CS109B/AC209B Final Project
Pavlos Protopapas, Mark Glickman

Recurrent Models for Detecting Cancer in 3D CT Volumes

May 2, 2018

Prepared by Group 29:
Luran He, lhe@college.harvard.edu
Alexander Lin, alexanderlin@college.harvard.edu
Amanda Zhang, amandazhang@college.harvard.edu

1 Problem Statement and Motivation

Problem Statement

Lung cancer is one of the leading causes of cancer-related deaths in the United States [2]. High resolution Computed Tomography (CT) scans are commonly used by radiologists to detect lung cancer nodules, but visual detection of lung nodules remains difficult and inefficient. In recent years, much attention has been directed to developing computer-aided diagnostic (CAD) methods for nodule segmentation, detection, classification, and quantitative assessment [3].

Our project aims improve nodule detection through deep learning methods. We chose to train our model on data from The Lung Image Database Consortium (LIDC), the largest public CT image database of lung nodules, for its size and extensive annotations. These annotations identify the location and radiological characteristics of the lesions and certain lung abnormalities [1]. Ultimately, by improving automated methods to detect lung nodules, we hope to contribute to the ongoing efforts to reduce medical screening costs and aid radiologists. In recent years, there have been reports of a looming shortage of radiologists not only in the UK [8], but in general [4]. An additional hope is that the development of sufficient CAD models can decrease the workload on the decreasing radiologist population and perhaps even help incentivize specialization in radiology.

Goals

The goal of this project was to build a CAD tool for pixel-level nodule detection by combining a fully convolutional network (FCN) for intraslice feature extraction with a recurrent neural network (RNN) for interslice information. To work up to this goal, we outlined three possible results we could achieve for a U-Net:

- Detect whether a slice has nodules.
- Detect whether each pixel is part of a nodule.

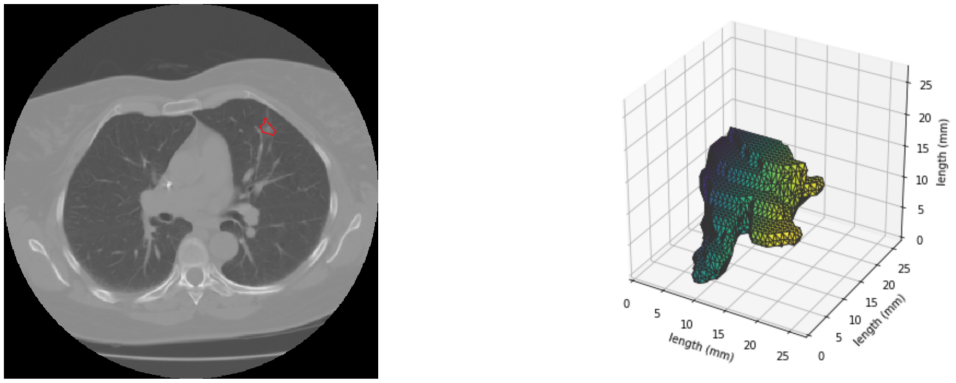
- Detect at each pixel both whether it is part of a nodule and the malignancy if it is part of a nodule.

From there, our main goal would be to build an RNN structure that improves U-Net predictions.

2 Description of the Data

As the largest public database for lung CT scans, LDIC is a valuable tool for the development, training, and evaluation of CAD systems for lung nodule detection. The data is 120 GB of primarily DICOM image files. There are 1018 scans from 1010 patients, each including a XML document with metadata and a few hundred or so DICOM (512x512) files representing horizontal (axial) cross-sections of the chest cavity. The metadata gives the positions of the annotated nodules, most of which are annotated several times over by multiple radiologists. The scans are classified as containing nodules $\geq$3 mm, nodules $<$3 mm, and non-nodules $\geq$3 mm. For each nodule $\geq$3 mm, radiologists also labeled its malignancy (unknown, benign, malignant-lung, malignant-metastatic).

