Construction of a Machine Learning Dataset through Collaboration: The RSNA 2019 Brain CT Hemorrhage Challenge

(RSNA-ASNR 2019 Brain Hemorrhage CT Annotators) Flanders, Adam E.; ... Mogensen, Monique; Morón, Fanny; Miller, Elka; Ikuta, Ichiro; Zohrabian, Vahe; McDonnell, Olivia; ... Nath, Jaya; Rumboldt, Zoran

Source / Izvornik: **Radiology: Artificial Intelligence, 2020, 2, 1 - 8**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

https://doi.org/10.1148/ryai.2020190211

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:184:272087

Rights / Prava: Attribution-NonCommercial-NoDerivatives 4.0 International/Imenovanje-Nekomercijalno-Bez prerada 4.0 međunarodna

Download date / Datum preuzimanja: **2024-06-26**

Repository / Repozitorij:

Repository of the University of Rijeka, Faculty of Medicine - FMRI Repository

UNIRI DIGITALNA KNJIŽNICA

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

Radiology: Artificial Intelligence

DATA RESOURCES

Construction of a Machine Learning Dataset through Collaboration: The RSNA 2019 Brain CT Hemorrhage Challenge

Adam E. Flanders, MD • Luciano M. Prevedello, MD • George Shih, MD • Safwan S. Halabi, MD • Jayashree Kalpathy-Cramer, PhD • Robyn Ball, PhD • John T. Mongan, MD • Anouk Stein, PhD • Felipe C. Kitamura, MD • Matthew P. Lungren, MD • Gagandeep Choudhary, MD • Lesley Cala, MD • Luiz Coelho, MD • Monique Mogensen, MD • Fanny Morón, MD • Elka Miller, MD • Ichiro Ikuta, MD, MMSc • Vahe Zohrabian, MD • Olivia McDonnell, MB, MCh, BaO • Christie Lincoln, MD • Lubdha Shah, MD • David Joyner, MD • Amit Agarwal, MD • Ryan K. Lee, MD • Jaya Nath, MD • For the RSNA-ASNR 2019 Brain Hemorrhage CT Annotators

From the Department of Radiology/Division of Neuroradiology, Thomas Jefferson University Hospital, 132 S Tenth St, Suite 1080B Main Building, Philadelphia, PA 19107 (A.E.F.); Department of Radiology, The Ohio State University, Columbus, Ohio (L.M.P.); Department of Radiology, Weill Cornell Medical College, New York, NY (G.S.); Department of Radiology, Stanford University, Stanford, Calif (S.S.H.); Department of Radiology and Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Mass (J.K.); Quantitative Sciences Unit, Stanford University, Stanford, Calif (R.B.); Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, Calif (J.T.M.); MD.ai, New York, NY (A.S.); Department of Diagnostic Imaging, Universidade Federal de São Paulo, São Paulo, Brazil (F.C.K.); Department of Radiology, Stanford University, Stanford, Calif (M.P.L.); Department of Radiology, University of Alabama at Birmingham, Birmingham, Ala (G.C.); Faculty of Health and Medical Sciences, University of Western Australia, Perth, Australia (L. Cala); Advanced Diagnostic Imaging, Clínica DAPI, Curitiba, Brazil (L. Coelho); Department of Radiology, University of Washington, Seattle, Wash (M.M.); Department of Radiology, Baylor College of Medicine, Houston, Tex (F.M., C.L.); Department of Radiology, University of Ottawa, Ottawa, Canada (E.M.); Department of Radiology & Biomedical Imaging, Yale University, New Haven, Conn (I.I., V.Z.); Department of Medical Imaging, Gold Coast University Hospital, Southport, Australia (O.M.); Department of Neuroradiology, University of Utah Health Sciences Center, Salt Lake City, Utah (L.S.); Department of Radiology and Medical Imaging, University of Virginia Health, Charlottesville, Va (D.J.); Division of Neuroradiology, University of Texas Southwestern Medical Center, Dallas, Tex (A.A.); Department of Radiology, Albert Einstein Healthcare Network, Philadelphia, Pa (R.K.L.); and Department of Radiology, SUNY Downstate Medical Center, Albany, NY (J.N.). Received December 9, 2019; revision requested January 10, 2020; revision received February 23; accepted April 7. **Address correspondence to** A.E.F. (e-mail: *adam.flanders@jefferson.edu*).

