

Classification of Hyper-scale Multimodal Imaging Datasets

Craig MacFadyen¹ Ajay Duraiswamy¹ David Harris-Birtill^{1*}

¹University of St Andrews, St Andrews, United Kingdom

Abstract

Algorithms that can classify hyper-scale multi-modal datasets, comprising of millions of images, into constituent modality types can help researchers quickly retrieve and classify diagnostic imaging data, accelerating clinical outcomes. This research aims to demonstrate that a deep neural network that is trained on a hyper-scale dataset (4.5 million images) composed of heterogeneous multi-modal data, can be used to obtain significant modality classification accuracy (96%). By combining 102 medical imaging datasets, a dataset of 4.5 million images was created. A ResNet-50, ResNet-18 and VGG16 were trained to classify these images by the imaging modality used to capture them (Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and X-ray) across many body locations. The classification accuracy of the models was then tested on unseen data.

The best performing model achieved classification accuracy of 96% on unseen data. The model achieved a balanced accuracy of 86%.

This research shows it is possible to train Deep Learning (DL) Convolutional Neural Networks (CNNs) with hyper-scale multimodal data-sets, composed of millions of images. The trained model can be used to classify images by modality, with the best performing model achieving a classification accuracy of 96%. Such models can find use in real-world applications with volumes of image data in the hyper-scale range, such as medical imaging repositories, or national healthcare institutions. Further research can expand this classification capability to include 3D-scans.

1 Introduction

With the proliferation of deep neural networks trained on heterogenous multimodal data to detect and predict diseases, there has been an explosion in the volume of diagnostic medical imaging data [1]. Clinicians often order multiple scans of the same patient in different modalities to gather evidence to make improved diagnosis/prognosis [2]. Algorithms that can accurately classify a large heterogenous dataset into its constituent

modalities can be beneficial to researchers and clinicians, allowing them to automatically segment a particular

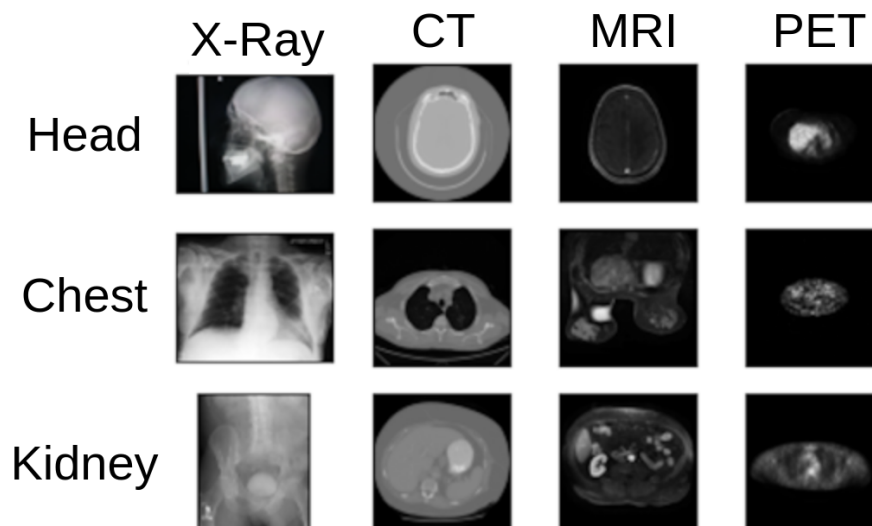
NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

*Corresponding author email: dcchb@st-andrews.ac.uk

29 type of modality for retrieval, archival, data balancing, and diagnostic purposes. Manual methods for classifying
 30 medical images are typically error-prone unless done by costly domain experts [3].

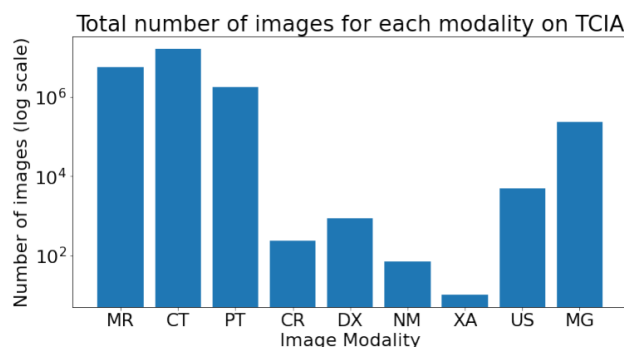
31 This paper outlines an deep neural network that accurately classifies a hyper-scale (4.5 million images),
 32 mixed-modality dataset into constituent modalities. The developed approach has significant benefit potential
 33 for researchers, clinicians, and imaging archives by helping effectively and efficiently classify diagnostic imaging
 34 data, in the magnitude of real-world volumes. While classification of hyperscale datasets have been attempted in
 35 other areas, such as Earth-science[4], including studies of plankton and marine snow [5], and XYZ, the proposed
 36 approach is novel in the field of classification of medical imaging modalities. This study aims to stimulate other
 37 hyper-scale projects in this area.

Figure 1: Visualisation of a spread of images from different locations in different modalities. Different modalities use different kinds of radiation, and these are absorbed to varying degrees by tissue in the human body. This leads to the same tissue looking different in each modality. Examples of modalities showing variation of the same tissue: [6] [7] [8] [9] [10] [11] [12] [13] [14] [14] [15]



38 Multiple open-access data sets were used to build the hyper-scale multimodal dataset of 4.5 million images
 39 from sources such as The Cancer Imaging Archive [16], Stanford ML Group[17] the largest of which contains
 40 262,000 chest X-ray images, and Kaggle [18] host labelled datasets.

Figure 2: XXX



41 The models trained on this hyper-scale multimodal dataset were a ResNet-18,23 ResNet-50 and a VGG16.
42 When these models were tested for classification accuracy, these results are in the high 90%'s across the train,
43 validate and test sets which shows that the models are able to classify with significant accuracy. Table 1 shows
44 the accuracy and balanced accuracy of each of the models on the test set. Figure 6 shows the confusion matrix
45 for the ResNet18, the best performing of the models tested in this study. The confusion matrix shows that the
46 model demonstrates significant classification performance (96%+) on classifying CT, MR and PET modalities.

47 1.1 Previous Literature

48 A number of research articles focus on deep learning systems to classify modalities in diagnostic imaging data.
49 However, to the best of our knowledge, there have not been any examples of a system that combines medical
50 imaging datasets at the hyper-scale (millions of images) level to perform modality classification.

51 Approaches to classifying medical imaging data by modality primarily take two forms (1) hand-crafted
52 features, and (2) Deep Learning.

53 The early approaches were based on hand-crafted features, such as picking a specific texture and colour[19],
54 SIFT descriptors[20], bag-of-colours[21] and then using SVM [22], KNN [23] as the classifier[24]. These ap-
55 proaches were limited by the choice of features, and limited accuracy[3]. Further, typically high computational
56 costs inherently limit the size of the datasets used.

57 A number of approaches using deep learning classifiers are seen in literature. However, all approaches
58 reviewed were seen to be utilising a limited dataset volumes with sizes in the lower order of magnitude, typically
59 hundreds to thousands ($10^2 - 10^3$) of images. Therefore, real-world classification performance of these algorithms
60 when operated on typical image-repository scales of millions of images is unknown.

61 Chiang et al. use a dataset of 2,878 images to train a CNN classifier on 4 modalities[25], Abdominal CT,
62 Brain CT, Lumbar Spine MRI, and Brain MRI, achieving an average validation accuracy of $> 99.5\%$. Cheng
63 et al. use a cascaded CNN to classify a bimodal dataset, comprised of MRI and PET images[26]. Using a
64 dataset in the order of 10^2 images, they achieved a classification accuracy of 89.6%. Yu et al. use a DNN, and
65 a dataset from the ImageCLEF database, comprising of 2,901 training and 2,582 test images to demonstrate
66 a best classification accuracy of 70%[27]. Sevakula et al. use transfer learning to compare performance of
67 seven DCNNs[28]. Using a curated dataset of 5,500 images from the Open-i Biomedical Image Search Engine,
68 they achieve a best classification accuracy of 99.45% on the Inception-V3 network. Finally, Trenta et al. use
69 a dataset comprised of 8,500 slices and a test set of 1,320 slices (split across 5 classes), and transfer learning
70 techniques to achieve an overall accuracy of upto 100% on specific modalities, on their pre-trained VGGNet
71 implementation[24].

72 Summarising classification performance figures reported in extant literature:

73 To summarise, two things are evident, (1) deep learning learning models present several advantages over
74 handcrafted, feature driven models, and (2) it is seen that the largest of the datasets in the literature reviewed
75 is in the order of 10^3 images. Given that image repositories are now typically in the hyper-scale order, and
76 growing rapidly, a suitably trained CNN capable of handling hyper-scale datasets is required.

Table 1: Dataset sizes vs Performance

Study	Dataset	Classifier
	Magnitude	Accuracy
Chiang et al.[25]	10^3	> 99.5%
Cheng et al.[26]	10^2	> 89.6%
Yu et al.[27]	10^3	70%
Sevakula et al.[28]	10^3	99.45%
Trenta et al.[24]	10^3	100%

2 Materials & Methods

2.1 Data

In total, 102 datasets were downloaded and combined to form an hyper-scale image dataset of 4.5 million images. The full list of datasets with citations is provided in Appendix A. Four modalities were selected as targets for the classification task; CT, MRI, X-ray and PET. Other modalities (e.g. ultrasound) were excluded from this study because of a lack of appreciable volumes of data. The main source of this data was the Cancer Imaging Archive (TCIA) [16]. The Cancer Imaging Archive provides a REST API that allows for programmatic retrieval of images which allowed data to be downloaded and combined easily, and in a reproducible way. However, because the Cancer Imaging Archive’s main purpose is to host datasets relating to cancer research it was important to seek out some extra datasets to augment the data TCIA provides. The full list of datasets can be found in Appendix A.

