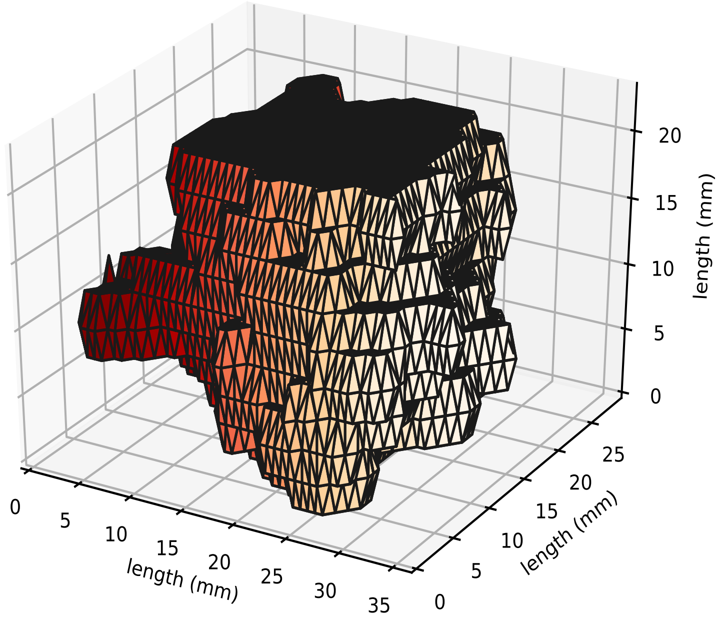
Predictive capabilities of statistical learning methods for lung nodule malignancy classification using diagnostic image features

Matthew C. Hancock, Jerry F. Magnan

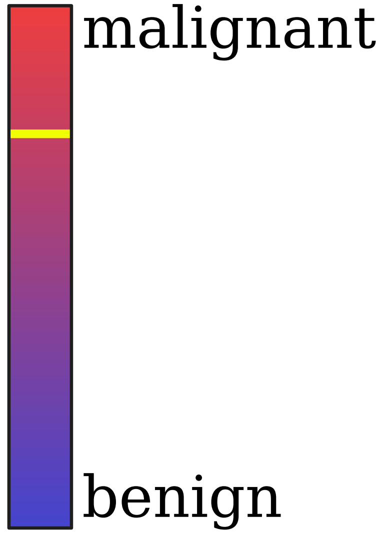
Dept. of Mathematics, Florida State University

SPIE Medical Imaging Symposium 2017

A Modular Approach to Computer-Aided Diagnosis (CAD)



volume
spiculation
calcification
margin
...