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2020; 2(3):e190211 • https://doi.org/10.1148/ryai.2020190211 • Content codes: IN CT NR • ©RSNA, 2020

Intracranial hemorrhage is a potentially life-threatening problem that has many direct and indirect causes. Accuracy in diagnosing the presence and type of intracranial hemorrhage is a critical part of effective treatment. Diagnosis is often an urgent procedure requiring review of medical images by highly trained specialists and sometimes necessitating confirmation through clinical history, vital signs, and laboratory examinations. The process is complicated and requires immediate identification for optimal treatment.

Intracranial hemorrhage is a relatively common condition that has many causes, including trauma, stroke, aneurysm, vascular malformation, high blood pressure, illicit drugs, and blood clotting disorders (1). Neurologic consequences can vary extensively from headache to death depending upon the size, type, and location of the hemorrhage. The role of the radiologist is to detect the hemorrhage, characterize the type and cause of the hemorrhage, and to determine if the hemorrhage could be jeopardizing critical areas of the brain that might require immediate surgery.

While all acute hemorrhages appear attenuated on CT images, the primary imaging features that help radiologists determine the cause of hemorrhage are the location, shape, and proximity to other structures. Intraparenchymal hemorrhage is blood that is located completely within the brain itself. Intraventricular or subarachnoid hemorrhage is blood that has leaked into the spaces of the brain that normally contain cerebrospinal fluid (the ventricles or subarachnoid cisterns, respectively). Extra-axial hemorrhage is blood that collects in the tissue coverings that surround the brain (eg, subdural or epidural subtypes). It is important to note that patients may exhibit more than one type of cerebral hemorrhage, which may appear on the same image or imaging study. Although small hemorrhages are typically less morbid than large hemorrhages, even a small hemorrhage can lead to death if it is in a critical location. Small hemorrhages also may herald future hemorrhages that could be fatal (eg, ruptured cerebral aneurysm). The presence or absence of hemorrhage may guide specific treatments (eg, stroke).

Detection of cerebral hemorrhage with brain CT is a popular clinical use case for machine learning (2–5). Many of these early successful investigations were based upon relatively small datasets (hundreds of examinations) from single institutions. Chilamkurthy et al created a diverse brain CT dataset that was selected from 20 geographically distinct centers in India (more than 21 000 unique examinations). This was used to create smaller randomly selected subsets for validation and testing on common acute brain abnormalities (6). The ability for machine learning algorithms to generalize to "real-world" clinical imaging data from disparate institutions is paramount to successful use in the clinical environment.

The intent for this challenge was to provide a large multi-institutional and multinational dataset to help develop machine learning algorithms that can assist in the detection and characterization of intracranial hemorrhage with brain CT. The following is a summary of how the dataset was collected, prepared, pooled, curated, and annotated.

Materials and Methods

Creation of the dataset for the 2019 Radiological Society of North America (RSNA) Machine Learning Challenge was inherently more complex than the prior two challenges for several reasons. First, the image da-

This copy is for personal use only. To order printed copies, contact reprints@rsna.org

Abbreviations

ASNR = American Society of Neuroradiology, RSNA = Radiological Society of North America

Summary

This dataset is composed of annotations of the five hemorrhage subtypes (subarachnoid, intraventricular, subdural, epidural, and intraparenchymal hemorrhage) typically encountered at brain CT.

