Feasibility Study of Convolutional Long Short-Term Memory Network for Pulmonary Movement Prediction in CT Images

Zahra Ghasemi (PhD Candidate)[1]*, Payam Samadi Miandoab (PhD Candidate)[2]

[1]Department of Energy Engineering and Physics, Amirkabir University of Technology, Tehran, Iran
[2]Department of Medical Radiation Engineering, Amirkabir University of Technology, Tehran, Iran

ABSTRACT

Background: During X-ray imaging, pulmonary movements can cause many image artifacts. To tackle this issue, several studies, including mathematical algorithms and 2D-3D image registration methods, have been presented. Recently, the application of deep artificial neural networks has been considered for image generation and prediction.

Objective: In this study, a convolutional long short-term memory (ConvLSTM) neural network is used to predict spatiotemporal 4DCT images.

Material and Methods: In this analytical analysis study, two ConvLSTM structures, consisting of stacked ConvLSTM models along with the hyperparameter optimizer algorithm and a new design of the ConvLSTM model are proposed. The hyperparameter optimizer algorithm in the conventional ConvLSTM includes the number of layers, number of filters, kernel size, epoch number, optimizer, and learning rate. The two ConvLSTM structures were also evaluated through six experiments based on Root Mean Square Error (RMSE) and structural similarity index (SSIM).

Results: Comparing the two networks demonstrates that the new design of the ConvLSTM network is faster, more accurate, and more reliable in comparison to the tuned-stacked ConvLSTM model. For all patients, the estimated RMSE and SSIM were 3.17 and 0.988, respectively, and a significant improvement can be observed in comparison to the previous studies.

Conclusion: Overall, the results of the new design of the ConvLSTM network show excellent performances in terms of RMSE and SSIM. Also, the generated CT images with the new design of the ConvLSTM model show a good consistency with the corresponding references regarding registration accuracy and robustness.

Citation: Ghasemi Z, Samadi Miandoab P. Feasibility Study of Convolutional Long Short Term Memory Network for Pulmonary Movement Prediction in CT Images. *J Biomed Phys Eng*. 2024;14(1):55-66. doi: 10.31661/jbpe.v0i0.2105-1339.

Keyword

Radiotherapy; Deep Learning; Convolutional Long Short-Term Memory; Pulmonary Movement; Prediction Model; 4DCT; Diagnostic Imaging; Hyperparameter Optimization

Introduction

In medical image analysis, image registration plays an important role in medical physics and radiation therapy applications. This technique has been extensively studied for external beam radiotherapy to deliver a therapeutic dose more efficiently to the target volume that moved with patient respiration [1]. In this regard, deformable image registration (DIR) applications are widely used to account for anatomical deformations in the field of intensity-modulated radiation therapy or volumetric modulated arc therapy [1-4]. In these methods, online

*Corresponding author:
Zahra Ghasemi
Department of Energy Engineering and Physics, Amirkabir University of Technology, Tehran, Iran
E-mail: ghasemi.za90@yahoo.com

Received: 19 May 2021
Accepted: 10 June 2021

imaging techniques, such as Computed tomography (CT) on the rail, cone beam computerized tomography (CBCT), integrated magnetic resonance imaging (MRI), and four-dimensional computed tomography (4DCT), have been used to account for anatomical deformations. Therefore, the number of data acquired using the X-ray imaging system has significantly increased in advanced radiotherapy in contrast to conformal radiotherapy. However, the patient's breathing as a significant impact might result in degrading the spatiotemporal resolution [5-7]. In this relation, the image registration algorithms, including intensity-based or feature-based methods can be used to address this issue [8-10]. In recent years, several approaches have been presented for the image processing of CT images throughout the breathing cycle. However, there are still some limitations, such as time-consuming, degraded spatiotemporal resolution, and error-prone, which required more consideration. To tackle this issue, deep convolutional neural networks have shown great success in image reconstruction and prediction of pulmonary movements in CT images [11-14].

Kai et al. performed a study to use the advantage of recurrent neural networks (RNN) for lung motion estimation. The RMSE value reported by the proposed approach was less than 1 mm in 3D space, which showed a better performance and result compared to the classical neural networks [11]. Nabavi et al. used the PredNet network, which is a type of CNN-LSTM model, to predict and generate pulmonary movements in CT images. The evaluation results by the proposed model in 4DCT images of six patients show a 0.943 structural similarity index [12]. A recent report from Yabo Fu et al. conducted a study to use a deep learning method, namely LungRegNet, to develop an accurate and unsupervised DIR method for predicting large lung motion 4DCT images. The LungRegNet model also comprises two subnetworks: CoarseNet and FineNet. Whereas the CoarseNet is used to predict large lung motion, the FineNet predicts the local lung motion. The mean value and standard deviation of target registration error were 1.00±0.53 mm for ten 4DCT datasets and 1.59±1.58 mm for ten DIRLAB datasets [14]. It should be noted that the state of the art deep learning techniques was recently used in different studies, such as predicting future frames in stock market prediction [15], traffic accident prediction [16], text recognition [17, 18], precipitation prediction [19], weather forecasting [20], ocean temperature [21], medical imaging [13, 14, 22], direction of slip detection [23], and travel demand prediction [24].