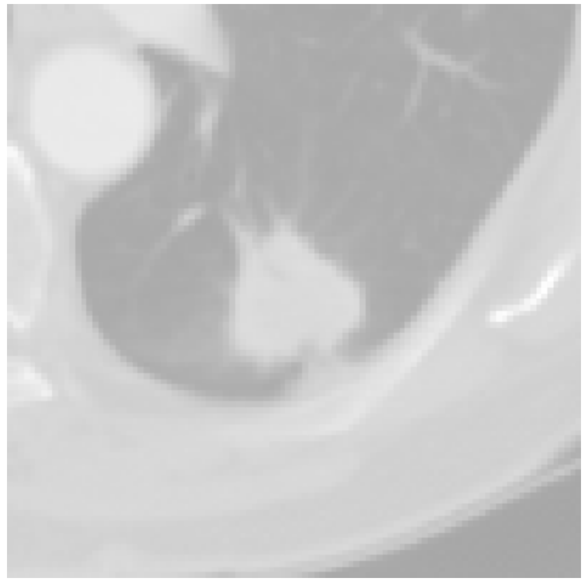
Region of interest

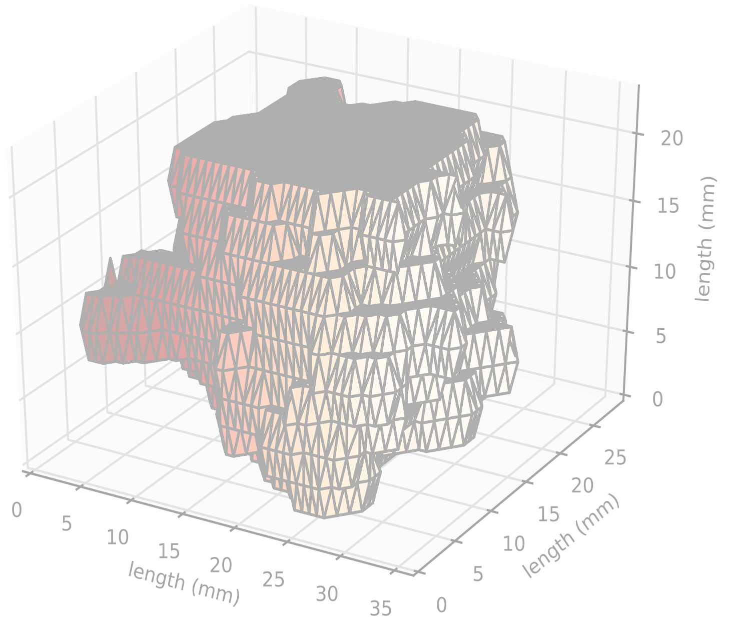
Segmentation

Features

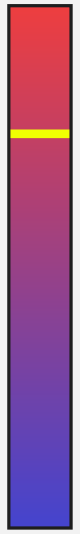
Diagnosis

A Modular Approach to Computer-Aided Diagnosis (CAD)





volume
spiculation
calcification
margin
…



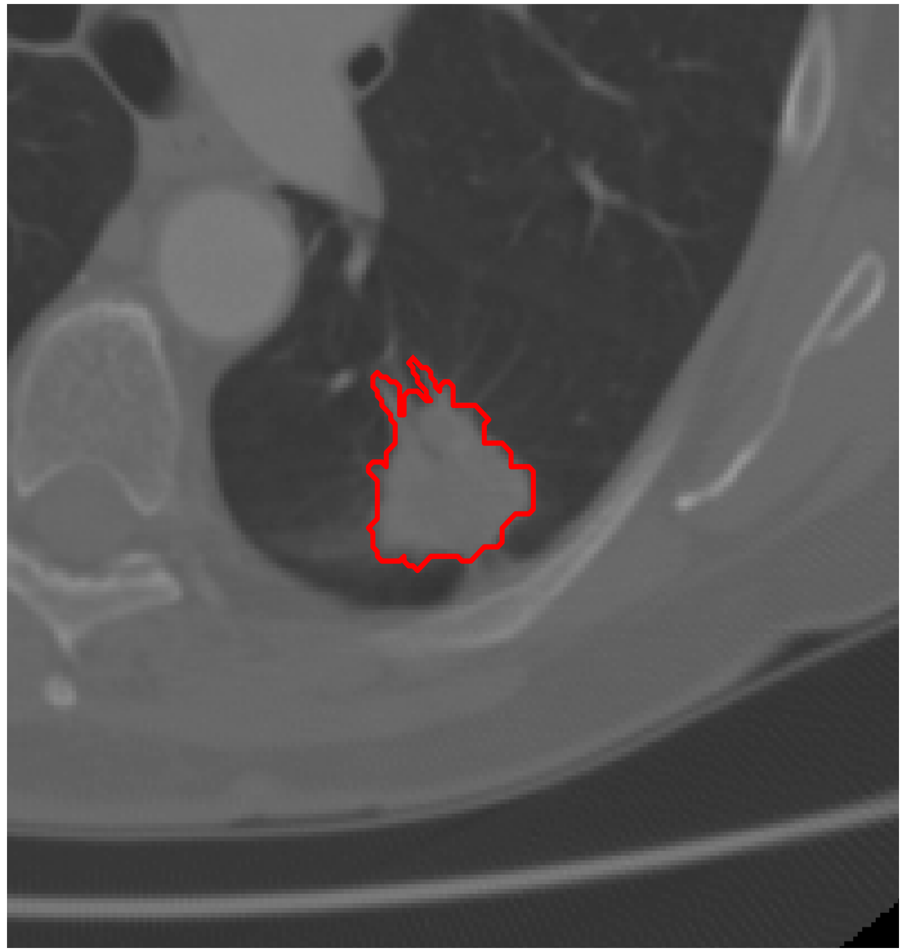
Region of interest

Segmentation

Features

Diagnosis

The LIDC Dataset



Feature	Numerical	Interpretation
Volume	7869.4	-
Diameter	32.2	-
Subtlety	5	Obvious
Internal Structure	1	Soft Tissue
Calcification	6	Absent
Sphericity	3	Ovoid
Margin	3	Medium
Lobulation	3	Medium Lobulation
Spiculation	4	Medium-High Spiculation
Texture	5	Solid Texture
Malignancy	5	High Malignancy

