EXPLAINING DEEP FEATURES USING RADIOLOGIST DEFINED SEMANTIC (and RADIOMICS) FEATURES

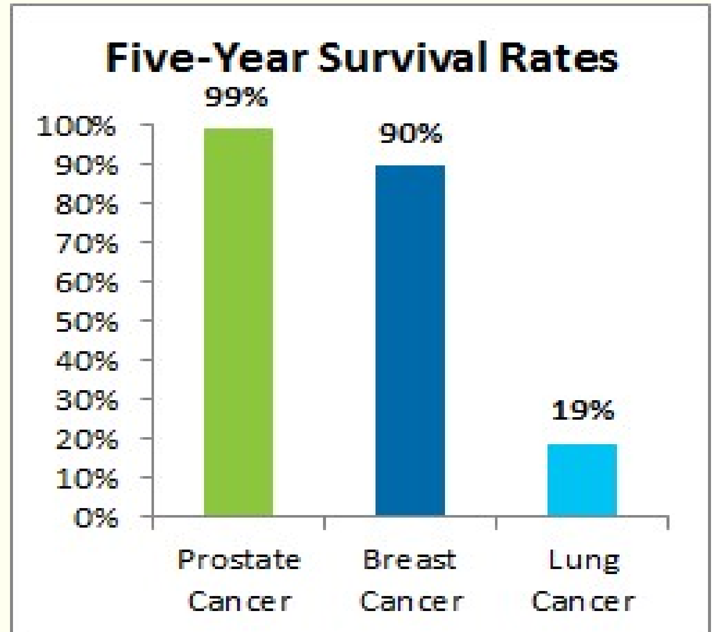
Dmitry Goldgof

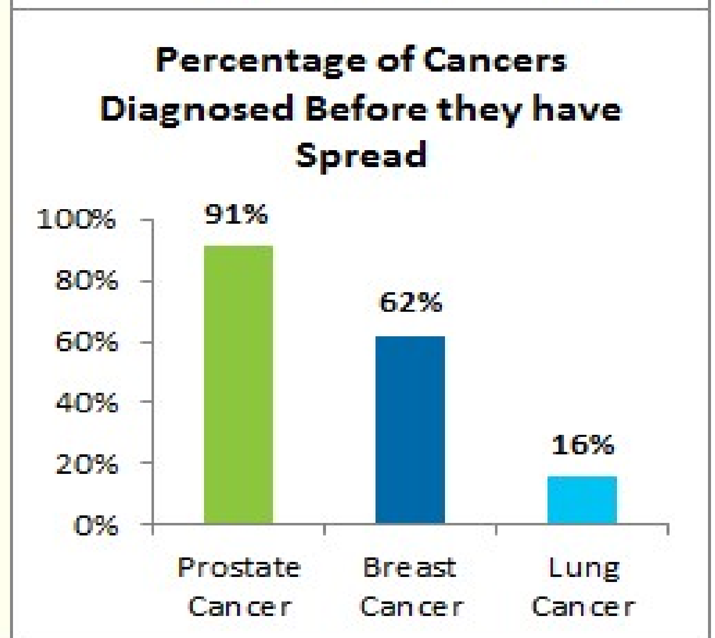
Department of Computer Science and Engineering,
University of South Florida, Tampa, FL

18th July 2020

Lung Cancer

- Lung cancer is one of the most common causes of malignancy worldwide.
- 19% 5-year survival rate in 2020.
- 230K (14%) new cases of lung cancer: 2nd most in the USA in 2020
- 135K deaths from lung cancer: the most in the USA in 2020
- Up to 10 million lung cancer deaths by 2030



