Nodule Generation of Lung CT Images using a 3D Convolutional LSTM Network

Kolawole Olulana, Pius Owolawi, Chunling Tu and Bolanle Abe

Nodule Generation of Lung CT Images using a 3D Convolutional LSTM Network

Kolawole Olulana[#1][0000-0001-9210-0989] Pius Owolawi[#2,] Chunling Tu[#3], and Bolanle Abe[*]

[#] Computer Systems Engineering Department, Tshwane University of Technology, Pretoria. South Africa
[1]kolawolegolulana@gmail.com
[2]owolawipa@tut.ac.za
[3]duc@tut.ac.za
[*] Electrical Engineering Department, Tshwane University of Technology, Pretoria
abebt@tut.ac.za

Abstract. In the US, the American Cancer Society report for 2020 estimates about 228,820 new cases which could result in 135,720 deaths which translates to 371 deaths per day compared to the overall daily cancer death of 1660. The Cancer Society of South Africa, (CANSA) reports that lung cancer and other chronic lung diseases are leading causes of death nationally. Research in this area is necessary in order to reduce the number of reported deaths through early detection and diagnosis. A number of studies have been done using datasets for Computed Tomography (CT) images in the diagnosis and prognosis by oncologists, radiologists and medical professionals in the healthcare sector and a number of machine learning methods are being developed using conventional neural networks (CNN) for feature extraction and binary classification with just a few researches making use of combined(hybrid) methods that have shown the capability to increase performance and accuracy in prediction and detection of early stage onset of lung cancer. In this paper, a combined model is proposed using 3D images as input to a combination of a CNN and long short-term memory (LSTM) network which is a type of recurrent neural network (RNN). The hybridization which often lead to increase need for computational resources will be adjusted by improving the nodule generation to focus only on the search space around the lung nodules, this proposed model requires less computation resources, avoiding the need to adding the whole 3D CT image into the network, therefore only the region of interest near candidate regions with nodules will be pre-processed. The results of previous traditional CNN architecture is compared to this combined 3D Convolutional LSTM for nodule generation. In the experiments, the proposed hybrid model overperforms the traditional CNN architecture which shows how much improvement a hybridization of suitable models can contribute to lung cancer research

Keywords: ConvLSTM, CNN, deep learning, hybrid network, lung cancer diagnosis, LSTM.

1   Introduction

According to the World Health Organization (WHO), amongst other types of cancer, lung cancer is responsible for 1.76 million deaths alone in 2018, the highest among any type of cancer. Early detection of lung cancer can significantly increase survival rate, many cases can be cured by early detection through treatment [1].

The screening of patients plays an important role in improving survival rate where improvement of current machine learning methods can also help in reducing cost, expenses, and help advance current methods used in analyzing CT images compared to present techniques where 2D images with limited information and lower accuracy in detection methods are being used for monitoring the growth of lung nodules [2], using a hybrid model will lead to better results, hence the motivation for the development and deployment of a model using Computed Aided Diagnosis (CAD) services with increased performance and accessibility will go a long way in improving chances of survival of lung cancer. There exists a number of previous studies that have explored the technique of using low-dose CT images to develop 3D images to achieve accurate nodule detection [3,4] using conventional neural networks alone, however, in this study, a hybrid network will be employed to increase prediction accuracy rate and early detection. Due to the nature of the convolutional long term short term memory (ConvLSTM) network which requires a larger 3D input compared to 2D traditional conventional neural network, the pre-processing step will take into consideration the large search space constraint which can lead to a strain on   limited available computation resources, so only the low-level features are selected around the nodules of interest within that search space to generate the image voxels, this would be an important step to reduce processing time and size of input of the network [5]. Use of this low-dose lung image scans make it easier to design the hybrid model with a higher performance than traditional convolutional network as seen in a previous studies where a similar hybrid network was used in weather forecasting, applying this deep learning method would bring more accurate diagnosis and efficiency for medical professionals to overcome the weaknesses seen in other models, it is a better model and more details about the spatial temporal characteristic of the model lead to better performance [6].

In literatures, where hybrid networks have been used in other areas of study [7], there is an observed improved performance, compared to using recurrent neural networks, LSTM models and traditional conventional neutral networks being used alone, hybrid approach can lead to better results in accuracy and early detection [8] .