Essentially, the review of prior studies in medical physics and radiation therapy shows that the convolutional long short-term memory (ConvLSTM) network as a powerful model can be used for image reconstruction, next frame prediction, and prediction of pulmonary movements in CT images. However, there are some challenges faced by researchers, including inaccurate results, high target registration error, insufficient sample size, and using default network parameters [11-14]. In this study, two networks, including the stacked ConvLSTM model along with the hyperparameter optimizer algorithm and a new design of the ConvLSTM network, were initially proposed. The hyperparameters for the stacked ConvLSTM network also included the number of layers, number of filters, kernel size, epoch number, optimizer, and learning rate. The new design of the ConvLSTM network, on the other hand, is equipped with multi kernels in input images accompanied with different filters. To propose a reliable and accurate model, a comparison between two networks was considered. To evaluate the model performance, 4DCT of six lung patients, which each dataset consisting of ten 3DCT frames along with the breathing cycle, was also used. Moreover, the Root Mean Square Error (RMSE) and structural similarity index measure (SSIM) are considered to evaluate the obtained results. Overall, the generated CT images with the new design of the

ConvLSTM model show a good consistency with the corresponding references in terms of registration accuracy and robustness.

Material and Methods

Database properties and pre-processing

In this analytical analysis study, six 4DCT images with pulmonary tumors acquired from Brilliance CT Big Bore (Philips with a 16 Slice) were used [25]. Also, each dataset consists of ten 3DCT frames along with the breathing cycle. More information about the patient's number, label, image dimensions, pixel spacing, and slice thickness is shown in Table 1. In a preprocessing step, all images were resized to 256×256, and then the pixel values were normalized to grayscale pixels by dividing each pixel value by the maximum pixel value. After fitting the network and predicting the test dataset, the predicted images are transformed into 0 to 255 to calculate RMSE and SSIM.

ConvLSTM Theory and Structure

ConvLSTM, which is an extension of the LSTM neural network, consists of two structures, convolutional and LSTM networks. Convolutional Neural Networks (CNNs) are used to transfer image data to an output variable, which are sufficient for prediction problems involving image data as an input. On the other hand, the Long Short Term Memory (LSTM) network, which is a type of recurrent neural network, is used to learn time-series data from temporal patterns, such as sequential data. The structure of the conventional ConvLSTM network is shown in Figure 1. Overall, the ConvLSTM is designed for 3-D input data, which is suitable for spatial sequence data, while the LSTM is used for one-dimensional input data. Note that the feed-forward method of the LSTM is changed from Hadamard product to convolution in the ConvLSTM to capture underlying spatial features by convolution operations in multiple-dimensional data. More information about the structure of the ConvLSTM models is described in detail elsewhere [26].

In this study, two ConvLSTM networks were proposed for image generation and prediction; a stacked ConvLSTM network accompanied with the hyperparameter optimization algorithm and a new structure based on the ConvLSTM network. The stacked ConvLSTM network structure, which is shown in Figure 2a, consists of a couple of the ConvLSTM layers along with each other, usually followed by a Conv3d layer as the output. BatchNormalization and Dropout layers can be used between consecutive ConvLSTM

Table 1: Patient's number, label, image dimensions, pixel spacing, and slice thickness.

Patient Number	Label	Image Dimensions (Coronal×Sagittal×Axial)	Pixel Spacing (mm)	Slice Thickness (mm)
Patient 1	4DCT	(512, 512, 141)	[0.9765625, 0.9765625]	2.00
Patient 2	4DCT	(512, 512, 169)	[0.9765625, 0.9765625]	2.00
Patient 3	4DCT	(512, 512, 170)	[0.87890625, 0.87890625]	2.00
Patient 4	4DCT	(512, 512, 187)	[0.78125, 0.78125]	2.00
Patient 5	4DCT	(512, 512, 139)	[1.171875, 1.171875]	2.00
Patient 6	4DCT	(512, 512, 161)	[1.171875, 1.171875]	2.00

4DCT: Four-Dimensional Computed Tomography

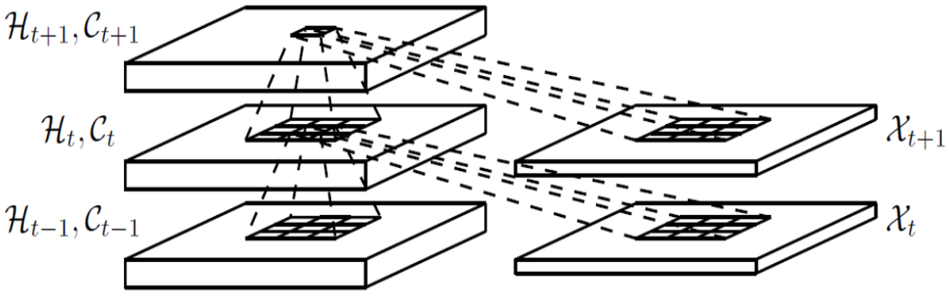

Figure 1: Inner structure of convolutional long short-term memory (ConvLSTM) model.