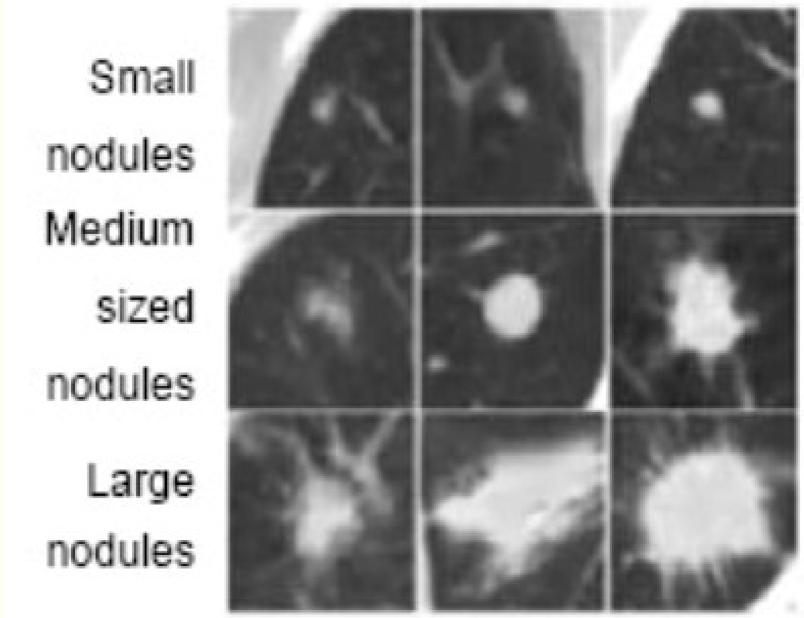


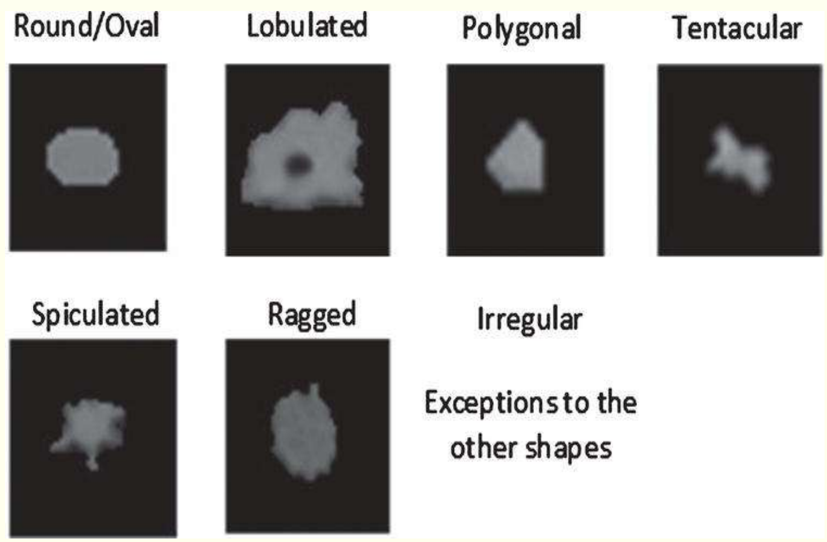
Detection of Lung Cancer

- 2 types of lung cancers: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC)
- Approximately 75% of the lung cancers patients are first diagnosed at the advanced stages (III/IV)
- Lung cancer detection : CT (Computed Tomography), Chest X-ray, Biopsy
- The National Lung Screening Trial (NLST) [1999-2001] : 54K patients
- Comparison of LDCT and chest X-ray by NLST.

Radiomics : Rationale

- To personalize treatment optimally we need to identify the biological differences between tumors – **Biomarkers**.
- The goal is to find a biomarker:
  - Easy to perform and calculate
  - Non-invasive
  - Low cost
  - Captures the 3D complexity of the tumors.

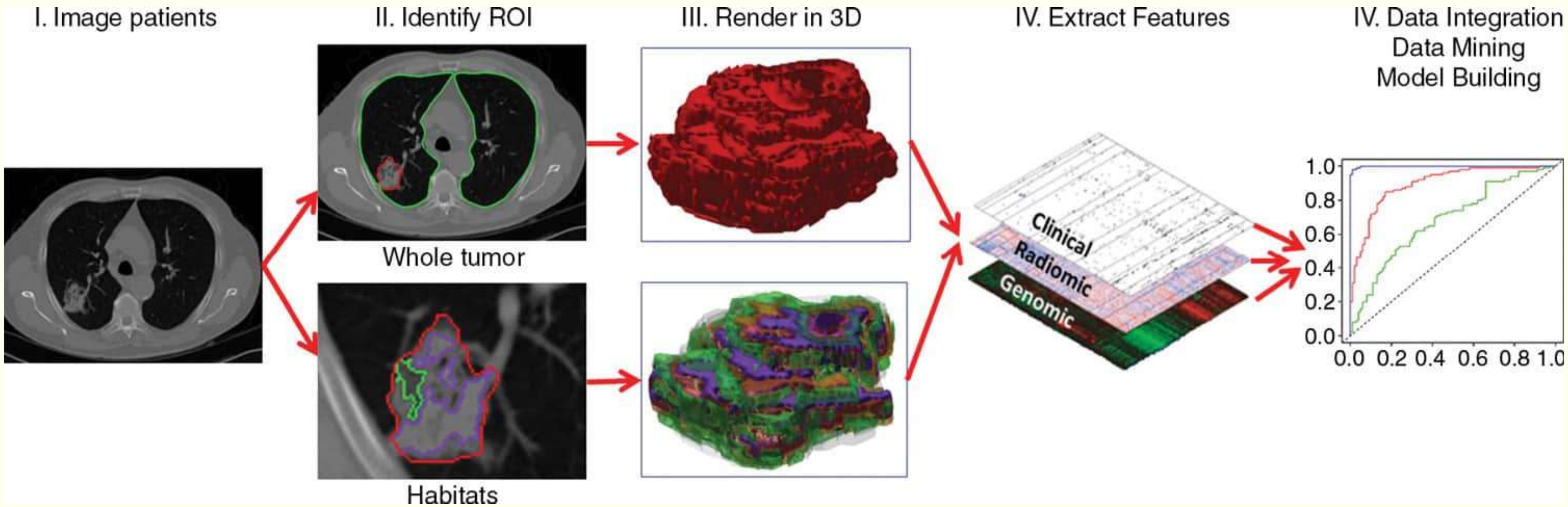




Kumar, et al. "Radiomics: the process and the challenges“, *Magnetic resonance imaging*, 30(9), 2012.