This project was approved by the University of St Andrews University Teaching and Research Ethics Committee (UTREC), approval code CS15171.

2.2 Train-Validate-Test Split

The downloaded data was split into three separate parts - train, validate and test. The train set was used to train the model, the validate set was used to evaluate the models between training runs, and the test set was used once to evaluate the final trained models. It was important to create the splits at the dataset level to prevent data-leakage. That is, all the images from a dataset were placed in the same split. Scans of the same patient in the same modality are likely to be similar, so if there is an image of the same patient in the train and test set then the test set does not contain completely unseen data. Putting each dataset into one of train, validate or test prevents this data leakage. Splitting the datasets like this also helps achieve the goal of demonstrating generalisation across datasets, because no dataset in the train set is represented in the test set.

The train-validate-test split was created manually to ensure as even a spread as possible of images for each modality and location in each split. The manual split ensured that there are at least two locations for each

101 modality in each of the train, validate and test split. The main difficulty for this was X-rays, because in the
102 TCIA datasets most X-rays are mammograms. This meant the non-TCIA datasets had to be carefully split.
103 Again, Appendix A shows the split each dataset was placed in. Figure 3c shows the number of images in the
104 train, validate and test set. TCIA hosts many CT and MR datasets and some of these datasets are very large.
105 For example, the CT Colonography dataset [29] has more than 900,000 CT images, which is more than the
106 total number of X-ray images across all datasets used in this study. To ensure the other modalities were not
107 completely dwarfed by these datasets, a maximum of 50,000 CT images and 100,000 MR images was taken from
108 each individual dataset. The images were selected in the order given by TCIA. This selection method was not
109 applied to the images from sources other than TCIA. After imbalance correction, the total number of images
110 in the dataset were 6,433,838 (6.4 million images), with a split of 4,104,184 in training, 936,347 in test, and
111 1,393,307 in validate datasets.

112 2.3 Preprocessing

113 In order for 2D and 3D scans to be used in the same study, the 3D scans (CT, MR and PT) were treated as a
114 collection of 2D images. These images are sometimes referred to as slices. The images were resized to 224×224
115 and rescaled between 0 and 1. Each image was rescaled using min-max normalisation with the maximum and
116 minimum values being the highest and lowest pixel values present within the image.

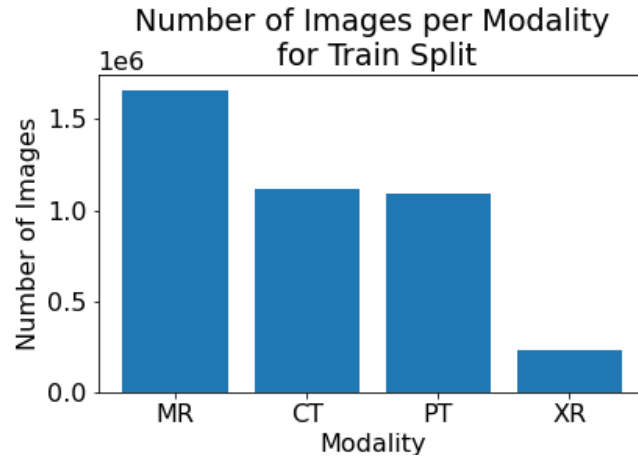
117 2.4 Network Architecture and Training

118 The models trained on this dataset were a ResNet-18 [30], ResNet-50 and a VGG16 [31]. The code used was
119 adapted from PyTorch's hosted versions of these models[32]. Changes were made to the channel depth of the
120 input layer, from three channels to one channel (grayscale). These three models were chosen because they have
121 all been shown to perform well when trained with large quantities of data on the ImageNet dataset [30, 31].
122 The code created as part of this research is open-source and hosted online at Github [33].

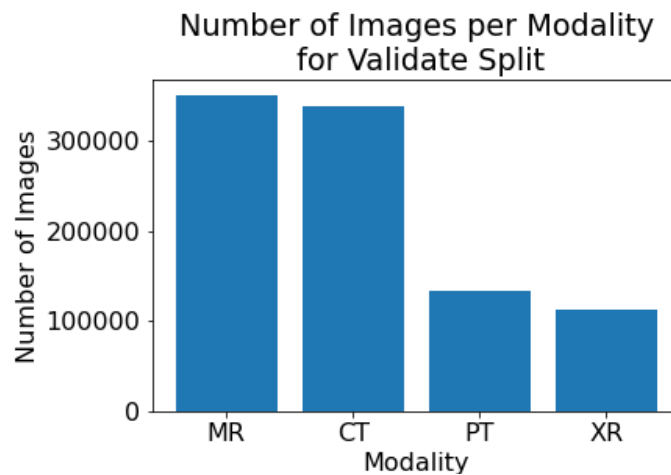
123 All models were trained for 10 epochs with a batch size of 128. The training set contained 2,954,097 (2.9×10^6)
124 samples and the validate set contained 704,685 samples. The models were optimised using stochastic gradient
125 descent, with a learning rate of 0.1 that was divided by 10 every time the loss plateaued, a momentum of 0.9
126 and an L2 weight decay penalty of 0.005. The models were trained on a machine with an Intel(R) Xeon(R)
127 CPU E5-1650 v4 @3.60GHz with 6 physical cores (12 threads), 250GB of RAM and two Nvidia GeForce GTX
128 1080Tis.

Figure 3: Figures showing the number of images for each modality in the created splits: a) train, b) validate and c) test. Note that each graph has a different scale, the purpose is to show the ratios of each class are similar. There are 73 datasets in the train set, 13 in the validate set and 16 in the test set.

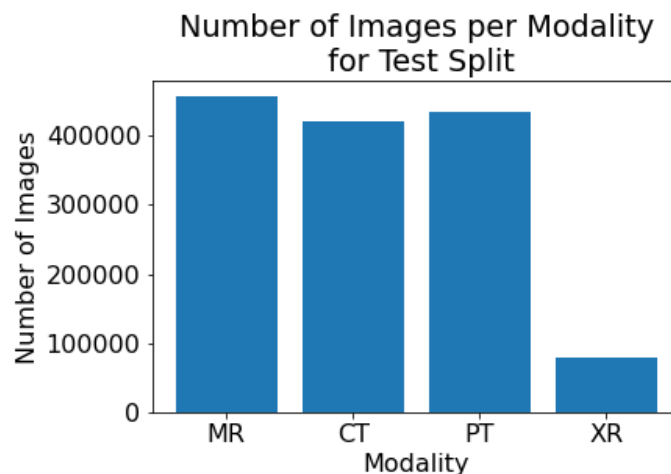
(a) Number of images for each modality in the created train set.



(b) Number of images for each modality in the created validate set.



(c) Number of images for each modality in the created test set.

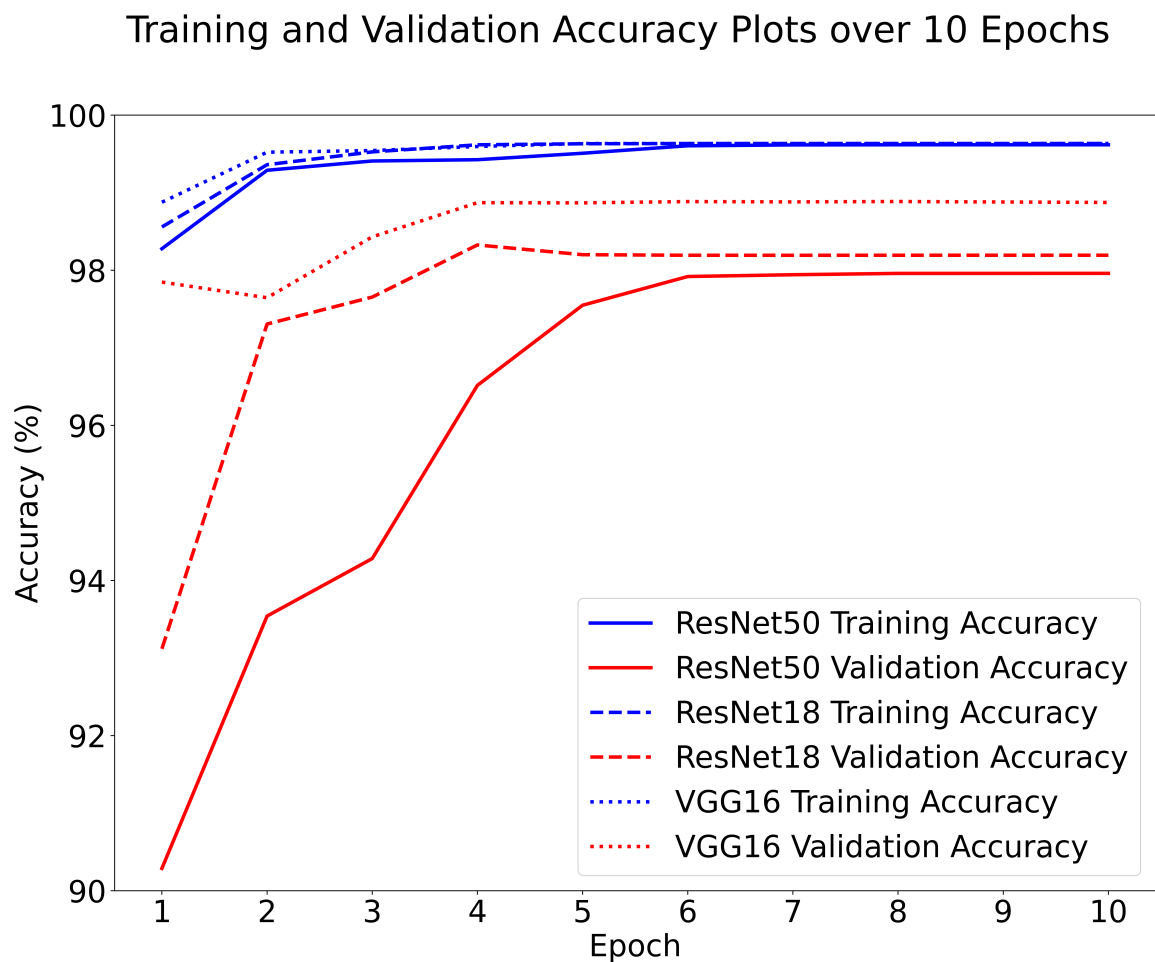


129 3 Results & Discussion

130 3.1 Training and Validation Accuracy

131 Figures 4, shows the training and validation accuracy curves for the ResNet50, ResNet18 and VGG16 models.
132 The small gap between the training and validation accuracies suggests that the models are not overfitting.
133 Figure 5 shows the time it took to train the models over the 10 epochs.

