

Article

ViSTA: A Novel Network Improving Lung Adenocarcinoma Invasiveness Prediction from Follow-Up CT Series

Wei Zhao ^{1,†}, Yingli Sun ^{2,†}, Kaiming Kuang ³, Jiancheng Yang ^{3,4}, Ge Li ⁵, Bingbing Ni ⁴, Yingjia Jiang ¹, Bo Jiang ¹, Jun Liu ^{1,6,*}  and Ming Li ^{2,7,*}

¹ Department of Radiology, The Second Xiangya Hospital, Central South University, Changsha 410011, China; wei.zhao@csu.edu.cn (W.Z.); moshangqingcheng@163.com (Y.J.); jiangbo@csu.edu.cn (B.J.)

² Department of Radiology, Huadong Hospital, Fudan University, Shanghai 200040, China; sunyingli208ok@126.com

³ Dianei Technology, Shanghai 200051, China; kaiming.kuang@dianei-ai.com (K.K.); jekyll4168@sjtu.edu.cn (J.Y.)

⁴ Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; nibingbing@sjtu.edu.cn

⁵ Department of Radiology, The Xiangya Hospital, Central South University, Changsha 410008, China; ligeanyi@126.com

⁶ Radiology Quality Control Center, Changsha 410011, China

⁷ Institute of Functional and Molecular Medical Imaging, Fudan University, Shanghai 200437, China

* Correspondence: junliu123@csu.edu.cn (J.L.); ming_li@fudan.edu.cn (M.L.); Tel.: +86-137-8708-5002 (J.L.); +86-138-1662-0371 (M.L.); Fax: +86-0731-85292116 (J.L.); +86-21-57643271 (M.L.)

† These authors contributed equally to this work.



Citation: Zhao, W.; Sun, Y.; Kuang, K.; Yang, J.; Li, G.; Ni, B.; Jiang, Y.; Jiang, B.; Liu, J.; Li, M. ViSTA: A Novel Network Improving Lung Adenocarcinoma Invasiveness Prediction from Follow-Up CT Series. *Cancers* **2022**, *14*, 3675. <https://doi.org/10.3390/cancers14153675>

Academic Editors: Hamid Khayyam, Ali Madani, Rahele Kafieh and Ali Hekmatnia

Received: 7 June 2022

Accepted: 20 July 2022

Published: 28 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Assessing follow-up computed tomography(CT) series is of great importance in clinical practice for lung nodule diagnosis. Deep learning is a thriving data mining method in medical imaging and has obtained surprising results. However, previous studies mostly focused on the analysis of single static time points instead of the entire follow-up series and required regular intervals between CT examinations. In the current study, we propose a new deep learning framework, named ViSTA, that can better evaluate tumor invasiveness using irregularly serial follow-up CT images to avoid aggressive procedures or delay diagnosis in clinical practice. ViSTA provides a new solution for irregularly sampled data. ViSTA delivers superior performance compared with other static or serial deep learning models. The proposed ViSTA framework is capable of improving performance close to the human level in the prediction of invasiveness of lung adenocarcinoma while being transferrable to other tasks analyzing serial medical data.

Abstract: To investigate the value of the deep learning method in predicting the invasiveness of early lung adenocarcinoma based on irregularly sampled follow-up computed tomography (CT) scans. In total, 351 nodules were enrolled in the study. A new deep learning network based on temporal attention, named Visual Simple Temporal Attention (ViSTA), was proposed to process irregularly sampled follow-up CT scans. We conducted substantial experiments to investigate the supplemental value in predicting the invasiveness using serial CTs. A test set composed of 69 lung nodules was reviewed by three radiologists. The performance of the model and radiologists were compared and analyzed. We also performed a visual investigation to explore the inherent growth pattern of the early adenocarcinomas. Among counterpart models, ViSTA showed the best performance (AUC: 86.4% vs. 60.6%, 75.9%, 66.9%, 73.9%, 76.5%, 78.3%). ViSTA also outperformed the model based on Volume Doubling Time (AUC: 60.6%). ViSTA scored higher than two junior radiologists (accuracy of 81.2% vs. 75.4% and 71.0%) and came close to the senior radiologist (85.5%). Our proposed model using irregularly sampled follow-up CT scans achieved promising accuracy in evaluating the invasiveness of the early stage lung adenocarcinoma. Its performance is comparable with senior experts and better than junior experts and traditional deep learning models. With further validation, it can potentially be applied in clinical practice.

Keywords: adenocarcinoma; invasiveness; X-ray computed tomography; deep learning; temporal attention

1. Introduction

Low-dose computed tomography (LDCT) is recommended for lung cancer screening in high-risk populations based on the National Lung Cancer Screening Trial (NLST) report, which is now included in US screening guidelines [1]. Owing to LDCT, more and more early stage lung adenocarcinomas are diagnosed and treated. In clinical practice, most people require follow-up CT scans due to the indeterminate diagnosis or low probability of malignancy on baseline CT. Assessing the changes in size, CT value, and other imaging features can substantially help the diagnosis and invasiveness evaluation of early stage lung adenocarcinomas. However, the evaluation process is tedious and lacks objectivity, which means that radiologists could be overwhelmed by numerous serial CT image evaluations. Moreover, the features indicating malignancy may not be present in the early stages of lung adenocarcinoma. As we know, biological changes may precede morphological changes. Therefore, an efficient tool for objectively evaluating the changes and mining the internal patterns of lung nodules on serial CTs is of great importance.