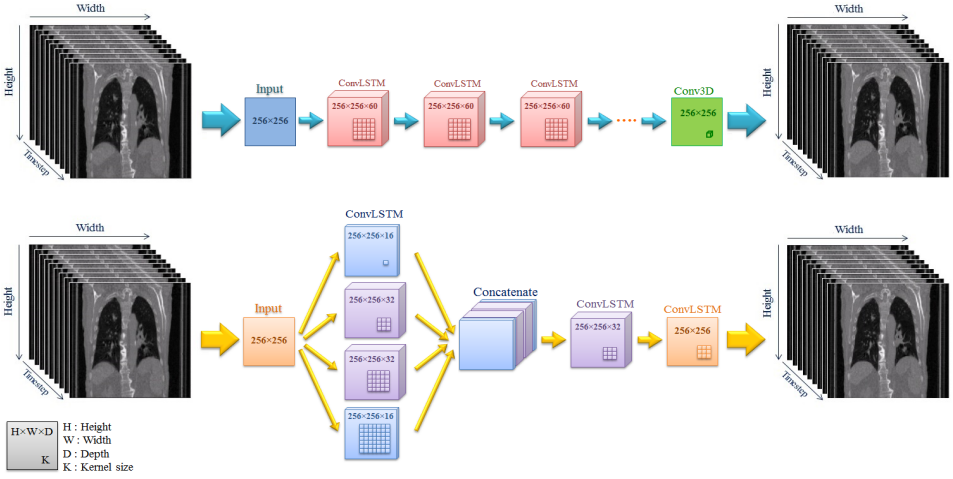

Figure 2: The simplified structure of the stacked convolutional long short-term memory (ConvLSTM) structure (upper image) and new design ConvLSTM model (lower image) for image reconstruction and prediction pulmonary movements in four-dimensional computed tomography (4DCT) images. For each ConvLSTM network, different high, width, depth, and kernel size are used.

layers to train speedup and prevent overfitting, respectively. In this study, the stacked ConvLSTM model has used the Relu activation function in the middle layers, while the Conv3D layer, which is the output layer, consisted of a single filter with kernel size (1×1×1) followed by a sigmoid activation function. In addition, MSE was used as the loss function. The optimization of the hyperparameters, in the stacked ConvLSTM network, which includes the number of layers, the number of filters, the kernel size, the epoch number, the optimizer, and the learning rate, is described in section 2.3. It should be noted that, in structures with more than one ConvLSTM layer, the number of filters and the kernel size were chosen to be the same in every layer.

The new design of the ConvLSTM base model, presented in Figure 2b, consists of four parallelized ConvLSTM layers with different filters and kernel sizes that receive a sequence of images during the respiratory cycle of the same patient. Note that in this study, a different kernel was used for each ConvLSTM layer. Overall, the kernel size revolves around the input data to extract features. In this series, different ranges of kernel size can be used; however, a common choice is to keep the kernel size at (3×3) or (5×5). In this study, four different kernel sizes, including (1×1), (3×3), (5×5), and (7×7), as shown in Figure 2b, were used. Then each of the kernels was connected

to different filters. Since the kernel size (3×3) or (5×5) are common and can extract more features, the number of filters along with these kernels is considered to be twice that of the other kernels. Finally, four ConvLSTM layers were concatenated and connected to the next ConvLSTM layer, which consisted of 32 filters with a kernel size (3×3). Finally, the last ConvLSTM layer, which included a single filter with a kernel size (3×3), was considered for image generation and prediction spatio-temporal datasets. Note that this configuration was obtained through trial and error based on preliminary experiments. For the optimization process, Adam's optimization algorithm with a learning rate of 0.005 was considered. Relu was selected as the activation function in the middle layers, while, in the last layer, the sigmoid activation function was used. In the presented model, ModelCheckpoint callback was also used to save the best-observed model during training, based on validation loss. The evaluations were performed for 200 epochs, selected based on the preliminary experiments on the EarlyStopping callback. In the preliminary experiments, EarlyStopping callback was specified to monitor the performance of the validation loss by setting the patience argument. Since the number of patience arguments depends on the variability of the dataset, the optimal number of patience was examined in the range of 1 to 20. The optimum number of patience was selected to be 10. Overall, the new structure led to better and more reliable results through increasing the SSIM, decreasing the number of epochs as well as the model runtime.

Hyperparameter Optimizer

Hyperparameter optimization (HPO) is a process to choose a set of optimal hyperparameters to achieve maximum accuracy, optimal training speed, and the best model configuration [27]. In machine learning, a hyperparameter is defined as a parameter, which has a direct and strong impact on network accuracy. In this study, the following parameters were examined for the stack ConvLSTM network.

1. Number of layers: the number of layers, determining the depth of the model, is defined as the number of hidden layers in the network structure.

2. Number of filters: the number of filters, which is a hyperparameter, refers to the number of neurons performing a different convolution to extract a suitable number of features from the input image. Also, there is a link between the number of features and the number of filters. The more wanted features lead to a higher number of required filters.

3. Kernel size: the kernel size refers to the filter size, revolving around the input image. In this study, the different kernel sizes (width×height) are considered as the feature extractors.

4. Number of epochs: the epoch number refers to the total processing times of the entire training dataset to update the internal parameters and minimize the network's error, properly.

5- Optimizer: optimizers are algorithms or methods responsible for minimizing the objective function by changing neural network attributes such as weights and learning rate to reduce losses.

6- Learning rate: the learning rate controls how quickly the model is adapted to the problem. It is usually in the range of 0.0 and 1.0. Setting a too low learning rate will result in very slow training, while setting a high a learning rate may result in undesirable divergent behavior.

Results

All experiments, including training, evaluation, and testing, were implemented in the Python (version 3.7) environment by using the high-level neural networks functional API Keras (version 2.4) and backend engine TensorFlow (version 2.4). To quantify the model performance, the root mean square error (RMSE) and structural similarity index

measure (SSIM) are considered. Whereas the RMSE is used to assess the metric performance of the model, the SSIM is considered to represent the rate of similarity or difference between the reference and predicted images. Moreover, the SSIM is a well-suited approach to quantify the differences perceptually in the human body. A detailed explanation of the SSIM analysis can be found in [28]. The mathematical expressions for these values are given in Equations 1 and 2, respectively.

$$RMSE\left(Y_i,\hat{Y}_i\right)=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i-\hat{Y}_i\right)^2} \quad (1)$$

$$SSIM\left(Y,\hat{Y}\right)=\frac{\left(2\mu_Y\mu_{\hat{Y}}+C_1\right)\left(2\sigma_{Y\hat{Y}}+C_2\right)}{\left(\mu_Y^2+\mu_{\hat{Y}}^2+C_1\right)\left(\sigma_Y^2+\sigma_{\hat{Y}}^2+C_2\right)} \quad (2)$$

Where $Y,\hat{Y},\mu_Y,\mu_{\hat{Y}},\sigma_Y^2,\sigma_{\hat{Y}}^2,\sigma_{Y\hat{Y}},C$ represent the original image, the predicted image, the average of the original image, the average of the predicted image, the variance of the original image, the variance of the predicted image, the covariance of the original and predicted images, and two variables to stabilize the division with weak denominator, respectively.