Key Points

- This 874 035-image, multi-institutional, and multinational brain hemorrhage CT dataset is the largest public collection of its kind that includes expert annotations from a large cohort of volunteer neuroradiologists for classifying intracranial hemorrhages.
- This dataset was used for the Radiological Society of North America (RSNA) 2019 Machine Learning Challenge.
- The curation of this dataset was a collaboration between the RSNA and the American Society of Neuroradiology and is made freely available to the machine learning research community for noncommercial use to create high-quality machine learning algorithms to help diagnose intracranial hemorrhage.

taset and annotations were created from scratch, whereas the prior two challenges were built around existing datasets (eg, pediatric hand radiographs in 2017 and the National Institutes of Health ChestNet-14 chest radiograph collection in 2018) (7,8). Second, this was the first RSNA competition to employ volumetric data (CT image series) rather than single images. This added complexity in data curation, label adjudication, and pooling of data from multiple sources, as well as an extra dimension of complexity reconciling image- and examination-based annotations. Moreover, the size and complexity of the dataset and use of image series required assembling, coordinating, training, and monitoring a large group of expert annotators to complete the project in the required time frame. Training, monitoring, and coordination of a large cadre of annotators added another factor of complexity to the process. All of these processes had to be staged, deployed, and completed in a relatively short time frame (Figs 1, 2).

The CT brain dataset was compiled from the clinical picture archiving and communication system archives from three institutions: Stanford University (Palo Alto, Calif), Universidade Federal de São Paulo (São Paulo, Brazil) and Thomas Jefferson University Hospital (Philadelphia, Pa). Each contributing institution secured approval of its institutional review board and institutional compliance officers. Methods for examination identification, extraction, and anonymization were unique to each contributing institution: Universidade Federal de São Paulo provided all brain CT examinations performed during a 1-year period, Stanford University preselected examinations based upon a normal versus abnormal assessment of radiology reports to provide a 50/50 sample of positive (for any abnormality) to negative examinations, and Thomas Jefferson University Hospital extracted cases using simple natural language processing on radiology reports mentioning specific hemorrhage subtypes. The image data collection process from each of the contributing institutions is outlined in Figure 1. As the methodology for discovery differed from each contributing site, there was a relative imbalance of feature types among datasets that needed to be reconciled and balanced during the curation and data pooling process. Original Digital Imaging and Communications in Medicine data were provided following local Health Insurance Portability and Accountability Act–compliant de-identification. In aggregate, 27 861 unique CT brain examinations (1 074 271 unique images) were submitted for the dataset. The entire process from data solicitation to collation of the final dataset is summarized in the workflow diagram in Figure 2.

Data Annotation

Annotation was performed using a commercial browser-based annotation platform (MD.ai, New York, NY), for which two of the authors (G.S. and A.S.) served as consultants. The annotation system allowed adjustment of the brightness, contrast,