Data for Approximating "Features → Diagnosis" Map

Feature	Numerical
Volume	7869.4
Diameter	32.2
Subtlety	5
Internal Structure	1
Calcification	6
Sphericity	3
Margin	3
Lobulation	3
Spiculation	4
Texture	5
Malignancy	5

$$\left(X^{(i)}, Y^{(i)} \right) \quad i = 1, \ldots, N = 2817$$

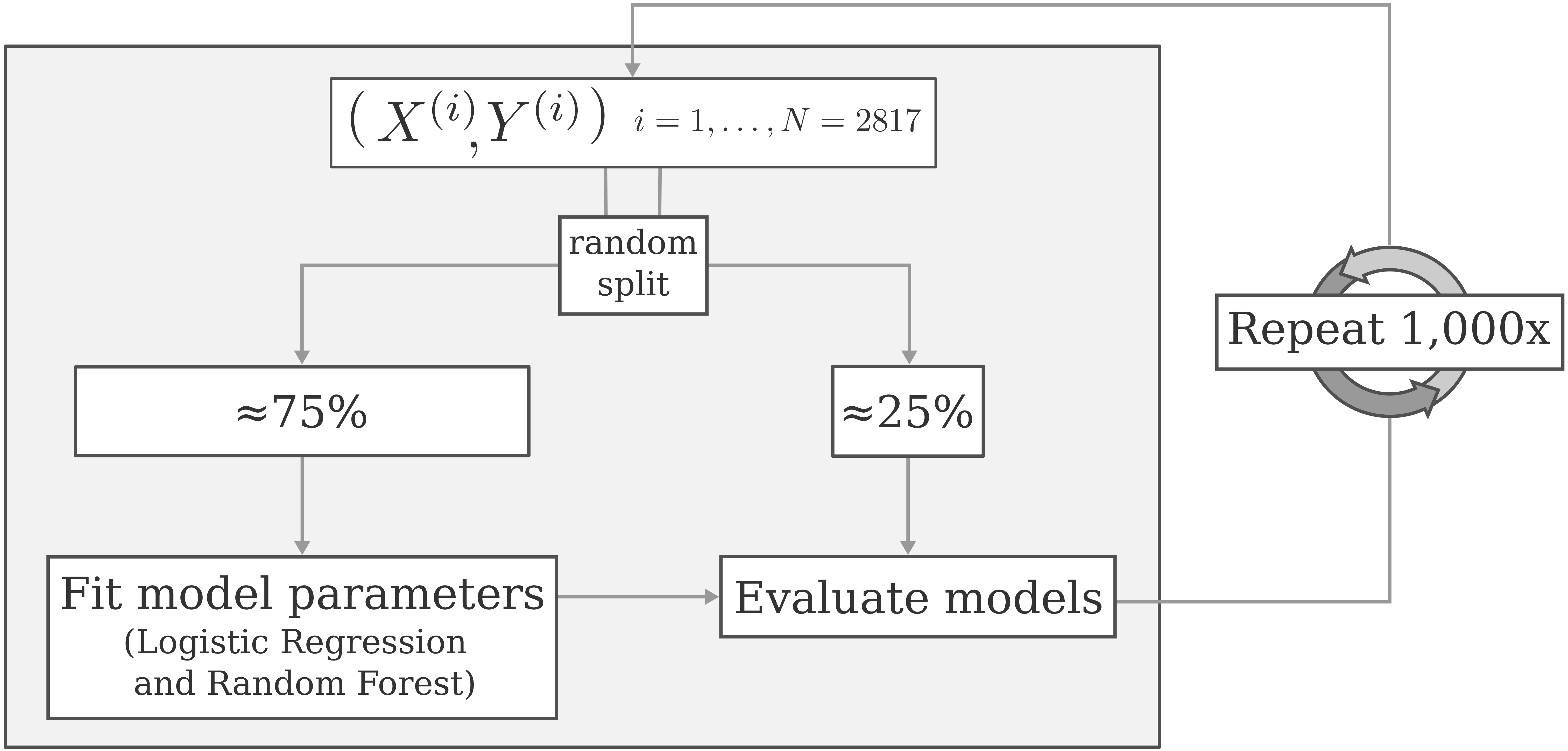
subset

1 if malignancy > 3
0 if malignancy < 3

Questions Addressed

- Accurate models for classification
- Upper bounds on classification accuracy