Radiomics

- **Radiomics:** High-throughput extraction of **quantitative** features for subsequent analysis
- It is a multi-stage technique that assists clinical diagnosis and prognosis using standard medical images.



R. Gillies, P. Kinahan, H. Hricak. "Radiomics: Images are more than Pictures, They are Data," Radiology, November 18, 2016.
https://pubs.rsna.org/doi/full/10.1148/radiol.2015151169

Radiomics via Deep Learning

- Deep learning is an emerging area for recognition, prediction and classification related tasks.
- For lung nodule analysis CNNs have been used effectively in recent years
- In the medical imaging field data is currently scarce (hence, transfer learning or augmentation).
- Deep Features are obtained from a trained network (input to the last layer)
- Feature obtained from a trained network don't have a "simple" mathematical or semantic meaning needed for explanation.

Goal

- Represent and explain deep features with respect to semantic/radiomics features

- This will provide a semantic and mathematical meaning for the deep features.

Paul, R., Schabath, M., Balagurunathan, Y., Liu, Y., Li, Q., Gillies, R., Hall, L.O. and Goldgof, D.B., 2019. Explaining deep features using Radiologist-Defined semantic features and traditional quantitative features. *Tomography*, *5*(1), p.192.

Paul, R., Liu, Y., Li, Q., Hall, L., Goldgof, D., Balagurunathan, Y., Schabath, M. and Gillies, R., 2018, July. Representation of Deep Features using Radiologist defined Semantic Features. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-7). IEEE.

