively. Because the predicted and actual respiration signals can have different peak positions, the time points of the 30% and 70% of the phases of the two respiration signals were different. Thus, the amplitude difference between two signals at the phases in the predicted signals was computed as an error. For the numerical validation experiment, T_s was set to 1000 ms. In the case of P_t , three different latencies (300, 500, and 700 ms) were used.

Ten representative cases for regular and irregular period (type 1) and irregular amplitude patterns (type 2) of motion are presented in Figs 6 and 7, respectively. The average errors of the three models were less than 0.78 mm and 1.38 mm with a latency of 300 ms and 500 ms,

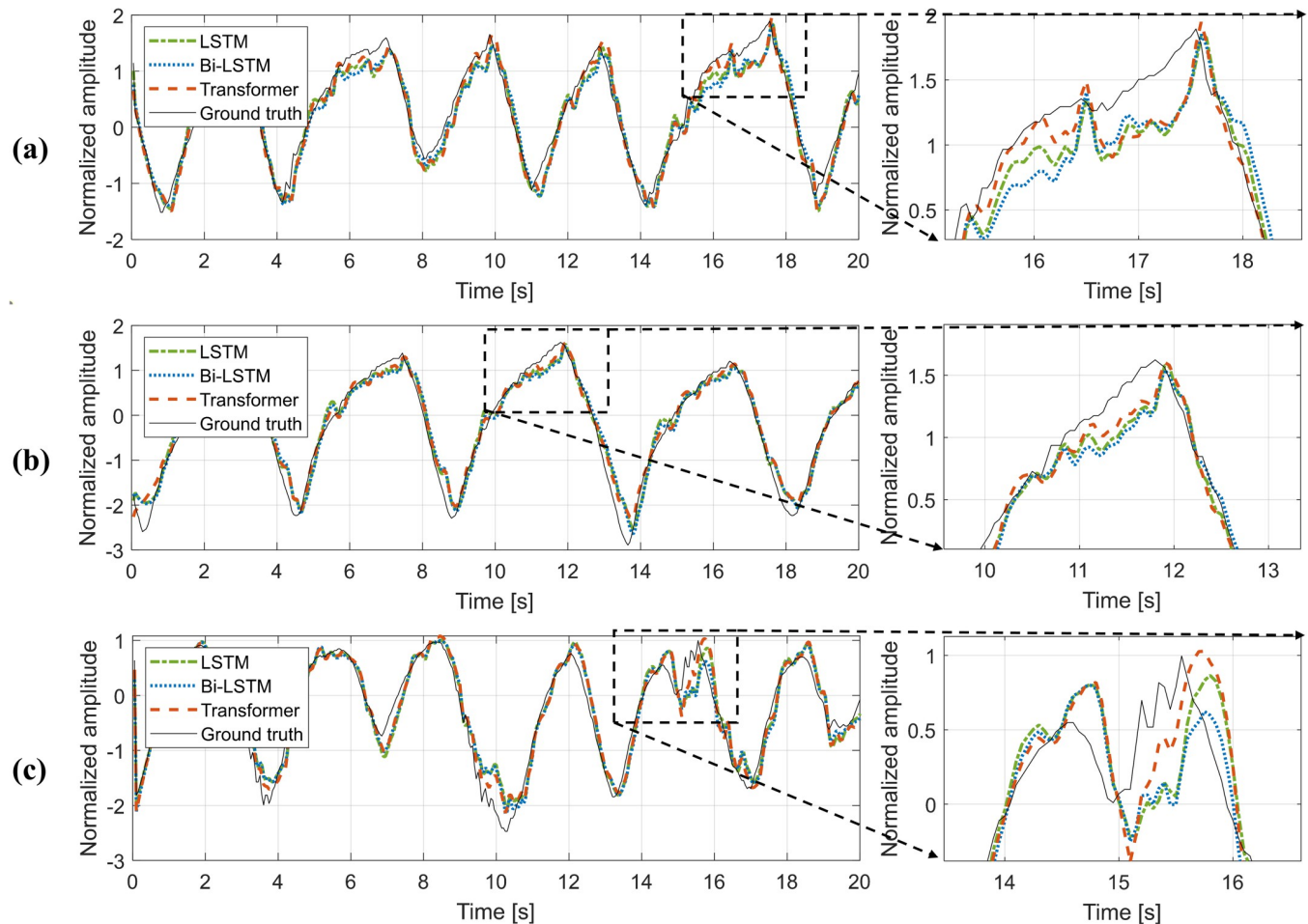


Fig 5. Actual and predicted respiratory signals for three different patterns of irregular respiration. (a) irregular pattern with periods, (b) irregular pattern with amplitude, and (c) irregular pattern with periods and amplitude.

<https://doi.org/10.1371/journal.pone.0275719.g005>

respectively. However, the average error was higher than 2.0 mm for 700 ms of latency. The maximum error of the prediction was 4.41 mm, 7.80 mm, and 9.17 mm at P_t values of 300, 500, and 700 ms, respectively.

4. Discussion

For precise 4DRT with a respiratory-gated system, a model that can compensate for the latency of the beam delivery system is essential. Thus, we compared the respiratory signal prediction performance of three deep learning-based prediction models: LSTM, Bi-LSTM, and Transformer. The Transformer achieved the best prediction accuracy under standard conditions in both the validation and test sets. In the test set, the performances of the LSTM, Bi-LSTM, and Transformer had CC values of 0.9689, 0.9661, and 0.9768, respectively.

Additionally, we analyzed the effects of T_s and P_t on the prediction accuracy because the beam on/off latency differs depending on manufacturer, motion detection device, and linac model. As the training length of T_s increased, the amount of information supplied to the model increased; thus, the prediction accuracy improved for both LSTM and the Transformer, but not for Bi-LSTM. In the case of P_t , as P_t increased, the prediction accuracy reduced for all

