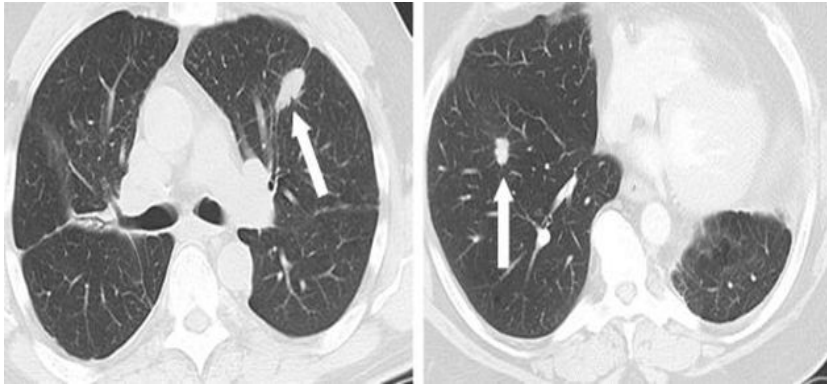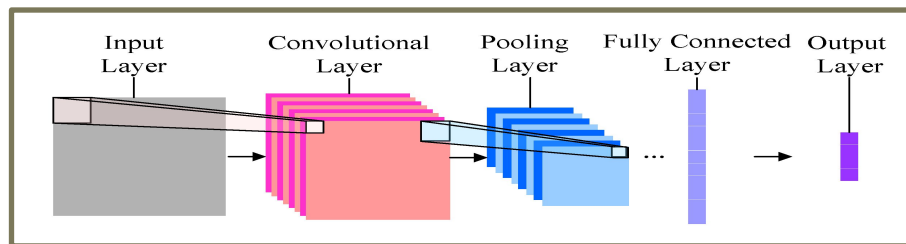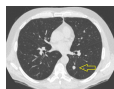# Lung Cancer Detection

By : Dhanachandra(1611CS10)
Rimjhim(1621CS17)