Figure 1 shows an example annotated 2D slice nodule and its 3D representation. We hope to be able to use a RNN to capture information about the 3D structure from a sequence of 2D slices.



(a) 2D slice

(b) 3D representation

Figure 1: Visualization of example nodule from Patient 1.

Figure 2 shows a crop of a nodule (bounding box) and its ground truth segmentation, both obtained from radiologist annotations. As can be seen, the radiologist segmentation is not a perfect outline of the nodule.

Because of the intractably large size of the dataset, we downsized the image slices from 512x512 pixels to a more manageable 256x256 dimension size. In the axial plane, we also cropped out the chest cavity region (50 slices) through visual inspection to standardize the input to the LSTM.

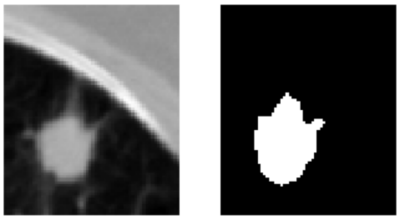


Figure 2: Crop of nodule and ground truth segmentation.

Exploratory Data Analysis

We have data on 1010 patients and 7371 nodules. 982 of the patients had nodules (one extreme individual had 119 nodules).

Figure 3 provides some information on the nodule sizes. Based off the similar distributions of nodule volume and nodule diameter, it seems that nodules in this dataset tend to be of the similar spherical form.

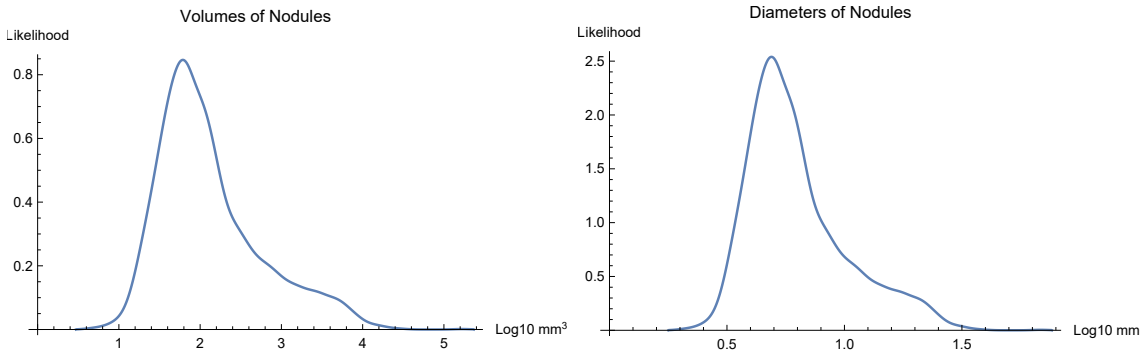


Figure 3: Distribution of nodule volumes and diameters

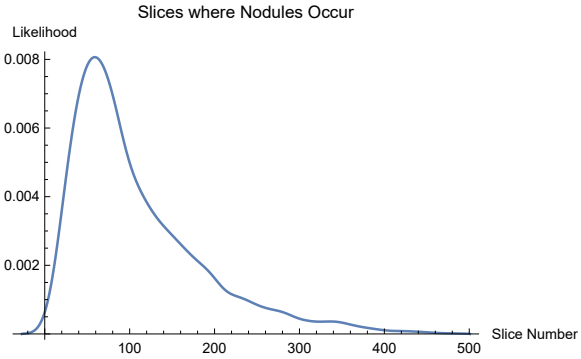
Figure 4 provides some information on the location in the body where the nodules are found. Most nodules are found in the beginning slices (slices 40-80). This reflects that the length of most CT scans are around 100-150 slices.

Figure 5 provides some information on the distribution of nodules per patient. In general, patients have a small number of nodules. Of patients who do have nodules, they tend to be small (in size) nodules.