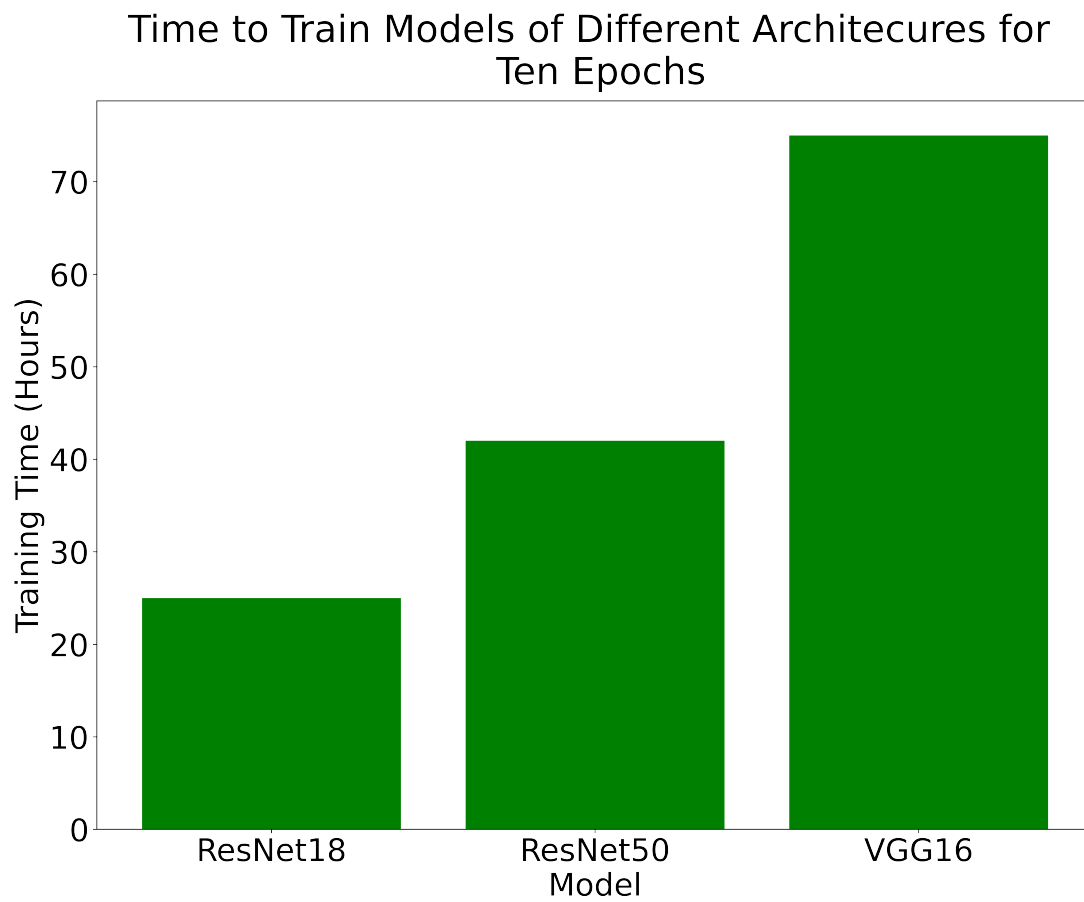
Figure 4: Training and validation accuracy each of the three networks, found at the end of each epoch. The small gap between the training and validation accuracies suggests that the models are not overfitting. Note the scale starts at 90%.



134 3.2 Test Set Accuracy

135 Figure 6 shows the accuracy of the three models. These results are in the high 90%'s across the train, validate
136 and test sets which shows that the models have all learned the problem well. Table 2 shows the accuracy and
137 balanced accuracy of each of the models on the test set. Figure 7 shows the confusion matrix for the ResNet18
138 model. The confusion matrix shows that the model performs very well on CT, MR and PET. Accuracy for
139 X-rays can be improved by adding additional X-ray images across a larger spread of locations.

Figure 5: Time in hours to train the models for 10 epochs. The training and validation accuracy both level-off around epochs 5–6 which shows that the models are able to fit the data.



Model	Accuracy	Balanced Accuracy
ResNet18	96.00%	86.17%
ResNet50	95.60%	85.65%
VGG16	94.58%	81.08%

Table 2: Table containing the accuracy and balanced accuracy of various models on the test set. Each model was trained for 10 epochs.

140 3.3 Dataset Level Results

141 Table 3 shows the accuracy of the model on each dataset in the test set for the ResNet18 model, chosen because
142 this model demonstrated superior classification performance over others tested in this study. It is interesting
143 to note that in both tables the X-ray performance is in the 80-90% range for the Cancer Imaging Archive
144 X-ray datasets, then drops for the MURA and Osteoarthritis Initiative datasets. This is likely because these
145 datasets are bone X-rays, and most of the datasets only contain chest X-rays. Therefore, a better spread of
146 X-ray datasets is needed for the performance of these models to be improved.

Figure 6: Accuracy of 3 models on the test set.

Comparing Accuracy of Models on the Train, Validate and Test Sets

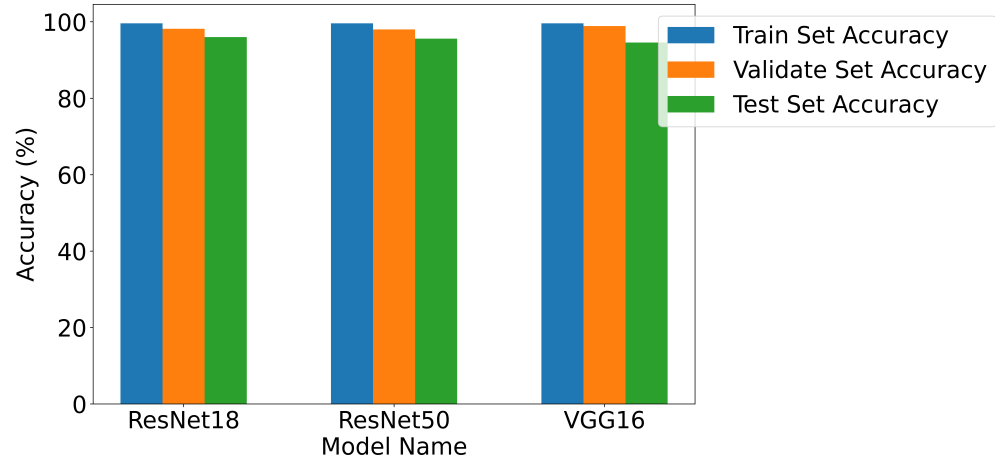


Figure 7: The confusion matrix for the ResNet18 on the test set. The model gains very high accuracy on the CT, MRI and PET. The ResNet18 results were chosen for this plot as this model achieved the highest accuracy and highest balanced accuracy.

Confusion Matrix - ResNet18

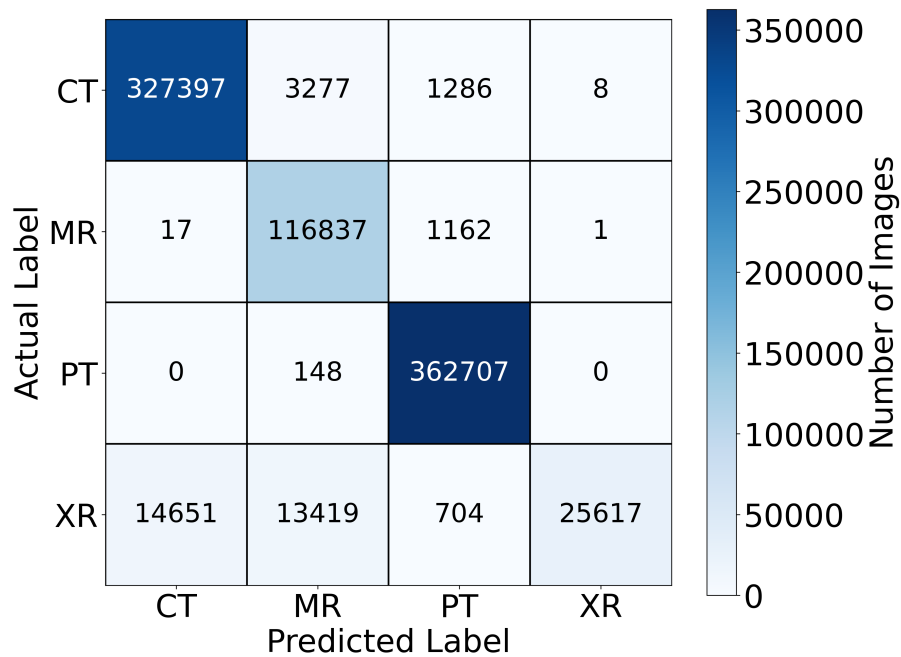


Table 3: Table containing the accuracy of the ResNet18 model on every dataset in the test set. Some datasets appear more than once in this table because they contain multiple image modalities.