The approach taken in hybrid networks takes advantage of the feature extraction characteristics and spatial temporal characteristics of recurrent neural networks and both combined networks.

In this paper, these layers are collectively referred to as Convolutional LSTM (ConvLSTM) to help distinguish the naming conventions in other architectures, all architectures in this layer have a minimum dimensionality of three, and this would account for samples taken over time in the LSTM aspect of the hybrid network; it can be increased from 3D to 4D and up to 5D if sequence time_steps are needed.

This model will be tested on the LUNA16 dataset where a region of interest within the search space will be selected based on the threshold of Hounsfield radiodensity units [9] of greater than 604(-400 HU).

2   Method

The proposed 3D ConvLSTM hybrid model for nodule candidate generation is as described in Fig.1. The inputs and previous hidden states are convolved to produce 3D tensors that flow through each cell. Changes to standard fully connected (LSTM-FC) are highlighted in red. This method covers the nodule generation step and it's accuracy and other performance metrics. The predicted mask is what would be needed in a classification task when pre-processing and nodule generation is completed.

The entire 3D search space is not of much consequence as interest only lies in the surrounding area of each nodule which is scaled and padded and resized from the original dimension of $(512 \times 512 \times 512)$ volume to $(64 \times 64 \times 64)$; to decrease the GPU memory usage, this adjustment is done to the input batch which is added to the nature in a grid pattern, the output batch is compared with an actual batch for the validation and testing set as discussed further below. The mathematical representation of the model is defined by [10] is as follows

$$\begin{matrix} \tilde{\mathbf{f}}^{(\tau)} \\ \tilde{\mathbf{i}}^{(\tau)} \\ \widetilde{\mathbf{o}}^{(\tau)} \end{matrix} = \mathbf{BN}\big(W_h * x^{(\tau)}; \ \boldsymbol{\gamma_x}\,\beta_{hx}\big) + \boldsymbol{W_{peep}} \odot \boldsymbol{C}^{(\tau-1)} + \boldsymbol{b} + \mathbf{BN}\big(W_h * \boldsymbol{h}^{(\tau-1)}; \ \boldsymbol{\gamma_h}\,\beta_h\big)$$

where

$$\widehat{\mathsf{C}}^{(\tau)} = \tan^{-1}(BN(W_{hc} * \boldsymbol{h}^{(\tau-1)}; \boldsymbol{\gamma_h}, \boldsymbol{\beta_h}) + BN(W_{xc} * \boldsymbol{x}^{(\tau)}; \boldsymbol{\gamma_x}, \boldsymbol{\beta_x}) + \boldsymbol{b_c})$$

$$\mathsf{C} = \sigma(\tilde{\mathbf{f}}^{(\tau)}) \odot \mathsf{C}^{(\tau-1)} + \sigma(\tilde{\mathbf{i}}^{(\tau)}) \odot \widehat{\mathsf{C}}^{(\tau-1)}$$

$$\boldsymbol{h}^{(\tau)} = \sigma(\quad) \odot \tan^{-1}(BN(\boldsymbol{C}^{(\tau)}; \boldsymbol{\gamma_c}, \boldsymbol{\beta_c})) \tag{1}$$
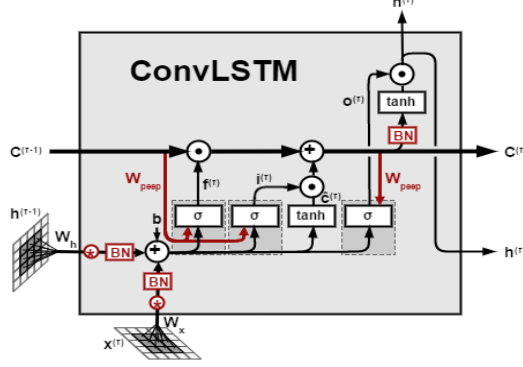
Fig. 1. ConvLSTM Cell.

2.1   Network Structure

In the LSTM cell, W_h is the shared weight matrix for the hidden-to-hidden transitions at time step τ, W_x is the shared weights for the input-to-hidden connections, as well as W_{peep} is the shared weights matrix for the peephole connections. The b is the bias, as well as C(0), h(0) are the initial states of the memory cell and the hidden state, respectively. Furthermore, one of the batch-normalization (BN) layers with its learned shift; γ and scale; β, are denoted by (x; γ, β) where the bias terms are denoted by b as did in [10].

