Lung Nodule Malignancy Prediction in Sequential CT Scans: *Summary of ISBI 2018 Challenge*
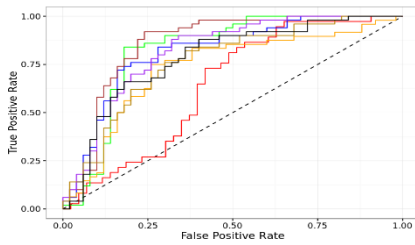
Yoganand Balagurunathan[1,*,#], Andrew Beers[2,#], Michael McNitt-Gray[3,#], Lubomir Hadjiiski[4,#], Sandy Napel[5,#], Dmitry Goldgof[6,#] (*IEEE Fellow*), Gustavo Perez[7], Pablo Arbelaez[7], Alireza Mehrtash[8,9], Tina Kapur[9], Ehwa Yang[10], Jung Won Moon[11], Gabriel Bernardino[12], Ricard Delgado-Gonzalo[13], M. Mehdi Farhangi[14,19], Amir A. Amini[14,15] (*IEEE Fellow*), Renkun Ni[16], Xue Feng[16,17], Aditya Bagari[18], Kiran Vaidhya[18], Benjamin Veasey[14,15], Wiem Safta[19], Hichem Frigui[19], Joseph Enguehard[20], Ali Gholipour[20], Laura Silvana Castillo[21], Laura Alexandra Daza[21], Paul Pinsky[22,#], Jayashree Kalpathy-Cramer[2,#], Keyvan Farahani[23,*,#].

[1]Dept. of Machine Learning, H Lee Moffitt Cancer Center (MCC), Tampa, FL. [2]Massachusetts General Hospital (MGH), MA. [3]University of California Los Angeles (UCLA), CA. [4]University of Michigan (UMICH), MI. [5]Dept. of Radiology, School of Medicine, Stanford University (SU), CA. [6]University of South Florida (USF), FL. [7]Biomedical computer vision lab (BCV), [21]Department of Biomedical Engineering, Universidad de los Andes, Colombia. [8]Robotics and Control Laboratory (RCL), Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC. [9]Surgical Planning Laboratory (SPL), Radiology Department, Brigham and Women's Hospital, Boston, MA, 02130. [10]Sungkyunkwan University School of Medicine, Seoul 06351, Korea. [11]Human Medical Imaging & Intervention Center, Seoul 06524, Korea. [12]Universitat Pompeu Fabra, Barcelona, Spain. [13]Centre Suisse d'Électronique et de Microtechnique, Neuchâtel, Switzerland. [14]Medical Imaging Laboratory, [15]Electrical and Computer Engineering Department, University of Louisville, Louisville, KY. USA. [16]Spingbok Inc. [17]Department of Biomedical Engineering, University of Virginia, Charlottesville. [18]Predible Health Inc, India. [19]Computer Engineering and Computer Science, University of Louisville. [20]Department of Radiology, Boston Children's Hospital, and Harvard Medical School. [21]Department of Biomedical Engineering, Universidad de los Andes, Bogota, Colombia. [22]Divsion of Cancer Prevention, [23]Center for Biomedical Informatics and Information Technology, National Cancer Institute (NCI), Washington DC.

[#]*Challenge organizing team*
Corresponding Authors: Drs. K. Farahani/ Y. Balagurunathan: farahank@mail.nih.gov/ yogab@moffitt.org.

Supplemental Figure SF1. Receiver operator characteristics of various estimates of change in volume and percent change in volume by different participants (reproduced from prior publication published in *Medical Phy. 45(3) 2018*).

Supplemental Figure SF2. ROC curve shows Volume change measurement from the teams compared to organizers estimates
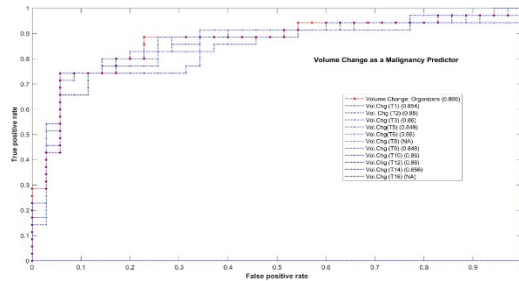

Table ST1. Nodule volume at individual time points (at time-t1 & t2) and change in volume as an estimator malignancy status. Performance of the team's volume estimates is compared to organizers volume and size measurements.

#	TEAM	AUC		
		Volume at time-t_1	Volume at time-t_2	Change in Volume (at time-t_1 & t_2)
1	Team T1	0.637	0.824	0.854
2	Team T2	0.674	0.843	0.85
3	Team T3	0.654	0.828	0.86
4	Team T5	0.675	0.834	0.849
5	Team T6	0.656	0.829	0.86
6	Team T8	0.793	0.793	**
7	Team T9	0.671	0.836	0.848
8	Team T10	0.656	0.829	0.86
9	Team T12	0.656	0.829	0.86
10	Team T14	0.674	0.838	0.856
11	Team T16	*	*	*
	Reference (Organizers)			
a	Volume	0.672	0.839	0.856
c	Nodule Size	0.694	0.824	0.76
d	**Volume Change** (or Vol. Chg) (Train on 30; Test on 70)			0.866

* Not reported
** Not available (same volume values reported for time -t1 & time-t2).