Contributors

Rahul Paul, Lawrence Hall , Dmitry Goldgof

Department of Computer Science
University of South Florida, Tampa, FL

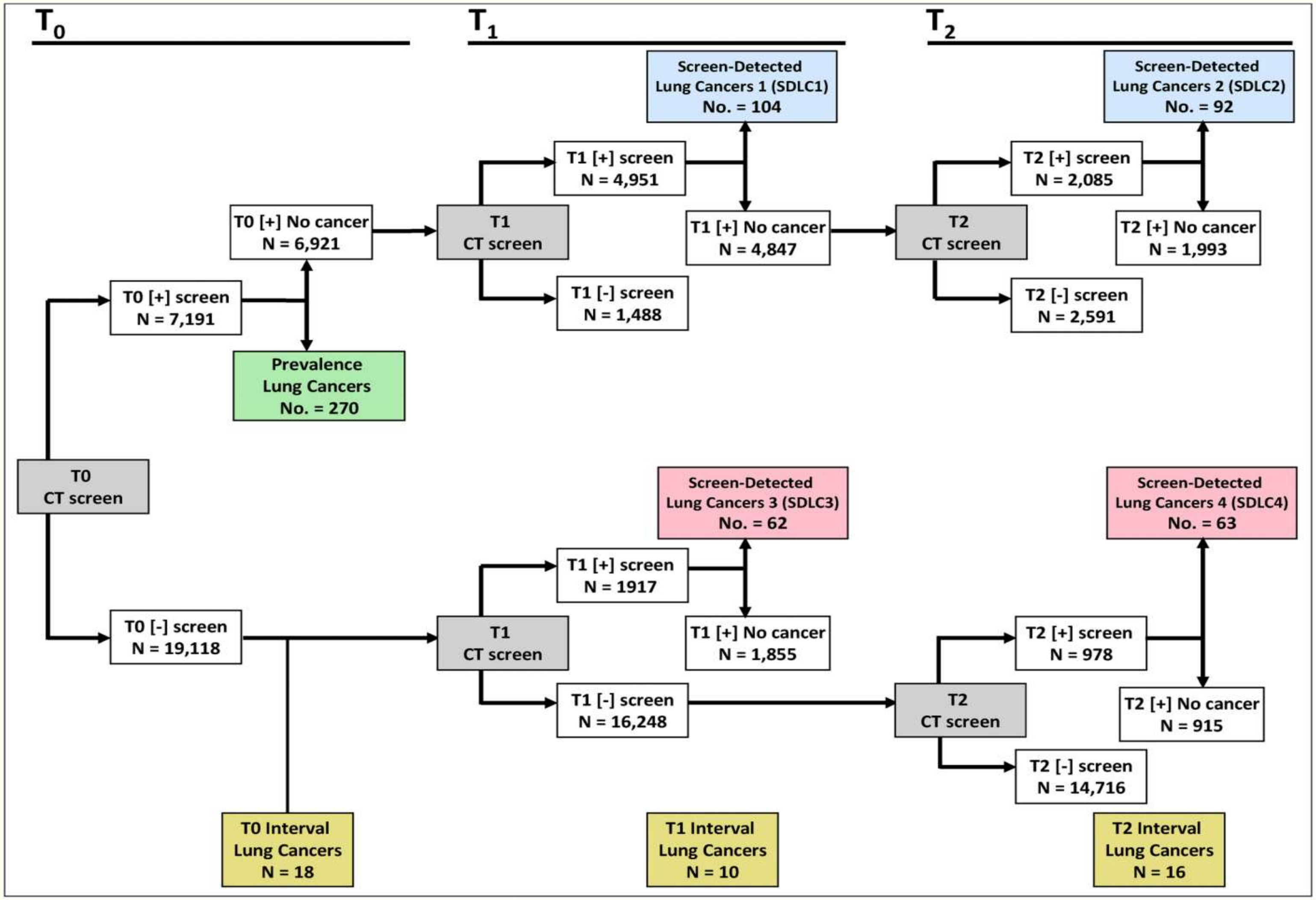
Ying Liu, Qian Li

Tianjin Medical University Cancer Institute and Hospital , China

Yoganand Balagurunathan, Matthew Schabath, and Robert Gillies

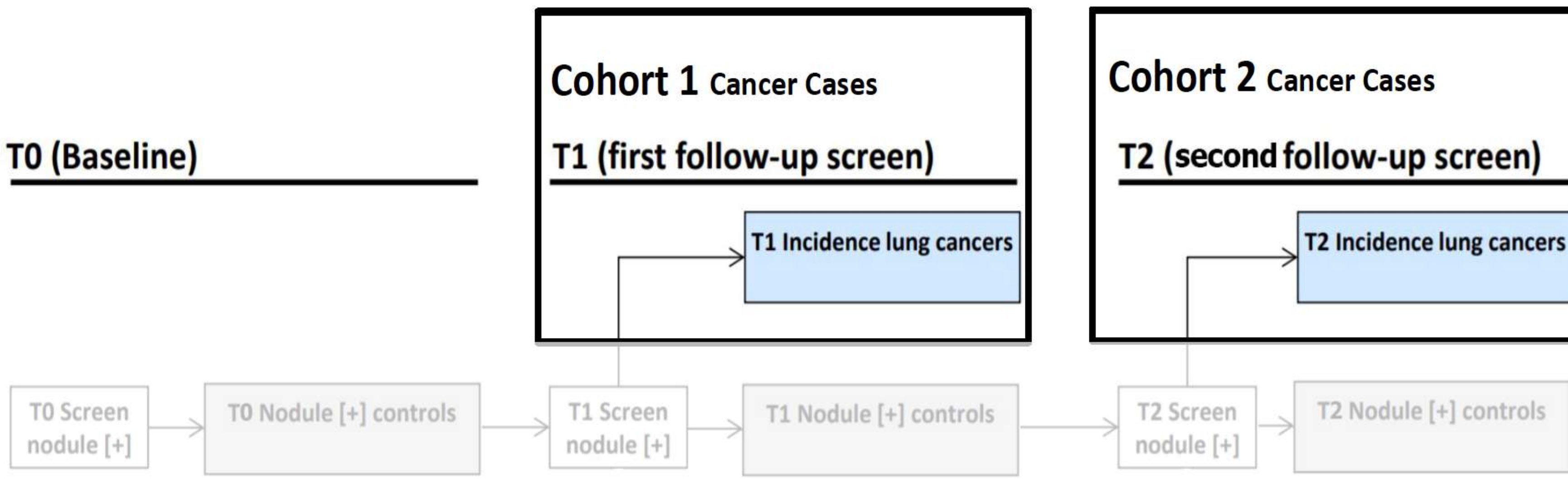
H. Lee. Moffitt Cancer Center, Tampa, FL

National Lung Screening Trial (NLST)



National Lung Screening Trial Research Team Writing committee:, Aberle, D.R., Adams, A.M., Berg, C.D., Clapp, J.D., Clingan, K.L., Gareen, I.F., Lynch, D.A., Marcus, P.M. and Pinsky, P.F., 2010. Baseline characteristics of participants in the randomized national lung screening trial. *Journal of the National Cancer Institute*, *102*(23), pp.1771-1779.

NLST Dataset



Schabath, M.B., Massion, P.P., Thompson, Z.J., Eschrich, S.A., Balagurunathan, Y., Goldof, D., Aberle, D.R. and Gillies, R.J., 2016. Differences in patient outcomes of prevalence, interval, and screen-detected lung cancers in the CT arm of the national lung screening trial. *PloS one*, *11*(8), p.e0159880. 3

- ❖ NLST Cohort 2* (T0): Test (58 malignant and 127 nodule positive control)

Semantic Features

- A radiologist can provide important individual information about one's lung tumor, which can provide useful guidance for prognosis and diagnosis: These unique characteristics are termed "Semantic" features