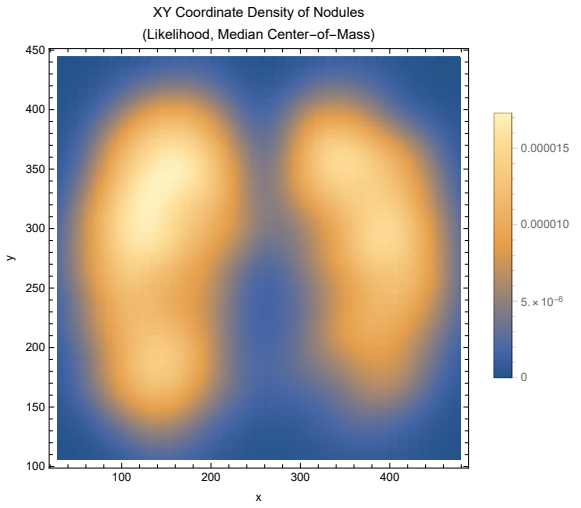
3 Literature Review

U-Net

Two of the goals of building a successful CAD is that there is high diagnosis accuracy and localization. The requirement of high diagnosis accuracy entails that the network can distinguish tumor from body in a grayscale image. Medical imaging is particularly hard for neural networks because of the grayscale format. There is less feature information in a monochrome image vs a normal color image. Coupled with low contrast, medical imaging



(a) Distribution of slices with nodules



(b) Nodule density

Figure 4: Information on location of nodules

proves to be a challenge, however a neural network architecture called the U-Net details with our requirements quite nicely. [5] The U-Net, originally designed for cell segmentation, takes advantage of pixel resolution data to localize. It's network is unique in that there are layers that upsample and downsample symmetrically. During the upsampling phase, the symmetrical downsample layer features are augmented to the input of the the symmetrical upsampling layer allowing for better localization. An additional advantage of this network is that it doesn't require a lot of training data to converge. The original authors of this paper did simple elastic deformations on the training samples.

We thought this network would be useful for our goals because of the useful properties of this network: localization and accuracy. It is not too far fetched to use a model originally for cell segmentation and instead use it for tumor detection. The only difficulty is that we have temporal data. A single patient is some number of slices. It would be naive to feed all slices of a person and treat them the same as every other slice. Instead we can improve on this idea and use a RNN U-Net. This is explored later.