Outline

- Introduction

- Network architecture

- Experimental Results

- Observations

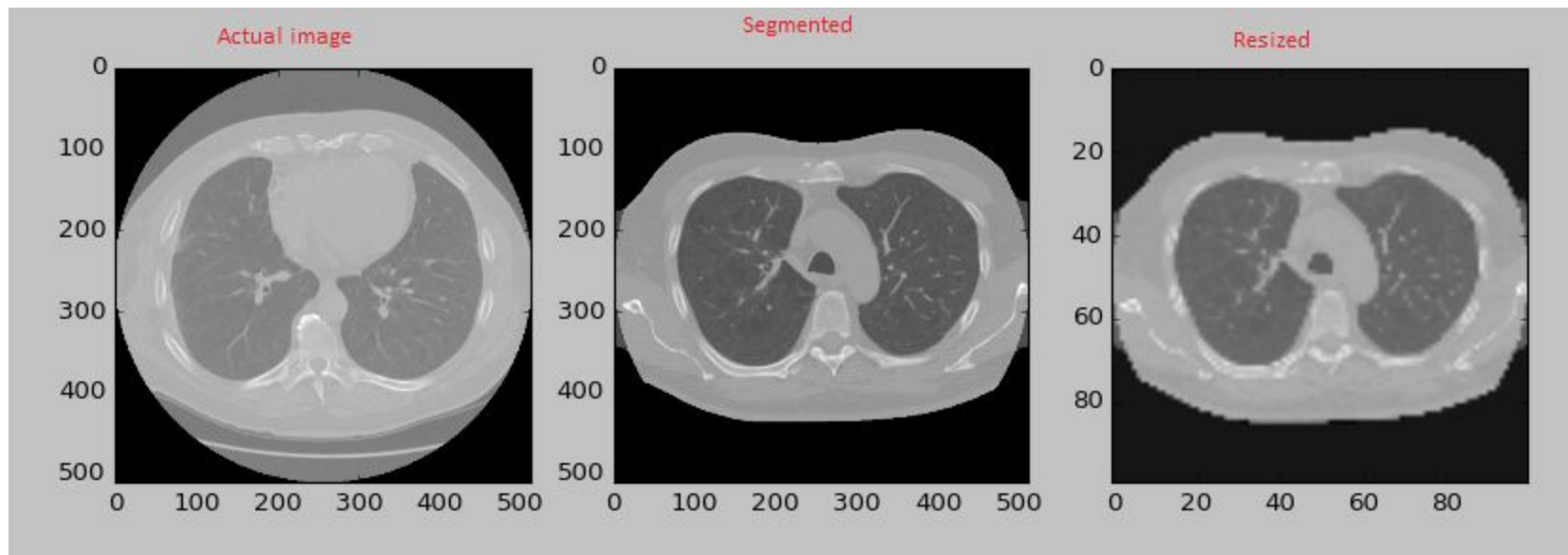- Conclusion

Problem Statement

- Given a CT scan image of lung, the system has to predict whether it is cancerous or not.
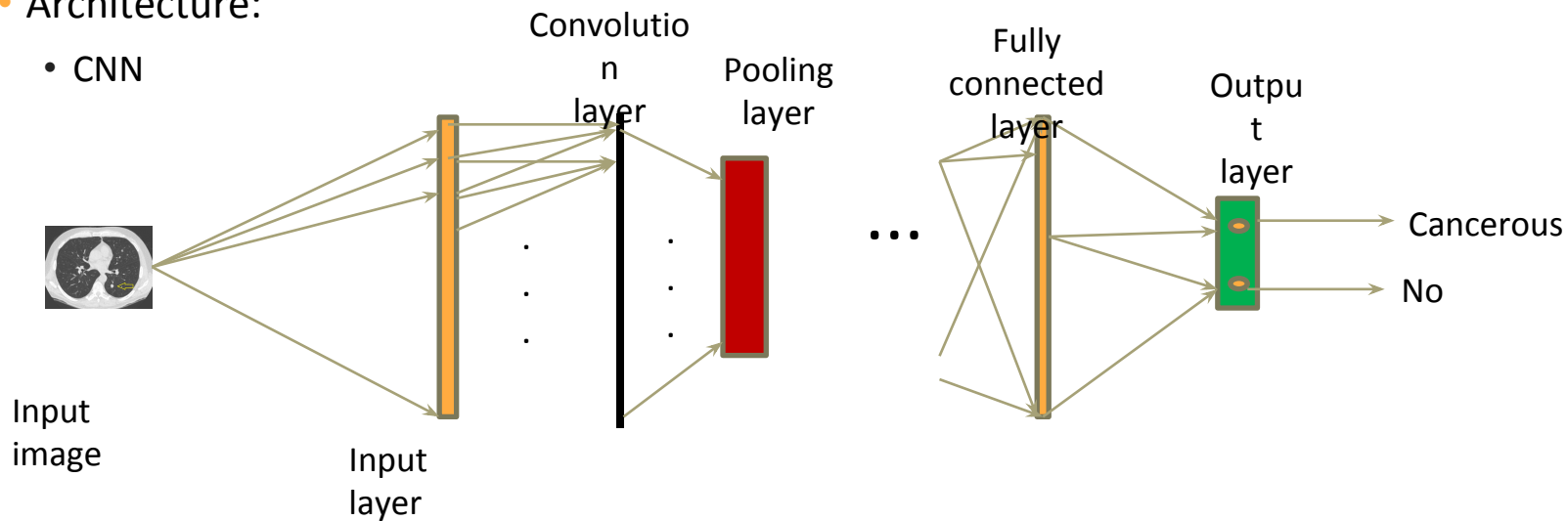
Cancer/Non-Cancer