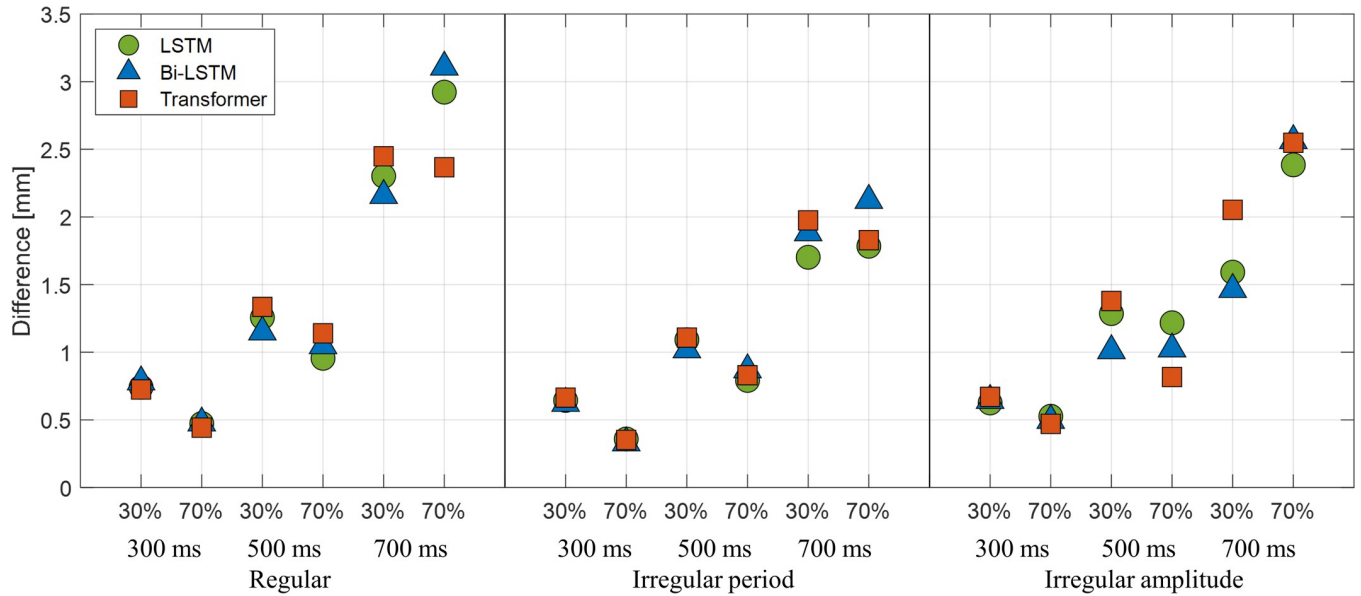


Fig 6. Average error results in millimeters. At 30% and 70% phases for regular and irregular patterns with three different prediction points: 300, 500, and 700 ms.

<https://doi.org/10.1371/journal.pone.0275719.g006>

three models. Therefore, the prediction accuracy was affected by the training data length (T_s) and the time distance to the prediction (P_t)

With regard to the prediction accuracy, when the beam on/off latency of a linac was less than or equal to 300 ms, LSTM, Bi-LSTM, and