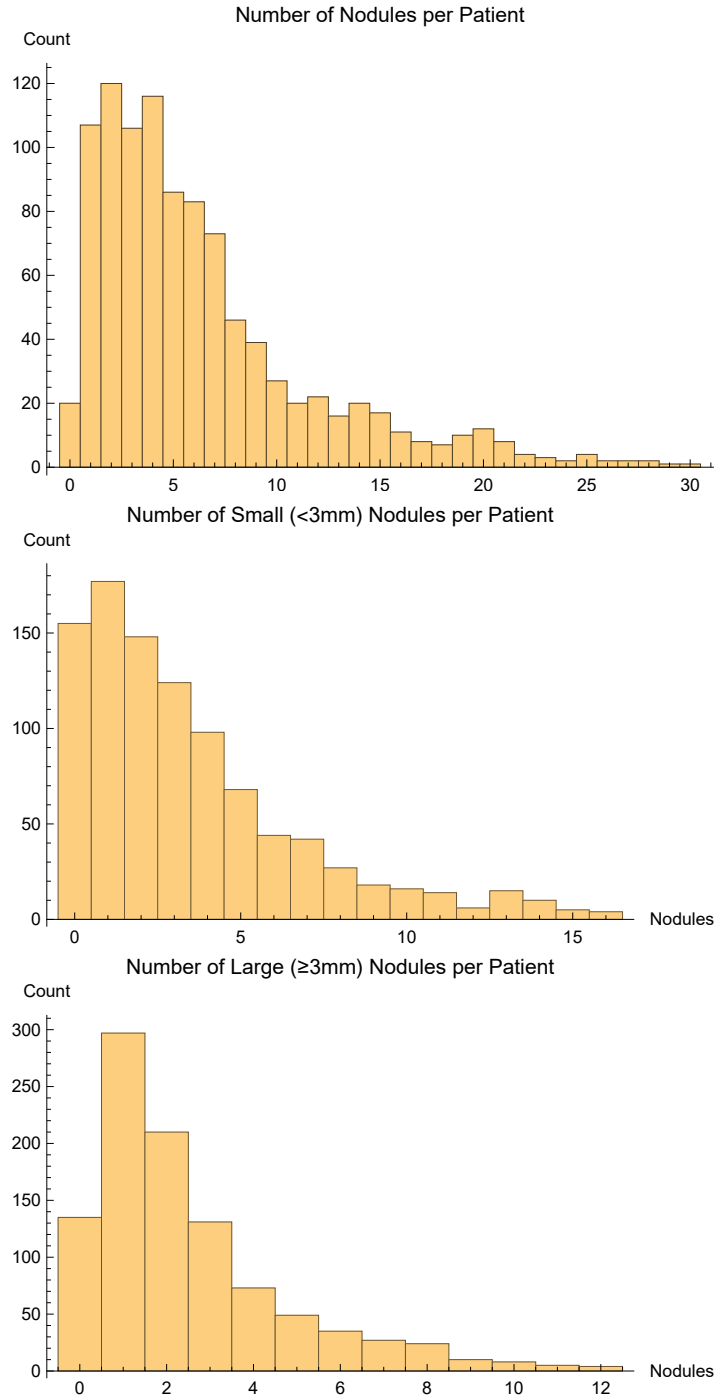


Figure 5: Nodules per patient (all nodules, small nodules, large nodules)

LSTM

LSTMs offer amazing potential for sequence to sequence learning; this paper looks at the success of one state-of-the-art model with language translation [7]. The methodology is

as follows: An encoder net and a decoder net work side-by-side. First the encoder is fed the entire input sequence. Then the decoder reads the internal state of the encoder and outputs the first element of the predicted sequence, before feeding the first element back in to produce the second element of the predicted sequence, and so forth. This methodology is very exciting because it's very cheap and does not involve too many parameters; for other neural net structures, such as dense layers, the input and output would gain an entire new dimension. The basic idea has tantalized researchers since the conception of the RNN, but it was not until the development of the LSTMs that long-term dependence became more effectively learnable, and that results such as this became possible.

We thought this idea would be useful because it would transform our 3-dimensional problem into a 2-dimensional sequence-learning problem. It makes sense to conceptualize the CT scan as a 2D sequence of image slices since there is different number of slices per scan. We would proceed exactly as this paper does, but instead of starting with 1-hot encoded input language sequences, we would start with sequences of feature vectors extracted by U-Net. Basically we would deal with the same setup, but with an extra U-Net in front of the encoder net.

4 Modeling Approach

We split the dataset by patients at a 75/25 ratio into train/validation sets. For simplicity, whenever at least one annotation marked a pixel as being part of a nodule, we counted it as being part of a nodule in the ground truth.

Baseline Model: U-Net

The baseline model for comparison is U-Net, which has repeatedly been shown to be a fast and accurate model for biomedical image segmentation [5]. Initially, we aimed to classify whether each slice had nodules or not. However, after much training, the U-Net was doing no better than random guessing. We then realized that U-Net was not the right choice for this classification because U-Net outputs a feature map, and we had been computing prediction by summing over that feature map. This was probably a surmountable problem, but since pixel-level detection of nodules was a more important task, we decided to switch over to that.

We initially trained the U-Net on all slices, but it did not train well due to the class imbalance (the nodules were very small relative to an entire chest cavity) so we trained only on slices that had nodules.

Measuring either training or validation performance for assessing over/underfitting is very difficult, given that accessing the data (on an Amazon S3 bucket) was an extremely time-consuming operation. We stuck to only training two epochs per random draw of a patient, and visual inspection of ROC curves on the validation set indicated that we did not perform too badly on validation (see Results).