Dataset (Location)	Modality	Accuracy (%)
CPTAC-LUAD (Chest)	CT	99
Pelvic-Reference-Data(Pelvis)	CT	81
C4KC-KiTS (Kidney)	CT	100
Anti-PD-1 Lung (Chest)	CT	97
CPTAC-PDA (Pancreas)	CT	100
NaF PROSTATE (Prostate)	CT	100
TCGA-READ (Kidney)	CT	100
QIN-HEADNECK (Head)	CT	100
CPTAC-LSCC (Chest)	CT	100
CPTAC-CCRCC (Kidney)	CT	100
CPTAC-LUAD (Chest)	MR	100
ISPY1 (Breast)	MR	99
Brain-Tumor-Progression (Head)	MR	92
REMBRANDT (Head)	MR	100
BraTS20 (Head)	MR	97
CPTAC-PDA (Pancreas)	MR	99
TCGA-READ (Kidney)	MR	98
CPTAC-CCRCC (Kidney)	MR	99
CPTAC-LUAD (Chest)	PT	100
Anti-PD-1 Lung (Chest)	PT	100
QIN-HEADNECK (Head)	PT	100
CPTAC-PDA (Pancreas)	PT	100
NaF PROSTATE (Prostate)	PT	100
CPTAC-LSCC (Chest)	PT	100
CPTAC-LUAD (Chest)	XR	100
CPTAC-PDA (Pancreas)	XR	96
CPTAC-LSCC (Chest)	XR	92
CPTAC-CCRCC (Kidney)	XR	100
MURA (Bone)	XR	28
Osteo-Arthritis Initiative (Bone)	XR	62

147 4 Conclusion

148 In this work, we proposed a hyper-scale classifier, capable of classifying diagnostic imaging data in the scale
149 of millions of medical images, with significant classification accuracy. We used a dataset comprised of 4.5
150 million images to train a ResNet-50, ResNet-18, and VGG16 CNN. The trained classifiers were then tested
151 for their classification accuracy on 4 modalities ((Computed Tomography (CT), Magnetic Resonance Imaging
152 (MRI), Positron Emission Tomography (PET), and X-ray). The best performing model, among the ones tested
153 demonstrated a classification accuracy of 96%. Our results show that hyper-scale classifiers are capable of
154 accurately classifying volumes of image data encountered in real-world applications, such as those contained in
155 image repositories or diagnostic imaging data collected by national healthcare institutions.

156 Future work on this topic will be to extend the scope of the hyper-scale modality classifier to work on 3D
157 scan modalities, such as CT, MR, PET.

158 A List of All Datasets Used

Name	Dataset Webpage	Citations	Split
ACRIN-DSC-MR-Brain	https://doi.org/10.7937/tcia.2019.zr1pjf4i	[7]	Train
Head-Neck-Radiomics-HN1	https://doi.org/10.7937/tcia.2019.8kap372n	[34]	Train
Lung-PET-CT-Dx	https://doi.org/10.7937/TCIA.2020.NNC2-0461	[10]	Train
AAPM RT-MAC Grand Ch. 2019	https://doi.org/10.7937/tcia.2019.bcfjqfqb	[35]	Train
COVID-19-AR	https://doi.org/10.7937/tcia.2020.py71-5978	[36]	Train
CPTAC-CM	https://doi.org/10.7937/K9/TCIA.2018.ODU24GZE	[13]	Train
CPTAC-HNSCC	https://doi.org/10.7937/K9/TCIA.2018.UW45NH81	[37]	Train
PDMR-997537-175-T	https://doi.org/10.7937/TCIA.2020.BRY9-4N29	[38]	Train
PDMR-292921-168-R	https://doi.org/10.7937/TCIA.2020.PCAK-8Z10	[38]	Train
PDMR-425362-245-T	https://doi.org/10.7937/TCIA.2020.7YRS-7J97	[38]	Train
HNSCC	https://doi.org/10.7937/k9/tcia.2020.a8sh-7363	[39, 40]	Train
DRO Toolkit	https://doi.org/10.7937/t062-8262	[41]	Train
QIN GBM Treatment Response	https://doi.org/10.7937/K9/TCIA.2016.nQF4gpn2	[42]	Train
CPTAC-GBM	https://doi.org/10.7937/K9/TCIA.2018.3RJE41Q1	[43]	Train
CPTAC-SAR	https://doi.org/10.7937/TCIA.2019.9bt23r95	[44]	Train
CPTAC-UCEC	https://doi.org/10.7937/K9/TCIA.2018.3R3JUISW	[45]	Train
OPC-Radiomics	https://doi.org/10.7937/tcia.2019.8dho2gls	[46]	Train
Acirin-FLT-Breast (ACRIN 6688)	https://doi.org/10.7937/K9/TCIA.2017.ol20zmxg	[8]	Train
QIN-Breast	https://doi.org/doi:10.7937/K9/TCIA.2016.21JUebH0	[47]	Train
Lung Fused-CT-Pathology	https://doi.org/10.7937/K9/TCIA.2018.SMT36LPN	[48]	Train
NSCLC-Radiomics	https://doi.org/10.7937/K9/TCIA.2015.PFOM9REI	[34]	Train
NSCLC-Radiomics-Interobserver1	https://doi.org/10.7937/tcia.2019.cwvlpd26	[49, 34]	Train
PDMR-BL0293-F563	https://doi.org/10.7937/tcia.2019.b6u7wmqw	[38]	Train

Name	Dataset Webpage	Citations	Split
QIN-BRAIN-DSC-MRI	https://doi.org/doi:10.7937/K9/TCIA.2016.5DI84Js8	[50]	Train
CC-Radiomics-Phantom	https://doi.org/10.7937/K9/TCIA.2017.zuzrm15b	[51]	Train
CC-Radiomics-Phantom-2	https://doi.org/10.7937/TCIA.2019.4124tz5g	[51]	Train
CC-Radiomics-Phantom-3	https://doi.org/10.7937/tcia.2019.j71i4fah	[51]	Train
LCTSC	https://doi.org/10.7937/K9/TCIA.2017.3r3fvz08	[52]	Train
Anti-PD-1 MELANOMA	https://doi.org/10.7937/tcia.2019.1ae0qtcu	[53]	Train
TCGA-UCEC	https://doi.org/10.7937/K9/TCIA.2016.GKJ0ZWAC	[54]	Train
TCGA-HNSC	https://doi.org/10.7937/K9/TCIA.2016.LXKQ47MS	[55]	Train
HNSCC-3DCT-RT	https://doi.org/10.7937/K9/TCIA.2018.13upr2xf	[56]	Train
MRI-DIR	https://doi.org/10.7937/K9/TCIA.2018.3f08iejt	[57]	Train
Head-Neck-PET-CT	https://doi.org/10.7937/K9/TCIA.2017.8oje5q00	[58]	Train
LGG-1p19qDeletion	https://doi.org/10.7937/K9/TCIA.2017.dwehtz9v	[59]	Train
CBIS-DDSM	https://doi.org/10.7937/K9/TCIA.2016.7002S9CY	[60]	Train
Phantom FDA	https://doi.org/10.7937/K9/TCIA.2015.ORBJKMUX	[61]	Train
QIN LUNG CT	https://doi.org/10.7937/K9/TCIA.2015.NPGZYBZ	[62]	Train
Mouse-Astrocytoma	https://doi.org/10.7937/K9TCIA.2017.SGW7CAQW	[63]	Train
TCGA-LUSC	https://doi.org/10.7937/K9/TCIA.2016.TYGKFKMQ	[64]	Train
TCGA-LUAD	https://doi.org/10.7937/K9/TCIA.2016.JGNIHEP5	[64]	Train
TCGA-KIRP	https://doi.org/10.7937/K9/TCIA.2016.ACWGBEF	[15]	Train
TCGA-LIHC	https://doi.org/10.7937/K9/TCIA.2016.IMMQW8UQ	[65]	Train
IvyGAP	https://doi.org/10.7937/K9/TCIA.2016.XLwaN6nL	[66]	Train
Prostate Fused-MRI-Pathology	https://doi.org/10.7937/K9/TCIA.2016.TLPMR1AM	[67, 68, 69, 70]	Train
TCGA-PRAD	https://doi.org/10.7937/K9/TCIA.2016.YXOGLM4Y	[71]	Train
Breast-MRI-NACT-Pilot	https://doi.org/10.7937/K9/TCIA.2016.QHSYHJKY	[72]	Train
RIDER Neuro MRI	https://doi.org/10.7937/K9/TCIA.2015.VOSN3HN1	[73]	Train
Soft-tissue-Sarcoma	https://doi.org/10.7937/K9/TCIA.2015.7G02GSKS	[74]	Train
Mouse-Mammary	https://doi.org/10.7937/K9/TCIA.2015.9P42KSE6	[75]	Train
TCGA-THCA	https://doi.org/10.7937/K9/TCIA.2016.9ZFRVF1B	[76]	Train
TCGA-SARC	https://doi.org/10.7937/K9/TCIA.2016.CX6YLSUX	[77]	Train
LungCT-Diagnosis	https://doi.org/10.7937/K9/TCIA.2015.A6V7JIWX	[78]	Train
TCGA-CESC	https://doi.org/10.7937/K9/TCIA.2016.SQ4M8YP4	[79]	Train
TCGA-OV	https://doi.org/10.7937/K9/TCIA.2016.ND01MDFQ	[80]	Train
TCGA-COAD	https://doi.org/10.7937/K9/TCIA.2016.HJJHBOXZ	[76]	Train
TCGA-KIRC	https://doi.org/10.7937/K9/TCIA.2016.V6PBVTDR	[14]	Train
TCGA-LGG	https://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK	[81]	Train
QIN PET Phantom	https://doi.org/10.7937/K9/TCIA.2015.ZPUKHCKB	[82]	Train
QIN Breast DCE-MRI	https://doi.org/10.7937/K9/TCIA.2014.A2N1IX0X	[83]	Train
NSCLC-Radiomics-Genomics	https://doi.org/10.7937/K9/TCIA.2015.L4FRET6Z	[34]	Train
Lung Phantom	https://doi.org/10.7937/K9/TCIA.2015.08A1IX00	[84]	Train
TCGA-KICH	https://doi.org/10.7937/K9/TCIA.2016.YU3RBCZN	[85]	Train
TCGA-GBM	https://doi.org/10.7937/K9/TCIA.2016.RNYFUYE9	[86]	Train
SPIE-AAPM Lung CT Challenge	https://doi.org/10.7937/K9/TCIA.2015.UZLSU3FL	[87]	Train
Prostate-3T	https://doi.org/10.7937/K9/TCIA.2015.QJTV5IL5	[88]	Train