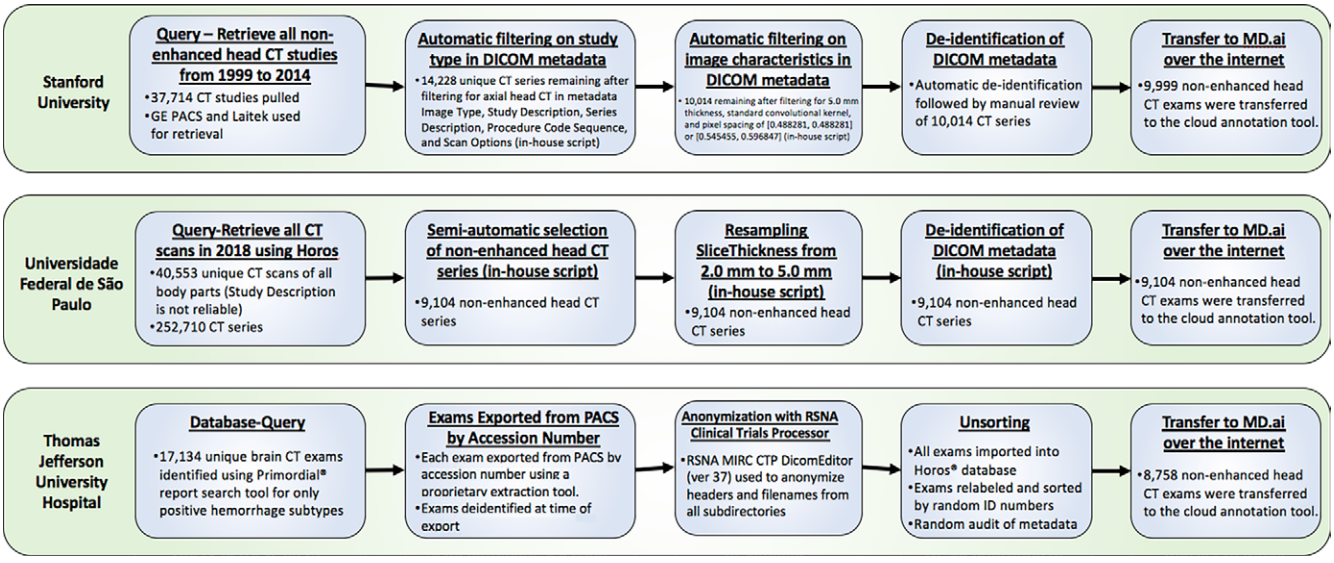

Figure 1: Workflow diagram for image data query, extraction, curation, anonymization, and exportation by the three contributing institutions. DICOM = Digital Imaging and Communications in Medicine, ID = identification, MIRC-CTP = Medical Image Resource Center-Clinical Trials Processor, PACS = picture archiving and communication system, RSNA = Radiological Society of North America.