.

Supplemental Table ST2. Participating team's performance comparison. (a)
Table shows predictors range, optimal statistics (sensitivity, specificity), AUC with CI.

	Teams	Malignancy Scores Range (min, max)	Sensitivity	Specificity	AUC with CI
		Supplemental Table ST2. (a)			
1	T1	0.592(0.405-0.996)	0.857	0.886	0.913 [0.844,0.982]
2	T2	0.896(0.104-1)	0.714	0.971	0.897 [0.822,0.972]
3	T3	1(0-1)	0.714	0.943	0.879 [0.797,0.962]
4	T5	1(0-1)	0.714	0.971	0.868 [0.778,0.958]
5	T6	1(0-1)	0.714	1	0.867 [0.778,0.956]
6	T8	0.874(0.123-0.998)	0.686	0.943	0.855 [0.761,0.949]
7	T9	0.906(0.001-0.907)	0.571	1	0.854 [0.764,0.943]
8	T10	0.989(0-0.989)	0.686	0.914	0.848 [0.756,0.941]
9	T12	0.92(0-0.92)	0.743	0.886	0.809 [0.697,0.92]
10	T14	0.97(0.021-0.991)	0.629	0.914	0.789 [0.678,0.901]
11	T16	0.992(0-0.992)	0.857	0.486	0.698 [0.572,0.823]
12	Volume Change		0.743	0.943	0.866 [0.773,0.96]

Supplemental Table- ST3. Significance testing of participants predictors assessed using Delong's statistics.

A.	Delong's P-value											
	T1	T2	T3	T5	T6	T8	T9	T10	T12	T14	T16	VolChg
T1	0	0.6481	0.4221	0.1987	0.2470	0.0842	0.0867	0.0754	0.0237	0.0049	0.0031	0.1932
T2	0	0	0.6385	0.2136	0.4005	0.1946	0.1865	0.1331	0.0590	0.0120	0.0056	0.4124
T3	0	0	0	0.8090	0.7483	0.5626	0.5213	0.4426	0.1782	0.0527	0.0161	0.8023
T5	0	0	0	0	0.9818	0.6919	0.6902	0.5612	0.2034	0.0610	0.0253	0.9679
T6	0	0	0	0	0	0.6647	0.6436	0.6369	0.1990	0.0791	0.0211	0.9867
T8	0	0	0	0	0	0	0.97693	0.8586	0.2941	0.0773	0.0558	0.7978
T9	0	0	0	0	0	0	0	0.8792	0.3049	0.1257	0.0336	0.8064
T10	0	0	0	0	0	0	0	0	0.4226	0.1186	0.0536	0.6315
T12	0	0	0	0	0	0	0	0	0	0.6948	0.1754	0.3098
T14	0	0	0	0	0	0	0	0	0	0	0.2685	0.1466
T16	0	0	0	0	0	0	0	0	0	0	0	0.0413
VolChg	0	0	0	0	0	0	0	0	0	0	0	0

Supplemental Table- ST4. Teams re-ranking based on TPR (sensitivity) and FPR(1-Specificity) at different cut off values. A) Teams with highest TPR (sensitivity) for a fixed FPR (1-specificity), B) Teams with lowest FPR (1-specificity) for a fixed TPR (sensitivity).

	A. Ranking based on '**Highest**' TPR (Sensitivity) for a set FPR (1-Specificity)							
Rank	**FPR (1-Specificity)**							
	0.01	0.03	0.05	0.1	0.15	0.20	0.25	0.30
1	T6	T2 (#)	T2 (#)	T1 (#)	T1	T1	T1	T1
2	T9	T5 (#)	T5 (#)	T5 (#)	Vol.Chg	Vol.Chg	Vol.Chg	T2 (#)
3	T5	T6 (#)	T6 (#)	T2 ($)	T2 (#)	T2	T2 (#)	T8(#)
4	T2	T3 ($)	T3 ($)	Vol.Chg ($)	T3 (#)	T3 (#)	T3(#)	Vol.Chg (#)
5	T3	T9 ($)	T9 ($)	T3	T5 (#)	T5 (#)	T8(#)	T3
6	T8	T8	T8	T6	T6	T8 (#)	T5	T6
7	T1	T1	T1	T8	T8	T12(#)	T12	T10
8	T14	Vol.Chg	Vol.Chg	T12	T12	T6	T6	T12
9	T10	T14	T14	T10	T10	T10	T9	T5
10	Vol.Chg	T10	T10	T14	T9	T9	T10	T9
11	T12	T16	T16	T9	T14	T14	T14	T14
12	T16	T12	T12	T16	T16	T16	T16	T16