The network's architecture is built using a recurrent neural network approach that passes the input grid images through a previous hidden state C(t−1) as a step to the next sequence, such that previous data the network has learnt is used in making decisions.

In the cases where time steps are implemented, the patient's previous CT images are compared. These progressions are used to keep true positives as accurate as possible.

The rest of the transition is shown in Table 1.

Table 1. The architecture of a ConvLSTM network

Layers	Params	Kernel Size
batch_normalization_1	4	None,None,None
conv3D_1 (Conv3D)	208	1,5,5
conv3D_2 (Conv3D)	1736	3,3,3
batch_normalization_2	32	None,None,None
bidirectional_1	27776	3,3
bidirectional_2	147712	3,3
conv3D_3	4616	1,3,3
conv3D_4	9	1,1,1
cropping3D_1 (Cropping3D)	0	1,2,2

2.2 Applying the Proposed Method in 3D Image Classification

As shown in Fig. 2 and discussed in [16,19], the input samples (spatial-temporal data) is fed into the algorithm network to learn how to follow a sequence of data like a physician reading through a number of images meaning that the order is important and takes notes of already noticeable patterns to make better decisions. Similar studies in video prediction and weather forecasting have been successfully modeled with high accuracy [15].
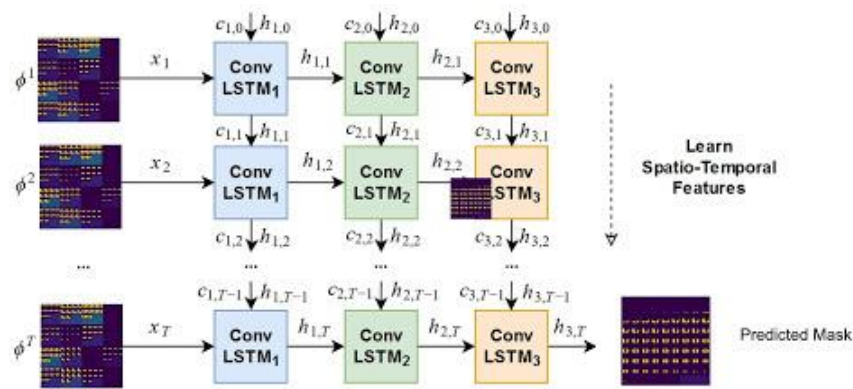

Fig. 2. Flow Diagram of Proposed ConvLSTM.

3 Experimental Results

The proposed model was evaluated against the publicly available LUNA16 dataset and compared with traditional conventional methods, specifically the DeepSeed approach in [3] and DL algorithm in [21].

3.1 Dataset

The LUNA16 dataset contains 226,589 test CT images from 888 studies where than 2.5 mm in a raw format. The annotation was modified accordingly as detailed in [4]. The pre-processing steps are done in preparation of the model for a 3D conventional neural network coupled with a LSTM model. Nodules are defined as non-nodule, nodule if diameter is less than 3 mm, and nodules are otherwise classified as nodules if greater than 3 mm. See this publication for the details of the annotation process.

The pre-processing steps included resizing and selecting lesion being classified as a nodule as defined by a consensus of at least three of the four radiologists involved in

classification of the image sets. This dataset was selected because of its 3D images which is the only type of input the hybrid layer will accept

3.2 Implementation Details

The model is compiled using an adam optimizer and binary cross entropy, then runs for 100 steps per epoch to train the parameters; it is trained in twenty epochs [12]. The model was trained on Kaggle using a Nvidia K80 GPU.

The training and validation sets are split with the ratio 80:20. Each input image has a binary segmentation mask to segment lungs. The image size is 64x64 with an input batch of sixteen with a slice count of ten as seen in Fig. 3, for the validation set the slice count was increased to 100 and batch size was reduced to just one (1) as detailed in Fig. 4.