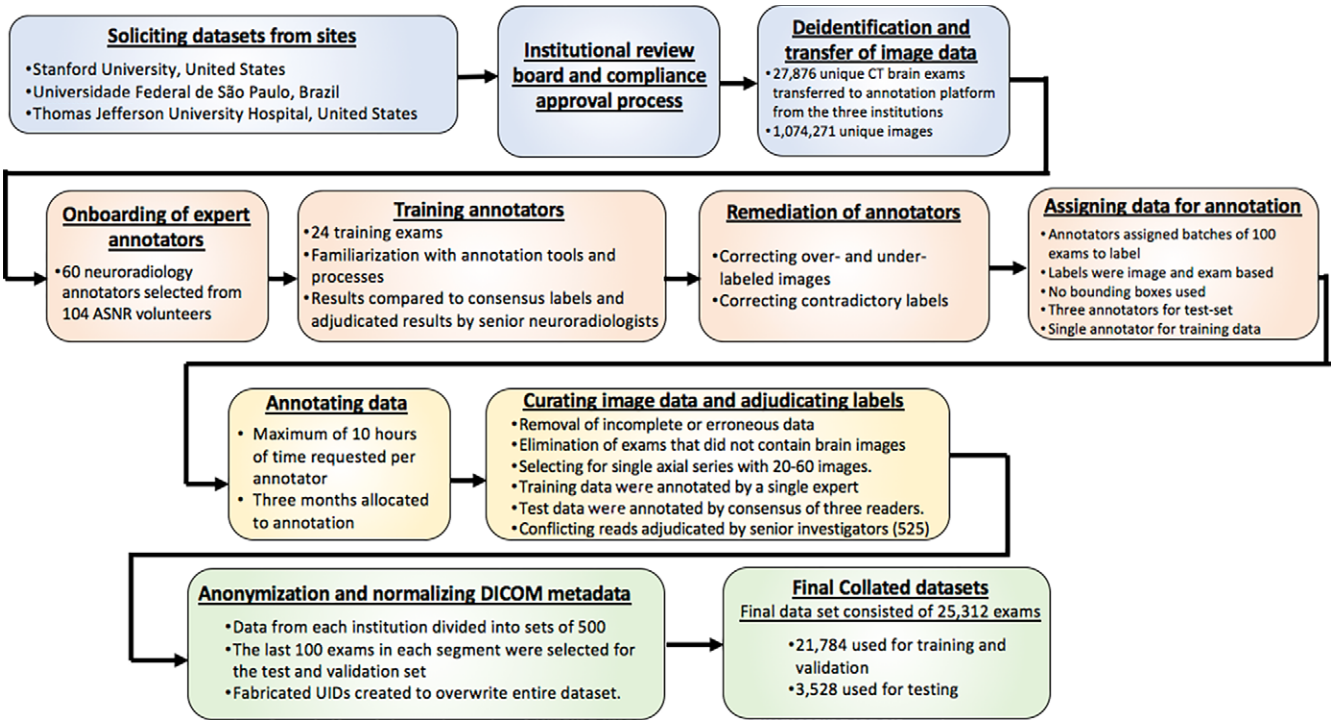

Figure 2: Workflow process diagram illustrates the steps to creation of the final brain CT hemorrhage dataset starting from solicitation from respective institutions to creation of the final collated and balanced datasets. ASNR = American Society of Neuroradiology, DICOM = Digital Imaging and Communications in Medicine, UIDs = unique identifiers.

and magnification of the images (Fig 3). Readers used personal computers to view and annotate the images using the annotation platform; diagnostic picture archiving and communication system was not used. Readers were blinded to the other readers' annotations. Adjudicators were able to see all annotations on an examination. Users were instructed to scroll through the provided series of axial images for each brain CT examination and select one or more of five hemorrhage subtype image labels: *(a)* subarachnoid, *(b)* intraventricular, *(c)* subdural, *(d)* epidural, or *(e)* intraparenchymal hemorrhage for each image. Users also were given the option to label the first and last slice of a specific hemorrhage occurrence and allow the user interface to interpolate the labels in between. When more than one hemorrhage subtype was identified, multiple appropriate labels were attached to the images in which the feature was annotated as present. If the entire examination exhibited no hemorrhagic features, the annotator could invoke one of two examination labels: *(a)* normal or *(b)* abnormal/not hemorrhage (eg, stroke, atrophy, white matter disease, hydrocephalus, tumor). An additional examination-based flag was available to the annotator to indicate an examination that might be incomplete or have an incorrect body part. This flag was used as part of the curation process to remove a particular examination from the final dataset.

The process of annotation of the dataset was undertaken as a collaborative effort between RSNA and the American Society of Neuroradiology (ASNR). With the approval and support of ASNR leadership, an open call was made for ASNR member volunteers to serve as expert annotators. Sixty individuals were selected from an initial cohort of 114, which consisted of both junior and senior members of the ASNR, including international representation. The top 15 annotators who successfully applied labels to 500 or more examinations were added as authors to this article. There was a broad spectrum of neuroradiologist volunteers with multinational representation (Table 1).

Recruited annotators were onboarded in groups of 10. Training materials and log-on credentials to the annotation tool were provided to each volunteer. Training cases were used for readers to become familiar with the annotation tool and the annotation process. Twenty-four training examinations were made available to each volunteer which they labeled using the annotation tool. Examination- and image-based annotations were recorded for each rater. Individual practice results were compared with the consensus labels and adjudicated results by senior neuroradiologists on the planning committee (L.M.P. and A.E.F., with 10 and 30 years of experience, respectively). As both image-based and examination-based labels were available to the annotators, particular attention was paid to not invoke conflicting labels in the image series as no algorithmic logic was present to prevent conflicting selections (eg, selection of an examination label of "normal" in addition to several hemorrhage image labels). Image hemorrhage subtype labels were to be invoked for any image with hemorrhage(s), or if no hemorrhage was identified, then one of the examination-based labels was to be chosen. Contradictory examination and hemorrhage image labels were either corrected through adjudication or removed as part of the curation process of the annotations. The most common user error observed was under-labeling, or inadvertent designation of a single image label to reflect the entire examination (eg, using a single hemorrhage image label regardless of the number of the images where hemorrhage was actually visible). The second most common error