In this study, the HPO method was used to find the optimal configurations for the stack ConvLSTM model. In the stack ConvLSTM model, the hyperparameter optimization algorithm includes the number of layers, the number of filters, the kernel size, the number of epochs, the optimizer, and the learning rate. The interaction of these parameters is also depicted in the box and whisker plot in Figure 3 to provide better concepts about the RMSE and SSIM values. It should be noted that the box and whisker plot is an effective graphical method for displaying variations in a set of data. It provides additional detail while allowing multiple sets of data to be displayed on the same graph.

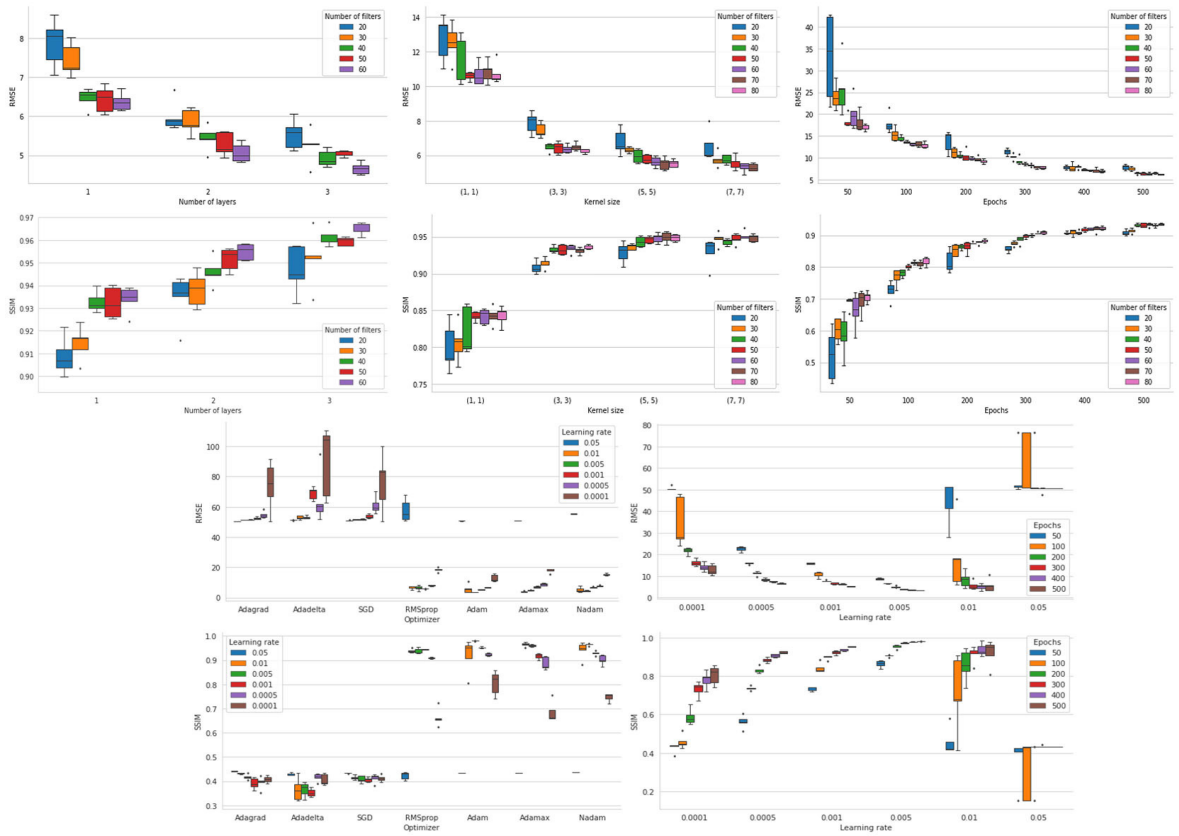

Figure 3: Impacts of different hyperparameters on the Root Mean Square Error (RMSE) and structural similarity index measure (SSIM) values.

A comparison between two proposed ConvLSTM models was considered in terms of RMSE, SSIM, and epoch numbers for all patients. In this relation, Table 2 represents that the new structure of the ConvLSTM base model is faster, more accurate, and more reliable compared to the stacked ConvLSTM model. Based on Table 2, the new ConvLSTM method also shows good consistency with the corresponding references in terms of registration accuracy, robustness, and model runtime. In other words, the new ConvLSTM model with fewer trainable parameters and epochs (lower runtime) can provide better results compared to the stacked ConvLSTM model. Note that we repeated each trial ten times and reported the mean value of the trials for all patients and variables.

Table 2: A comparison between two proposed convolutional long short-term memory (ConvLSTM) models; Stacked ConvLSTM and proposed ConvLSTM structure.