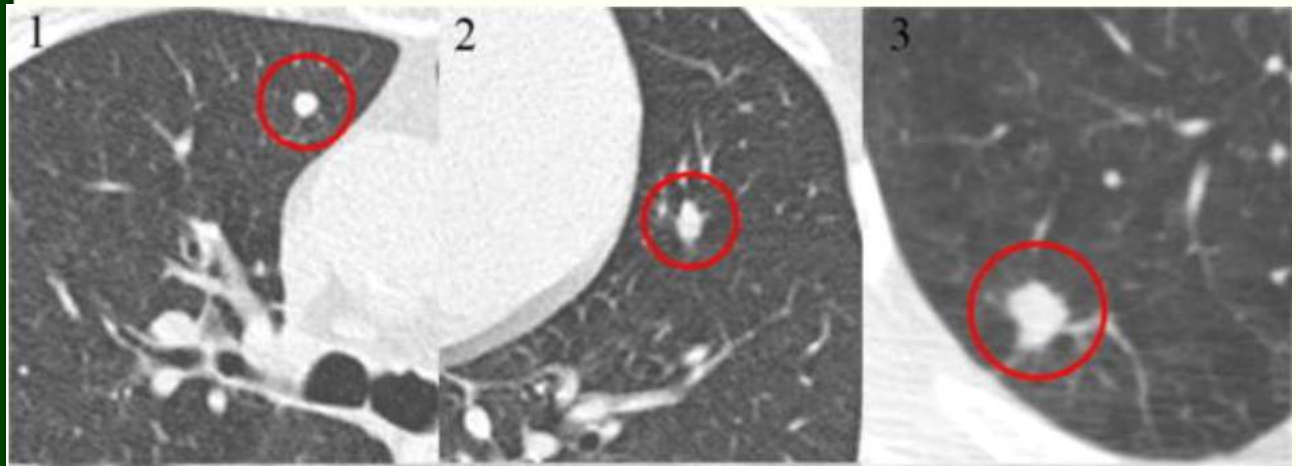
- 20 semantic features were described by an experienced radiologist (7 years of experience)

- These features can be subdivided into location, size, shape, margin, attenuation, external and associated findings

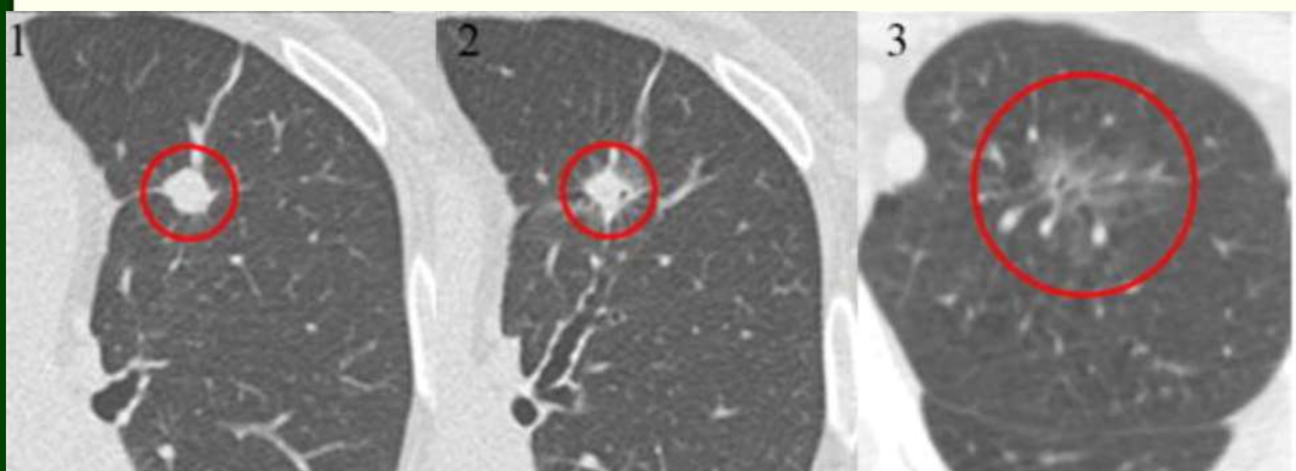
Li, Q., Balagurunathan, Y., Liu, Y., Qi, J., Schabath, M.B., Ye, Z. and Gillies, R.J., 2018. comparison between radiological semantic features and lung-rads in predicting malignancy of screen-detected lung nodules in the National Lung Screening Trial. *Clinical lung cancer*, *19*(2), pp.148-156.

Liu, Y., Balagurunathan, Y., Atwater, T., Antic, S., Li, Q., Walker, R.C., Smith, G.T., Massion, P.P., Schabath, M.B. and Gillies, R.J., 2017. Radiological image traits predictive of cancer status in pulmonary nodules. *Clinical Cancer Research*, *23*(6), pp.1442-1449.

Semantic Features: Example



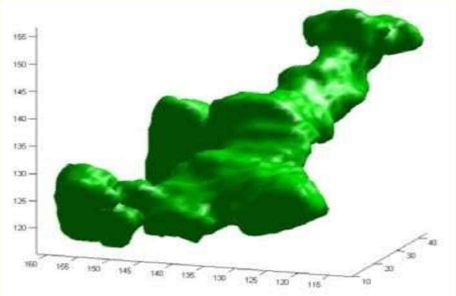
Examples of contour (1, round; 2, oval; 3, irregular)

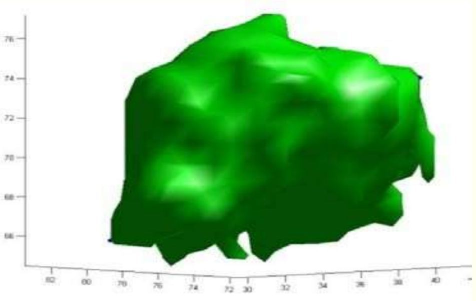


Examples of border definition (1, well defined; 2, slight poorly; 3, poorly defined)