Preprocessing

Approach

- Preprocessing:
  - Segmentation
  - Resize the image
- Architecture:
  - CNN

Experiments

Model 1: 4 2-D convolutional layers + 2 Fully connected layers + output layer

Model 2: 2 2-D convolutional layers + 1 Fully connected layers + output layer

S.No.	Experiments	Model Number	Dataset size	Number of images	Cancer images in test data	Non cancer images in test data	Accuracy
1	Exp1	1	2gb	3408	192	486	71.26%
2	Exp2	1	8gb	14245	594	2068	80.43%
3	Exp3	2	10gb	19200	704	2894	98.61%
4	Exp4	2	16gb	30407	1224	4406	96.13%
5	Exp3	1	10gb	19200	704	2894	97.72%
6	Exp4	1	16gb	30407	1224	4406	93.00%

Exp1. Results

Accuracy: 71%

	precision	recall	f1-score	support
CANCER	0.00	0.00	0.00	196
NON_CANCER	0.71	1.00	0.83	486
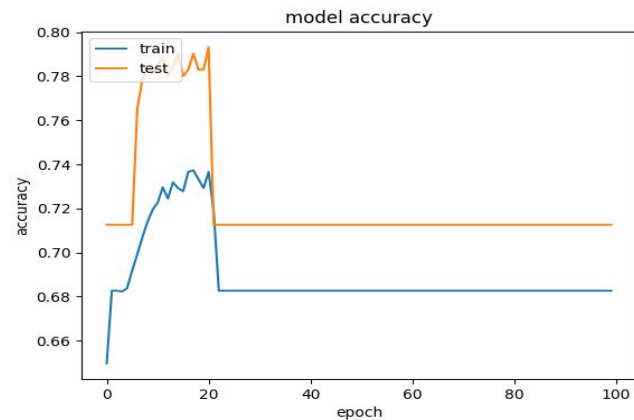
Accuracy plot 1.

	cancer	non -cancer
cancer	0	196
Non-cancer	0	486

Confusion Matrix:
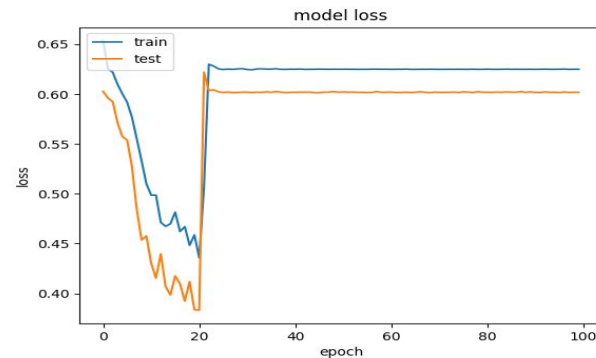
Loss plot 1.

Exp2. Results

Accuracy: 83.36%

	Precision	recall	f1-score	support
(CANCER)	0.78	0.36	0.49	594
(NON_CANCER)	0.84	0.97	0.90	2068

	cancer	noncancer
cancer	212	382
noncancer	61	2007

Confusion Matrix:
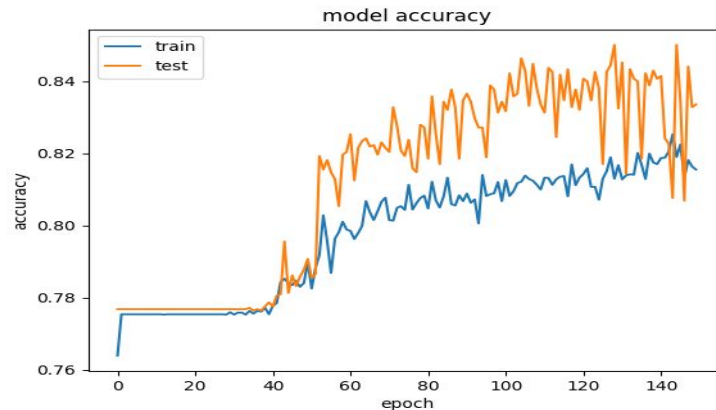
Accuracy plot 2.
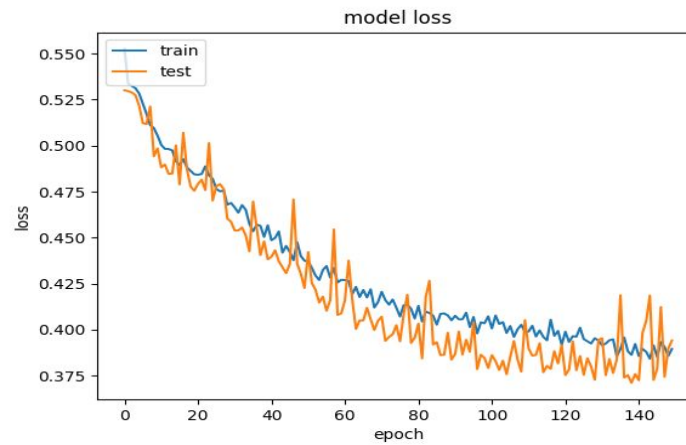
Loss plot 2.