Deep learning is a thriving data mining method in medical imaging and has obtained surprising results [2–4]. It can efficiently and automatically process medical images and has achieved promising performances on par with clinicians on various clinical tasks, including disease classification, medical image registration, and organ segmentation [5–9]. Previous studies have shown that deep learning could aid clinical decision-making for early lung cancer in disease management and invasiveness prediction [10–14]. However, most prior studies only included single-time CT scan images, while serial CT scan images were not fully investigated. Several powerful deep learning methods have been invented to process serial data, e.g., Long Short-term Memory, Gated Recurrent Unit Network, and Transformer [15–17]. Equipped with the aforementioned tools, a deep learning system can include serial images, better evaluate the biological behavior and changes, and then better predict different clinical events, such as prognosis, therapeutic effect, and subsequent growth patterns [18].

Serial deep learning models have achieved great success in serial data domains, including natural language processing, video classification, and speech recognition [15,19,20]. Nonetheless, it is important to notice that medical serial data such as electronic health records [21] or medical examinations are almost always sampled irregularly in time, separating them from the aforementioned modalities. Since the progression of the disease is strongly correlated with the time intervals between two time points, the asynchronous (sampled irregularly) nature of medical data requires special treatment. For example, by limiting sampling time intervals to 1, 3, and 6 months, deep learning methods proved effective in integrating multiple time points and improving the prediction of lung cancer treatment response [22]. However, this restriction on time intervals still limits the usage of the deep learning method in processing clinical serial data, especially for irregularly serial data.

In this article, we propose ViSTA (Visual Simple Temporal Attention), a deep learning framework capable of predicting the tumor invasiveness of pulmonary adenocarcinomas from Follow-up CT Series. The main contributions are three-fold: First, by introducing a simple temporal attention mechanism, we propose a new deep learning network, named ViSTA, to evaluate the invasiveness of early stage lung adenocarcinoma using irregularly serial CT scans images. ViSTA is able to gather information throughout the entire series and improve the prediction performance. Compared with serial analysis using traditional recurrent neural networks [22], ViSTA is not limited by different time intervals and can process completely irregularly sampled serial data. ViSTA was trained and validated on a dataset of 1121 CT scans from 282 follow-up series and evaluated on a hold-out test set of 113 CT scans from 69 follow-up series. Second, ViSTA delivers superior performances

compared with other static or serial deep learning models. ViSTA also outperforms size-based predictive methods (Volume Doubling Time [23]) by a large margin. Third, ViSTA was proven to achieve higher scores than two junior radiologists and came close to one senior radiologist in the observer study. Our results prove ViSTA's superiority in terms of processing irregularly sampled series and its great potential of being put into clinical practice in reality. Additionally, ViSTA is completely transferrable to other medical imaging tasks where analyzing serial data should yield better performances.

2. Materials and Methods

2.1. Data Collection

From January 2011 to October 2017, a search of the electronic medical records and the radiology information systems of the hospital was performed by one author (Yingli Sun). The inclusion criteria are as follows: (1) two or more available CT examinations with thin-slice (≤ 1.5 mm) images before resection. If there were only two CT examinations, the interval between two scans should be over 30 or more days. (2) Complete pathologic reports. The exclusion criteria for this analysis were: (1) prior treatment before surgery; (2) poor quality CT images; (3) lesions that were difficult to clearly delineate. Finally, a total of 351 nodules from 347 patients (mean age, 58.41 years ± 11.79 (SD); range, 22–84 years) were enrolled in the study. Among the 351 lung nodules, 191 nodules were pathologically identified as preinvasive lesions, including 1 atypical adenomatous hyperplasia (AAH), 39 adenocarcinomas in situ (AIS), and 151 minimally invasive adenocarcinoma (MIA); whereas 160 nodules were identified as invasive adenocarcinoma (IA). In total, 1234 serials CT scans of the 351 nodules were enrolled in this study. The median interval between the first and the last CT examinations was 366 ± 500 days (range, 30–2813 days; interquartile range, 165–852 days). The 351 nodules were randomly separated into a training set (245 nodules), validation set (37 nodules), and test set (69 nodules) (see Table 1).

Table 1. Number of CT scans/nodules in training, validation, and test set.

Pathological Type		No. CT Scans/Nodules			
		Training	Validation	Test	Total
Non-IA	AAH	5/1	0/0	0/0	5/1
	AIS	98/29	9/4	19/6	126/39
	MIA	383/104	40/16	114/31	537/151
	Total	486/134	49/20	133/37	668/191
	IA	398/111	64/17	104/32	566/160
	Total	884/245	113/37	237/69	1234/351

2.2. CT Scanning Parameters

Preoperative chest CT in our department was performed using the following four scanners: GE Discovery CT750 HD, 64-slice LightSpeed VCT (GE Medical Systems, Chicago, IL, USA); Somatom Definition flash, Somatom Sensation-16 (Siemens Medical Solutions, Erlangen, Germany) with the following parameters: 120 kVp; 100–200 mAs; pitch, 0.75–1.5; and collimation, 1–1.5 mm, respectively. All imaging data were reconstructed using a medium sharp reconstruction algorithm with a thickness of 1–1.5 mm.