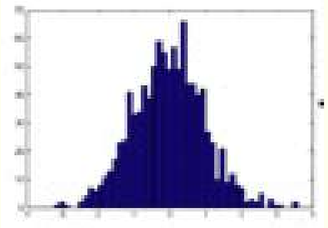
Traditional Radiomics Features

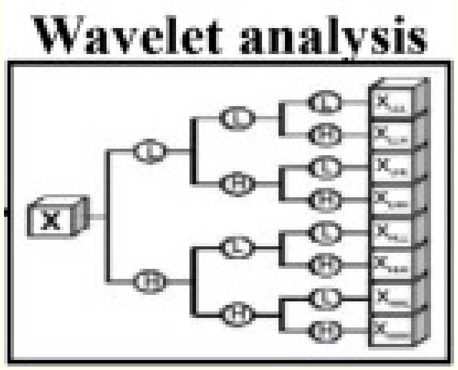
Shape feature



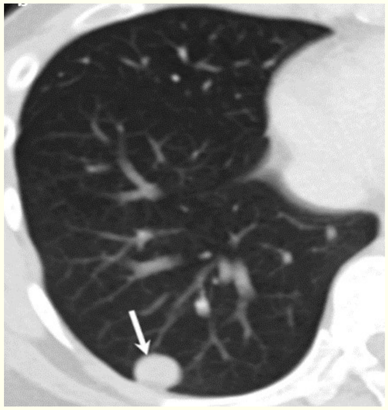


Histogram feature

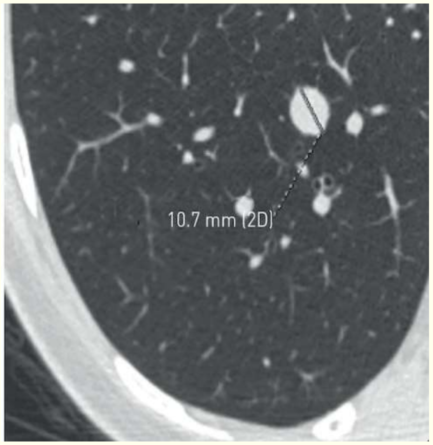




Location feature



Size feature



Hawkins, S., Wang, H., Liu, Y., Garcia, A., Stringfield, O., Krewer, H., Li, Q., Cherezov, D., Gatenby, R.A., Balagurunathan, Y. and Goldgof, D., 2016. Predicting malignant nodules from screening CT scans. *Journal of Thoracic Oncology*, *11*(12), pp.2120-2128.

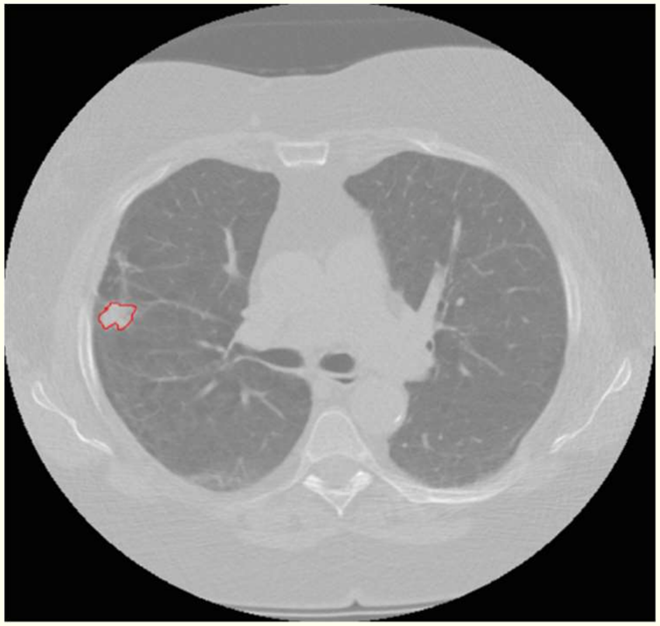
Radiomics Features

- Quantitative radiomics features (219) extracted from lung nodules
- 23 Rider (test/retest) stable features were used for this study.

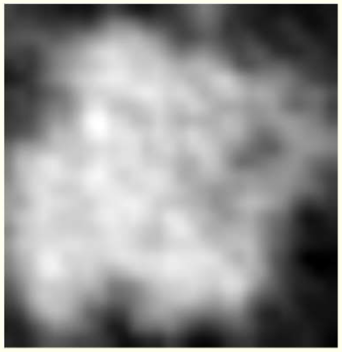
Characteristic	Features
Size	1. Long-axis Diameter 2. Short-axis Diameter 3. Long axis Diameter* Short axis diameter 4. Volume (cm) 5. Volume (pixel) 6. Number of pixels 7. Length/width
Pixel Intensity Histogram	8. Mean (HU) 9. Stand Deviation (HU)
Tumor Location	10. 8a_3D_ is attached to pleural wall 11. 8b_3D Relative border to lung 12. 8c_3D_Relative Border to pleural wall 13. 9e_3D_Standard deviation _COG to border 14. 9g_3D_max_Dist_COG to border
Tumor Shape (Roundness)	15. 9b-3D circularity 16. 5a_3D- MacSpic 17. Asymmetry 18. Roundness
Run-length and co-occurrence	19. Avg_RLN
Law's Texture feature	20. E5 E5 L5 layer 1 21. E5 E5 R5 layer 1 22. E5 W5 L5 layer 1 23. L5 W5 L5 layer 1

Deep Learning

- 2-D approach
- Chose the slice with largest nodule area for each patient case.
- Nodule region was extracted by putting a rectangular box around the nodule that completely fit the nodule.