From U-Net to RNN

We then fed U-Net feature map results on 50 consecutive slices into a Convolutional LSTM layer, a design first implemented for precipitation now-casting that has shown to work well on spatial temporal data [6]. Convolution LSTM layers have convolution transformations

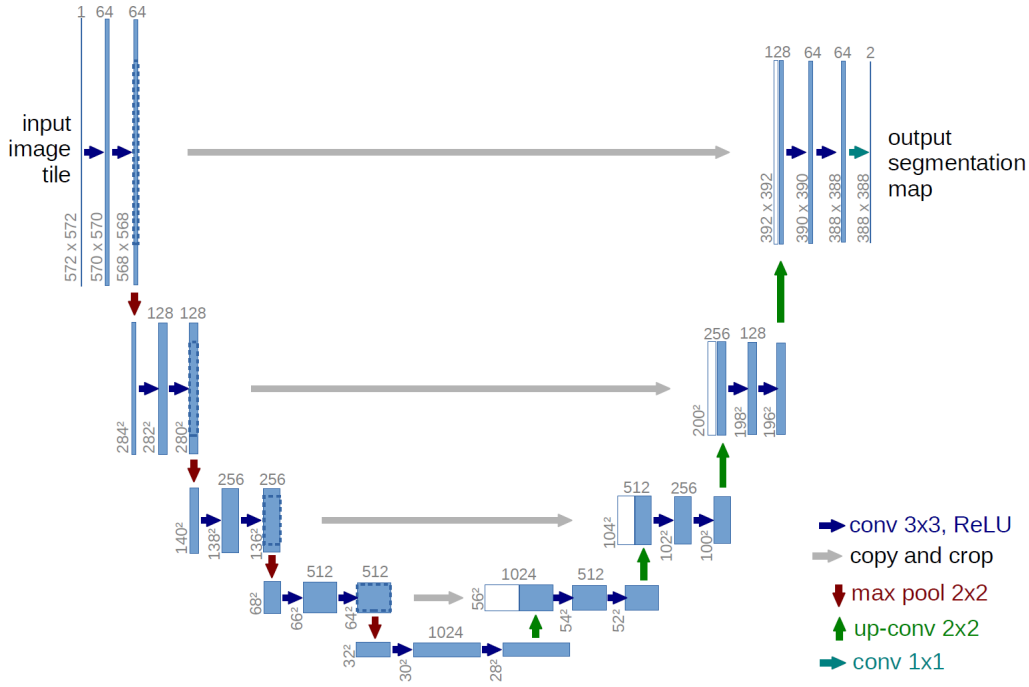


Figure 6: U-Net + RNN

at every time-step of the LSTM, an advantage which we hoped would infer localization information from slice to slice.

For one RNN model, we fed the 50 slices through once in one z direction (axial plane), and for another we fed it through in both z directions. To account for class imbalance, for every patient randomly drawn from the training set, if the patient had any nodules, we ensured that the sequence of slices would contain at least one scan with a nodule.

5 Project Trajectory, Results, Interpretation

Project Trajectory

Our original trajectory included slice level predictions obtained from the output of the LSTM using U-Net feature map inputs. However, we strayed from this idea after seeing U-Net perform so well on the pixel level. We focused instead on improving the pixel level definition with an Convolution LSTM but with lackluster results.

Due to computational difficulties, we were unable to create a combined model that would combine our U-Net outputs to our LSTM. This would have been beneficial since all errors would then propagate across the whole model instead of just through the U-Net and LSTM. This limitation was mainly due to the lack of GPU memory probably due to the high number of parameters we had in our model. Despite this, we reasoned that training two models separately and using the output of a model that does reasonably well as input for the second model would generate an overall pipeline that would yield decent predictions.