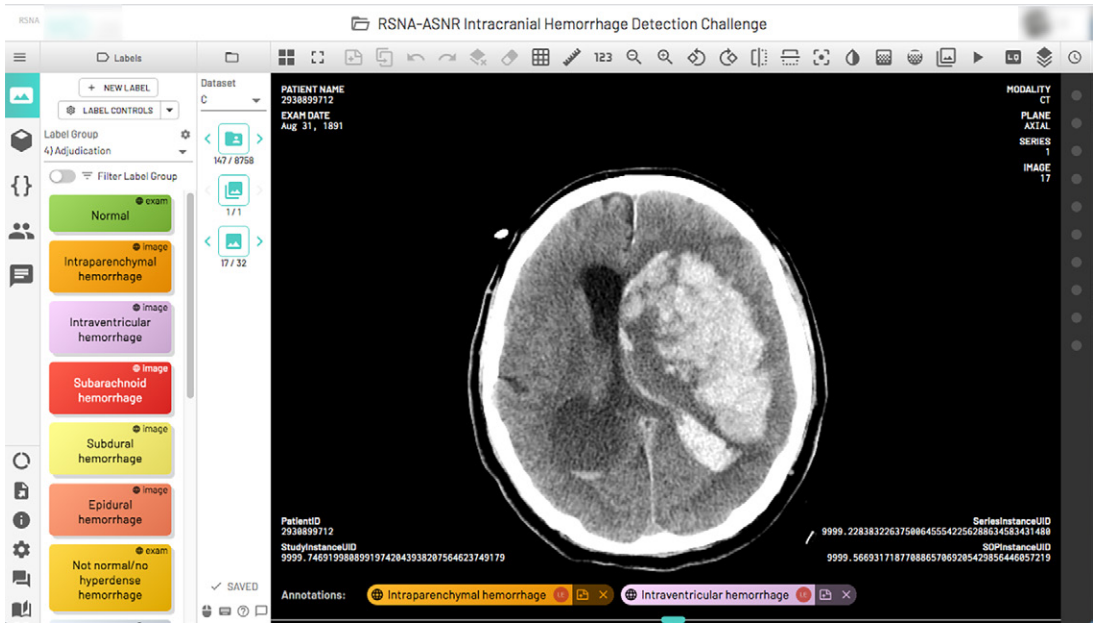

Figure 3: A complex multicompartmental cerebral hemorrhage on a single axial CT image displayed using the annotation tool in a single portal window. Hemorrhage labels (left column) relevant to the image display on the bottom of the image once selected. ASNR = American Society of Neuroradiology, RSNA =Radiological Society of North America.

was over-labeling by applying hemorrhage labels to images that extended beyond the visible feature. Misclassification of hemorrhage subtype was the third most common error identified. Remediation focused principally on correcting the under-labeling error because this would have the largest potential impact on the training and testing data. Over-labeling and subtype misclassification were treated as vagaries of the individual readers, and no attempts were made to re-educate prior to assessment of the actual dataset. Both over-labeling and subtype misclassification were addressed for scoring the validation and test sets.

The incorporation of image markup through bounding boxes, arrows, or regions of interest was considered but was abandoned due to the inherent practical limitation to circumscribe some of the more poorly defined features and overlapping features. Moreover, the process of circumscribing the image-based features, even with the use of a semiautomated segmentation tool, was determined to be far too time-consuming, inaccurate, and not reproducible. The addition of an image-based coordinate also added another dimension of complexity for assessment in these three-dimensional datasets for the purpose of scoring the challenge. It was reasoned that even in the absence of markings to localize the abnormalities, a model could "learn" the abnormal feature (ie, hemorrhage subtype) through inference from an image label alone and techniques such as saliency maps could be used for localization.