	B. Ranking based on '**Lowest**' FPR (1-Specificity) for a set TPR (sensitivity)

Rank	TPR (Sensitivity)							
	0.99	0.97	0.95	0.90	0.85	0.8	0.75	0.7
1	T1	T2	T2	T1	T1	T1	T1 (#)	T6
2	T6	T1	T1	T2	T2	Vol.Chg	T5 (#)	T2 (#)
3	T10	T5 (#)	T5 (#)	T3	T3 (#)	T8	T2 ($)	T5 (#)
4	T2	T9 (#)	T9 (#)	T9	T9 (#)	T2	T3 ($)	T1 ($)
5	T3 (#)	T3 ($)	T3 ($)	T10	T10 (#)	T3 ($)	Vol.Chg	T3 ($)
6	T8 (#)	T16 ($)	T16 ($)	Vol.Chg	Vol.Chg	T10 ($)	T8	Vol.Chg
7	T9 (#)	T10	T10	T5	T5	T5	T12	T8
8	T14 (#)	T6	T6	T6	T8	T9	T6	T12
9	T16 (#)	Vol.Chg	Vol.Chg	T16	T6	T6	T9	T10
10	T5	T8	T12	T12	T12	T12	T10	T9
11	T12	T14	T8	T8	T16	T14	T14	T14
12	Vol.Chg	T12	T14	T14	T14	T16	T16	T16

__Note:__ We highlight same predictive values (TPR or FPR) on the column side, in the tables using a symbol in the list {#, $}. Where teams with same symbol type in the parenthesis will have same predictive values, in their respective columns. Not to be compared across the columns.

Supplemental Table- ST5. Partial AUC to compare the performance of the participating team's predictor contrasted with the volume change. A) different intervals of FPR B) different intervals of TPR.

a)

Teams	Partial Area (*pAUC*)							
	FPR intervals with TPR [0,1]							
	≤0.05	≤0.10	≤0.15	≤0.20	≤0.25	≤0.3	≤0.35	≤0.4
T1	0.0106	0.0465	0.0931	0.142	0.1665	0.2171	0.271	0.3249
T2	0.0139	0.0555	0.0988	0.1429	0.1657	0.2114	0.2612	0.3135
T3	0.0131	0.0498	0.0922	0.1363	0.1584	0.2041	0.2506	0.3012
T5	0.0155	0.0571	0.1012	0.1453	0.1673	0.2114	0.2555	0.3037
T6	0.0204	0.0612	0.1029	0.1453	0.1665	0.2098	0.2555	0.302
T8	0.0131	0.0482	0.0898	0.1322	0.1543	0.2008	0.2482	0.2963
T9	0.0163	0.0498	0.0841	0.1224	0.1429	0.1861	0.231	0.2808
T10	0.0057	0.0343	0.0735	0.1135	0.1347	0.1771	0.2237	0.2735
T12	0.0008	0.0188	0.06	0.1024	0.1245	0.1686	0.2139	0.2367
T14	0.0082	0.0384	0.0743	0.111	0.1314	0.1731	0.2155	0.258
T16	0	0.0139	0.0343	0.058	0.0702	0.0963	0.1257	0.1576
Volume Change	0.0041	0.04	0.0824	0.1282	0.1518	0.1992	0.2465	0.2963

b)

Teams	Partial Area (*pAUC*)							
	TPR intervals with FPR [0,1]							
	≥0.95	≥0.90	≥0.85	≥0.80	≥0.75	≥0.7	≥0.65	≥0.6
T1	0.0188	0.0527	0.0912	0.1355	0.1804	0.2273	0.2745	0.3216
T2	0.0153	0.0449	0.0792	0.1143	0.1561	0.2033	0.2518	0.3004
T3	0.0108	0.0384	0.07	0.1037	0.1447	0.1898	0.2369	0.2841
T5	0.009	0.0294	0.0543	0.0865	0.1249	0.1722	0.2208	0.2694
T6	0.0114	0.0318	0.0561	0.0857	0.122	0.1669	0.2169	0.2669
T8	0.0084	0.0233	0.0486	0.0833	0.1224	0.1665	0.2133	0.2604
T9	0.0114	0.0331	0.0643	0.0971	0.1327	0.1714	0.2127	0.2555
T10	0.0118	0.0359	0.0659	0.0996	0.1353	0.1751	0.2196	0.2653
T12	0.0049	0.0202	0.04	0.069	0.1061	0.15	0.1955	0.2412
T14	0.0071	0.0167	0.0339	0.0563	0.082	0.1184	0.1586	0.2024
T16	0.0108	0.029	0.0484	0.0727	0.0969	0.1224	0.151	0.1804
Volume Change	0.0057	0.0208	0.0504	0.0873	0.1302	0.1767	0.2239	0.0057

(Complete Suppl, section version: : https://dx.doi.org/10.25833/8818-b129 or https://ncihub.org/publications/274/1)