2.3. Nodule Labeling, Segmentation and Imaging Preprocessing

A medical image processing and navigation software 3D Slicer (v4.8.0, Brigham and Women's Hospital, Boston, MA, USA) was used to manually delineate the volume of interest (VOI) of the included nodules at the voxel level by one radiologist (Yingli Sun, with 5 years of experience in chest CT interpretation), then the VOI was confirmed by another radiologist (Ming Li, with 12 years of experience in chest CT interpretation). Large vessels and bronchioles were excluded as much as possible from the volume of the nodule. The lung CT DICOM (Digital Imaging and Communications in Medicine) format images were imported into the software for delineation, and then the images with VOI information were

extracted with NII format for next step analysis. Each segmented nodule was attributed a specific pathological label (AAH, AIS, MIA, IA), according to the detailed pathological report. Two steps were performed to preprocess CT images before path extraction. First, the whole-volume CT image was resampled to the spacing of 1 mm in all three dimensions to guarantee isotropy. Second, HU values were clipped to the range of $(-1000, 400)$ and normalized to $(0, 1)$ using minimum–maximum normalization. Normalization can accelerate the convergence in the training of the deep learning model and improve its generalization ability.

2.4. Development of the Deep Learning Model

We developed a deep learning model named ViSTA to classify IA/non-IA lung nodules. The overall architecture of ViSTA is presented in Figure 1. ViSTA first extracts features from CT image patches using a CNN backbone and then integrates information from time series using a lightweight attention module named SimTA [24], which is designed specifically for analyzing asynchronous time series. Details regarding the architecture of ViSTA are provided in Supplementary Section S1, and a single SimTA layer was shown in Figure S1. To avoid overoptimization, we did not heavily tune the hyperparameters of our deep learning model and simply adopted common settings. ViSTA and all its counterparts are trained end-to-end for 100 epochs using the AdamW optimizer [25]. We used a cosine decay learning schedule from 10^{-3} to 10^{-6} . The batch size of each update was 32. The drop-out probability and weight decay were set at 0.2 and 0.01 to avoid overfitting.

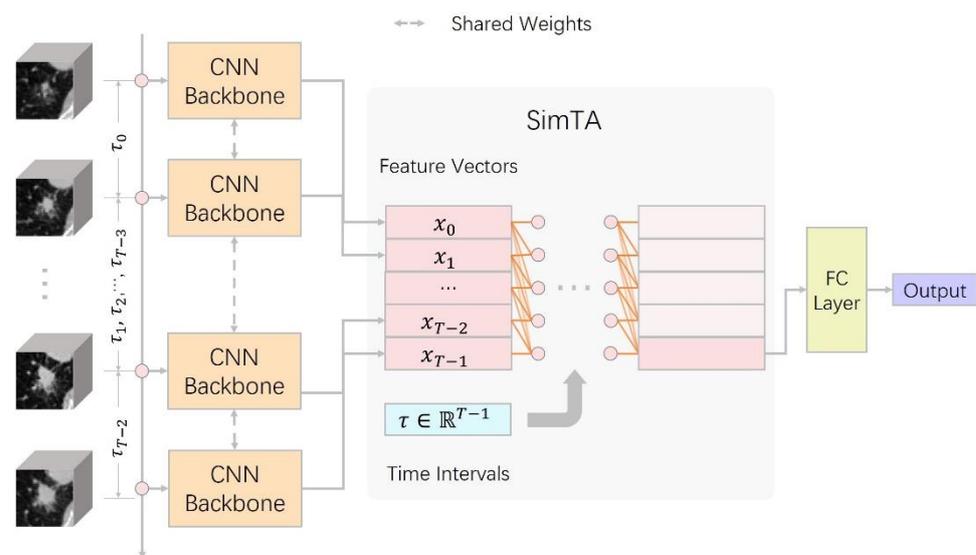


Figure 1. The model overview of the proposed ViSTA. It consists of a CNN backbone followed by the SimTA module made up of several SimTA layers.

2.5. Counterpart Methods

For comparison with ViSTA, we conducted experiments on a few of its counterparts:

- VDT (Volume Doubling Time). VDT is an important volumetric indicator used in follow-up examinations. It represents the time it takes for a nodule to double its volume. The formula of VDT is provided in Supplementary Section S1. Nodules with $VDT < 400$ days are considered fast-growing and are more likely to be malignant [23]. In this research, we evaluated VDT’s metrics under two different thresholds: 400 days and the cutoff that provides the best Youden index on the validation set. Youden index’s formula is presented in Supplementary Section S1;
- CNN (Convolutional Neural Network): To compare ViSTA against static models, we introduced CNN as a counterpart. We conducted the following experiments to further investigate the source of performance difference between ViSTA and CNN;

- CNN-last: This experiment was conducted to train and validate CNN only on the last time point of each follow-up series. It is obvious that the last time point is most relevant to the final diagnosis;
- CNN-first: This experiment was conducted to train and validate CNN only on the first time point of each follow-up series. In this experiment, the first steps were treated as if they had the same label as the last one. This setting was used to confirm that earlier time points convey less information than later ones;
- CNN-all: This experiment was conducted to train CNN on all time points of each follow-up series and validate it on the first and last time point separately (named CNN-all-first and CNN-all-last, respectively). This was used to investigate if ViSTA's superior performance only comes from the larger data size it enjoys;
- CNN+LSTM (Long Short-term Memory) [16]: LSTM is a subtype of RNN (Recurrent Neural Network) designed to analyze serial data and capture long-term relations. This setting is quite similar to previous research which combined CNN and RNN to predict lung cancer treatment response [22]. However, we did not limit time intervals to specific values so that we could fairly compare ViSTA and RNN-based methods. One major difference between CNN+LSTM and ViSTA is that CNN+LSTM treats all time points as if they had the same interval (synchronous). By comparing the previous two methods, we would like to see if ViSTA is more suitable for analyzing irregularly sampled time series.