Exp3 with Model2 result

Accuracy: 98.61%

	precision	recall	f1-score	support
(CANCER)	0.99	0.94	0.96	704
(NON_CANCER)	0.99	1.00	0.99	2894

	cancer	noncancer
cancer	662	42
noncancer	8	2886
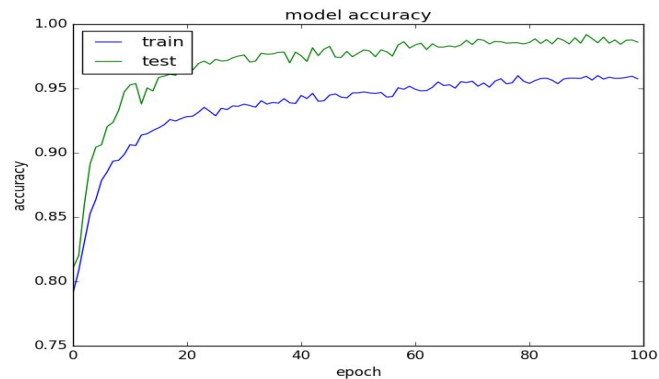
Accuracy plot 4.
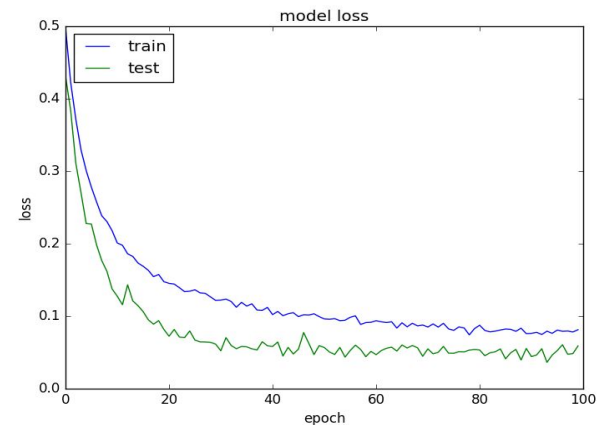
Loss plot 4.

Exp4 with model 2 Result

Accuracy: 96.13%

	precision	recall	f1-score	support
(CANCER)	0.98	0.83	0.90	1224
(NON_CANCER)	0.96	1.00	0.98	4406