	RMSE	SSIM	Epoch
Stacked ConvLSTM	4.7678	0.9594	500
Proposed ConvLSTM structure	3.17	0.988	200

RMSE: Root Mean Square Error, SSIM: Structural Similarity Index Measure

The obtained results from the new design of the ConvLSTM network in each of the six experiments throughout the respiratory cycle in terms of the RMSE and SSIM are presented in Table 3. Also, less RMSE represents an excellent performance in terms of prediction accuracy and robustness. Note that the SSIM value is ranged from −1 to 1, which the superior

Table 3: Quantitative evaluation of the proposed new design of the convolutional long short-term memory (ConvLSTM) network in terms of the Root Mean Square Error (RMSE) and Structural Similarity Index Measure (SSIM) through different phases.

Patient Number	View	Phase20		Phase 40		Phase 60		Phase 80		All Phases	
		RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM
Patient 1	Sagittal	1.8354	0.9954	1.5193	0.9956	2.0348	0.9905	2.1207	0.9893	2.3509	0.9919
	Coronal	1.5279	0.9964	1.1561	0.9978	1.1561	0.9982	1.1561	0.9981	2.1297	0.9962
	Axial	2.391	0.955	1.4228	0.9973	1.5894	0.9966	1.5289	0.9968	2.3232	0.994
Patient 2	Sagittal	2.6531	0.9858	2.2707	0.9861	2.0377	0.9888	2.2086	0.9882	3.3176	0.984
	Coronal	2.6145	0.9879	1.9974	0.9916	1.7559	0.9932	1.9379	0.9923	4.417	0.9842
	Axial	2.9563	0.9684	2.5875	0.9821	3.0332	0.974	3.2543	0.9729	4.4297	0.9743
Patient 3	Sagittal	2.5586	0.9814	1.758	0.9935	1.5641	0.9938	1.5822	0.9934	2.4434	0.9918
	Coronal	2.242	0.9922	1.6559	0.9948	1.3477	0.9965	1.4973	0.9959	2.7708	0.9924
	Axial	4.2566	0.9323	4.3586	0.9787	3.5988	0.9816	4.9469	0.9709	5.0797	0.9745
Patient 4	Sagittal	3.5906	0.972	2.9012	0.9846	2.741	0.9858	2.9059	0.9852	3.763	0.9823
	Coronal	2.3395	0.9891	1.8863	0.9912	1.6815	0.9933	1.7385	0.9929	2.9959	0.9881
	Axial	3.2867	0.962	3.1984	0.9731	2.9051	0.9789	2.7945	0.982	3.7308	0.9743
Patient 5	Sagittal	2.4922	0.9668	1.7447	0.9925	1.7477	0.9913	1.9797	0.9901	2.5762	0.9889
	Coronal	1.885	0.9942	1.4018	0.9966	1.1486	0.9977	1.2346	0.9975	2.6695	0.9943
	Axial	2.9789	0.9744	2.0227	0.9925	1.5752	0.9938	1.6192	0.9936	2.7721	0.9918
Patient 6	Sagittal	2.1951	0.9927	1.7065	0.9932	1.4621	0.9951	1.5299	0.9946	2.6019	0.9925
	Coronal	1.7111	0.9938	1.209	0.9971	1.0728	0.9974	1.1124	0.9974	2.7464	0.9936
	Axial	2.8102	0.9605	1.8461	0.9923	1.6344	0.994	1.6684	0.994	3.2081	0.9896

RMSE: Root Mean Square Error, SSIM: Structural Similarity Index Measure

values indicate significant similarity between the actual output image and the predicted image. In this study, the k-fold cross-validation method was used to evaluate the prediction models. Overall, in this method, the dataset was divided into M subsets, and M−1 subsets were used for training and validation, while one of the subsets was used to perform the test of the model. An elaborate description is in [29].

The results of predicted pulmonary motion during the respiratory cycle by the new design of the ConvLSTM network are presented in Figure 4 including the predicted image, the difference between the input image and actual output image, and the difference between actual output image and predicted image, in all three directions. The results of the proposed model show excellent performances in terms of registration accuracy, robustness, and similarity. The predicted image through the new design of the ConvLSTM model provides the ability to generate the desired frames when the 4DCT image suffers from image artifacts.

Discussion

Most radiation therapy techniques are planned to use a 3DCT scan; however, organ motion due to patient respiration would result in under/overdosage in the junction region. Therefore, a 4DCT system is required in radiation therapy planning to achieve a three-dimensional (3D) uniform dose delivery [30]. Also, the 4DCT aims to provide an ability to visualize the temporal dynamics with a sufficient spatiotemporal resolution. In this regard, the review of earlier literature shows that some 4DCT images suffer from image artifacts, such as streaks, rings, metal artifacts, blurring, incomplete structure, duplicated structure, and overlapping structure [5-7]. To tackle this issue, several methods have been proposed; in this study, the state-of-the-art ConvLSTM network was considered to develop a prediction model for image reconstruction or generation of the next slice (time frame). In this series, two ConvLSTM structures consisting of stacked ConvLSTM models accompanied with the hyperparameter optimization algorithm and a new design of the ConvLSTM model were proposed. To find the best configurations for the stacked ConvLSTM network, the HPO algorithm was proposed to improve training speed and prediction accuracy.