2.6. Evaluation and Statistical Analysis

We evaluated the proposed ViSTA model both quantitatively and qualitatively. To evaluate each method's performance, we used a variety of metrics, including accuracy, precision, sensitivity, F1 score, and AUC. Formulas of evaluation metrics are presented in Supplementary Section S1.

To explore the visual representation and interpretability of ViSTA, we followed Simonyan, K. et al. [26] and plotted our model's saliency maps through backpropagation, and investigated the mechanism under ViSTA and where it directed its attention.

2.7. Observer Study

To further evaluate the performance of ViSTA, we conducted an observer study to compare the performance of radiologists in the same task against other models. In the observer study, all 69 CT series in the test set were evaluated by three radiologists. One is a senior radiologist with 22 years of experience, and the other two are junior radiologists with 5 and 3 years of experience, respectively. Radiologists gave the results based on the evaluation of all available serials CTs. The reviewed results were analyzed and compared with the performance of our proposed model. Radiologists' performances were evaluated using accuracy, sensitivity, precision, and F1 score.

3. Results

3.1. Performance of Deep Learning Models in Predicting the Invasiveness of Early Lung Adenocarcinoma

To validate the effectiveness of ViSTA in predicting IA/non-IA nodules, we evaluated its performance using a variety of metrics against its counterparts: VDT (cutoff value set at best Youden index or 400 days), CNN (including CNN-last, CNN-first, CNN-all-first and CNN-all-last), and CNN+LSTM.

Tables S1 and S2 show their performances on the training dataset and validation dataset. Figure 2 provide the ROC curves of all models on the test dataset. Our proposed model outperformed all deep learning models and VDT-based methods in every metric by considerable margins (best among models are highlighted with an underscore). It is worth noting that VDT is far from effective in terms of invasiveness classification. It underperformed almost all deep learning models in terms of AUC, accuracy, and F1 score. Secondly, sequential models (ViSTA and CNN+LSTM) delivered better performances

than CNN models that utilize static data points. ViSTA outperformed CNN+LSTM by considerable margins in all metrics. This performance gap can be attributed to ViSTA's suitability to analyze asynchronous time series. Unlike CNN+LSTM, which treats all time points as if they were regularly sampled, ViSTA takes time intervals into account and is better at processing follow-up series. Furthermore, we trained CNN on all time points (CNN all-first and CNN all-last) to investigate if sequential models gain superiority over larger training datasets. It turned out that ViSTA and CNN+LSTM still outperformed CNN even when it was trained on all data.

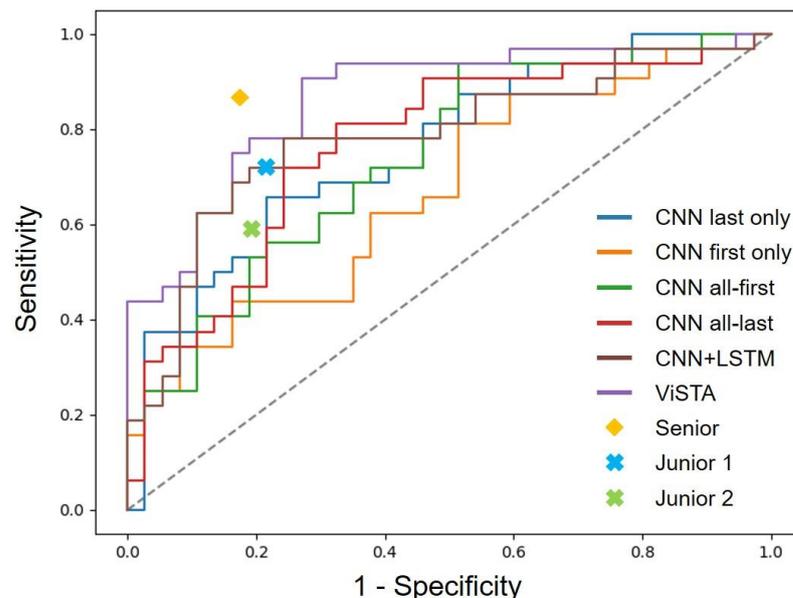


Figure 2. ROC curves of different models compared with performances of radiologists. The gray dotted line indicates the performance of a random classifier with no predictive ability.

3.2. Performance Comparison against Radiologists

In the observer study, we compared the performances of ViSTA and its counterparts against three radiologists (Table 2). One is a senior radiologist with 22 years of experience, and the other two are junior radiologists with 5 and 3 years of experience, respectively. All 69 follow-up series from the test set were included in the observer study. We evaluated radiologists' performances using accuracy, sensitivity and precision, and F1 score and compared them against the proposed model. In terms of metrics that require specifying threshold, we chose the threshold that delivers the best Youden Index on the validation set as the cutoff value. Figure 2 plot deep learning models' ROC curves against radiologists' metrics. In terms of accuracy and F1 score, ViSTA scored higher than the two junior radiologists (accuracy of 81.2% vs. 75.4% and 71.0%; F1 score of 81.7% vs. 73.0% and 65.5%) and came close to the senior radiologist (accuracy of 81.2% vs. 85.5%; F1 score of 81.7% vs. 84.8%).