	cancer	noncancer
cancer	1022	202
noncancer	16	4390
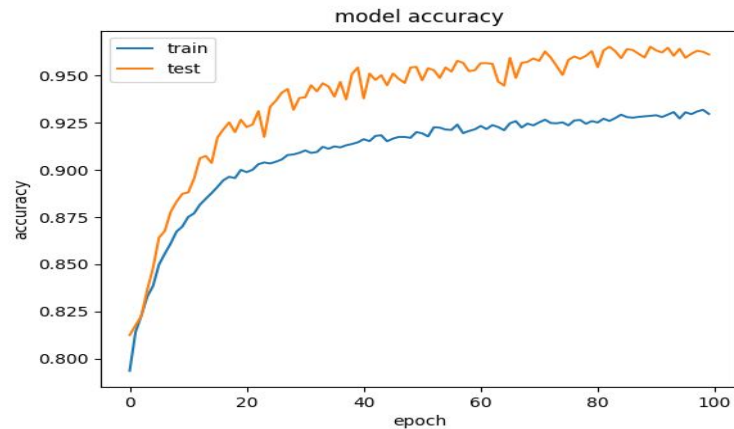
Accuracy plot 6.
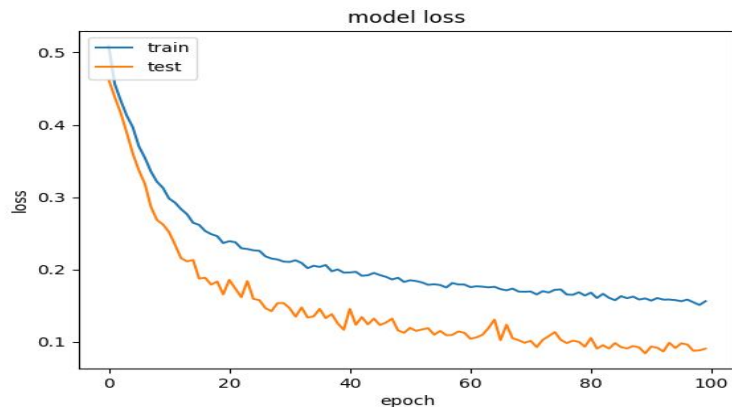
Loss plot 6.

Exp3 with Model1 result

Accuracy: 97.72%

	precision	recall	f1-score	support
(CANCER)	0.96	0.93	0.94	704
(NON_CANCER)	0.98	0.99	0.99	2894
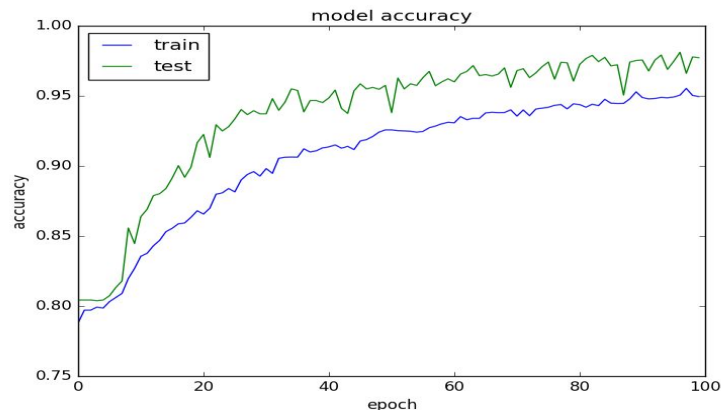
	cancer	noncancer
cancer	652	52
noncancer	30	2864
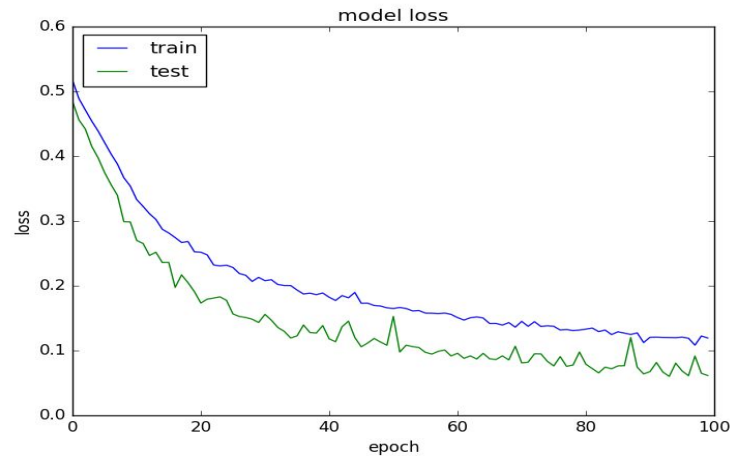
Accuracy plot 3.

Loss plot 3.