In this study, the 3D network structure emphasized on the memory features of RNN, the memory, M in LSTM to develop a neutral by augmenting the 3D input images in the pre-processing stages to select only the regions of interest by generating cutting 3D voxels around this regions as indicated by the image dataset. Yu Gu et al. [11] have established a dot enhancement filter method to generate 3D matrix of voxel data which was annotated in LUNA16 dataset where the radius of each nodule is used to generate the binary mask.

Once the 3D voxels are generated, each mask is fed into the network input layer, where it goes through the layers starting with the batch normalization (BN) layer where the dimensionality of the input is set to None when BN is the first layer. This layer is followed by the first conventional 3D (conv3D_1) layer; four convolution layers are employed with the same activation and padding but different kernel sizes. It is followed by the second of four conventional layers, conv3D_2 with a RELU activation [17,18]. The first LSTM bidirectional layer (bidirectional_1) follows the conv3D_2 to account for the recurrent layer for the sequence classification layer as shown in Fig.1, as labeled as h(t+1).

At the end of 20 epochs, the binary accuracy converged to 0.9317 and average binary accuracy of 0.9866, average mean square error of 0.0087and average loss of 0.0709
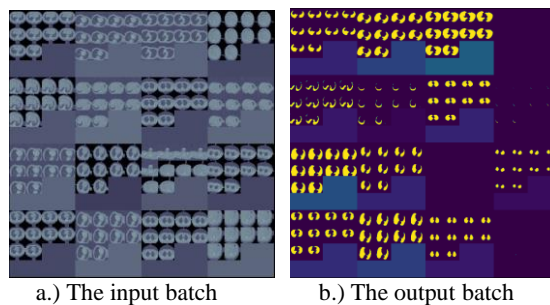

a.) The input batch   b.) The output batch

Fig.3. Augmentation of 3D images into little chunks.

The output of the prediction step was however output in a single batch due to the small nature of the 3D masks provided by the dataset, it is a resource intensive process to expand the dataset 2D images into 3D images [20]. More work still needs to be done in this area.
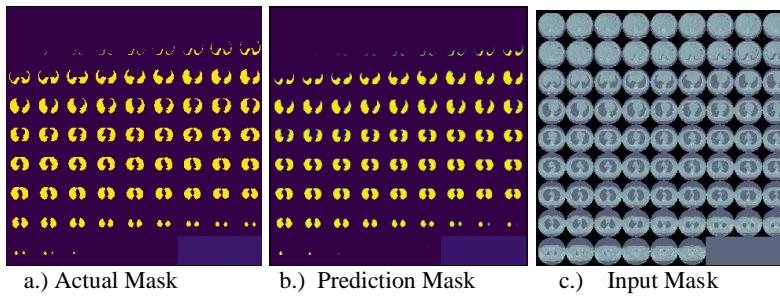

|a.) Actual Mask | b.) Prediction Mask | c.) Input Mask |

Fig.4. The Input mask validation mask and actual mask

3.3 Performance Comparison top Traditional CNNs

In comparison to 3D DeepSeed [3] and DL algorithm [21], there is a noticeable improved accuracy based on the LUNA16 dataset as shown in Table 3. The ConvLSTM accuracy peaked at 0.972 and its loss converged around 0.0709. It outperformed the DeepSeed and DL Algorithm approach

Table 2. Comparison of Metrics.

Network	Precision	Recall	Accuracy
ConvLSTM (ours)	0.9820	0.948	0.972
3D CNN Deep Seed	-	0.862	0.862
DL Algorithm	0.82	0.82	0.90

4 Conclusion

In this study, the focus was on improving the accuracy and early detection of the progression of lung cancer progress through the use of hybrid models are able in generation of nodules from low dose lung CT scan images for pre-processing step using the proposed ConvLSTM that combines the capabilities of a recurrent neural network using LSTM network coupled with a conventional CNNs. The expected results of increased accuracies were achieved on the LUNA 16 dataset which consisted of low-dose CT scan images by generating 3D image voxels around the region-of-interest by feeding them into the network.

A 10% improvement on the accuracy was observed compared to the DeepSeed approach.

