APPENDIX

Mean and Standard Deviation of Different Cartridges:

Appendix Table 1. HU Mean and Standard Deviation Value from Different Cartridges

Cartridge	Scanners	Mean	Standard Deviation
Rubber	GE STE (GE1)	-62.16	110.44
	GE 32 (GE2)	-66.67	90.50
	Definition AS (S1)	-80.66	103.13
	Siemens Sensation 64 (S2)	-83.18	105.04
	Siemens Sensation 40 (S3)	-84.71	107.90
	Siemens Sensation 16 (S4)	-76.23	102.03
	Philips Big Bore (P1)	-67.05	95.44
	Philips Brilliance 64 (P2)	-73.59	87.93
Dense-Cork	GE STE (GE1)	-677.88	35.45
	GE 32 (GE2)	-681.32	28.30
	Definition AS (S1)	-686.15	33.10
	Siemens Sensation 64 (S2)	-688.76	34.50
	Siemens Sensation 40 (S3)	-689.23	34.94
	Siemens Sensation 16 (S4)	-682.43	32.83
	Philips Big Bore (P1)	-677.02	32.10
	Philips Brilliance 64 (P2)	-679.35	29.75
Normal-Cork	GE STE (GE1)	-507.20	49.65
	GE 32 (GE2)	-508.74	43.51
	Definition AS (S1)	-512.36	47.19
	Siemens Sensation 64 (S2)	-516.28	47.58
	Siemens Sensation 40 (S3)	-518.11	48.29
	Siemens Sensation 16 (S4)	-511.47	46.12
	Philips Big Bore (P1)	-506.15	44.98
	Philips Brilliance 64 (P2)	-506.60	42.96

Prognostic Dataset Description:

De-identified data from the National Lung Screening Trial (NLST) was obtained via data access system of National Cancer Institute. NLST was a multi-institutional based lung cancer screening trial which spanned for three years: a baseline san (T0) and two follow-up scans one year apart. A nested case-control procedure with matched demographics from the NLSTs LDCT arm was used for this study, that included screen detected lung cancers (SDLC) and nodule positive controls (NPC).

170 SDLCs were deduced with a positive nodule at baseline (T0) and then the lung cancer was diagnosed at T1 (85 cases) and T2 (85 cases). A 2:1 nested case control was used to choose 328 NPCs which had nodules (never diagnosed as cancer) that were followed from T0 to T2 and had similar demographics as SDLCs. The SDLC and NPC were then divided into two Cohorts: Cohort1 (to train the machine learning model) and Cohort2 (a separate set to evaluate the trained model). Cohort1 had 85 SDLCs and 176 NPCs, whereas Cohort2 consisted of 85 SDLCs and 152 NPCs. The SDLCs which had a positive baseline (T0) screen and the positively screened nodule became diagnosed with lung cancer at the first follow-up screen (T1). While, in Cohort 2, the positively screened nodules during the baseline scan (T0) became malignant after two years (T2). Figure 10 describes the NLST study timeline as well as the criteria for dividing the SDLCs and NPCs into Cohort1 and Cohort2. More details about the dataset can be found in [23, 24].

Appendix Table 2. Comparing Different Classifiers for Malignancy prediction using NLST dataset

Classifiers	# of Features	Original Deep Features (ALL 4096 Features) *	Stable deep features-normalized by pixel area #
Naïve Bayes	Top 5	57.8 (0.61)	55.27 (0.68)
	Top 10	50.2 (0.6)	54.85 (0.65)
	Top 15	50.6 (0.57)	54.4 (0.64)
	Top 20	51 (0.57)	54.85 (0.65)
	Top 49 or All 49	45.14 (0.5)	54.85 (0.65)
Nearest Neighbor	Top 5	64.1 (0.6)	66.244 (0.62)
	Top 10	64.1 (0.6)	61.6 (0.62)
	Top 15	63.7 (0.6)	64.13 (0.62)
	Top 20	67.1 (0.62)	63.2 (0.62)
	Top 49 or All 49	64.97 (0.61)	65.4 (0.64)
Decision Tree	Top 5	57.4 (0.51)	64.55 (0.63)
	Top 10	64.1 (0.5	64.55 (0.63)
	Top 15	62.9 (0.52)	64.55 (0.64)
	Top 20	62.9 (0.52)	64.97 (0.6)
	Top 49 or All 49	62.86 (0.55)	63.29 (0.58)

*Top 49 features were chosen here using feature selector.

#All 49 features were stable deep features obtained from rubber cartridge.