Name	Dataset Webpage	Citations	Split
Prostate-Diagnosis	https://doi.org/10.7937/K9/TCIA.2015.FOQEUJVT	[89]	Train
RIDER Phantom PET-CT	https://doi.org/10.7937/K9/TCIA.2015.8WG2KN4W	[90]	Train
RIDER Lung CT	https://doi.org/10.7937/K9/TCIA.2015.U1X8A5NR	[91]	Train
RIDER Phantom MRI	https://doi.org/10.7937/K9/TCIA.2015.MI4QDDHU	[92]	Train
RIDER Breast MRI	https://doi.org/10.7937/K9/TCIA.2015.H1SXNUXL	[93]	Train
CT Colonography (ACRIN 6664)	https://doi.org/10.7937/K9/TCIA.2015.NWTESAY1	[29]	Train
Chexpert	https://stanfordmlgroup.github.io/competitions/chexpert/	[9]	Train
RSNA Bone Age	https://www.kaggle.com/kmader/rsna-bone-age	None	Train
TCGA-BRCA	https://doi.org/10.7937/K9/TCIA.2016.AB2NAZRP	[11]	Validate
Acir-FMISO-Brain (ACRIN 6684)	https://doi.org/10.7937/K9/TCIA.2018.vohlekok	[6]	Validate
TCGA-BLCA	https://doi.org/10.7937/K9/TCIA.2016.8LNG8XDR	[76]	Validate
Pancreas-CT	https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU	[94]	Validate
CT Lymph Nodes	https://doi.org/10.7937/K9/TCIA.2015.AQIIDCNM	[95]	Validate
TCGA-ESCA	https://doi.org/10.7937/K9/TCIA.2016.VPTNRGFY	[96]	Validate
TCGA-STAD	https://doi.org/10.7937/K9/TCIA.2016.GDHL9KIM	[97]	Validate
LIDC-IDRI	https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX	[98]	Validate
QIBA CT-1C	https://doi.org/10.7937/K9/TCIA.2016.YxgR4b1U	[99]	Validate
RIDER Lung PET-CT	https://doi.org/10.7937/K9/TCIA.2015.OFIP7TVM	[12]	Validate
Prostate-MRI	https://doi.org/10.7937/K9/TCIA.2016.6046GUDv	[100]	Validate
NIH 100000 Chest X-ray	https://nihcc.app.box.com/v/ChestXray-NIHCC	[101]	Validate
MRNet: Knee MRIs	https://stanfordmlgroup.github.io/competitions/mrnet/	[102]	Validate
CPTAC-CCRCC	https://doi.org/10.7937/K9/TCIA.2018.OBLAMN27	[103]	Test
C4KC-KiTS	https://doi.org/10.7937/TCIA.2019.IX49E8NX	[104]	Test
CPTAC-LSCC	https://doi.org/10.7937/K9/TCIA.2018.6EMUB5L2	[105]	Test
LDCT-and-Projection-data	https://doi.org/10.7937/9npb-2637	[106]	Test
CPTAC-LUAD	https://doi.org/10.7937/K9/TCIA.2018.PAT12TBS	[107]	Test
CPTAC-PDA	https://doi.org/10.7937/K9/TCIA.2018.SC20F018	[108]	Test
Pelvic-Reference-Data	https://doi.org/10.7937/TCIA.2019.woskq5oo	[109]	Test
Anti-PD-1 Lung	https://doi.org/10.7937/tcia.2019.zjjwb9ip	[110]	Test
ISPY1 (ACRIN 6657)	https://doi.org/10.7937/K9/TCIA.2016.HdHpgJLK	[111]	Test
QIN-HeadNeck	https://doi.org/10.7937/K9/TCIA.2015.KOF5CGLI	[112]	Test
TCGA-READ	https://doi.org/10.7937/K9/TCIA.2016.F7PPNPNU	[113]	Test
NaF Prostate	https://doi.org/10.7937/K9/TCIA.2015.ISOQTHKO	[114]	Test
REMBRANDT	https://doi.org/10.7937/K9/TCIA.2015.5880ZUZB	[115]	Test
MURA	https://stanfordmlgroup.github.io/competitions/mura/	[116]	Test
NDA Osteoarthritis Initiative	https://nda.nih.gov/oai/	None	Test
BraTS20	http://braintumorsegmentation.org/	None	Test

160 Code, Data, & Materials Availability

161 The data used in this study comes from multiple sources, as indicated in Appendix A. If not specified here then
162 the data can be accessed with no special permission required by the data provider. Permission was required to
163 access MURA, ChexPERT, MRNet, the NDA Osteo-Arthritis Initiative and BraTS20.

164 Code created for this research is hosted on GitHub with an MIT Licence[33].

165 References

- 166 [1] Hafizović, L., Čaušević, A., Deumić, A., Bećirović, L. S., Pokvić, L. G., and Badnjević, A., “The use of
167 artificial intelligence in diagnostic medical imaging: Systematic literature review,” in [2021 IEEE 21st
168 International Conference on Bioinformatics and Bioengineering (BIBE)], 1–6, IEEE (2021).
- 169 [2] Bhatnagar, G., Wu, Q. J., and Liu, Z., “A new contrast based multimodal medical image fusion frame-
170 work,” *Neurocomputing* **157**, 143–152 (2015).
- 171 [3] Hassan, M., Ali, S., Alquhayz, H., and Safdar, K., “Developing intelligent medical image modality classi-
172 fication system using deep transfer learning and lda,” *Scientific reports* **10**(1), 1–14 (2020).
- 173 [4] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al., “Deep learning
174 and process understanding for data-driven earth system science,” *Nature* **566**(7743), 195–204 (2019).
- 175 [5] Irisson, J.-O., Ayata, S.-D., Lindsay, D. J., Karp-Boss, L., and Stemmann, L., “Machine learning for the
176 study of plankton and marine snow from images,” *Ann. Rev. Mar. Sci* **14**, 277–301 (2022).
- 177 [6] Gerstner, E. R., Zhang, Z., Fink, J. R., Muzi, M., Hanna, L., Greco, E., Prah, M., Schmainda, K. M.,
178 Mintz, A., Kostakoglu, L., et al., “Acrin 6684: assessment of tumor hypoxia in newly diagnosed glioblas-
179 toma using 18f-fmiso pet and mri,” *Clinical Cancer Research* **22**(20), 5079–5086 (2016).
- 180 [7] Boxerman, J. L., Zhang, Z., Safriel, Y., Larvie, M., Snyder, B. S., Jain, R., Chi, T. L., Sorensen, A. G.,
181 Gilbert, M. R., and Barboriak, D. P., “Early post-bevacizumab progression on contrast-enhanced MRI as
182 a prognostic marker for overall survival in recurrent glioblastoma: results from the ACRIN 6677/RTOG
183 0625 Central Reader Study,” *Neuro-Oncology* **15**, 945–954 (07 2013).
- 184 [8] Kostakoglu, L., Duan, F., Idowu, M. O., Jolles, P. R., Bear, H. D., Muzi, M., Cormack, J., Muzi, J. P.,
185 Pryma, D. A., Specht, J. M., et al., “A phase ii study of 3'-deoxy-3'-18f-fluorothymidine pet in the
186 assessment of early response of breast cancer to neoadjuvant chemotherapy: results from acrin 6688,”
187 *Journal of Nuclear Medicine* **56**(11), 1681–1689 (2015).
- 188 [9] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball,
189 R. L., Shpanskaya, K. S., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson,
190 D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., and Ng, A. Y., “Chexpert: A large chest radiograph
191 dataset with uncertainty labels and expert comparison,” *CoRR* **abs/1901.07031** (2019).