We also decided to test two LSTM models with different parameters. One LSTM had one

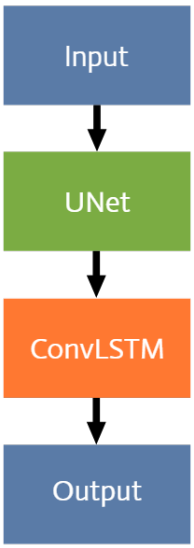


Figure 7: U-Net + RNN

Conv2dLSTM layer with a Dense output on every timestep, whereas another had, in addition, two layers as well as the option to feeding in data in order and reversed (both z directions). This was due to the results of applying a rudimentary LSTM model being unsatisfactory prompting us to more heavily tune the LSTM model in hopes of a performance boost.

Results

Figure 8 shows the ground truth (y-labels), feature map outputs from U-Net, and LSTM outputs for some example slices. While we succeeded in building a well-performing U-Net (see ROC curves), from visual inspection, the U-Net seemed unable to consistently detect nodules. For example, looking at row 1, the bright spot in the ground truth image is not captured by the U-Net nor the LSTM.

It was not feasible to compute results over the entire test dataset, so Figure 9 shows some ROC curves for random patients (false positives on the x-axis, true positives on the y-axis). As can be seen in the ROC curves, we succeeded in our first goal to reproduce U-Net and adapt it for detecting lung nodules in the LIDC dataset. However, the RNN models did quite poorly, usually much worse than U-Net. There are some cases (not shown here) where the LSTM models actually improved the U-Net predictions, however this was rare - the displayed ROC curves are more representative of the norm. From the ROC curves it is apparent that U-Net fares rather well on its own without the "help" of an additional LSTM model. We also observe that the LSTM model that included the feeding of patient data in both z directions ("LSTM for-bac") did better than other LSTM model with only one Conv2DLSTM layer.