3.3. Visual Presentation Investigation

To investigate the mechanism of ViSTA, we used a neural network visualization technique [26] to visualize the attention heatmap of the model, which was mostly attributed to the predicted results and potentially correlated to the biological behavior (Figure 3). We took the absolute value of the raw heatmap and clipped it to the range of (0, 0.01) for better visualization and interpretation. In view of the created heatmaps, we can see that the "attention" of the deep learning system was mostly focused on the nodule. Areas surrounding the nodule draw the attention of ViSTA as well, meaning that they also carry valuable information as the nodule does (Figure 3A,B). Figure 3A show a long follow-up series of 11 time points. We observed that heatmaps stay blank in the first half of the

series, during which both nodule volume and IA probability remain relatively stable. In the latter half, heatmaps begin to show along with significant increases in nodule volume and IA probability. Heatmaps are sometimes only lit up at the last time point (Figure 3B). We contribute this to the sudden increase of nodule volume between the third and the fourth time point, which provides sufficient information for the model. This argument is supported by the spike of IA probability at the fourth time point. In some rare cases, heatmaps on all time points are close to invisible (Figure 3C). We conjecture that this is because the lung nodule had almost no progression, which was proven by the fact that both nodule volume and IA probability stayed almost unchanged throughout the entire series.

Table 2. The performance of different models and radiologists on the test dataset. The highest among all is highlighted in bold, and the highest among models and VDT (Volume Doubling Time)-based methods is highlighted with an underscore.

	AUC	Acc.	Prec.	Sens.	F1
Senior	-	85.5%	82.4%	87.5%	84.8%
Junior 1	-	75.4%	74.2%	71.9%	73.0%
Junior 2	-	71.0%	73.1%	59.4%	65.5%
1/VDT (best Youden index)	60.6%	62.3%	56.3%	84.4%	67.5%
1/VDT (400 days)	60.6%	58.0%	71.4%	15.6%	25.6%
CNN last only	75.9%	72.5%	72.4%	65.6%	68.9%
CNN first only	66.9%	65.2%	70.0%	43.8%	64.3%
CNN all-first	73.9%	65.2%	60.5%	71.9%	65.7%
CNN all-last	76.5%	73.9%	71.9%	71.9%	71.9%
CNN+LSTM	78.3%	76.8%	73.5%	78.1%	75.8%
ViSTA	86.4%	81.2%	74.4%	90.6%	81.7%

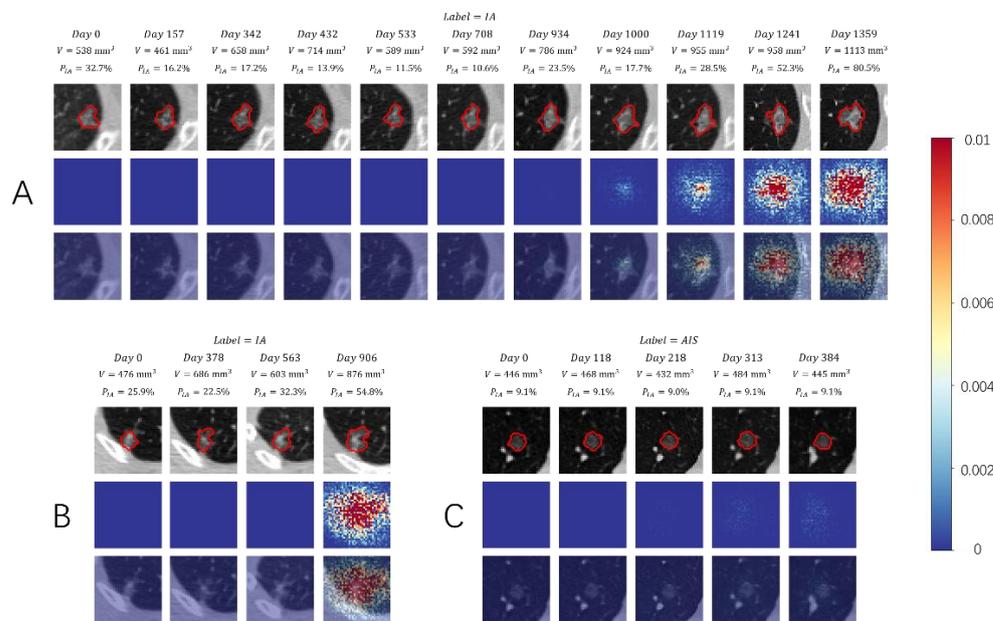


Figure 3. Visualization investigation of ViSTA. The top row shows CT slices of each time point in the follow-up series. The middle row shows attention heatmaps extracted using the technique proposed by Simonyan, K. et al. [26]. The bottom row masks heatmaps on top of CT slices. (A) Attention gradually grew along with the nodule volume and IA probability as the nodule progressed to the end of the series. (B) The heatmap only lit up at the last time point as it is considered the one carrying valuable information. (C) All time points are allocated with little to no attention, which may be caused by the slow progress of the nodule.

4. Discussion

In the current study, we proposed a deep learning framework named ViSTA to predict the invasiveness of lung adenocarcinomas using serial CT images. Our results showed that models fed with serial CT images substantially and consistently outperformed models fed with single CT images. Moreover, our proposed model can effectively process asynchronous time series and outperform the traditional serial network, i.e., LSTM. Our models achieved an AUC of 86.4% and an F1 score of 81.7% in the test dataset, which were higher than those of all its counterparts. In the observer study, ViSTA achieved higher accuracy and F1 scores than two junior radiologists (accuracy of 81.2% vs. 75.4% and 71.0%, F1 score of 81.7% vs. 73.0% and 65.5%). When compared with the senior radiologist, our proposed model delivered close performance (accuracy of 81.2% vs. 85.5%, F1 score of 81.7% vs. 84.8%).