To adhere to the tight schedule for the 2019 RSNA Challenge, 3 months were allocated to perform annotations. Annotators were assigned batches of 100 examinations to review. Institutional data sources were rotated in each batch. A work commitment from our volunteer force was set at no more than 10 hours of aggregate effort per annotator, recognizing that there would be a wide range in performance per individual. Efficiency estimates using a variety of examination types and the annotation tools demonstrated that, on average, an examination could be accurately reviewed and labeled in a minute or less. On the basis of these estimates, it was projected that the 60 annotators could potentially evaluate and effectively label 36 000 examinations at a rate of one per minute for a maximum of 10 hours of effort. This provided a buffer of 11 000 potential annotations, recognizing that efficiency was subject to variation and that the validation and test sets required multiple evaluations.

Table 1: Geographic Representation of the Top 15 Neuroradiologist Annotators

The University of Alabama at Birmingham
University of Western Australia
Clínica DAPI–Diagnóstico Avançado por Imagem, Curitiba, PR, Brazil
University of Washington Medical Center
Baylor College of Medicine
University of Ottawa
Yale School of Medicine
Gold Coast University Hospital, Southport, Queensland, Australia
University of Utah
UVA Health
UT Southwestern Medical Center
Albert Einstein Healthcare Network
SUNY Downstate Medical Center

Assumptions

Because the annotation process took place in a nonclinical setting and in the absence of medical history, information regarding acuity of symptoms and findings, age of the patient, and

Table 2: Distribution of Hemorrhage Label Subtypes by Examination and Images for Both the Training and Test Sets

Subtype	Training Set Images	Test Set Images	Training Set Examinations	Test Set Examinations
Any hemorrhage type	107 933	15 902	8889	1243
Epidural	3145	208	354	23
Intraparenchymal	36 118	5468	5324	758
Intraventricular	26 205	4546	3692	616
Subarachnoid	35 675	4908	3936	528
Subdural	47 166	6555	3814	503
None	644 870	105 330	12 895	2285
Total	752 803	121 232	21 784	3528

Note.—The number of labels exceeds the actual number of examinations and images because more than one label may have been applied.

prior imaging was lacking. Without knowledge of patient age or sex, it was impossible to assess for some conditions (eg, age-appropriate volume loss or white matter disease) that would have aided in designating an examination as "no hemorrhage/not normal." The training dataset contained serial imaging on abnormal examinations, and the temporal sequence of the evolution of the abnormalities was not needed for labeling. However, in some instances, the hemorrhagic feature may have been related to an intervention (eg, postthrombolytic intraparenchymal hemorrhage or postoperative extra-axial hemorrhage). Annotators also encountered difficulty in classifying a postoperative hemorrhage especially after a craniotomy or craniectomy had been performed. Reconciling the finding against whether the feature was expected after surgery (eg, postoperative extra-axial hemorrhage after craniectomy) created some disparity for annotators. In addition, hemorrhage outside of the brain from scalp hematoma or facial and/or orbital injury when present, while important from a clinical reporting standpoint, was not a task required for this challenge.

Adjudication

Because of the large size of the combined dataset (25 312 distinct examinations), the limited time allotted to create expert labels (3 months), and the expected variation in productivity with our volunteer force of neuroradiology annotators, it was decided that the training examinations would be labeled by a single expert annotator. A total of 3528 examinations were read by three neuroradiologists in a blinded fashion. A majority vote (two of three) served as consensus for image- and examination-level labels for the triple-read examinations. A total of 525 examinations were adjudicated by one of two senior neuroradiologists if there was lack of consensus for an image label subtype category or a contradictory examination label (eg, normal or no finding vs hemorrhage).

Resulting Datasets

Although the three contributing institutions provided a similar number of unique CT examinations, the distribution of labels among sites differed due to site sampling bias during local data discovery and extraction. To address this imbalance, the distribution of the labels for each contributing site was employed to normalize distribution of examinations in the pooled data for the test and validation sets. Data from each institution were divided into sets of 500 examinations, and the last 100 examinations in each segment were selected for the test and validation sets, which were reviewed independently by an additional two neuroradiologists. The training, test, and validation sets were disjoint at the patient level. Institutional representation was equally balanced, and hemorrhage subtypes were appropriately represented in the validation and test sets. The goal was to utilize 15% of the triple-reviewed examinations for the final test set with the remainder (5%) used for validation.