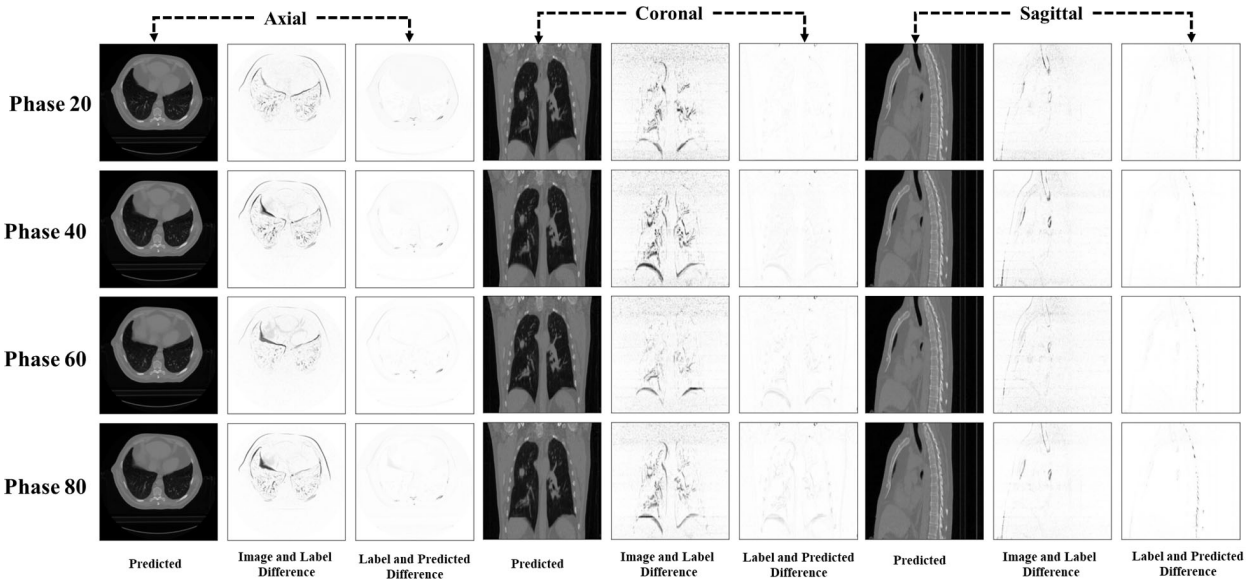

Figure 4: The predicted pulmonary motion of patient two, including the predicted image, the difference between the input image and actual output image, and difference between actual output image and predicted image, during the respiratory cycle in all three directions.

Moreover, the hyperparameters include the number of layers, number of filters, kernel size, epoch number, optimizer, and learning rate. The obtained results from the proposed HPO method are reported in Figure 3 and discussed in detail in the following.

1. Number of layers: in this study, the impact of different layers, including one, two, and three layers, on the performance and accuracy of the model was investigated. The obtained results indicate that increasing the number of layers improves the performance of the model, while it might result in increasing model runtime as well as the probability of overfitting. Therefore, a compromise is required between the number of layers and model runtime. Based on Figure 3, there is also a relation between the number of filters and the number of layers. The deeper nets, the higher the number of filters. Note that, in each layer, increasing the filter number beyond a particular number has a minor impact on accuracy and performance.

2. Number of filters: filter is a suitable approach to extract different features from the image. In this study, different ranges of the filter size, such as 20, 30, 40, 50, 60, 70, and 80, were considered to extract a suitable number of features from the input image. From Figure 3, the kernel size is an important hyperparameter to the filter size.

3. Kernel size: in this study, different kernel sizes, including (1×1), (3×3), (5×5), and (7×7), were proposed. The obtained results in Figure 3 represent that increasing the kernel size through a larger filter size would improve the accuracy and performance of the model. Also, the kernel (1×1), known as feature pooling, is used for dimensionality reduction and reduces the number of features. Note that the deeper models required the optimal kernel size, sufficient number of filters, and number of layers. It should be noted that, for each layer and kernel size, increasing the number of filters beyond a certain number has no or minor impact on the accuracy of the network, while model runtime increases.

4. Number of epochs: different epoch numbers, including 50, 100, 200, 300, 400, and 500, were investigated in this study. The obtained results indicate that increasing the number of epochs results in increasing the similarity and reducing the RMSE value. However, increasing the epoch number results in increasing the model runtime. As seen in Figure 3, there is no role in determining the numbers of epochs and filters used to construct stacked ConvLSTM models. It should be noted that training the network with low or large numbers of epochs would result in overfitting or underfitting.

5. Optimizer: in this study, we evaluated different optimizers, including stochastic gradient descent (SGD), RMSprop, Adam, Adamax, Nadam, Adagrad, and Adadelta for different learning rates. The performance of different optimizers is represented in the box and whisker plot in Figure 3. Based on Figure 3, RMSprop, Adam, Adamax, and Nadam optimizers show good performances with learning rates between 0.01 and 0.0005, while SGD, Adagrad, and Adadelta had a much higher variance and RMSE.

6. Learning rate: as mentioned in the optimizer part, different learning rates, including 0.05, 0.01, 0.005, 0.001, 0.0005, and 0.0001, were studied for different optimizers. Also, several epochs for different learning rates in the Adam optimizer were investigated, and results were presented in Figure 3. Figure 3 also reveals that using the optimal learning rate leads to faster convergence, smaller variance, and better performance.

A comparison between the two proposed networks was also considered in terms of RMSE, SSIM, and epoch number. Based on Table 2, the new ConvLSTM network was more reliable and accurate regarding RMSE, SSIM, and runtime for image generation and prediction. Therefore, the state-of-the-art of this study is to propose a new ConvLSTM structure network to predict and generate future frames in CT images during the breathing cycle. In this relation, the RMSE and SSIM values for

different patients are reported in Table 3. The obtained results demonstrate a promising accuracy and excessive similarity through the previous studies [12, 31]. Samadi et al. used standard 2D and 3D interpolation methods to generate future CT images during the breathing cycle. They report 5.49, 5.54, 5.63, 5.59, and 7.23 RMSE for 2D nearest, 2D linear, 2D spline, 2D cubic, and 3D DIRART software, respectively [31]. For the same input image data, Nabavi et al. recently used the PredNet model for the generation of a sequence of CT images throughout the respiratory cycle. They report 0.943 estimated SSIM for all patients, respectively [12]. Our estimated RMSE (3.17) and SSIM (0.988) show a significant improvement for all patients compared to the cited studies.