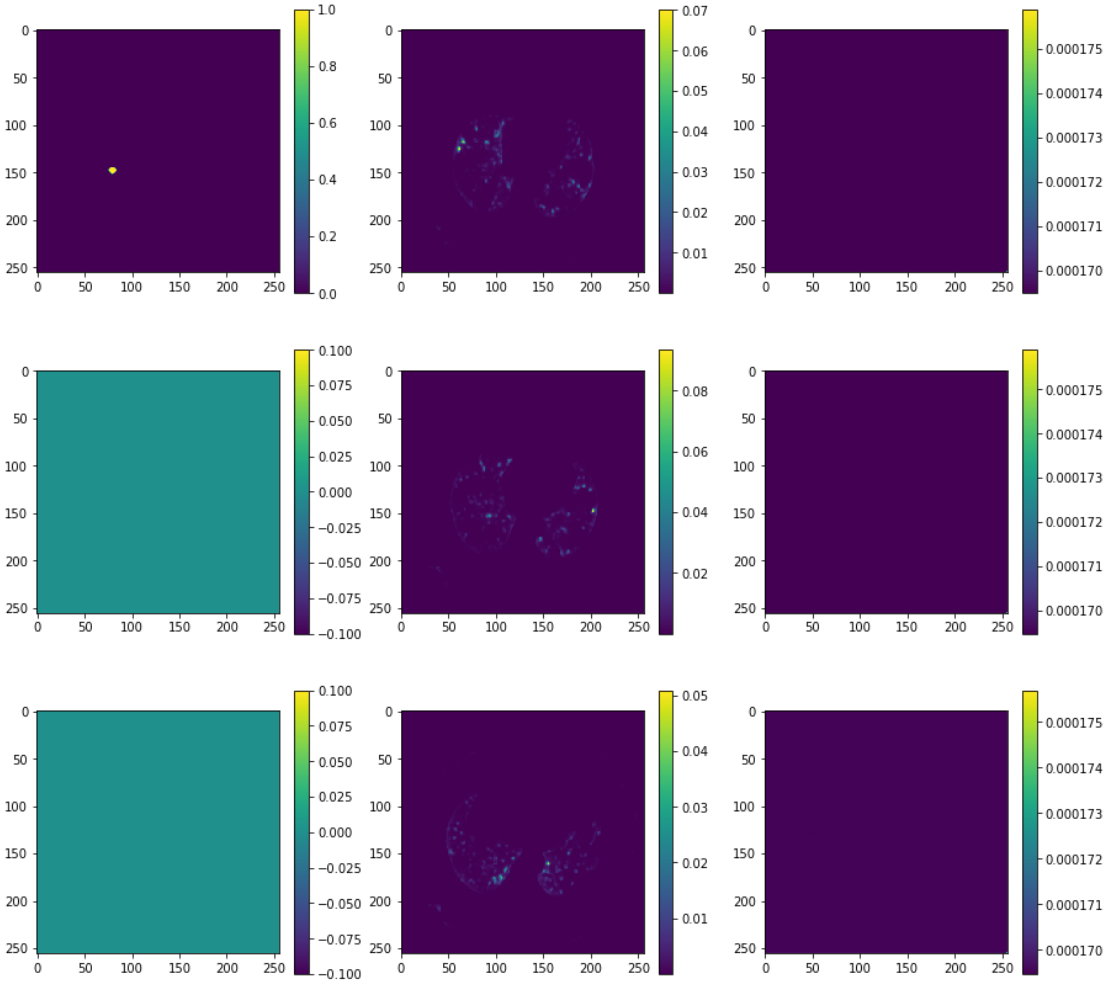


Figure 8: Ground truth, U-Net output, and LSTM output

Interpretation

We suspect that the RNNs did not train well because they were much worse at dealing with class imbalance than the U-Net. It is unfortunate that the RNN structure did not generalize U-Net to 3 dimensions as easily as we would have hoped. Instead what it seems like it did was recognize was that the ground truth is mostly uniform. The LSTM seems to have learned to "flatten" probabilities in low value pixels and raise high value pixels in order to match the ground truth uniformity. This is reflected in the uniform probabilities in the third column of Figure 8. Optimistically, it seems that the bi-directional LSTM model outperformed the single-directional LSTM model. We suspect that feeding in the data in both directions allows the LSTM learns a more accurate block-like structure of nodules rather than a temporal structure.