- 192 [10] Li, P., Wang, S., Li, T., Lu, J., HuangFu, Y., and Wang, D., “A large-scale ct and pet/ct dataset for lung
193 cancer diagnosis [data set].,” *The Cancer Imaging Archive* (2020).
- 194 [11] Lingle, W., Erickson, B., Zuley, M., Jarosz, R., Bonaccio, E., Filippini, J., and Gruszauskas, N., “Radi-
195 ology data from the cancer genome atlas breast invasive carcinoma [tcga-brca] collection,” *The Cancer*
196 *Imaging Archive* (2016).
- 197 [12] Muzi, P., Wanner, M., and Kinahan, P., “Data from rider lung pet-ct,” *The Cancer Imaging Archive*
198 (2015).
- 199 [13] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
200 cutaneous melanoma [cptac-cm] collection,” *The Cancer Imaging Archive* (2018).
- 201 [14] Akin, O., Elnajjar, P., Heller, M., Jarosz, R., Erickson, B., Kirk, S., and Filippini, J., “Radiology data
202 from the cancer genome atlas kidney renal clear cell carcinoma [tcga-kirc] collection,” *The Cancer Imaging*
203 *Archive* (2016).
- 204 [15] Linehan, M., Gautam, R., Kirk, S., Lee, Y., Roche, C., Bonaccio, E., and Jarosz, R., “Radiology data from
205 the cancer genome atlas cervical kidney renal papillary cell carcinoma [kirp] collection,” *Cancer Imaging*
206 *Arch* (2016).
- 207 [16] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D.,
208 Pringle, M., et al., “The cancer imaging archive (tcia): maintaining and operating a public information
209 repository,” *Journal of digital imaging* **26**(6), 1045–1057 (2013).
- 210 [17] Stanford, “Stanford ml group,” (2023). <https://stanfordmlgroup.github.io/>, Last accessed on 2023-
211 01-02.
- 212 [18] Kaggle, “Kaggle,” (2023). <https://www.kaggle.com/>, Last accessed on 2023-01-02.
- 213 [19] Kalpathy-Cramer, J., Hersh, W., et al., “Automatic image modality based classification and annotation
214 to improve medical image retrieval,” in [*Medinfo*], 1334–1338 (2007).
- 215 [20] Kitanovski, I., Dimitrovski, I., and Loshkovska, S., “Fcse at medical tasks of imageclef 2013,” (2013).
- 216 [21] Valavanis, L., Stathopoulos, S., and Kalamboukis, T., “Ipl at clef 2016 medical task.,” *CLEF (Working*
217 *Notes)* **1609**, 413–420 (2016).
- 218 [22] Cao, J., Wang, M., Li, Y., and Zhang, Q., “Improved support vector machine classification algorithm
219 based on adaptive feature weight updating in the hadoop cluster environment,” *PloS one* **14**(4), e0215136
220 (2019).
- 221 [23] Markonis, D., de Herrera, A. G. S., Eggel, I., and Müller, H., “Multi-scale visual words for hierarchical
222 medical image categorisation,” in [*Medical Imaging 2012: Advanced PACS-based Imaging Informatics and*
223 *Therapeutic Applications*], **8319**, 99–109, SPIE (2012).
- 224 [24] Trenta, F., Battiato, S., and Ravi, D., “An explainable medical imaging framework for modality classi-
225 fications trained using small datasets,” in [*International Conference on Image Analysis and Processing*],
226 358–367, Springer (2022).

- 227 [25] Chiang, C.-H., Weng, C.-L., and Chiu, H.-W., “Automatic classification of medical image modality and
228 anatomical location using convolutional neural network,” *Plos one* **16**(6), e0253205 (2021).
- 229 [26] Cheng, D. and Liu, M., “Classification of alzheimer’s disease by cascaded convolutional neural networks us-
230 ing pet images,” in [*International Workshop on Machine Learning in Medical Imaging*], 106–113, Springer
231 (2017).
- 232 [27] Yu, Y., Lin, H., Yu, Q., Meng, J., Zhao, Z., Li, Y., and Zuo, L., “Modality classification for medical images
233 using multiple deep convolutional neural networks,” *J. Comput. Inf. Syst* **11**(15), 5403–5413 (2015).
- 234 [28] Sevakula, R. K., Singh, V., Verma, N. K., Kumar, C., and Cui, Y., “Transfer learning for molecular
235 cancer classification using deep neural networks,” *IEEE/ACM transactions on computational biology and
236 bioinformatics* **16**(6), 2089–2100 (2018).
- 237 [29] Johnson, C. D., Chen, M.-H., Toledano, A. Y., Heiken, J. P., Dachman, A., Kuo, M. D., Menias, C. O.,
238 Siewert, B., Cheema, J. I., Obregon, R. G., et al., “Accuracy of ct colonography for detection of large
239 adenomas and cancers,” *New England Journal of Medicine* **359**(12), 1207–1217 (2008).
- 240 [30] He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,”
241 *CoRR* **abs/1512.03385** (2015).
- 242 [31] Simonyan, K. and Zisserman, A., “Very deep convolutional networks for large-scale image recognition,”
243 *arXiv preprint arXiv:1409.1556* (2014).
- 244 [32] Pytorch, “Pytorch,” (2023). <https://pypi.org/project/torchvision/>, Last accessed on 2023-01-02.
- 245 [33] Craig, MacFadyen and David, Harris-Birtill, “Research code,” (2023). [https://github.com/
246 cdmacfadyen/classify-modality/](https://github.com/cdmacfadyen/classify-modality/), Last accessed on 2023-01-02.
- 247 [34] Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., Bussink,
248 J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., et al., “Decoding tumour phenotype by noninvasive
249 imaging using a quantitative radiomics approach,” *Nature communications* **5**(1), 1–9 (2014).
- 250 [35] Cardenas, C. E., Mohamed, A. S., Yang, J., Gooding, M., Veeraraghavan, H., Kalpathy-Cramer, J., Ng,
251 S. P., Ding, Y., Wang, J., Lai, S. Y., et al., “Head and neck cancer patient images for determining auto-
252 segmentation accuracy in t2-weighted magnetic resonance imaging through expert manual segmentations,”
253 *Medical Physics* **47**(5), 2317–2322 (2020).
- 254 [36] Desai, S., Baghal, A., Wongsurawat, T., Jenjaroenpun, P., Powell, T., Al-Shukri, S., Gates, K., Farmer,
255 P., Rutherford, M., Blake, G., et al., “Chest imaging representing a covid-19 positive rural us population,”
256 *Scientific data* **7**(1), 1–6 (2020).
- 257 [37] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
258 head and neck squamous cell carcinoma [cptac-hnscc] collection,” *The Cancer Imaging Archive* (2018).
- 259 [38] Tatum, J. L., Kalen, J. D., Ileva, L. V., Riffle, L. A., Keita, S., Patel, N., Jacobs, P. M., Sanders,
260 C., James, A., Difilippantonio, S., Thang, L., Hollingshead, M. G., Phillips, J., Evrard, Y., Clunie, D.
261 A. and Liu, Y., Suloway, C., Smith, K. E., Wagner, U., and Doroshov, J. H., “Imaging characterization of

- 262 a metastatic patient derived model of adenocarcinoma colon: Pdmr-997537-175-t [data set].,” *The Cancer*
263 *Imaging Archive* (2020).
- 264 [39] Grossberg, A. J., Mohamed, A. S., Elhalawani, H., Bennett, W. C., Smith, K. E., Nolan, T. S., Williams,
265 B., Chamchod, S., Heukelom, J., Kantor, M. E., et al., “Imaging and clinical data archive for head and
266 neck squamous cell carcinoma patients treated with radiotherapy,” *Scientific data* **5**, 180173 (2018).
- 267 [40] Elhalawani, H., Mohamed, A. S., White, A. L., Zafereo, J., Wong, A. J., Berends, J. E., AboHashem,
268 S., Williams, B., Aymard, J. M., Kanwar, A., et al., “Matched computed tomography segmentation and
269 demographic data for oropharyngeal cancer radiomics challenges,” *Scientific data* **4**, 170077 (2017).
- 270 [41] Jaggi, A., Mattonen, S. A., McNitt-Gray, M., and Napel, S., “Stanford dro toolkit: digital reference
271 objects for standardization of radiomic features,” *Tomography* **6**(2), 111 (2020).
- 272 [42] Prah, M., Stufflebeam, S., Paulson, E., Kalpathy-Cramer, J., Gerstner, E., Batchelor, T., Barboriak, D.,
273 Rosen, B., and Schmainda, K., “Repeatability of standardized and normalized relative cbv in patients
274 with newly diagnosed glioblastoma,” *American Journal of Neuroradiology* **36**(9), 1654–1661 (2015).
- 275 [43] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
276 glioblastoma multiforme [cptac-gbm] collection [data set].,” *The Cancer Imaging Archive* (2018).
- 277 [44] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
278 sarcomas [cptac-sar] collection [data set].,” *The Cancer Imaging Archive* (2018).
- 279 [45] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
280 uterine corpus endometrial carcinoma [cptac-ucec] collection [data set].,” *The Cancer Imaging Archive*
281 (2018).
- 282 [46] Kwan, J. Y. Y., Su, J., Huang, S. H., Ghoraie, L. S., Xu, W., Chan, B., Yip, K. W., Giuliani, M.,
283 Bayley, A., Kim, J., et al., “Radiomic biomarkers to refine risk models for distant metastasis in hpv-
284 related oropharyngeal carcinoma,” *International Journal of Radiation Oncology* Biology* Physics* **102**(4),
285 1107–1116 (2018).
- 286 [47] Li, X., Abramson, R. G., Arlinghaus, L. R., Kang, H., Chakravarthy, A. B., Abramson, V. G., Farley,
287 J., Mayer, I. A., Kelley, M. C., Meszoely, I. M., et al., “Multiparametric magnetic resonance imaging
288 for predicting pathological response after the first cycle of neoadjuvant chemotherapy in breast cancer,”
289 *Investigative radiology* **50**(4), 195–204 (2015).
- 290 [48] Rusu, M., Rajiah, P., Gilkeson, R., Yang, M., Donatelli, C., Thawani, R., Jacono, F. J., Linden, P.,
291 and Madabhushi, A., “Co-registration of pre-operative ct with ex vivo surgically excised ground glass
292 nodules to define spatial extent of invasive adenocarcinoma on in vivo imaging: a proof-of-concept study,”
293 *European radiology* **27**(10), 4209–4217 (2017).
- 294 [49] Kalendralis, P., Shi, Z., Traverso, A., Choudhury, A., Sloep, M., Zhovannik, I., Starmans, M. P., Grittner,
295 D., Feltens, P., Monshouwer, R., et al., “Fair-compliant clinical, radiomics and dicom metadata of rider,
296 interobserver, lung1 and head-neck1 tcia collections,” *Medical Physics* (2020).