Lung image with nodule inside outlined by red and extracted nodule



Paul, Rahul, et al. "Convolutional Neural Network ensembles for accurate lung nodule malignancy prediction 2 years in the future." *Computers in Biology and Medicine* (2020): 103882.

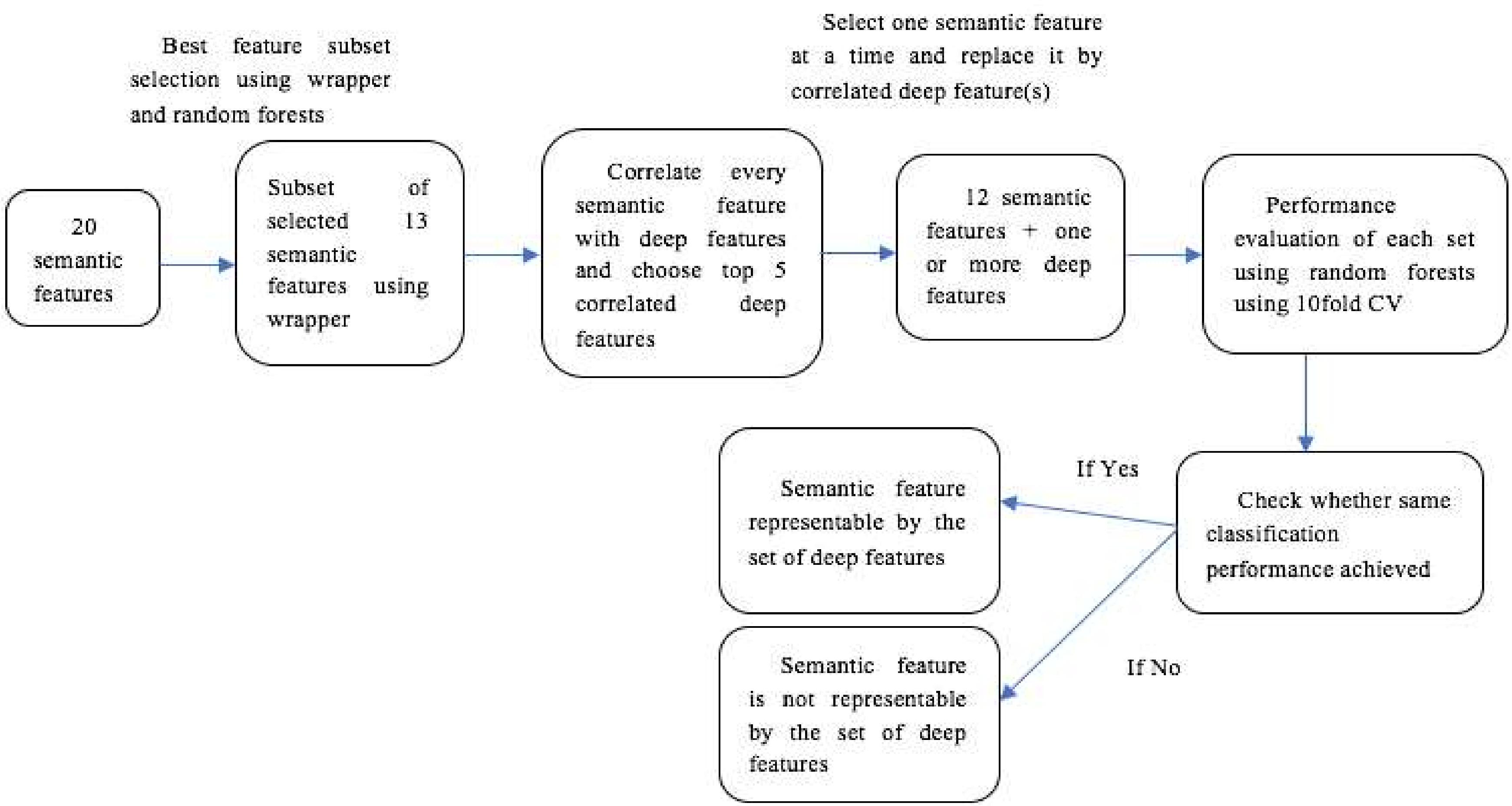
Deep Features

- 2 pre-trained networks were used for extracting deep features

- Vgg-S pre-trained CNN (Feature vector size : 4096) – trained on ImageNet dataset

- Our trained CNN architecture (**Feature Vector size: 1024**) – trained on lung nodules from NLST Cohort 1 T0.