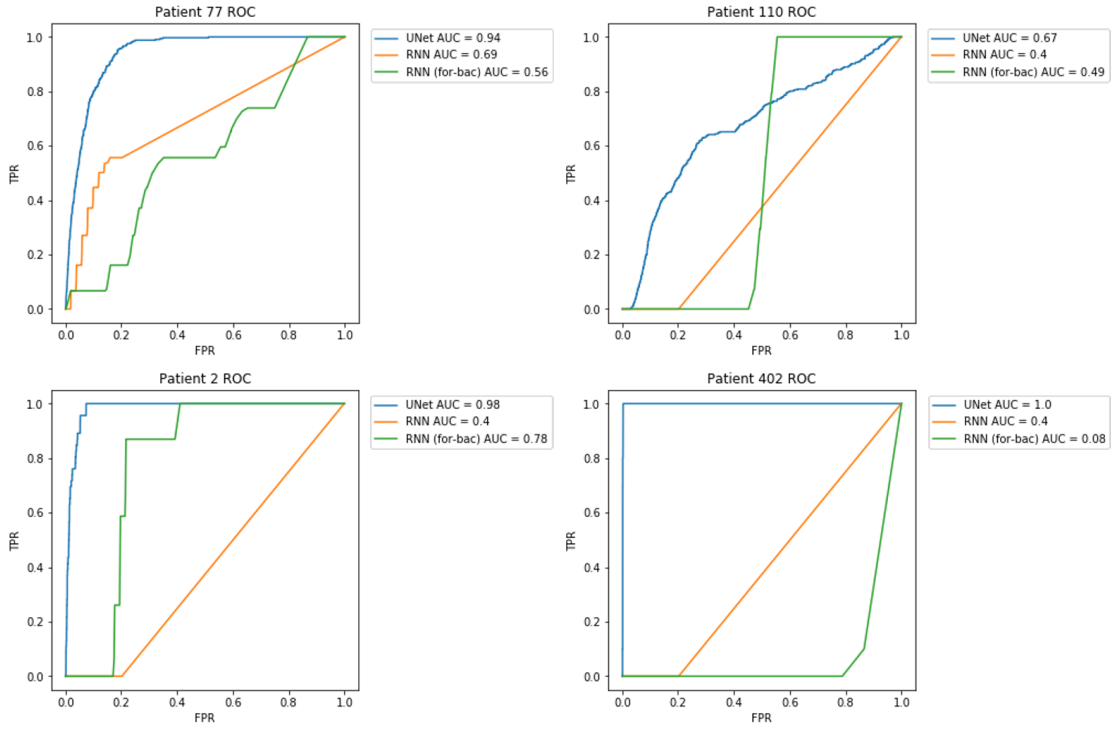


Figure 9: ROC curves for 4 randomly selected patients

6 Conclusions and Possible Future Work

Project Summary

In this project, we sought to build a CAD tool for pixel-level nodule detection. In particular, we hoped to improve on existing 2D models by combining a fully convolutional network (FCN) for intraslice feature extraction with a recurrent neural network (RNN) for interslice information. While we succeeded in reproducing good results in our U-Net model, our efforts to improve those predictions using an additional LSTM model failed - at least with our parameters. This was probably due to insufficient efforts to fully tackle the class imbalance problem which plagues prediction. While our predictions fell short of our goals, there were cases where our model pipeline improved the results of the U-Net output. This seems to suggest that the LSTM model was able to improve the predictions that the U-Net model was unsure about. This may hint at a promising future research; though more research is required, the results seem promising on this front.

Possible Future Work

Dealing with Class Imbalance

We learned that class imbalance is a significant challenge for medical image classification. To address this challenge in the future for RNN training, we can use image augmentation and undersampling to procure a dataset that has relatively balanced classes. This plays very

heavily on our aim to control the precision and recall of our final model. In the context of this project, false positives are preferred to false negatives for several obvious reasons; fixing the class imbalance will help us with this step.

Additionally we could switch to Tversky loss (with Dice coefficient) in order to further tune our specificity and recall rates. This would be more advantageous since accuracy as a metric here means very because of the large class imbalance.

Improving Training

A way to decrease the input size of our model, thus cutting the number of parameters, while also dealing with the class imbalance problem, would be to train on smaller windows of slices rather than entire 256x256 slices. In this design, slice segments would be the inputs to the U-Net and LSTM. The smaller the window, the less of a problem of class imbalance, however we would lose the U-Net's ability to globally localize since segments are so granular. This would be a balancing game, which could be aided by an LSTM. Feeding in the segments in order into the LSTM may recover the spacial information lost by the preprocessing.

While our LSTM did not improve upon the predictions provided by this setup, a promising idea would be to give the LSTM more input than just the final U-Net feature map. The final U-Net feature map output in Figure 8 is very sparse. We see that there are a few values that are non-zero. It isn't hard to imagine why our LSTM didn't do that well since the input we are passing still has the class imbalance problem. Instead, we could have taken the output of second to last layer of our U-Net intermediate layers as input for our LSTM (with some data processing). This might allow the LSTM to train better since it now has more granularity and information about the kinds of output the U-Net is predicting, and thus might be able to improve on it.

References

[1] LIDC-IDRI.

[2] American Cancer Society. Key statistics for lung cancer.

[3] El-Baz, A., Beache, G. M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., Soliman, A., and Abdollahi, B. Computer-aided diagnosis systems for lung cancer: challenges and methodologies. *International journal of biomedical imaging 2013* (2013).

[4] Kaplan, D. A. Is radiology part of the physician shortage?, Aug 2016.

[5] Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241.

[6] Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., and Woo, W. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *CoRR abs/1506.04214* (2015).

[7] Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (2014), pp. 3104–3112.

[8] Thakar, S. Uk experiencing 'desperate' shortage of radiologists, Oct 2017.