Timely and accurately assessing the biological behavior of early stage lung adenocarcinomas has been a continuous focus of attention in clinical practice. In contrast to traditional radiographic features and handcraft features, deeper and higher dimension level features mined by the deep learning method present promising advantages in many tasks, including predicting the invasiveness of the early lung adenocarcinoma. Kim et al. performed a comparison study and revealed that the predictive accuracy of the deep learning method was superior to those of the size-based logistic model [11]. We also analyzed the predictive value of VDT [27], a size-based key parameter in the differentiation of aggressive tumors from slow-growing tumors in clinical practice [24]. Not surprisingly, the performance of our proposed model substantially exceeded that of the VDT-based methods. It indirectly verified the conjecture that a deep learning system could extract and learn deeper and more valuable features, then better discover the biological behavior of the tumors and predict the invasiveness of early stage lung adenocarcinoma.

Although the deep learning method can obtain better performance, most previous studies only used single CT scan data prior to the surgery for training and extracting features, which cannot reveal and learn the internal growth pattern of the nodules. In clinical scenarios, internal growth is a vital component of Lung-RADS, a guideline to standardize image interpretation by radiologists and dictate management recommendations. Including serial CTs can facilitate medical tasks, such as differentiating benign tumors from malignant ones [28] and monitoring and predicting treatment response [22,29]. The discovery of our study supports this. By modelling serial CTs, the predictive performance of ViSTA substantially surpassed its counterparts in analyzing static data. In clinical practice, sequential medical data is generally sampled irregularly, i.e., with different follow-up time intervals. To address the irregular sampling issue, we adopted SimTA in our proposed model to process irregularly sampled time series. This lightweight module enables modeling sequential information in an efficient way. It turned out that the proposed ViSTA significantly outperformed the standard serial framework, i.e., CNN+LSTM, with considerably fewer parameters and less computation and memory footprint. ViSTA can better take advantage of the complete information of all time point CTs by modelling simple yet effective exponentially decay attention in time series. This was proved by our experiments comparing ViSTA, CNN+LSTM, and pure CNN models trained with all time point CTs (CNN-all). ViSTA's superiority over CNN-all proved that its performance gain does not come from a larger training dataset.

In the visualization analysis, we found that ViSTA can drive its attention on the nodule and the surrounding tissue and drop more attention when the probability of invasiveness increases. It can partly explain the mechanism of the deep learning system. We also found some cases where the model appeared to use features close to the nodule, such as the vasculature and parenchyma surrounding the nodule. In fact, peritumoral tissue may possess valuable information, such as tumor-infiltrating status. Features extracted from the peritumoral tissues can improve the efficiency of intramodular radiomic analysis [30,31]. However, we still cannot fully interpret whether the model incorporates other abnormalities such as background emphysema in its predictions. Further investigation using more comprehensive model attribution techniques may allow clinicians to take advantage of the

same visual features used by the model to assess the biological status of tumors. It is worth noting that some of the heatmaps in the time series are completely blue, meaning that the deep learning model allocated close to zero attention to these time points. Though this phenomenon is not completely interpretable, we argue that it can be attributed to these two facts: these time points are too far from the current one, and they lack findings informative for the deep learning model.

Even though the proposed ViSTA proved effective in processing irregularly sampled CT series in our experiments, there are several limitations left untouched. First, due to the difficulty of collecting complete lung nodule follow-up series, we only included data from a single center in this study. In clinical practice, it is preferable if the proposed method generalizes to multiple data domains. Furthermore, it is possible that a single follow-up series contains CT scans from different centers, which would be an important challenge to solve if the proposed model were to be put into clinical usage. In future studies, we will include CT series from external centers to validate the generalization performance of ViSTA. Second, the SimTA module in ViSTA models a simple temporal attention mechanism that monotonically increases weights as the time point gets closer to the current time. However, it is viable to model more complicated attention relations using deep learning models such as Transformer or Informer [15,32]. These temporal models enable capturing non-monotonic and dynamic temporal attention that could be useful in predicting invasiveness. Last but not least, even though we conducted a visual investigation on ViSTA, the interpretation of deep learning model predictions still remains a major challenge. Additionally, the final clinical decision is still up to clinicians to date. In our future research, we will further investigate the underlying mechanism of ViSTA or other similar attention mechanisms.

5. Conclusions

To summarize, we designed a deep learning model processing irregularly sampled CT series to predict the invasiveness of early stage lung adenocarcinoma from follow-up CT scans. The model achieved promising accuracy comparable with senior experts and better than junior experts and its counterparts. With further validation, the proposed model could better evaluate the invasiveness of early stage lung adenocarcinoma, avoiding aggressive procedures or delayed diagnosis and helping precise management in clinical practice.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers14153675/s1>. References [24,33–36]. Section S1: Supplementary Methods. Figure S1: The illustration of a single SimTA layer. The colormap of the attention matrix indicates the magnitude of attention allocated in a particular timepoint. Gray represents zero attention. The darker the orange color is, the more attention the timepoint gets; Table S1: Different models' performances in predicting invasiveness on the training dataset. The highest among all is highlighted in bold; Table S2: Different models' performances in predicting invasiveness on the validation dataset. The highest among all is highlighted in bold.