- Important features for accurate classification

Process to Determine Model Capabilities



Degenerate Groups and Maximum Accuracy

Example "degenerate group" of distinct annotations

X		Y
1, 1, 6, 4, 1, 1, 1, 1	0	
1, 1, 6, 4, 1, 1, 1, 1	0	
1, 1, 6, 4, 1, 1, 1, 1	0	
1, 1, 6, 4, 1, 1, 1, 1	0	
1, 1, 6, 4, 1, 1, 1, 1	1	

Behavior of Ideal Classifier on Test Data

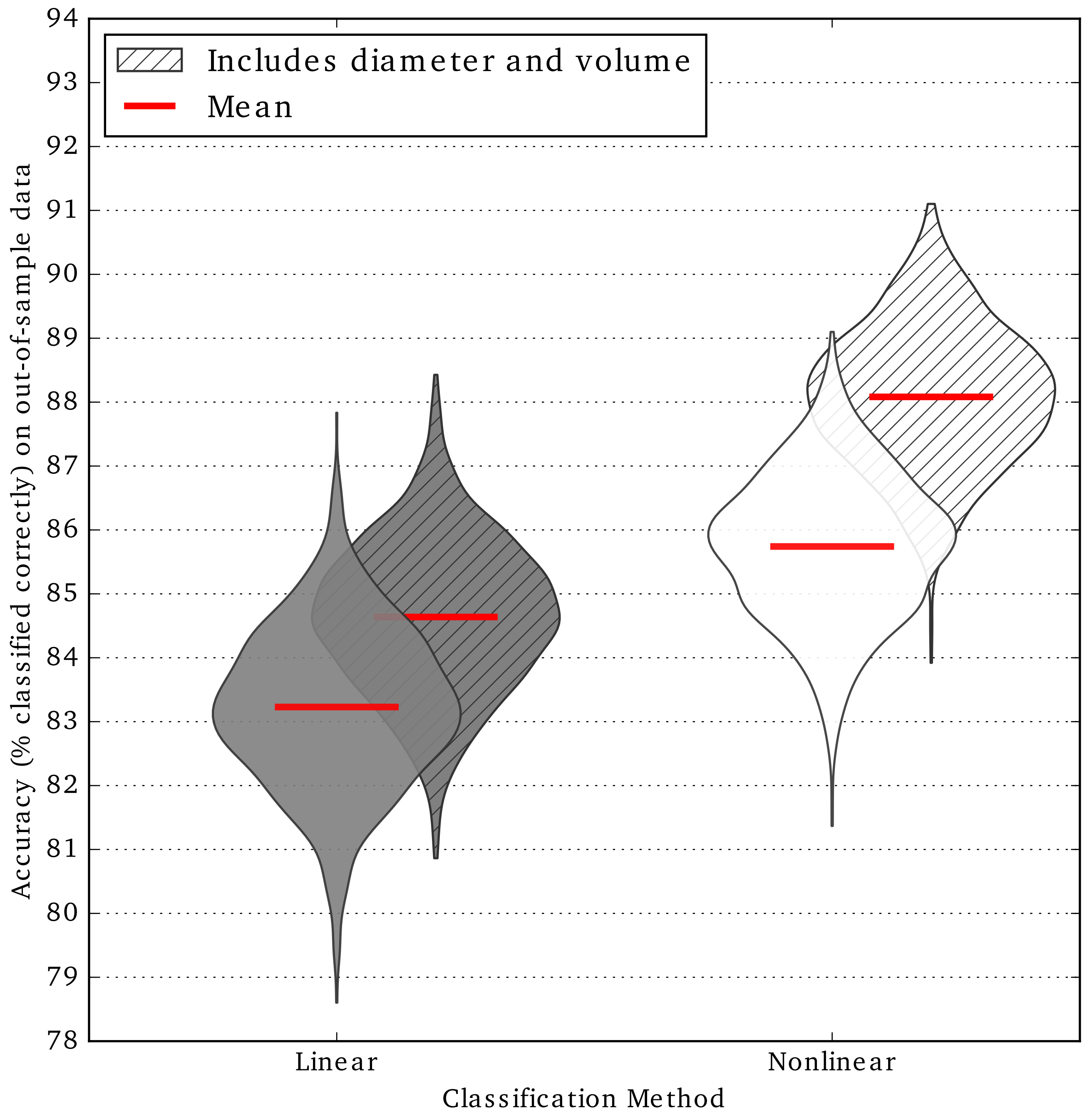
1. Classifies all non-degenerate examples correctly