the Transformer exhibited similar performances, with CC values of 0.9943, 0.9942, and 0.9934, respectively. However, when the beam on/off latency exceeded 500 ms, the prediction achieved with the Transformer was superior to that of LSTM and Bi-LSTM. In particular, when the beam on/off latency was 900 ms, LSTM,

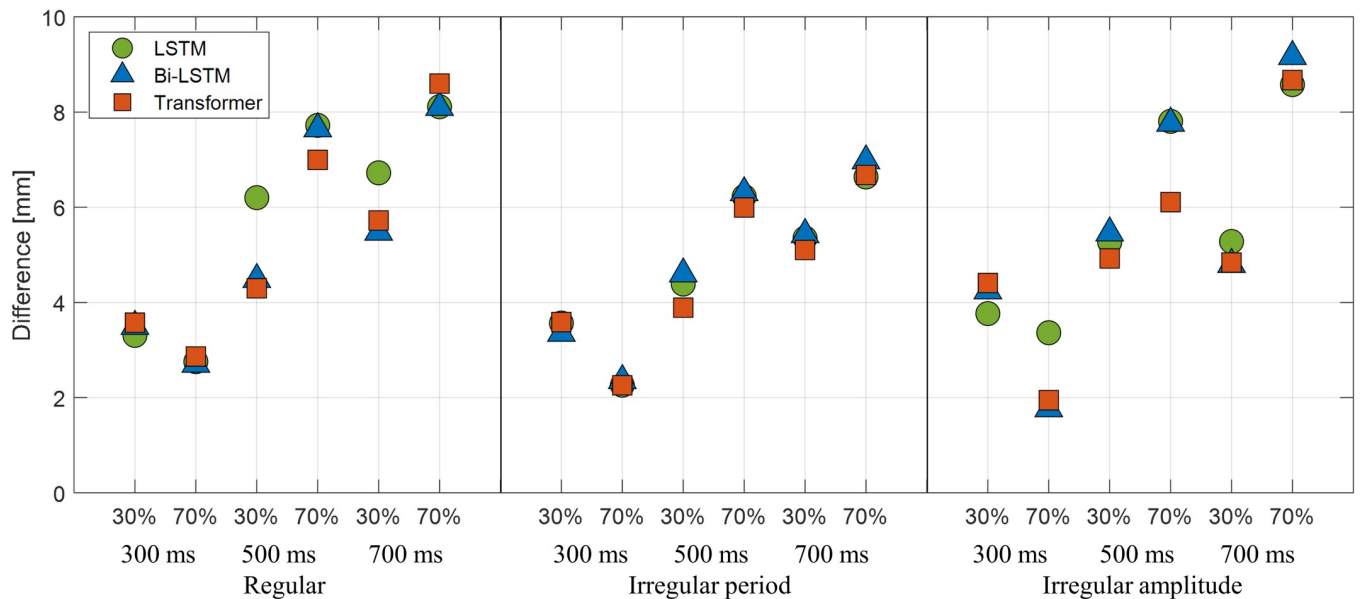


Fig 7. Maximum error results in millimeters. At 30% and 70% phases for regular and irregular patterns with three different prediction points: 300, 500, and 700 ms.