It is interesting to note that the patient's respiration, which consists of inhaling and exhaling, affects the accuracy of the model in different phases. Based on Table 3, the accuracy of the model was degraded when the air entered the lungs and the diaphragm moved, while the predicted results improve during exhalation. Figure 4 provides details about the results of pulmonary motion prediction based on the new design of the ConvLSTM model. The predicted image, the difference between the input image and target image, and the difference between the target image and predicted image in all three directions during the breathing cycle were shown in Figure 4. As shown in Figure 4, the ConvLSTM network can predict pulmonary motion in a specific area (deformed region) with significant accuracy and similarity in all three views. Overall, the new design of the ConvLSTM network can be used in different radiation therapy applications as follows: 1) the proposed network used for 4DCT images suffer from rings, metals, streaks, and blurring artifacts; therefore, the patient has kept away from re-scanning and radiation delivery, 2) the generated sequence of CT images throughout the respiratory cycle can be used for margin delineation in radiotherapy treatment planning, and 3) the proposed method can be used for real-time tumor tracking along with surrogate signals.

Conclusion

In recent years, several studies have reported a sequence of CT images during the respiratory cycle based on deep artificial neural network architectures. In this study, the advantage of the ConvLSTM network is used to predict and generate CT images during the patient's breathing cycle. In this relation, two structures consisting of stacked ConvLSTM models associated with the hyperparameter optimizer algorithm and a new design of the ConvLSTM model were proposed. The stacked ConvLSTM model constructed with three layers, including 60 filters with (5×5) kernel size shows superior performance compared to other stacked ConvLSTM structures investigated in this study. The new design of the ConvLSTM model, on the other hand, was a combination of four parallelized ConvLSTM layers with a different number of filters and kernel sizes, concatenated together and associated with the ConvLSTM layer, including 32 filters with kernel size (3×3). The ConvLSTM network was also evaluated on a dataset that included all three views. The obtained results of the suggested ConvLSTM structure demonstrate an improvement of about 34% in the term of RMSE in comparison to the best stacked ConvLSTM model while using 5.2 times fewer parameters. Suggested ConvLSTM structure demonstrates that the generated CT images are more consistent with the corresponding references. As a further study, other deep learning networks, such as Bidirectional LSTM or Bidirectional ConvLSTM networks, would be presented to provide a comparative study.

Acknowledgment

The authors would like to thank the Léon Bérard Cancer Center and CREATIS Laboratory, Lyon, France, for sharing their imaging data. The authors would kindly also thank

Mansour Esmaeili Sanjavanmareh for his helpful remarks.

Authors' Contribution

Ghasemi Z and Samadi P, developed the original idea, performed the experiments, and wrote the paper. All the authors read, modified, and approved the final version of the manuscript.

Ethical Approval

This research study was conducted retrospectively using non-identifiable human subject data. The POPI study has been approved by the local research ethics committee and data have been anonymously open to the community since 2007. Applicable law and standards of ethics have been respected in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements.

Informed Consent

Informed consent for all patients was conducted by relevant guidelines and regulations.

Conflict of Interest

None

References

1. De Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal.* 2019;**52**:128-43. doi: 10.1016/j.media.2018.11.010. PubMed PMID: 30579222.
2. Oh S, Kim S. Deformable image registration in radiation therapy. *Radiat Oncol J.* 2017;**35**(2):101-11. doi: 10.3857/roj.2017.00325. PubMed PMID: 28712282. PubMed PMCID: PMC5518453.
3. Rigaud B, Simon A, Castelli J, Lafond C, Acosta O, Haigron P, Cazoulat G, et al. Deformable image registration for radiation therapy: principle, methods, applications and evaluation. *Acta Oncol.* 2019;**58**(9):1225-37. doi: 10.1080/0284186X.2019.1620331. PubMed PMID: 31155990.
4. Samadi Miandoab P, Esmaili Torshabi A, Parandeh S. Calculation of inter-and intra-fraction motion errors at external radiotherapy using a markerless strategy based on image registration combined with correlation model. *Iranian Journal of Medical Physics.* 2019;**16**(3):224-31. doi: 10.22038/IJMP.2018.30477.1348.
5. Watkins WT, Li R, Lewis J, Park JC, Sandhu A, Jiang SB, et al. Patient-specific motion artifacts in 4DCT. *Med Phys.* 2010;**37**(6):2855-61. doi: 10.1118/1.3432615. PubMed PMID: 20632597.
6. Kwong Y, Mel AO, Wheeler G, Troupis JM. Four-dimensional computed tomography (4DCT): A review of the current status and applications. *J Med Imaging Radiat Oncol.* 2015;**59**(5):545-54. doi: 10.1111/1754-9485.12326. PubMed PMID: 26041442.
7. Keall PJ, Vedam SS, George R, Williamson JF. Respiratory regularity gated 4D CT acquisition: concepts and proof of principle. *Australas Phys Eng Sci Med.* 2007;**30**(3):211-20. doi: 10.1007/BF03178428. PubMed PMID: 18044305.
8. Sarrut D, Clippe S. Fast DRR generation for intensity-based 2D/3D image registration in radiotherapy. *LIRIS UMR.* 2003;5205.
9. Xie Y, Xing L, Gu J, Liu W. Tissue feature-based intra-fractional motion tracking for stereoscopic x-ray image guided radiotherapy. *Phys Med Biol.* 2013;**58**(11):3615-30. doi: 10.1088/0031-9155/58/11/3615. PubMed PMID: 23648334.
10. Miandoab PS, Torshabi AE, Nankali S. Extraction of respiratory signal based on image clustering and intensity parameters at radiotherapy with external beam: A comparative study. *J Biomed Phys Eng.* 2016;**6**(4):253-64. PubMed PMID: 28144595. PubMed PMCID: PMCPMC5219576.
11. Kai J, Fujii F, Shiinoki T. Prediction of Lung Tumor Motion Based on Recurrent Neural Network. International Conference on Mechatronics and Automation (ICMA); Changchun, China: IEEE; 2018. p. 1093-9. doi: 10.1109/ICMA.2018.8484575.
12. Nabavi S, Abdoos M, Moghaddam ME, Mohammadi M. Respiratory Motion Prediction Using Deep Convolutional Long Short-Term Memory Network. *J Med Signals Sens.* 2020;**10**(2):69-75. doi: 10.4103/jmss.JMSS_38_19. PubMed PMID: 32676442. PubMed PMCID: PMC7359959.
13. Van De Leemput SC, Prokop M, Van Ginneken B, Manniesing R. Stacked Bidirectional Convolutional LSTMs for Deriving 3D Non-Contrast CT From Spatiotemporal 4D CT. *IEEE Trans Med Imaging.* 2020;**39**(4):985-96. doi: 10.1109/TMI.2019.2939044. PubMed PMID: 31484111.
14. Fu Y, Lei Y, Wang T, Higgins K, Bradley JD, Curran WJ, et al. LungRegNet: An unsupervised deform-