Author Contributions: Conceptualization, K.K. and J.L.; Data curation, W.Z., Y.S. and M.L.; Formal analysis, W.Z., Y.S. and J.Y.; Funding acquisition, M.L.; Investigation, W.Z., K.K., Y.S., G.L., B.N., Y.J. and B.J.; Methodology, W.Z., K.K., Y.S., J.Y., Y.J. and B.J.; Project administration, M.L.; Resources, Y.S.; Software, K.K., B.N. and M.L.; Supervision, J.Y., J.L. and M.L.; Validation, G.L.; Visualization, B.N.; Writing—original draft, W.Z. and Y.S.; Writing—review and editing, W.Z., J.L. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China 82102157 (Wei Zhao), 61976238 (Ming Li), Hunan Provincial Natural Science Foundation for Excellent Young Scholars 2022JJ20089 (Wei Zhao), Hunan Provincial Natural Science Foundation of China 2021JJ40895 (Wei Zhao), the Research Project of Postgraduate Education and Teaching Reform of Central South University 2021JGB147 (Jun Liu), 2022JGB117 (Wei Zhao), the Clinical Research Center For Medical Imaging In Hunan Province 2020SK4001 (Jun Liu), the science and technology innovation program of Hunan Province 2021RC4016 (Jun Liu), and the Clinical Medical Technology Innovation Guidance Project in Hunan Province 2020SK53423 (Wei Zhao), Science and Technology Planning Project of Shanghai Science and Technology Commission 22Y11910700 (Ming Li), Science and Technology

Planning Project of Shanghai Science and Technology Commission 20Y11902900 (Ming Li), Shanghai “Rising Stars of Medical Talent” Youth Development Program “Outstanding Youth Medical Talents” SHWJRS [2021]-99 (Ming Li).

Institutional Review Board Statement: This retrospective study was approved by the institutional review board of Huadong Hospital, affiliated with Fudan University (No. 2019K134).

Informed Consent Statement: The institutional review board waived the requirement for patients’ written informed consent.

Data Availability Statement: Data are available for bona fide researchers who request it from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ettinger, D.S.; Wood, D.E.; Aisner, D.L.; Akerley, W.; Bauman, J.R.; Bharat, A.; Bruno, D.S.; Chang, J.Y.; Chirieac, L.R.; D’Amico, T.A.; et al. Non-Small Cell Lung Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Cancer Netw.* **2022**, *20*, 497–530. [[CrossRef](#)] [[PubMed](#)]
2. Shen, D.; Wu, G.; Suk, H.I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)] [[PubMed](#)]
3. Krittanawong, C.; Johnson, K.W.; Rosenson, R.S.; Wang, Z.; Aydar, M.; Baber, U.; Min, J.K.; Tang, W.H.W.; Halperin, J.L.; Narayan, S.M. Deep learning for cardiovascular medicine: A practical primer. *Eur. Heart J.* **2019**, *40*, 2058–2073. [[CrossRef](#)] [[PubMed](#)]
4. Bi, W.L.; Hosny, A.; Schabath, M.B.; Giger, M.L.; Birkbak, N.; Mehrtash, A.; Allison, T.; Arnaout, O.; Abbosh, C.; Dunn, I.F.; et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J. Clin.* **2019**, *69*, 127–157. [[CrossRef](#)]
5. Bulten, W.; Pinckaers, H.; van Boven, H.; Vink, R.; De Bel, T.; Van Ginneken, B.; Van der Laak, J.; Hulsbergen-van de Kaa, C.; Litjens, G. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study. *Lancet Oncol.* **2020**, *21*, 233–241. [[CrossRef](#)]
6. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Guttag, J.; Dalca, A.V. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Trans. Med. Imaging* **2019**, *38*, 103–111. [[CrossRef](#)]
7. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [[CrossRef](#)]
8. Ahmad, M.; Qadri, S.F.; Ashraf, M.U.; Subhi, K.; Khan, S.; Zareen, S.S.; Qadri, S. Efficient Liver Segmentation from Computed Tomography Images Using Deep Learning. *Comput. Intell. Neurosci.* **2022**, *2022*, 2665283. [[CrossRef](#)]
9. Qadri, S.F.; Shen, L.; Ahmad, M.; Qadri, S.; Zareen, S.S.; Akbar, M.A. SVseg: Stacked Sparse Autoencoder-Based Patch Classification Modeling for Vertebrae Segmentation. *Mathematics* **2022**, *10*, 796. [[CrossRef](#)]
10. Massion, P.P.; Antic, S.; Ather, S.; Arteta, C.; Brabec, J.; Chen, H.; Declerck, J.; Dufek, D.; Hickes, W.; Kadir, T.; et al. Assessing the Accuracy of a Deep Learning Method to Risk Stratify Indeterminate Pulmonary Nodules. *Am. J. Respir. Crit. Care Med.* **2020**, *202*, 241–249. [[CrossRef](#)]
11. Kim, H.; Lee, D.; Cho, W.S.; Lee, J.C.; Goo, J.M.; Kim, H.C.; Park, C.M. CT-based deep learning model to differentiate invasive pulmonary adenocarcinomas appearing as subsolid nodules among surgical candidates: Comparison of the diagnostic performance with a size-based logistic model and radiologists. *Eur. Radiol.* **2020**, *30*, 3295–3305. [[CrossRef](#)]
12. Zhao, W.; Yang, J.; Sun, Y.; Li, C.; Wu, W.; Jin, L.; Yang, Z.; Ni, B.; Gao, P.; Wang, P.; et al. 3D Deep Learning from CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas. *Cancer Res.* **2018**, *78*, 6881–6889. [[CrossRef](#)]
13. De Margerie-Mellon, C.; Gill, R.R.; Salazar, P.; Oikonomou, A.; Nguyen, E.T.; Heidinger, B.H.; Medina, M.A.; VanderLaan, P.A.; Bankier, A.A. Assessing invasiveness of subsolid lung adenocarcinomas with combined attenuation and geometric feature models. *Sci. Rep.* **2020**, *10*, 14585. [[CrossRef](#)]
14. Wang, X.; Li, Q.; Cai, J.; Wang, W.; Xu, P.; Zhang, Y.; Fang, Q.; Fu, C.; Fan, L.; Xiao, Y.; et al. Predicting the invasiveness of lung adenocarcinomas appearing as ground-glass nodule on CT scan using multi-task learning and deep radiomics. *Transl. Lung Cancer Res.* **2020**, *9*, 1397–1406. [[CrossRef](#)]
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
16. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
17. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. In Proceedings of the SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014; pp. 103–111. [[CrossRef](#)]
18. Li, Y.; Yang, J.; Xu, Y.; Xu, J.; Ye, X.; Tao, G.; Xie, X.; Liu, G. Learning Tumor Growth via Follow-Up Volume Prediction for Lung Nodules. *arXiv* **2020**, arXiv:2006.13890. [[CrossRef](#)]