- 297 [50] Schmainda, K., Prah, M., Connelly, J., and Rand, S., “Glioma dsc-mri perfusion data with standard
298 imaging and rois,” *The Cancer Imaging Archive*. <http://doi.org/10.7937/K9> (2016).
- 299 [51] Mackin, D., Fave, X., Zhang, L., Fried, D., Yang, J., Taylor, B., Rodriguez-Rivera, E., Dodge, C., and
300 Jones, A., “Court 1,” *Data from Credence Cartridge Radiomics Phantom CT Scans*. *The Cancer Imaging*
301 *Archive* (2017).
- 302 [52] Yang, J., Veeraraghavan, H., Armato III, S. G., Farahani, K., Kirby, J. S., Kalpathy-Kramer, J., van
303 Elmpt, W., Dekker, A., Han, X., Feng, X., et al., “Autosegmentation for thoracic radiation treatment
304 planning: A grand challenge at aapm 2017,” *Medical physics* **45**(10), 4568–4581 (2018).
- 305 [53] Patnana, M., Patel, S., and Tsao, A., “Anti-pd-1 immunotherapy melanoma dataset [data set].,” *The*
306 *Cancer Imaging Archive* (2019).
- 307 [54] Erickson, B. J., Mutch, D., Lippmann, L., and Jarosz, R., “Radiology data from the cancer genome atlas
308 uterine corpus endometrial carcinoma (tcga-ucec) collection,” *The Cancer Imaging Archive* (2016).
- 309 [55] Zuley, M. L., Jarosz, R., Kirk, S., Lee, Y., Colen, R., Garcia, K., and Aredes, N. D., “Radiology data from
310 the cancer genome atlas head-neck squamous cell carcinoma [tcga-hnsc] collection,” *The Cancer Imaging*
311 *Archive* (2016).
- 312 [56] Bejarano, T., De Ornelas-Couto, M., and Mihaylov, I. B., “Longitudinal fan-beam computed tomography
313 dataset for head-and-neck squamous cell carcinoma patients,” *Medical physics* **46**(5), 2526–2537 (2019).
- 314 [57] Ger, R. B., Yang, J., Ding, Y., Jacobsen, M. C., Cardenas, C. E., Fuller, C. D., Howell, R. M., Li, H.,
315 Stafford, R. J., Zhou, S., et al., “Synthetic head and neck and phantom images for determining deformable
316 image registration accuracy in magnetic resonance imaging,” *Medical physics* **45**(9), 4315–4321 (2018).
- 317 [58] Vallieres, M., Kay-Rivest, E., Perrin, L. J., Liem, X., Furstoss, C., Aerts, H. J., Khaouam, N., Nguyen-
318 Tan, P. F., Wang, C.-S., Sultanem, K., et al., “Radiomics strategies for risk assessment of tumour failure
319 in head-and-neck cancer,” *Scientific reports* **7**(1), 1–14 (2017).
- 320 [59] Akkus, Z., Ali, I., Sedlář, J., Agrawal, J. P., Parney, I. F., Giannini, C., and Erickson, B. J., “Predicting
321 deletion of chromosomal arms 1p/19q in low-grade gliomas from mr images using machine intelligence,”
322 *Journal of digital imaging* **30**(4), 469–476 (2017).
- 323 [60] Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., and Rubin, D. L., “A curated mam-
324 mography data set for use in computer-aided detection and diagnosis research,” *Scientific data* **4**, 170177
325 (2017).
- 326 [61] Gavrielides, M. A., Kinnard, L. M., Myers, K. J., Peregoy, J., Pritchard, W. F., Zeng, R., Esparza, J.,
327 Karanian, J., and Petrick, N., “A resource for the assessment of lung nodule size estimation methods:
328 database of thoracic ct scans of an anthropomorphic phantom,” *Optics express* **18**(14), 15244 (2010).
- 329 [62] Kalpathy-Cramer, J., Napel, S., Goldgof, D., and Zhao, B., “Qin multi-site collection of lung ct data with
330 nodule segmentations,” *Cancer Imaging Arch* **10**, K9 (2015).
- 331 [63] Jansen, S. and Van Dyke, T., “Tcia mouse-astrocytoma collection,” *The Cancer Imaging Archive* (2015).

- 332 [64] Albertina, B., Watson, M., Holback, C., Jarosz, R., Kirk, S., Lee, Y., and Lemmerman, J., “Radiology
333 data from the cancer genome atlas lung adenocarcinoma [tcga-luad] collection,” *The Cancer Imaging*
334 *Archive* (2016).
- 335 [65] Erickson, B., Kirk, S., Lee, Y., Bathe, O., Kearns, M., Gerdes, C., Rieger-Christ, K., and Lemmerman,
336 J., “Radiology data from the cancer genome atlas liver hepatocellular carcinoma [tcga-lihc] collectionthe,”
337 *Cancer Imaging Archive* (2016).
- 338 [66] Puchalski, R. B., Shah, N., Miller, J., Dalley, R., Nomura, S. R., Yoon, J.-G., Smith, K. A., Lankerovich,
339 M., Bertagnolli, D., Bickley, K., et al., “An anatomic transcriptional atlas of human glioblastoma,” *Sci-*
340 *ence* **360**(6389), 660–663 (2018).
- 341 [67] Singanamalli, A., Rusu, M., Sparks, R. E., Shih, N. N., Ziober, A., Wang, L.-P., Tomaszewski, J., Rosen,
342 M., Feldman, M., and Madabhushi, A., “Identifying in vivo dce mri markers associated with microves-
- 343 sel architecture and gleason grades of prostate cancer,” *Journal of Magnetic Resonance Imaging* **43**(1),
344 149–158 (2016).
- 345 [68] Toth, R. J., Shih, N., Tomaszewski, J. E., Feldman, M. D., Kutter, O., Yu, D. N., Paulus Jr, J. C.,
346 Paladini, G., and Madabhushi, A., “Histostitcher™: An informatics software platform for reconstructing
347 whole-mount prostate histology using the extensible imaging platform framework,” *Journal of Pathology*
348 *Informatics* **5** (2014).
- 349 [69] Xiao, G., Bloch, B. N., Chappelow, J., Genega, E. M., Rofsky, N. M., Lenkinski, R. E., Tomaszewski,
350 J., Feldman, M. D., Rosen, M., and Madabhushi, A., “Determining histology-mri slice correspondences
351 for defining mri-based disease signatures of prostate cancer,” *Computerized Medical Imaging and Graph-*
352 *ics* **35**(7-8), 568–578 (2011).
- 353 [70] Chappelow, J., Bloch, B. N., Rofsky, N., Genega, E., Lenkinski, R., DeWolf, W., and Madabhushi,
354 A., “Elastic registration of multimodal prostate mri and histology via multiattribute combined mutual
355 information,” *Medical Physics* **38**(4), 2005–2018 (2011).
- 356 [71] Zuley, M., Jarosz, R., Drake, B., et al., “Radiology data from the cancer genome atlas prostate adeno-
357 carcinoma [tcga-prad] collection,” *The Cancer Imaging Archive*. Available online: <http://doi.org/10.7937>
358 *K 9* (2016).
- 359 [72] Newitt, D. and Hylton, N., “Single site breast dce-mri data and segmentations from patients undergoing
360 neoadjuvant chemotherapy,” *The Cancer Imaging Archive* **2** (2016).
- 361 [73] Barboriak, D., “Data from rider neur mri. the cancer imaging archive,” (2015).
- 362 [74] Vallières, M., Freeman, C. R., Skamene, S. R., and El Naqa, I., “A radiomics model from joint fdg-pet
363 and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities,”
364 *Physics in Medicine and Biology* **60**(14), 5471 (2015).
- 365 [75] Jansen, S., Ileva, L., Lu, L., and Van Dyke, T., “Tcia mouse-mammary collection,” *The Cancer Imaging*
366 *Archive* (2015).