able image registration method for 4D-CT lung. *Med Phys*. 2020;**47**(4):1763-74. doi: 10.1002/mp.14065. PubMed PMID: 32017141. PubMed PMCID: PMC7165051.

15. Lee SW, Kim HY. Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation. *Expert Systems with Applications.* 2020;**161**:113704. doi: 10.1016/j.eswa.2020.113704. X.
16. Yuan Z, Zhou X, Yang T. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; New York: Association for Computing Machinery; 2018. p. 84-92. doi: 10.1145/3219819.3219922.
17. Ray A, Rajeswar S, Chaudhury S. Text recognition using deep BLSTM networks. International conference on advances in pattern recognition (ICAPR); Kolkata, India: IEEE; 2015. doi: 10.1109/ICAPR.2015.7050699.
18. Balci B, Saadati D, Shiferaw D. Handwritten text recognition using deep learning. CS23 1n: Convolutional Neural Networks for Visual Recognition; Stanford University: Spring; 2017. p. 752-9.
19. Kim S, Hong S, Joh M, Song SK. Deeprain: Convlstm network for precipitation prediction using multichannel radar data[Internet]. arXiv [Preprint]. 2017 [cited 2017 November 7]. Available from: https://arxiv.org/abs/1711.02316.
20. Salman AG, Heryadi Y, Abdurahman E, Suparta W. Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting. *Procedia Computer Science*. 2018;**135**:89-98. doi: 10.1016/j.procs.2018.08.153.
21. Zhang K, Geng X, Yan XH. Prediction of 3-D ocean temperature by multilayer convolutional LSTM. *IEEE Geoscience and Remote Sensing Letters*. 2020;**17**(8):1303-7. doi: 10.1109/lgrs.2019.2947170.
22. Dehkordi AN, Sina S, Khodadadi F. A Comparison of Deep Learning and Pharmacokinetic Model Selection Methods in Segmentation of High-Grade Glioma. *Frontiers in Biomedical Technologies.* 2021;**8**(1);50-60. doi: 10.18502/fbt.v8i1.5858.
23. Zapata-Impata BS, Gil P, Torres F. Learning Spatio Temporal Tactile Features with a ConvLSTM for the Direction Of Slip Detection. *Sensors (Basel).* 2019;**19**(3):523. doi: 10.3390/s19030523. PubMed PMID: 30691197. PubMed PMCID: PMC6387284.
24. Wang D, Yang Y, Ning S. Deepstcl: A deep spatio-temporal convlstm for travel demand prediction. International joint conference on neural networks (IJCNN); Brazil: IEEE; 2018.
25. Vandemeulebroucke J, Sarrut D, Clarysse P. The POPI-model, a point-validated pixel-based breathing thorax model. XVth international conference on the use of computers in radiation therapy; Toronto, Ontario, Canada: ICCR; 2007. p. 195-99.
26. Xingjian SH, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [Internet]. arXiv [Preprint]. 2015 [cited 2015 June 13]. Available from: https://arxiv.org/abs/1506.04214.
27. Yu T, Zhu H. Hyper-parameter optimization: A review of algorithms and applications. *ArXiv:200305689.* 2020.
28. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* 2004;**13**(4):600-12. doi: 10.1109/tip.2003.819861. PubMed PMID: 15376593.
29. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. 14th international joint conference on Artificial intelligence; Montreal, Canada: Ijcai; 1995. p. 1137-43.
30. Purdie TG, Moseley DJ, Bissonnette JP, et al. Respiration correlated cone-beam computed tomography and 4DCT for evaluating target motion in Stereotactic Lung Radiation Therapy. *Acta Oncol.* 2006;**45**(7):915-22. doi: 10.1080/02841860600907345. PubMed PMID: 16982558.
31. Samadi-Miyandoab P, Torshabi A, Nankali S. 2D and 3D Optical Flow Based Interpolation of the 4DCT ImageSequences in the External Beam Radiotherapy. *Frontiers in Biomedical Technologies.* 2015;**2**(2):93-102.