<https://doi.org/10.1371/journal.pone.0275719.g007>

Bi-LSTM, and the Transformer exhibited significant performance differences, with CC values of 0.7797, 0.7794, and 0.8212, respectively. In the analysis of T_s and P_p , Transformer outperformed LSTM and Bi-LSTM.

Based on clinical observations, numerous patients do not have perfectly regular respiration patterns. Respiration patterns are affected by daily conditions, stress level, and the severity of the disease. Thus, the prediction accuracy of the models is important in the case of irregular respiration patterns. A comparison of the irregular respiration signals revealed that irregularities in respiration amplitude have the greatest impact on prediction accuracy. Therefore, when training and preparing a patient for respiratory-gated radiation therapy, efforts should be made to minimize variation in the patients' respiration amplitude.

Among the RNN-based deep learning models, Bi-LSTM perform better than LSTM in stock market prediction studies [36]; however, LSTMs are more suitable for reservoir inflow prediction studies [37]. In the analysis performed in this study, the predictive accuracy of LSTM was better than that of Bi-LSTM, although the difference in the evaluation metrics was small. There is a difference in performance between LSTM and Bi-LSTM depending on the data used; therefore, verification is required when predicting respiration using LSTM and Bi-LSTM.

Because the accuracy evaluation in this study deals with the average errors in predicting respiratory signals, it does not clearly present the associated errors in clinical practice. To evaluate the clinical impact of a predictive model for gated radiation therapy, three representative patterns of respiration signals were used to mimic tumor movement, the patient's respiration was managed to limit tumor motion to 10 mm, and the prediction error was calculated in millimeters. The results in Fig 6 reveal that when the delay time was 300 ms, all three models exhibited an average error of less than 0.78 mm, which is considered acceptable for patient treatment. The Transformer exhibited the lowest error in three out of six in the average error evaluations, but the maximum observed error was 4.41 mm during motion with an irregular amplitude. When the delay time was 500 ms, the average error was less than 1.38 mm, the maximum value of the error increased in all models, and the maximum error of 7.80 mm was exhibited by LSTM. The Transformer exhibited the lowest maximum error among the three respiratory signal patterns. When the delay time was 700 ms, the maximum error was 9.17 mm, and the average error was approximately 3.11 mm. Therefore, with the beam delivery system with delay times of 300 and 500 ms, the respiratory prediction model can achieve an acceptable performance (< 1.38 mm on average for 10 mm of tumor motion), and with the beam delivery system having a beam latency of 700 ms, the respiratory prediction model potentially generates an error larger than 3.11 mm on average for 10 mm of tumor motion.

Although the movement of the external respiration signal was assumed to be quantitatively identical to the movement of the tumor, such a scenario rarely occurs in real clinical situations. Nevertheless, the hypothesis enables us to estimate and determine the error range when applying deep learning-based prediction models to compensate for the beam delay time of a radiation therapy device. As the performances of the models were compared based on the relative error using a normalized respiration signal, quantitative evaluation in clinical practice was necessary. In a future study, we plan to use one-dimensional respiratory signals to predict three-dimensional tumor motion over time through an analysis of respiratory signals and tumor motion.

5. Conclusion

We successfully evaluated the clinical feasibility of LSTM, Bi-LSTM, and Transformer models for respiratory signal prediction. Among the deep learning-based models, the Transformer model was superior to the LSTM and Bi-LSTM models. Prediction accuracy was found to be

affected by the training data length and the time distance to the prediction, and was considerably affected by irregular amplitude patterns. Thus, the feasibility of using a respiratory prediction deep learning model for a clinical application depends on the beam on/off latencies of the radiation therapy equipment. In addition, patient respiration management strategies are fundamentally important factors in 4DRT.