Algorithm



Best feature subset selection using wrapper and random forests
Select one semantic feature at a time and replace it by correlated deep feature(s)
20 semantic features
Subset of selected 13 semantic features using wrapper
Correlate every semantic feature with deep features and choose top 5 correlated deep features
12 semantic features + one or more deep features
Performance evaluation of each set using random forests using 10fold CV
Check whether same classification performance achieved
If Yes
Semantic feature representable by the set of deep features
If No
Semantic feature is not representable by the set of deep features

Algorithm

- Use wrapper with random forests classifier (10 fold cross validation) to choose the best feature subset from the 20 semantic/23 radiomics features
- Calculate correlation of every semantic/radiomics features of best feature subset with deep features
- Choose top 5 correlated deep features for every semantic/radiomics feature
- Select one semantic/radiomics feature at a time and replace it by correlated deep feature(s). Our objective was to substitute each semantic/radiomics feature by one or more deep features and achieve the same classification result using random forests classifier with 10 fold cv
- Check whether the same classification performance achieved after replacing semantic/radiomics features, if yes, then the deep feature(s) can be represented by semantic/radiomics feature.

Results: Explaining deep features by semantic features

- Subset of 13 semantic features was chosen using the wrapper approach resulting in an accuracy of 83.78% (AUC 0.837).

- Subset of 13 features: Location, Long axis diameter, Short axis diameter, Lobulation, Concavity, Border definition, Spiculation, Texture, Cavitation, Vascular-convergence, Vessel attachment, Peri-nodule fibrosis, Nodules in primary tumor lobe.

- From Vgg-s pre-trained CNN : 8 semantic features (long axis diameter, lobulation, concavity, spiculation, texture, cavitation, vascular convergence, peripheral fibrosis) explained 15 deep features;

- From our trained CNN : 4 semantic features (long axis diameter, concavity, cavitation, nodules in primary tumor lobe) explained 6 deep features

Results Analysis: VGG-S features explained by Semantic features

Deep Features from VGG-S	Equivalent Semantic Feature with correlation value in brackets
3353	Long axis diameter (0.43), Cavitation (0.39)
2135	Long axis diameter (0.42)
3534	Lobulation (0.57), Concavity (0.5)
1372	Lobulation (0.56), Concavity (0.48)
2111	Lobulation (0.55), Concavity (0.475)
2975	Lobulation (0.56), Concavity (0.484)
3246	Concavity (0.46)
1201	Texture (-0.3119)
3350	Texture (0.2936)
526	Cavitation (0.36)
1464	Vascular Convergence (0.7052)
2115	Vascular Convergence (0.7)
3305	Peripheral Fibrosis (0.2076)
3064	Peripheral Fibrosis (0.2043)
2811	Spiculation (0.411)

Results Analysis: Our CNN features explained by Semantic features

Deep Features from our CNN	Equivalent Semantic Feature with correlation value in brackets
230	Long axis diameter (0.3035)
547	Concavity (0.1776)
440	Concavity (0.1514)
395	Cavitation (0.28)
425	Nodules in primary lobe (0.19)
57	Nodules in primary lobe (0.18)

Results: Explaining deep features by radiomics features

- **The 9 traditional quantitative features that enabled the best accuracy were: Mean (HU), is_attached to pleural wall, Relative border to pleural wall, circularity, Asymmetry, Roundness, Volume, E5W5L5, and L5W5L5**

- Subset of 9 radiomics features was chosen using the wrapper approach resulting in an accuracy of 84.32% (AUC 0.84).

- From vggs features: 3 radiomics features (circularity, roundness, and L5W5L5**)** explained 10 deep features

- From our trained CNN features: 3 radiomics features **circ**ularity, roundness, and L5W5L5) explained 6 deep features

Results Analysis: VGG-S features explained by Radiomics features

Deep Features from VGG-S	Equivalent Radiomics Feature with correlation value in brackets
1395	Roundness (0.3), Circularity (0.24)
2510	Roundness (0.27)
1757	Circularity (-0.234)
3401	Circularity (-0.2069)
2777	Circularity (-0.2069)
51	L5W5L5 (0.77)
66	L5W5L5 (0.75)
163	L5W5L5 (0.69)
476	L5W5L5 (0.69)
928	L5W5L5 (0.69)

Results Analysis: Our CNN features explained by Radiomics features

Deep Features from our CNN	Equivalent Radiomics Feature with correlation value in brackets
160	Roundness (0.16), Circularity (0.13)
20	Roundness (0.14), Circularity (0.13)
547	L5W5L5 (0.28)
169	L5W5L5 (0.27)
265	L5W5L5 (0.26)
309	L5W5L5 (0.26)

Discussions

- Deep learning provides higher predictive performance.
- Correlated semantic and radiomics features provides insight into deep features
- Often explanation of CNN means visualization
- To our knowledge, this is the first work to explain deep features with respect to traditional quantitative features and semantic features extracted from a lung nodule.

- To explain we need more GT then “just” to predict
- Too much explanation? Deep learning models leaking training data

THANK YOU!!

QUESTIONS?