Supplementary Materials

S1 Datasets

S1.1 RSNA

The RSNA dataset was released for the 2019 RSNA Intracranial Hemorrhage Detection Challenge [27]. We used the training set of the first stage of the data challenge in this study, which contains 674257 CT slices of 17079 patients. The slices were labeled as 7 classes, normal, intracranial hemorrhage, and five subclasses of intracranial hemorrhage. We used this dataset as a binary classification task (normal/abnormal). The dataset was randomly partitioned into training and testing datasets with a 4 : 1 ratio on the patient-level by us.

S1.2 DDSM

The DDSM dataset contains 2620 well-labeled cases, including 10480 digitized screen-film mammography images [6]. The dataset has been almost 20 years old, which was initially constructed in 1999. DDSM is the largest publicly available mammography dataset and widely used for developing deep learning models. We chose the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [17], which is an updated and standardized version of the original DDSM, for our study. We used the training and testing sets provided by the data provider for training and testing respectively. We used the provided training and testing sets in this study.

S1.3 Mendeley V2

The Mendeley V2 [13] dataset contains both of the optical coherence tomography (OCT) images of the retina and pediatric chest X-ray images. We used the pediatric chest X-ray images in this study. The dataset includes 4273 pneumonia images and 1583 normal images. We used the provided training and testing sets in this study.

S1.4 Kather 5000

The Kather 5000 [12] dataset contains 5000 histological images of 150×150 pixels. Each image belongs to exactly one of eight tissue categories: tumour epithelium, simple stroma, complex stroma, immune cells, debris, normal mucosal glands, adipose tissue, background (no tissue). All images are RGB, $0.495\mu m$ per pixel, digitized with an Aperio ScanScope (Aperio/Leica biosystems), magnification $20\times$. Histological samples are fully anonymized images of formalin-fixed paraffin-embedded human colorectal adenocarcinomas (primary tumors) from the Institute of Pathology, University Medical Center Mannheim, Heidelberg University, Mannheim, Germany). The dataset was randomly partitioned into training and testing datasets with a 4 : 1 ratio by us.

S2 CNN Models

In this study, we used AlexNet [15], ResNet-50 [5], DenseNet-121 [8], and SqueezeNet 1-1 [9] were as the feature extractor to build our CNN models. More specifically, we firstly

Table S1: Datasets used in this study.

Name	Modality	# of Images	# of Classes
RSNA	Head CT	674257	2
DDSM	Mammography	10480	2
Mendeley V2	Chest X-ray	5856	2
Kather 5000	Histological	5000	8

pre-trained these four networks on the ImageNet dataset [1]. Then, the fully connected (FC) layers and the pooling layer before the FC layers were removed. We froze the parameters of the remaining convolutional (Conv) layers of each network and used them as feature extractors. A shallow CNN classifier was trained on top of each feature extractor. The classifier contained one Conv layer and two FC layers. The Conv layer included convolution, batch normalization [10], leaky ReLU [36], and max pooling [15]. Max pooling had a 2×2 receptive field with stride 1. Weighted cross-entropy loss was used in training. Adam optimizer [14] with a learning rate of 0.0001 was used as the optimizer. Dropout [30] with a rate of 0.5 was applied to the FC layers. For the same architecture, all the hyper-parameters were maintained the same among different datasets, except batch sizes. The batch size of AlexNet on DDSM, Mendeley V2, and Kather 5000 datasets was set as 64, and for RSNA was 512. The batch sizes were set as half of the AlexNet with the corresponding datasets for the rest of the architectures.