Author Contributions

Conceptualization: Wonjoong Cheon, Youngyih Han.

Data curation: Sangwoon Jeong, Sungkoo Cho.

Funding acquisition: Youngyih Han.

Investigation: Wonjoong Cheon.

Methodology: Sangwoon Jeong.

Project administration: Youngyih Han.

Software: Sangwoon Jeong, Wonjoong Cheon.

Validation: Sungkoo Cho, Youngyih Han.

Writing – original draft: Sangwoon Jeong, Wonjoong Cheon.

Writing – review & editing: Youngyih Han.

References

1. Paganetti H, Niemierko A, Ancukiewicz M, Gerweck LE, Goitein M, Loeffler JS, et al. Relative biological effectiveness (RBE) values for proton beam therapy. *International Journal of Radiation Oncology* Biology* Physics*. 2002; 53(2):407–21. [https://doi.org/10.1016/s0360-3016\(02\)02754-2](https://doi.org/10.1016/s0360-3016(02)02754-2) PMID: 12023146
2. Giovannini G, Böhlen T, Cabal G, Bauer J, Tessonnier T, Frey K, et al. Variable RBE in proton therapy: comparison of different model predictions and their influence on clinical-like scenarios. *Radiation Oncology*. 2016; 11(1):1–16. <https://doi.org/10.1186/s13014-016-0642-6> PMID: 27185038
3. Engelsman M, Damen EM, De Jaeger K, van Ingen KM, Mijnheer BJ. The effect of breathing and set-up errors on the cumulative dose to a lung tumor. *Radiotherapy and Oncology*. 2001; 60(1):95–105. [https://doi.org/10.1016/s0167-8140\(01\)00349-8](https://doi.org/10.1016/s0167-8140(01)00349-8) PMID: 11410310
4. Hector C, Webb S, Evans P. The dosimetric consequences of inter-fractional patient movement on conventional and intensity-modulated breast radiotherapy treatments. *Radiotherapy and Oncology*. 2000; 54(1):57–64. [https://doi.org/10.1016/s0167-8140\(99\)00167-x](https://doi.org/10.1016/s0167-8140(99)00167-x) PMID: 10719700
5. Dowdell S, Grassberger C, Sharp G, Paganetti H. Interplay effects in proton scanning for lung: a 4D Monte Carlo study assessing the impact of tumor and beam delivery parameters. *Physics in Medicine & Biology*. 2013; 58(12):4137. <https://doi.org/10.1088/0031-9155/58/12/4137> PMID: 23689035
6. Bussels B, Goethals L, Feron M, Bielen D, Dymarkowski S, Suetens P, et al. Respiration-induced movement of the upper abdominal organs: a pitfall for the three-dimensional conformal radiation treatment of pancreatic cancer. *Radiotherapy and Oncology*. 2003; 68(1):69–74. [https://doi.org/10.1016/s0167-8140\(03\)00133-6](https://doi.org/10.1016/s0167-8140(03)00133-6) PMID: 12885454
7. Udrescu C, Jalade P, de Bari B, Michel-Amadry G, Chapet O. Evaluation of the respiratory prostate motion with four-dimensional computed tomography scan acquisitions using three implanted markers. *Radiotherapy and Oncology*. 2012; 103(2):266–9. <https://doi.org/10.1016/j.radonc.2012.03.016> PMID: 22521750
8. Latty D, Stuart KE, Wang W, Ahern V. Review of deep inspiration breath-hold techniques for the treatment of breast cancer. *Journal of Medical Radiation Sciences*. 2015; 62(1):74–81. <https://doi.org/10.1002/jmrs.96> PMID: 26229670
9. Brandner ED, Chetty IJ, Giaddui TG, Xiao Y, Huq MS. Motion management strategies and technical issues associated with stereotactic body radiotherapy of thoracic and upper abdominal tumors: a review from NRG oncology. *Medical physics*. 2017; 44(6):2595–612. <https://doi.org/10.1002/mp.12227> PMID: 28317123

