		General		
	Number	**Name**	**Link**	**Comment**
Viable				
	d3	COVID-19 Lung CT Scans	https://www.kaggle.com/luisblanche/covidct	Parsed from Papers
	d5	SARS-COV-2 Ct-Scan Dataset	https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset	
	d7.1	HUST-19 Slice Based Dataset	http://ictcf.biocuckoo.cn/HUST-19.php	Same as d7
	d8.1	COVID-CTset : A Large COVID-19 CT Scans dataset		Small version of 8
Variants				
	d7	CT Scans for COVID-19 Classification	https://www.kaggle.com/azaemon/preprocessed-ct-scans-for-covid19	Same as d7.1
	d7 pp	CT Scans for COVID-19 Classification (Pre Processed)		d7 but only lung is visible (other areas are blacked out)
	d7.2	HUST-19 All Raw Scans	http://ictcf.biocuckoo.cn/HUST-19.php	Raw Data not labeled for individual slices
	d8	COVID-CTset : A Large COVID-19 CT Scans dataset	https://www.kaggle.com/mohammadrahimzadeh/covidctset-a-large-covid19-ct-scans-dataset	
Out				
	d1	COVID-19 CT scans	https://www.kaggle.com/andrewmvd/covid19-ct-scans	
	d2	CT Medical Images	https://www.kaggle.com/kmader/siim-medical-images	
	d4	OSIC Pulmonary Fibrosis Progression	https://www.kaggle.com/c/osic-pulmonary-fibrosis-progression/data	Competition Data, needs registration
	d6	Finding and Measuring Lungs in CT Data	https://www.kaggle.com/kmader/finding-lungs-in-ct-data	Just for segmentation or measurements

		Image Data					*Data Type Metrics*			*Expert Assessment of Images*				
	Number	**# Images**	**# Individual Scans**	**#Patients**	**Number of Classes**	**Class Names**	**Raw or Recons**	**File Type**	**Resolution**	**Sample Size**	**Sample %**	**Flawed Images**	**Flawed %**	**Image Quality (0-10)**
Viable														
	d3	746		216	2	covid, no_covid	reconstructed	jpg	mixed	150	20.11%	71	47.33%	3 [1]
	d5	2482		120	2	covid, no_covid	reconstructed	png	mixed	294	11.85%	19	6.46%	7 [4]
	d7.1	19685		104	3	covid, no_covid, no_lung	reconstructed	jpg	512x512	985	5.00%	23	2.34%	8 [7]
	d8.1	12058			2	covid, no_covid	raw	tiff, 16bit uint	512x512	603	5.00%	75	12.44%	6 [11]
Variants														
	d7	19685		104	3	covid, no_covid, no_lung	reconstructed	jpg	512x512					
	d7 pp	19685		104	3	covid, no_covid, no_lung	reconstructed	jpg	512x512					
	d7.2	288498		1521	2	control, type_i, type_ii	raw	dcm						
	d8	63849	377	377	2	covid, no_covid	raw	tiff, 16bit uint	512x512					
Out														
	d1	6020	20	20?	1	corona_lung	reconstructed	Nifti	512x512					
	d2	100	475	69	2	contrast, no_contrast		dcm, tiff						
	d4													
	d6				1									

						Additional Features			*Dataset Usability*			**Reference Model**
	Number	**Wrong Labels**	**Wrong Label %**	**Label Quality (0-10)**	**Consistency**	**Segmentation**	**Patient Metadata**	**Patient History**	**Dataset Description**	**Paper**	**Sample Code**	Availability
Viable												
	d3	26	17.33%	6 [2]	8 [3]	no	no	no	yes	article	yes	yes
	d5	4	1.36%	8 [5]	8 [6]	no	no	no	yes	yes	yes	yes
	d7.1	7	0.71%	7 [8]	8 [9]	no	no	no	yes	yes	yes	yes
	d8.1	0	0.00%	10 [12]	8 [13]	no	age, sex	no	yes	yes	yes	yes
Variants												
	d7					no	no	no	yes	yes, about data		
	d7 pp											
	d7.2					no	yes	some	yes	kinda	kinda	kinda
	d8					no	age, sex	no	yes	yes	yes	yes
Out												
	d1					Lung Mask, Infection M	no	no	yes	no	yes	no
	d2					Age, Contrast Type	yes	no	yes	no	no	yes
	d4											
	d6											

							Verdict		
	Number	Model Architecture	Code	LoC	Used Framework		Usable	**Score**	***Reason***
Viable									
	d3	yes	yes	717	pytorch		yes	2	Low Quality Images, but heterogeneous
	d5	xDNN	yes		matlab		yes	3	Code in Matlab, xDNN is not standard
	d7.1	VGG [10]	yes	100-300	keras, tensorflow, scikit-learn		yes	7	Model peforms at around 80-90 % accuracy
	d8.1	ResNet50v2+FPN, VGG [1	yes	420	tensorflow, keras		yes	8	Model performs at around 95-98%
Variants									
	d7						yes	7	
	d7 pp								
	d7.2	kinda					yes, but additional work required	7	Raw Labels need to be added
	d8	yes [15]	yes	420	tensorflow, keras		yes	8	
Out									
	d1						no		Not suitable for classification
	d2	yes	yes		keras		no		Classification for Contrast is not disease classification
	d4						no		Not public
	d6						no		Only suitable for a segmentation use case

[1] Out of 150 randomly selected images from the complete dataset, 48 contain external markings such as arrows.
20 are extremely blurry and 3 contain heavy metal artifacts.

Score 3/10

[2] Of 150 randomly selected images, 26 out contain abnormal lung parenchyma but are unclear Covid.

Score 6/10

[3] Different Images Sizes, minor diffrerences in Patient positioning. Overall examination identical

Score 8/10

[4] Out of 294 randomly selected images 4 contain external markings like arrows, 11 are noticeably blurry and 4 have extreme image contrast.

Score 7/10

[5] 4 out of 294 randomly selected images are labeled positive but contain no traces of Covid.

Score 8/10

[6] Different Images Sizes, minor diffrerences in Patient positioning. Overall examination identical

Score 8/10

[7] Out of 985 randomly selected images, 23 are slightly noisy and 28 contain external metal objects but without image artifacts.

Score 8/10

[8] 5 out of 985 randomly selected images are labeled NiCT (no or not enough lung parenchyma) but show lung tissue as well as traces of Covid (Score -2 because of this alone). 2 are labeled positive but show no traces.

Score 7/10

[9] Different Images Sizes, minor diffrerences in Patient positioning. Overall examination identical

Score 8/10

[10] Slice centered: VGG
Patient centered: HUST19, Inception, ChexNet

[11] Out of 603 selected images of a random sample 35 are noisy and 40 contain external metal objects, that lead to minor metal artifacts.

Score 6/10

[12] None of the 603 selected images were labeled incorrectly

Score 10/10

[13] Different Images Sizes, minor diffrerences in Patient positioning. Overall examination identical

Score 8/10

[14] ResNet50v2+FPN
ResNet50v2
VGG

[15] ResNet50v2+FPN
ResNet50v2
VGG