- 367 [76] Kirk, S., Lee, Y., Roche, C., Bonaccio, E., Filippini, J., and Jarosz, R., “Radiology data from the cancer
368 genome atlas thyroid cancer [tcga-thca] collection,” *Cancer Imaging Archive*. doi **10**, K9 (2016).
- 369 [77] Roche, C., Bonaccio, E., and Filippini, J., “cited 2019 18/01/2019,” *Radiology data from The Cancer
370 Genome Atlas Sarcoma collection. The Cancer Imaging Archive 2016* (2016).
- 371 [78] Grove, O., Berglund, A. E., Schabath, M. B., Aerts, H. J., Dekker, A., Wang, H., Velazquez, E. R., Lambin,
372 P., Gu, Y., Balagurunathan, Y., et al., “Quantitative computed tomographic descriptors associate tumor
373 shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma,” *PloS one* **10**(3),
374 e0118261 (2015).
- 375 [79] Lucchesi, F. and Aredes, N., “Radiology data from the cancer genome atlas cervical squamous cell carci-
376 noma and endocervical adenocarcinoma (tcga-cesc) collection. the cancer imaging archive,” (2016).
- 377 [80] Holback, C., Jarosz, R., Prior, F., Mutch, D. G., Bhosale, P., Garcia, K., and Erickson, B. J., “Radiology
378 data from the cancer genome atlas ovarian cancer [tcga-ov] collection,” (2016).
- 379 [81] Pedano, N., Flanders, A. E., Scarpace, L., Mikkelsen, T., Eschbacher, J., Hermes, B., and Ostrom, Q.,
380 “Radiology data from the cancer genome atlas low grade glioma [tcga-egg] collection,” *The Cancer Imaging
381 Archive* **2** (2016).
- 382 [82] Beichel, R. R., Ulrich, E. J., Bauer, C., Byrd, D. W., Muzi, J. P., Muzi, M., and Buatti, J. M., “Data
383 from qin pet phantom,” *The Cancer Imaging Archive* (2015).
- 384 [83] Huang, W., Li, X., Chen, Y., Li, X., Chang, M.-C., Oborski, M. J., Malyarenko, D. I., Muzi, M., Ja-
385 jamovich, G. H., Fedorov, A., et al., “Variations of dynamic contrast-enhanced magnetic resonance imag-
386 ing in evaluation of breast cancer therapy response: a multicenter data analysis challenge,” *Translational
387 oncology* **7**(1), 153 (2014).
- 388 [84] Zhao, B., “Data from lung phantom,” *The Cancer Imaging Archive* (2015).
- 389 [85] Linehan, M. W., Gautam, R., Sadow, C. A., and Levine, S., “Radiology data from the cancer genome
390 atlas kidney chromophobe [tcga-kich] collection,” *The Cancer Imaging Archive* (2016).
- 391 [86] L, S., T, M., Cha and, R. S., S, T., D, G., JH, S., BJ, E., N, P., AE, F., J, B.-S., Q, O., D, B., and LJ,
392 P., “Radiology data from the cancer genome atlas glioblastoma multiforme [tcga-gbm] collection,” *The
393 Cancer Imaging Archive* (2016).
- 394 [87] Armato III, S. G., Drukker, K., Li, F., Hadjiiski, L., Tourassi, G. D., Engelmann, R. M., Giger, M. L., Red-
395 mond, G., Farahani, K., Kirby, J. S., et al., “Lungx challenge for computerized lung nodule classification,”
396 *Journal of Medical Imaging* **3**(4) (2016).
- 397 [88] Litjens, G., Futterer, J., and Huisman, H., “Data from prostate-3t: the cancer imaging archive,” (2015).
- 398 [89] Bloch, B. N., Jain, A., and Jaffe, C. C., “Data from prostate-diagnosis,” *The Cancer Imaging Archive.
399 Available online: <http://doi.org/10.7937> K 9* (2015).
- 400 [90] Muzi, P., Wanner, M., and Kinahan, P., “Data from rider phantom pet-ct,” *The Cancer Imaging Archive
401* (2015).

- 402 [91] Zhao, B., James, L. P., Moskowitz, C. S., Guo, P., Ginsberg, M. S., Lefkowitz, R. A., Qin, Y., Riely, G. J.,
403 Kris, M. G., and Schwartz, L. H., “Evaluating variability in tumor measurements from same-day repeat
404 ct scans of patients with non-small cell lung cancer,” *Radiology* **252**(1), 263–272 (2009).
- 405 [92] Jackson, E. F., Barboriak, D. P., Bidaut, L. M., and Meyer, C. R., “Magnetic resonance assessment of
406 response to therapy: tumor change measurement, truth data and error sources,” *Translational Oncol-*
407 *ogy* **2**(4), 211 (2009).
- 408 [93] Meyer, C. R., Chenevert, T. L., Galbán, C. J., Johnson, T. D., Hamstra, D. A., Rehemtulla, A., and
409 Ross, B. D., “Data from rider-breast-mri. the cancer imaging archive,” (2015).
- 410 [94] Roth, H. R., Lu, L., Farag, A., Shin, H.-C., Liu, J., Turkbey, E. B., and Summers, R. M., “Deepor-
411 gan: Multi-level deep convolutional networks for automated pancreas segmentation,” in [*International*
412 *conference on medical image computing and computer-assisted intervention*], 556–564, Springer (2015).
- 413 [95] Roth, H. R., Lu, L., Seff, A., Cherry, K. M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., and Summers,
414 R. M., “A new 2.5 d representation for lymph node detection using random sets of deep convolutional
415 neural network observations,” in [*International conference on medical image computing and computer-*
416 *assisted intervention*], 520–527, Springer (2014).
- 417 [96] Lucchesi, F. R. and Aredes, N. D., “Radiology data from the cancer genome atlas esophageal carcinoma
418 [tcga-esca] collection,” *The Cancer Imaging Archive* (2016).
- 419 [97] Lucchesi, F. and Aredes, N., “Radiology data from the cancer genome atlas stomach adenocarcinoma
420 [tcga-stad] collection, 2016,” *The Cancer Imaging Archive* **10**, K9.
- 421 [98] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B.,
422 Aberle, D. R., Henschke, C. I., Hoffman, E. A., et al., “The lung image database consortium (lidc) and
423 image database resource initiative (idri): a completed reference database of lung nodules on ct scans,”
424 *Medical physics* **38**(2), 915–931 (2011).
- 425 [99] Fenimore, C., McNitt-Gray, M. F., Clunie, D., Gavrielides, M. A., Petrick, N., Samei, E., and Slazak, K.,
426 “Data from qiba ct-1c,” *The Cancer Imaging Archive* (2016).
- 427 [100] P, C., B, T., P, P., M, M., and B., W., “Data from prostate-mri,” *The Cancer Imaging Archive* (2016).
- 428 [101] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R. M., “Chestx-ray8: Hospital-scale chest
429 x-ray database and benchmarks on weakly-supervised classification and localization of common thorax
430 diseases,” in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 2097–2106
431 (2017).
- 432 [102] Bien, N., Rajpurkar, P., Ball, R. L., Irvin, J., Park, A., Jones, E., Bereket, M., Patel, B. N., Yeom,
433 K. W., Shpanskaya, K., et al., “Deep-learning-assisted diagnosis for knee magnetic resonance imaging:
434 development and retrospective validation of mrnet,” *PLoS medicine* **15**(11), e1002699 (2018).
- 435 [103] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
436 clear cell renal cell carcinoma [cptac-crcce] collection [data set].,” *The Cancer Imaging Archive* (2018).

- 437 [104] Heller, N., Isensee, F., Maier-Hein, K. H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M.,
438 et al., “The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging:
439 Results of the kits19 challenge,” *Medical Image Analysis* **67**, 101821 (2019).
- 440 [105] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
441 lung squamous cell carcinoma [cptac-lsccl] collection [data set].,” *The Cancer Imaging Archive* (2018).
- 442 [106] Moen, T. R., Chen, B., Holmes III, D. R., Duan, X., Yu, Z., Yu, L., Leng, S., Fletcher, J. G., and
443 McCollough, C. H., “Low dose ct image and projection dataset,” *Medical Physics* (2020).
- 444 [107] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
445 lung adenocarcinoma [cptac-luad] collection [data set].,” *The Cancer Imaging Archive* (2018).
- 446 [108] Consortium, N. C. I. C. P. T. A., “Radiology data from the clinical proteomic tumor analysis consortium
447 pancreatic ductal adenocarcinoma [cptac-pda] collection [data set].,” *The Cancer Imaging Archive* (2018).
- 448 [109] Yorke, A., Sala, I., Solis, D., and Guerrero, T., “A statistically characterized reference data set for image
449 registration of pelvis using combinatorial affine registration optimization,” in [*MEDICAL PHYSICS*],
450 **46**(6), E340–E340, WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA (2019).
- 451 [110] Madhavi, P., Patel, S., and Tsao, A. S., “Data from anti-pd-1 immunotherapy lung [data set].,” *The
452 Cancer Imaging Archive* (2019).
- 453 [111] Hylton, N. M., Gatsonis, C. A., Rosen, M. A., Lehman, C. D., Newitt, D. C., Partridge, S. C., Bernreuter,
454 W. K., Pisano, E. D., Morris, E. A., Weatherall, P. T., et al., “Neoadjuvant chemotherapy for breast
455 cancer: functional tumor volume by mr imaging predicts recurrence-free survival—results from the acrin
456 6657/calgb 150007 i-spy 1 trial,” *Radiology* **279**(1), 44–55 (2016).
- 457 [112] Fedorov, A., Clunie, D., Ulrich, E., Bauer, C., Wahle, A., Brown, B., Onken, M., Riesmeier, J., Pieper, S.,
458 Kikinis, R., et al., “Dicom for quantitative imaging biomarker development: a standards based approach
459 to sharing clinical data and structured pet/ct analysis results in head and neck cancer research,” *PeerJ* **4**,
460 e2057 (2016).
- 461 [113] Kirk, S., Lee, Y., Sadow, C. A., and Levine, S., “Radiology data from the cancer genome atlas rectum
462 adenocarcinoma [tcga-read] collection,” *The Cancer Imaging Archive* (2016).
- 463 [114] Kurdziel, K. A., Shih, J. H., Apolo, A. B., Lindenberg, L., Mena, E., McKinney, Y. Y., Adler, S. S.,
464 Turkbey, B., Dahut, W., Gulley, J. L., et al., “The kinetics and reproducibility of 18f-sodium fluoride for
465 oncology using current pet camera technology,” *Journal of Nuclear Medicine* **53**(8), 1175–1184 (2012).
- 466 [115] Scarpace, L., Flanders, A. E., Jain, R., Mikkelsen, T., and Andrews, D. W., “Data from rembrandt,” *The
467 Cancer Imaging Archive* **10**, K9 (2015).
- 468 [116] Rajpurkar, P., Irvin, J., Bagul, A., Ding, D., Duan, T., Mehta, H., Yang, B., Zhu, K., Laird, D., Ball,
469 R. L., et al., “Mura: Large dataset for abnormality detection in musculoskeletal radiographs,” *arXiv
470 preprint arXiv:1712.06957* (2017).