5   References

1. Rebecca L. Siegel, Kimberly D. Miller, Ahmedin Jemal. Cancer statistics. pp 8 - (2019).
2. Nigel Hawkes. Cancer survival data emphasis importance of early diagnosis. pp 364: l408 (2019)
3. Yuemeng Li, Yong Fan. DeepSeed: 3D Squeeze-and-Excitation Encoder-Decoder Convolutional Neural Networks for Pulmonary Nodule Detection, pp 2-4, (2019)
4. Wentao Zhu,Chaochun Liu, Wei Fan, Xiaohui Xie .Deeplung: Deep 3D dual path nets for automated pulmonary nodule detection and classification. pp. 2–7 (2018)
5. Wei Li, Peng Cao, Dazhe Zhao, Junbo Wang: Pulmonary Nodule Classification with Deep Convolutional Neural Networks on Computed Tomography Images, pp 1-3 (2016)
6. Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, Wang-chun Woo: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, pp 2-8. (2015)
7. Borja Seijo-Pardo, Iago Porto-Díaz, Verónica Bolón-Canedo, Amparo Alonso-Betanzos. Ensemble feature selection: Homogeneous and heterogeneous approaches. pp 6 - 27, (2017).
8. Riqiang Gao, Yuankai Huo, Shunxing Bao, Yucheng Tang, Sanja L. Antic, Emily S. Epstein, Aneri B. Balar, Steve Deppen, Alexis B. Paulson, Kim L. Sandler, Pierre P. Massion, Bennett A. Landman: Distanced LSTM: Time-Distanced Gates in Long Short-Term Memory Models for Lung Cancer Detection, pp 2 -5. (2019)
9. Kingsley Kuan, Mathieu Ravaut, Gaurav Manek, Huiling Chen, Jie Lin, Babar Nazir, Cen Chen, Tse Chiang Howe, Zeng Zeng, Vijay Chandrasekhar. Deep Learning for Lung Cancer Detection: Tackling the Kaggle Data Science Bowl 2017 Challenge. Pp.2-5 (2017)
10. Benjamin Sautermeister. Deep Learning Approaches to Predict Future Frames in Videos. Pp 45 - 46 (2016)
11. Gu, Y., Lu, X., Zhang, B., Zhao, Y., Yu, D., Gao, L., Cui, G., Wu, L., & Zhou, T. Automatic lung nodule detection using multi-scale dot nodule-enhancement filter and weighted support vector machines in chest computed tomography.pp 2- 5 (2019)
12. Onur Ozdemir, Rebecca L. Russell, d Andrew A. Berlin. A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans. Pp 6 - 11. 2020
13. Wei Li, Peng Cao, Dazhe Zhao, Junbo Wang: Pulmonary Nodule Classification with Deep Convolutional Neural Networks on Computed Tomography Images, pp 1-3 (2016)
14. Wen-Shuai Hu, Feature Extraction and Classification Based on Spatial-Spectral ConvLSTM Neural Network for Hyperspectral Images
15. S. Oprea, P. Martinez-Gonzalez, A. Garcia-Garcia, J.A. Castro-Vargas, S. Orts-Escolano, J. Garcia-Rodriguez, A. Argyros. A Review on Deep Learning Techniques for Video Prediction. Pp 9-14 (2020).
16. Brayan S, Zapata-Impata, Pablo Gil, Fernando Torres. Learning Spatio Temporal Tactile Features with a ConvLSTM for the Direction of Slip Detection. pp 6-7 (2019).
17. Yan Li, Majid Sarvi,Kourosh Khoshelham, Milad Haghani. Predicting traffic congestion maps using convolutional long short-term memory. Pp 4-6 (2019).
18. SyedAshiqur Rahman & DonaldA.Adjeroh. Deep Learning using Convolutional LSTM estimates Biological Age from Physical Activity. Pp 10 (2019).
19. Yang Li, Huahu Xu, Minjie Bian, Junsheng Xiao. Attention Based CNN-ConvLSTM for Pedestrian Attribute Recognition. Pp 4-5. (2020)
20. Shuo Feng, Huiyu Zhou, Hongbiao Dong. Using deep neural network with small dataset to predict material defects. Pp 302 (2019).

21. Sijia Cui, Shuai Ming, Yi Lin, Fanghong Chen, Qiang Shen, Hui Li, Gen Chen,Xiangyang Gong,Haochu Wang1. Development and clinical application of deep learning model for lung nodules screening on CT images. Pp 4 (2020).