19. Wu, C.-Y.; Girshick, R.; He, K.; Feichtenhofer, C.; Krahenbuhl, P. A Multigrid Method for Efficiently Training Video Models. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
20. Saon, G.; Kurata, G.; Sercu, T.; Audhkhasi, K.; Thomas, S.; Dimitriadis, D.; Cui, X.; Ramabhadran, B.; Picheny, M.; Lim, L.-L.; et al. English Conversational Telephone Speech Recognition by Humans and Machines. In Proceedings of the INTERSPEECH 2017: Conference of the International Speech Communication Association, Stockholm, Sweden, 20–24 August 2017.
21. Rajkomar, A.; Oren, E.; Chen, K.; Dai, A.M.; Hajaj, N.; Hardt, M.; Liu, P.J.; Liu, X.; Marcus, J.; Sun, M.; et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit. Med.* **2018**, *1*, 18. [[CrossRef](#)]
22. Xu, Y.; Hosny, A.; Zeleznik, R.; Parmar, C.; Coroller, T.; Franco, I.; Mak, R.H.; Aerts, H.J. Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clin. Cancer Res.* **2019**, *25*, 3266–3275. [[CrossRef](#)]
23. Heuvelmans, M.A.; Oudkerk, M.; De Bock, G.H.; De Koning, H.J.; Xie, X.; Van Ooijen, P.M.A.; Greuter, M.; De Jong, P.A.; Groen, H.J.M.; Vliegthart, R. Optimisation of volume-doubling time cutoff for fast-growing lung nodules in CT lung cancer screening reduces false-positive referrals. *Eur. Radiol.* **2013**, *23*, 1836–1845. [[CrossRef](#)]
24. Yang, J.; Chen, J.; Kuang, K.; Lin, T.; He, J.; Ni, B. MIA-Prognosis: A Deep Learning Framework to Predict Therapy Response. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020.
25. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2017**, arXiv:1711.05101.
26. Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv* **2013**, arXiv:1312.6034. [[CrossRef](#)]
27. Park, S.; Lee, S.M.; Kim, S.; Lee, J.-G.; Choi, S.; Do, K.-H.; Seo, J.B. Volume Doubling Times of Lung Adenocarcinomas: Correlation with Predominant Histologic Subtypes and Prognosis. *Radiology* **2020**, *295*, 703–712. [[CrossRef](#)] [[PubMed](#)]
28. Ardila, D.; Kiraly, A.P.; Bharadwaj, S.; Choi, B.; Reicher, J.J.; Peng, L.; Tse, D.; Etemadi, M.; Ye, W.; Corrado, G.; et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **2019**, *25*, 954–961. [[CrossRef](#)]
29. Trebeschi, S.; Bodalal, Z.; Boellaard, T.N.; Bucho, T.M.T.; Drago, S.G.; Kurilova, I.; Calin-Vainak, A.M.; Pizzi, A.D.; Muller, M.; Hummelink, K.; et al. Prognostic Value of Deep Learning-Mediated Treatment Monitoring in Lung Cancer Patients Receiving Immunotherapy. *Front. Oncol.* **2021**, *11*, 609054. [[CrossRef](#)]
30. Choi, Y.; Ahn, K.-J.; Nam, Y.; Jang, J.; Shin, N.-Y.; Choi, H.S.; Jung, S.-L.; Kim, B.-S. Analysis of heterogeneity of peritumoral T2 hyperintensity in patients with pretreatment glioblastoma: Prognostic value of MRI-based radiomics. *Eur. J. Radiol.* **2019**, *120*, 108642. [[CrossRef](#)]
31. Beig, N.; Khorrani, M.; Alilou, M.; Prasanna, P.; Braman, N.; Orooji, M.; Rakshit, S.; Bera, K.; Rajiah, P.; Ginsberg, J.; et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. *Radiology* **2019**, *290*, 783–792. [[CrossRef](#)]
32. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2020**, *35*, 11106–11115.
33. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
34. Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *arXiv* **2015**, arXiv:1511.07289. [[CrossRef](#)]
35. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer Normalization. *arXiv* **2016**, arXiv:1607.06450.
36. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.