10. Bouilhol G, Ayadi M, Rit S, Thengumpallil S, Schaerer J, Vandemeulebroucke J, et al. Is abdominal compression useful in lung stereotactic body radiation therapy? A 4DCT and dosimetric lobe-dependent study. *Physica Medica*. 2013; 29(4):333–40. <https://doi.org/10.1016/j.ejmp.2012.04.006> PMID: 22617761
11. Hanley J, Debois MM, Mah D, Mageras GS, Raben A, Rosenzweig K, et al. Deep inspiration breath-hold technique for lung tumors: the potential value of target immobilization and reduced lung density in dose escalation. *International Journal of Radiation Oncology* Biology* Physics*. 1999; 45(3):603–11. [https://doi.org/10.1016/s0360-3016\(99\)00154-6](https://doi.org/10.1016/s0360-3016(99)00154-6) PMID: 10524412
12. Mampuya WA, Matsuo Y, Ueki N, Nakamura M, Mukumoto N, Nakamura A, et al. The impact of abdominal compression on outcome in patients treated with stereotactic body radiotherapy for primary lung cancer. *Journal of radiation research*. 2014; 55(5):934–9. <https://doi.org/10.1093/jrr/rru028> PMID: 24801474
13. Bruzzaniti V, Abate A, Pinnarò P, D'Andrea M, Infusino E, Landoni V, et al. Dosimetric and clinical advantages of deep inspiration breath-hold (DIBH) during radiotherapy of breast cancer. *Journal of Experimental & Clinical Cancer Research*. 2013; 32(1):1–7. <https://doi.org/10.1186/1756-9966-32-88> PMID: 24423396
14. Gu C, Li R, Jiang SB, Li C, editors. A multi-radar wireless system for respiratory gating and accurate tumor tracking in lung cancer radiotherapy. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2011: IEEE.
15. Giraud P, Houle A. Respiratory gating for radiotherapy: main technical aspects and clinical benefits. *International Scholarly Research Notices*. 2013;2013.
16. Bertholet J, Knopf A, Eiben B, McClelland J, Grimwood A, Harris E, et al. Real-time intrafraction motion monitoring in external beam radiotherapy. *Physics in Medicine & Biology*. 2019; 64(15):15TR01. <https://doi.org/10.1088/1361-6560/ab2ba8> PMID: 31226704
17. Chen L, Bai S, Li G, Li Z, Xiao Q, Bai L, et al. Accuracy of real-time respiratory motion tracking and time delay of gating radiotherapy based on optical surface imaging technique. *Radiation Oncology*. 2020; 15(1):1–9. <https://doi.org/10.1186/s13014-020-01611-6> PMID: 32650819
18. John H, Saito M, Onishi H. Prediction-based compensation for gate on/off latency during respiratory-gated radiotherapy. *Computational and mathematical methods in medicine*. 2018;2018. <https://doi.org/10.1155/2018/5919467> PMID: 30622625
19. Cho B, Poulsen PR, Sawant A, Ruan D, Keall PJ. Real-time target position estimation using stereoscopic kilovoltage/megavoltage imaging and external respiratory monitoring for dynamic multileaf collimator tracking. *International Journal of Radiation Oncology* Biology* Physics*. 2011; 79(1):269–78.
20. Makridakis S, Hibon M. ARMA models and the Box–Jenkins methodology. *Journal of forecasting*. 1997; 16(3):147–63.
21. McCall K, Jeraj R. Dual-component model of respiratory motion based on the periodic autoregressive moving average (periodic ARMA) method. *Physics in Medicine & Biology*. 2007; 52(12):3455. <https://doi.org/10.1088/0031-9155/52/12/009> PMID: 17664554
22. Wu H, Sharp GC, Salzberg B, Kaeli D, Shirato H, Jiang SB. A finite state model for respiratory motion analysis in image guided radiation therapy. *Physics in Medicine & Biology*. 2004; 49(23):5357. <https://doi.org/10.1088/0031-9155/49/23/012> PMID: 15656283
23. Vedam S, Keall P, Docef A, Todor D, Kini V, Mohan R. Predicting respiratory motion for four-dimensional radiotherapy. *Medical physics*. 2004; 31(8):2274–83. <https://doi.org/10.1118/1.1771931> PMID: 15377094
24. Wang F, Balakrishnan V. Robust steady-state filtering for systems with deterministic and stochastic uncertainties. *IEEE Transactions on Signal Processing*. 2003; 51(10):2550–8.
25. Putra D, Haas O, Mills JA, Burnham KJ. Prediction of tumour motion using interacting multiple model filter. 2006.
26. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*. 2018; 13(3):e0194889. <https://doi.org/10.1371/journal.pone.0194889> PMID: 29584784
27. Rangapuram SS, Seeger MW, Gasthaus J, Stella L, Wang Y, Januschowski T. Deep state space models for time series forecasting. *Advances in neural information processing systems*. 2018; 31:7785–94.
28. Kai J, Fujii F, Shiinoki T, editors. Prediction of lung tumor motion based on recurrent neural network. 2018 IEEE International Conference on Mechatronics and Automation (ICMA); 2018: IEEE.
29. Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 1998; 6(02):107–16.