2. Decides majority class for degenerate groups with members only in testing set

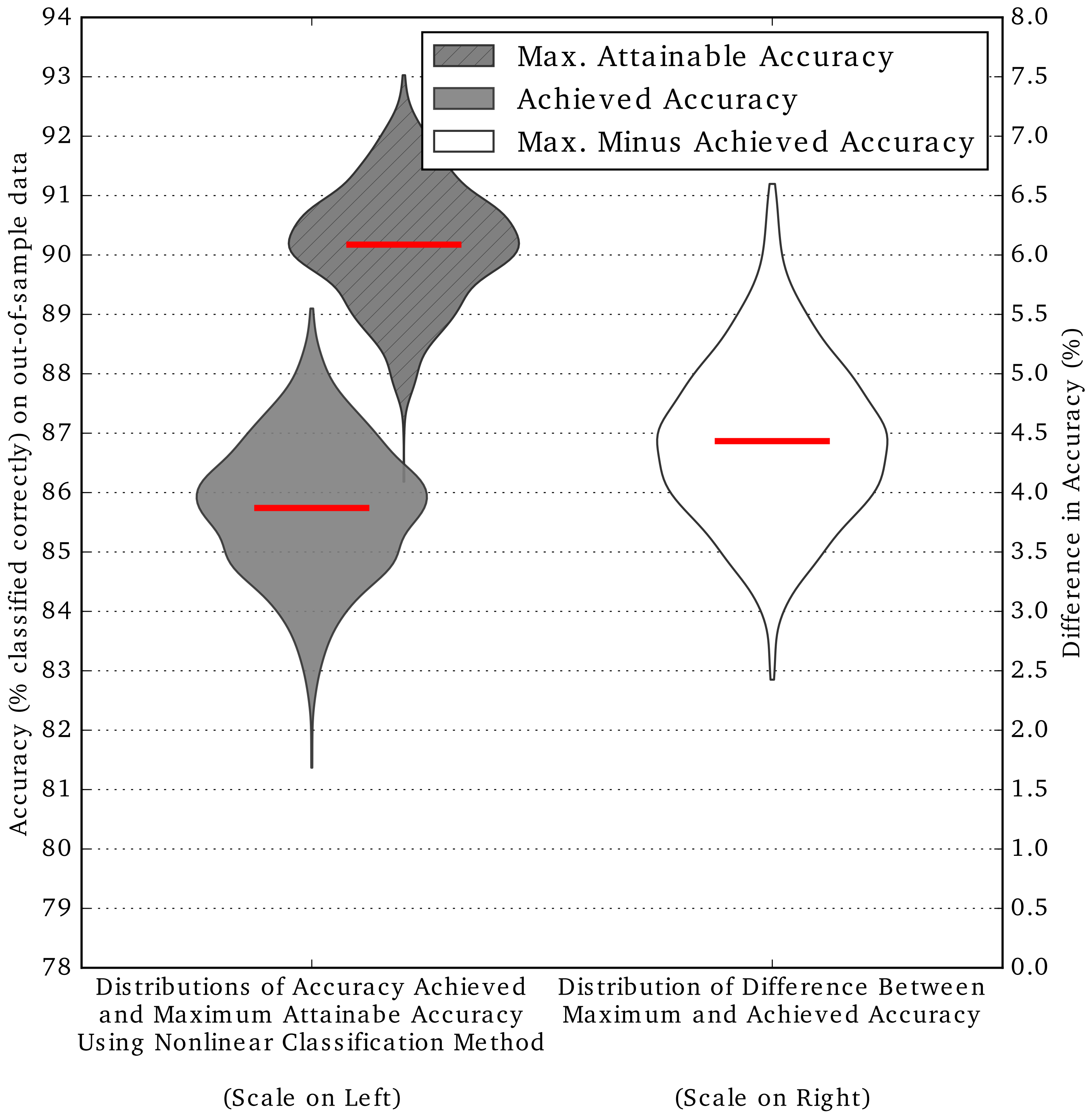
3. Decision restricted by training data for degenerate groups with members in training and test