Exp 4 with model 1 result

Accuracy: 93.00%

	precision	recall	f1-score	support
(CANCER)	0.97	0.70	0.81	1224
(NON_CANCER)	0.92	0.99	0.96	4406

	cancer	noncancer
cancer	857	367
noncancer	27	4379
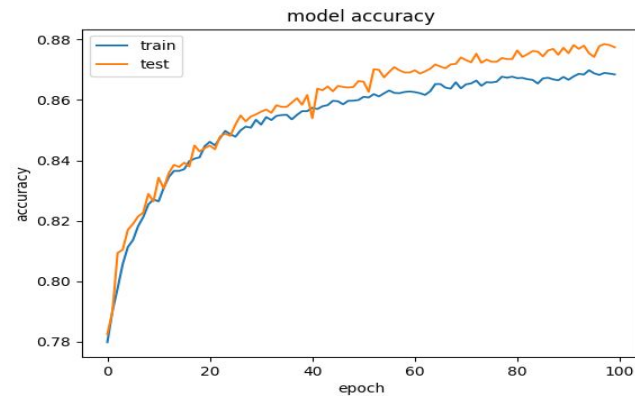
Accuracy plot 5.
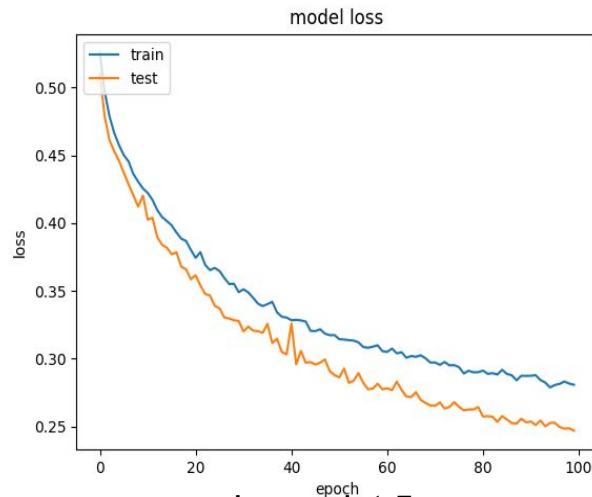
Loss plot 5.

Different Optimizers

Stochastic Gradient Descent

Model :Model 2: 2 2-D convolutional layers + 1 Fully connected layers + output layer

- Data size : 2 Gb, accuracy : 71%

- Data size: 10gb accuracy : 98.61%

- Data size: 16gb accuracy : 96.13%

RMSprop

Model: Model 2: 2 2-D convolutional layers + 1 Fully connected layers + output layer

- Data size: 2Gb, accuracy : 91%

- Data size: 10GB,accuracy : 96%

- Data size:16gb ,accuracy : 91.13%

Observations

Model	Exp1	Exp3	Exp4
1	Accuracy: 71%	Accuracy: 97.72% F-1 score: 96.5%	Accuracy: 93.00% F-1 score: 88.5%
2	--	Accuracy: 98.61% F-1 score: 97.5%	Accuracy: 96.13% F-1 score: 94%

Accuracy Wise	Exp 1 gives lowes accuracy whereas Exp3 with model 2 gives highest accuracy
Model Wise	Model 2 performs better.

Conclusion

- Model with lesser hidden layers performs better.

- With 'sgd' as optimizer, the results vary with change in data size.

- With 'rmsprop', the accuracy increases with increase in data.

- Accuracy increases with increase in data.

  - Accuracy increases enormous form 2gb dataset to 8 gb and from 8gb to 10gb i.e 71 to 80 to 98. But Accuracy on 10gb gives better accuracy gives better results than 16gb.
  - Accuracy also depends on distribution of data

References

- https://www.kaggle.com/c/data-science-bowl-2017#tutorial
- https://keras.io/getting-started/sequential-model-guide
- http://docs.opencv.org/3.0-beta/doc/py_tutorials/py_gui/py_image_display/py_image_display.html
- Jinsa Kuruvilla and K. Gunavathi. 2014. Lung cancer classification using neural networks for CT images. Comput. Methods Prog. Biomed. 113, 1 (January 2014), 202-209.
- R. Golan, C. Jacob and J. Denzinger, "Lung nodule detection in CT images using deep convolutional neural networks," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver,BC, 2016, pp. 243-250.
- B. van Ginneken, A. A. A. Setio, C. Jacobs and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, NY, 2015, pp. 286-289

Thank you