10.1148/ryai.2019180066

Appendix E1. GrayNet: Data Preparation and Pretrained Model Generation

GrayNet contains CT scans from 322 patients (158 F, 164 M) who underwent contrast-enhanced of head, neck, chest, abdomen and pelvis. on two scanners-Discovery ST, GE Healthcare (April 2016-July 2017) and Biograph 64, Siemens Healthcare (May-July 2012). These scans comprised of 120,349 axial sections. A radiologist (R2; J.H.L., 5 years of experience) established 23 overlapping anatomic labels and annotated all axial sections (Fig E1). Details on distribution of dataset in GrayNet are tabulated in Table E1. Noncompressed, full resolution (512 × 512 pixels), 8-bit grayscale images were created at three window levels (WL) and window widths (WW) that were encoded into three different channels: abdomen window (WL = 40HU, WW = 400 HU) for the red channel, bone window (WL = 300HU, WW = 1500HU) for the green channel, and lung window (WL =-500HU, WW = 1400HU) for the blue channel. To improve model generalization, training data were augmented by applying the same geometric transformations used for developing stone detection classification models described in Materials & Methods (subsection: "Network Training"). Additionally, to enhance model variance to changes in contrast enhancement, mathematically-processed unenhanced images were generated from corresponding contrast-enhanced CT images by suppressing the contrast enhanced pixel values using a density-dependent conversion function (Fig E2). The conversion function was generated by fitting a set of original HU values X = {150, 200, 201, 202, …, 400, 450} to a set of target HU values Y = {150, 90, 91, 92, ..., 290, 450} with a cubic spline interpolation using a function named interpolation.interp1d available in scipy v1.1.0 (41), a scientific and technical computing Python library. Though not ideal, this approach was used as a viable option to augment data without affecting anatomic labeling. To generate the GrayNet pretrained model, 40 cases (10 from each scanner and gender) were selected for use as validation dataset and the rest were used as training dataset. GrayNet models were trained for a total of 60 epochs using an ADAM optimizer with a mini batch size of 64 and a weight decay of 5 × 10–5. A base learning rate of 0.001 was decayed by a factor of 10 three times to minimize training loss, which was the sum of cross-entropy losses for 23 binary classification outputs. The best model was selected based on validation loss and reserved as a pretrained model to be used for weight initialization of the models for other tasks.

Appendix E2. Data distribution from 535 patients for CNN-1 (identification of urinary tract region) and CNN-2 (detection of stone) in terms of number of 2-D axial sections

No. of sections (CNN-1) *Total No. = 80,620*	**Scanner 1**			**Scanner 2**		
	Development	**Test**	**Total**	**Development**	**Test**	**Total**
Outside urinary tract (% number of total)	5,580 (6.9%)	1,086 (1.3%)	6,666 (8.3%)	14,720 (18.3%)	3,644 (4.5%)	18,364 (22.8%)
Within urinary tract (% number of total)	14,417 (17.9%)	2,964 (3.7%)	17,381 (21.6%)	30,322 (37.6%)	7,887 (9.8%)	38,209 (47.4%)
Total (% number of total)	19,997 (24.8%)	4,050 (5.0%)	24,047 (29.8%)	45,042 (55.9%)	11,531 (14.3%)	56,573 (70.2%)

No. of sections (CNN-2) *Total No = 25,978*	Scanner 1			Scanner 2		
	Development	**Test**	**Total**	**Development**	**Test**	**Total**
No stone (% number of total)	7,991 (30.8%)	1,473 (5.7%)	9,464 (36.4%)	10,976 (42.3%)	3,831 (14.7%)	14,807 (57.0%)
Stone (% number of total)	494 (1.9%)	139 (0.5%)	633 (2.4%)	866 (3.3%)	208 (0.8%)	1,074 (4.1%)
Total (% number of total)	8,485 (3.3%)	1,612 (6.2%)	10,097 (38.9%)	11,842 (45.6%)	4,039 (15.5%)	15,881 (61.1%)

Appendix E3. Overview of the preprocessing pipeline is illustrated in Figure E3.

Appendix E4. Distribution of patients according to presence of phleboliths in each dataset group.

	Scanner 1				Scanner 2			
	Train	Validation	Test	Total	Train	Validation	Test	Total
No Stone	63/118	7/15	18/25	88/158	32/58	9/15	14/25	55/98
Stone	45/91	13/15	16/25	74/131	64/108	10/15	11/25	85/148
Total	108/209	20/30	34/50	162/289	96/166	19/30	25/50	140/246

Note.—Numerator depicts number of patients with phleboliths and denominator depicts total number of patients.

Reference

41. Jones E, Oliphant T, Peterson P. {SciPy}: Open source scientific tools for {Python}. http://www.scipy.org. Published 2001. Accessed February 14, 2019

Table E1. Data characteristics and anatomic label distribution of scans in GrayNet

	Discovery ST (GE Healthcare)				Biograph 64 (Siemens Healthcare)			
Period	April 2016-July 2017				May-July 2012			
No. of cases	151				171			
No. of sections	48,784				71,565			
Patient Sex	Female: *n* = 72		Male: *n* = 79		Female: *n* = 86		Male: *n* = 85	
Patient Age (Mean ± SD)	62 ± 15 years				61 ± 15 years			
Anatomic Label Distribution and *No. of sections (% number of total)*	L1	3,393 (7.0%)	L2	2,189 (4.5%)	L1	5,232 (7.3%)	L2	3,428 (4.8%)
	L3	1,316 (2.7%)	L4	1,717 (3.5%)	L3	1,954 (2.7%)	L4	2,383 (3.3%)
	L5	1,852 (3.8%)	L6	6,080 (12.5%)	L5	2,926 (4.1%)	L6	8,408 (11.7%)
	L7	2,228 (4.6%)	L8	1,649 (3.4%)	L7	2,745 (3.8%)	L8	2,249 (3.1%)
	L9	1,730 (3.5%)	L10	5,669 (11.6%)	L9	2,440 (3.4%)	L10	8,054 (11.3%)
	L11	15,758 (32.3%)	L12	4,622 (9.5%)	L11	21,156 (29.6%)	L12	6,983 (9.8%)
	L13	5,469 (11.2%)	L14	6,313 (12.9%)	L13	7,514 (10.5%)	L14	9,384 (13.1%)
	L15	6,216 (12.7%)	L16	5,540 (11.4%)	L15	8,894 (12.4%)	L16	8,072 (11.3%)
	L17	2,813 (5.8%)	L18	4,132 (8.5%)	L17	4,608 (6.4%)	L18	5,458 (7.6%)
	L19	13,062 (26.8%)	L20	2,097 (4.3%)	L19	18,398 (25.7%)	L20	3,217 (4.5%)
	L21	1,833 (3.8%)	L22	3,903 (8.0%)	L21	2,420 (3.4%)	L22	4,789 (6.7%)
	L23	6,722 (13.8%)			L23	10,794 (15.1%)		