Each randomly chosen test dataset has corresponding upper bound on accuracy due to degeneracy

Results



Includes diameter and volume
Mean
Accuracy (% classified correctly) on out-of-sample data
94
93
92
91
90
89
88
87
86
85
84
83
82
81
80
79
78
Linear
Nonlinear
Classification Method



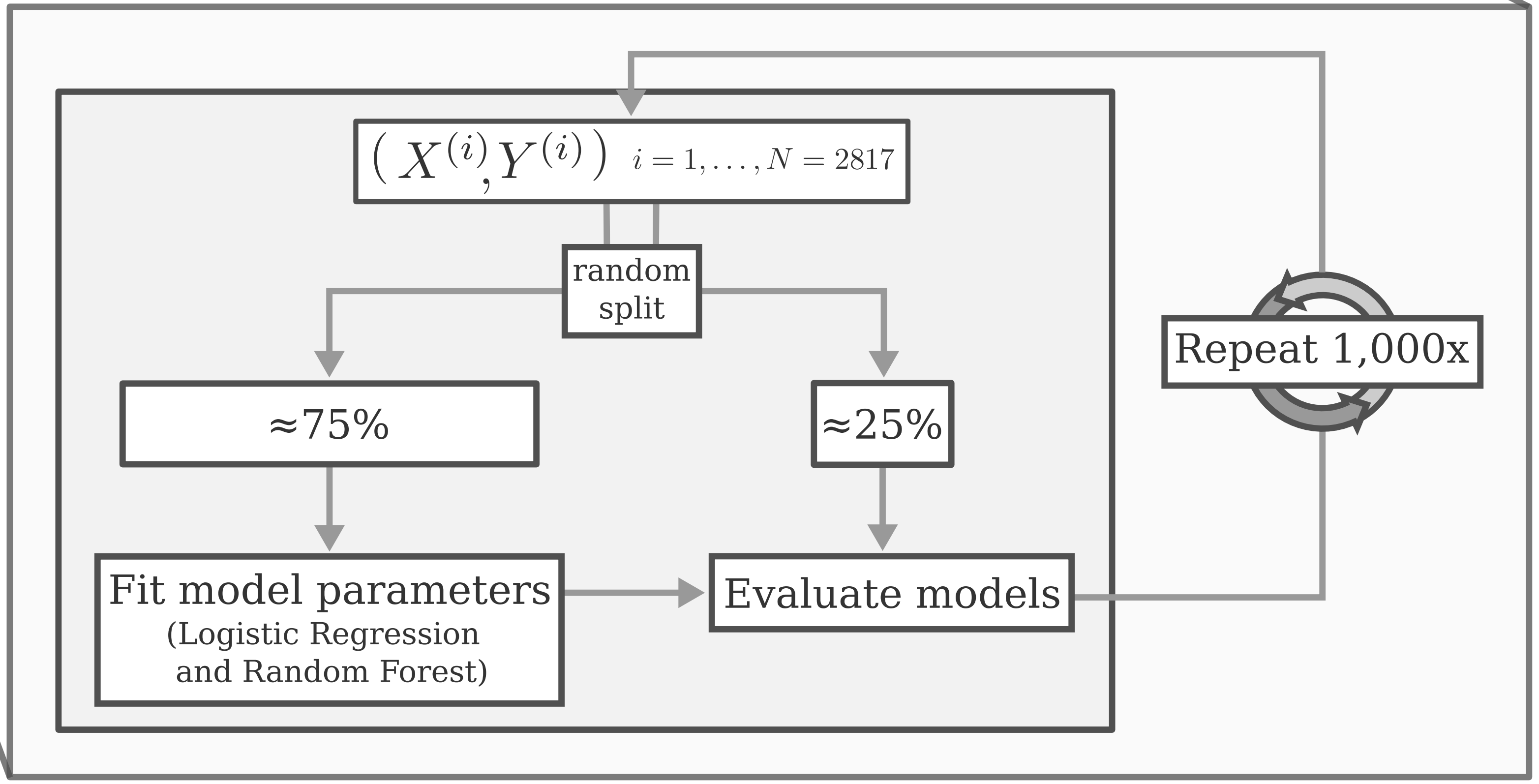
Max. Attainable Accuracy
Achieved Accuracy
Max. Minus Achieved Accuracy
Accuracy (% classified correctly) on out-of-sample data
Difference in Accuracy (%)
94
93
92
91
90
89
88
87
86
85
84
83
82
81
80
79
78
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Distributions of Accuracy Achieved and Maximum Attainabe Accuracy Using Nonlinear Classification Method
Distribution of Difference Between Maximum and Achieved Accuracy
(Scale on Left)
(Scale on Right)