30. Squartini S, Hussain A, Piazza F, editors. Preprocessing based solution for the vanishing gradient problem in recurrent neural networks. *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003 ISCAS'03*; 2003: IEEE.
31. Yadav A, Jha C, Sharan A. Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science*. 2020; 167:2091–100.
32. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation*. 1997; 9(8):1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735> PMID: 9377276
33. 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. *Physics in Medicine & Biology*. 2019; 64(8):085010. <https://doi.org/10.1088/1361-6560/ab13fa> PMID: 30917344
34. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks*. 2005; 18(5–6):602–10. <https://doi.org/10.1016/j.neunet.2005.06.042> PMID: 16112549
35. Wang R, Liang X, Zhu X, Xie Y. A feasibility of respiration prediction based on deep Bi-LSTM for real-time tumor tracking. *IEEE Access*. 2018; 6:51262–8.
36. Siami-Namini S, Tavakoli N, Namin AS, editors. The performance of LSTM and BiLSTM in forecasting time series. *2019 IEEE International Conference on Big Data (Big Data)*; 2019: IEEE.
37. Apaydin H, Feizi H, Sattari MT, Colak MS, Shamsirband S, Chau K-W. Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. *Water*. 2020; 12(5):1500.
38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al., editors. Attention is all you need. *Advances in neural information processing systems*; 2017.
39. Karita S, Chen N, Hayashi T, Hori T, Inaguma H, Jiang Z, et al., editors. A comparative study on transformer vs rnn in speech applications. *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*; 2019: IEEE.
40. Wu N, Green B, Ben X, O'Banion S. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:200108317*. 2020.
41. Mizuno H, Saito O, Tajiri M, Kimura T, Kuroiwa D, Shirai T, et al. Commissioning of a respiratory gating system involving a pressure sensor in carbon-ion scanning radiotherapy. *Journal of applied clinical medical physics*. 2019; 20(1):37–42. <https://doi.org/10.1002/acm2.12463> PMID: 30387271
42. Keall PJ, Mageras GS, Balter JM, Emery RS, Forster KM, Jiang SB, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76 a. *Medical physics*. 2006; 33(10):3874–900.
43. Mohabeer H, Soyjaudah KS, Pavaday N, editors. Enhancing The Performance Of Neural Network Classifiers Using Selected Biometric Features. *Proc 5th International Conference on Sensor Technologies and Applications, French Riviera, Nice/Saint Laurent du Var, France*; 2011.
44. Press WH, Teukolsky SA. Savitzky-Golay smoothing filters. *Computers in Physics*. 1990; 4(6):669–72.
45. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980*. 2014.