More Results

	Accuracy $\left(t = \frac{1}{2}\right)$	TPR $\left(t = \frac{1}{2}\right)$	AUC
Linear classifier, diameter and volume features **excluded**	83.23 ($\pm$1.252)%	0.8013 ($\pm$0.0216)	0.9164 ($\pm$0.0087)
Linear classifier, diameter and volume features **included**	84.64 ($\pm$1.184)%	0.7906 ($\pm$0.0218)	0.9302 ($\pm$0.0079)
Nonlinear classifier, diameter and volume features **excluded**	85.74 ($\pm$1.141)%	0.8430 ($\pm$0.0239)	0.9322 ($\pm$0.0123)
Nonlinear classifier, diameter and volume features **included**	88.08 ($\pm$1.109)%	0.8461 ($\pm$0.0218)	0.9492 ($\pm$0.0070)

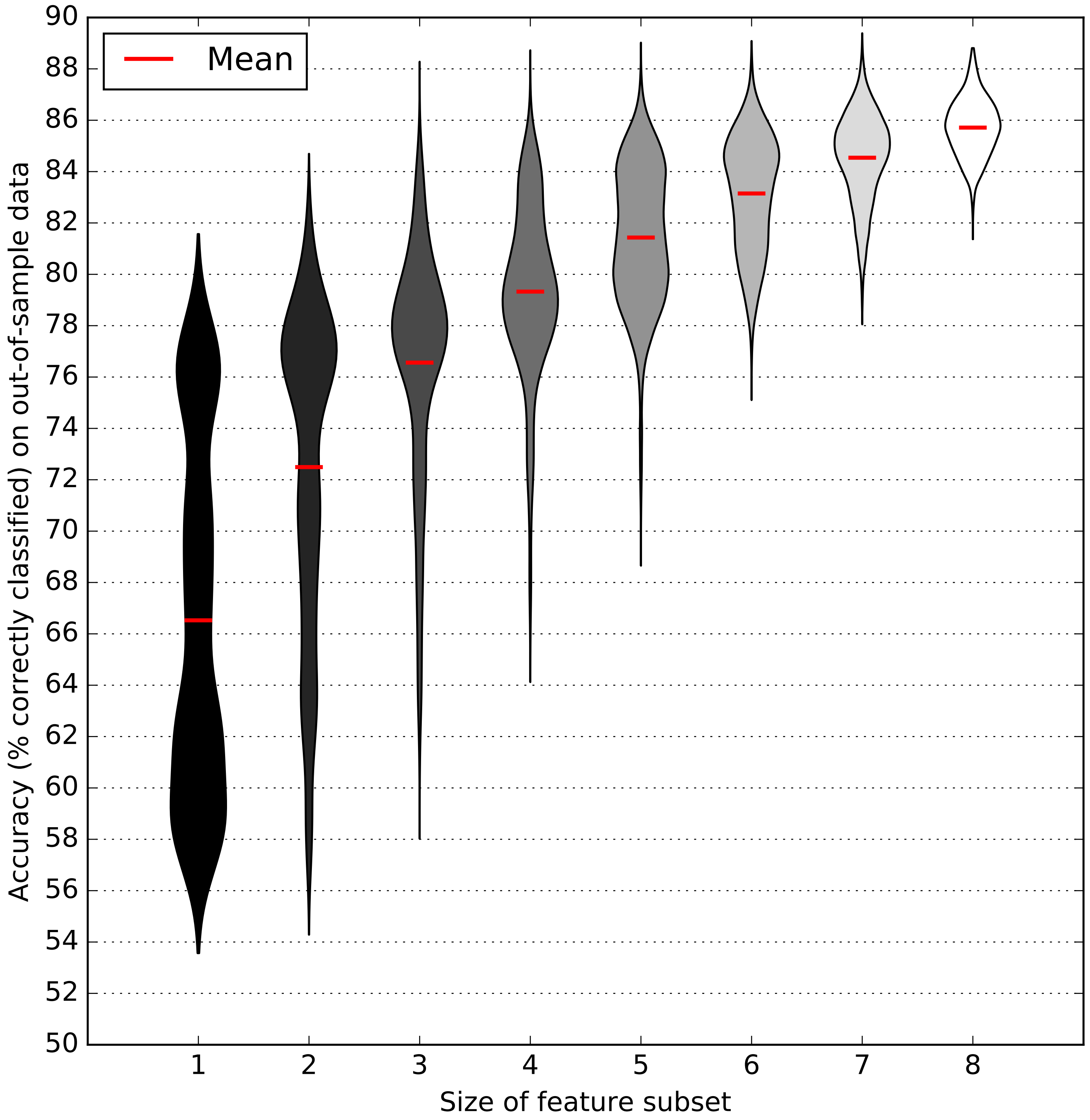
Process to Determine Feature Importance

<u>Brute force search</u>: For each possible feature subset*, repeat first experimental process with this subset



*(excluding volume and diameter features)

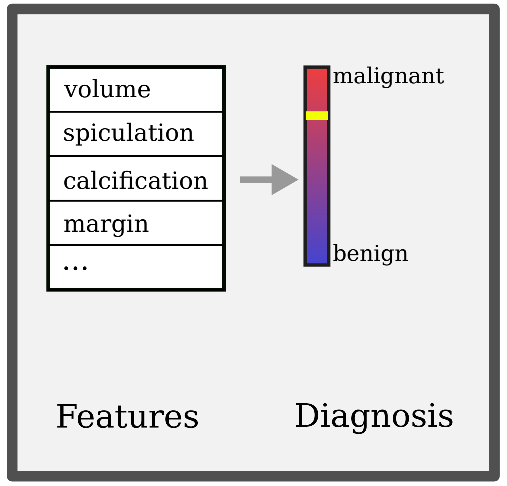
Results from Feature Importance Search



More Results from Feature Importance Search

	Single-Feature Accuracy	Percent Feature Significance	Geometric Mean	RF Feature-Importance
Best	(77.12%) Spiculation	(100.00%) Subtlety	(87.12%) Spiculation	(0.2173) Subtlety
	(75.56%) Lobulation	(99.21%) Calcification	(86.24%) Lobulation	(0.2147) Spiculation
	(70.90%) Margin	(98.43%) Spiculation	(82.04%) Subtlety	(0.1818) Lobulation
	(67.30%) Subtlety	(98.43%) Lobulation	(75.72%) Calcification	(0.1737) Calcification
	(63.01%) Texture	(83.46%) Texture	(75.46%) Margin	(0.1116) Margin
	(61.27%) Sphericity	(80.31%) Margin	(72.52%) Texture	(0.0529) Sphericity
	(59.26%) Internal Structure	(71.65%) Sphericity	(66.26%) Sphericity	(0.0437) Texture
Worst	(57.79%) Calcification	(62.20%) Internal Structure	(60.71%) Internal Structure	(0.0044) Internalstructure

Conclusions and Next Steps



- Last component of modular CAD approach is feasible (average AUC = 0.9492)
- Certain features more important for accurate algorithmic diagnosis (spiculation, lobulation, subtlety, calcification)
- Future work: 1. Robust lung nodule segmentation
  2. Medical image feature quantification