Data curation also included removal of erroneous data that were deemed incorrect, incomplete, or inappropriate. Some examinations were incorrectly labeled as brain CT but actually contained another body part and had to be eliminated from the dataset. Some of the data extracts contained multiple series using different reconstruction algorithms or slice thicknesses. In this event, an annotator could flag the inappropriate series for removal and apply labels to the most appropriate series. If more than one series was inadvertently labeled and were otherwise equivalent, the series with the most labels was retained. The remaining series were removed from the final datasets. The series in the final dataset(s) had between 20 and 60 axial images that were 3–5-mm thick. Disagreements between global (examination-based) labels and image labels were adjudicated by a senior reviewer. The curation process resulted in a final dataset size of 25 312 examinations, with 21 784 for training/validation and 3528 for test. The labels for normal and "not normal/no hyperdense hemorrhage" were combined into one "No Bleed" category. Additionally, all images without labels on series with hemorrhage were labeled as "No Bleed" in the final dataset. The distribution of labels in the training and test datasets is provided in Table 2 and in Figures 4 and 5.

Limitations and Future Directions

The amount of volunteer labor required to compile, curate, and annotate a large complex dataset of this type was substantial. Even though the use case was limited to hemorrhage labels alone, it took thousands of radiologist-hours to produce a final working dataset in the stipulated time period. To optimally mitigate against misclassification in the training data, the training, validation, and test datasets should have employed multiple reviewers. The size of the final dataset and the narrow time frame to deliver it prohibited multiple evaluations for all of the available examinations. The auditing mechanism employed for training new annotators showed that the most common error produced was under-labeling of data, namely tagging an entire examination with a single image label. Raising awareness of this error early in the process before the annotators began working on the actual data helped to reduce the frequency of this error and improve consistency of the single evaluations.

Random auditing of the annotations in the training, validation, and test partitions showed that variations in hemorrhage subtype classes still occurred among observers, as did a variability in extent of hemorrhage (reflected in the number of hemorrhage subtype labels in a given series). This reflects the general variation in perceptual skills among observers and is not easily corrected through further training. Although hemorrhage subtype was the primary objective for this challenge, there was a large variety of cerebral disease states represented in the data that were never specifically classified. The deliberate exclusion of coordinates or regions of interest for any of the hemorrhage subtypes is also a limitation in the final dataset. These limitations notwithstanding, the resultant dataset achieved the stated objectives of complexity and heterogeneity.

The process of assembling a single dataset derived from multiple disparate institutions helped to identify several key opportunities for improvement in the data procurement process. Broadly, these can be divided into refinements for data discovery, curation, and anonymization. The data imbalance problem was principally related to variations in data discovery methods employed at the contributing sites. Each contributing site devised a different method for data extraction from their respective archives; the proportions of hemorrhage subclasses varied considerably by site. Deployment of a single query/retrieval solution at each contributing site potentially may have produced similar proportions of subclasses from each site thereby requiring less curation when assembling the final datasets. Automated methods to curate data prior to leaving the institutions would also have diminished the need to perform quality control tasks by the annotators and adjudicators. Finally, each contributing site employed different de-identification and anonymization solutions prior to the data leaving each respective site. This produced site-specific anonymization "signatures" that could be used to discriminate one site from another, hence creating another intrinsic bias in the data. This required another round of anonymization and synthetic unique identifier generation to normalize the examination metadata. Deployment of a single solution at each contributing site for de-identification and anonymization would have eliminated these additional steps.

As this is a public dataset, it is available for further enhancement and use including the possibility of adding multiple readers for all studies, performance of detailed segmentations, performance of federated learning on the separate datasets, and evaluation of the examinations for disease entities beyond hemorrhage.

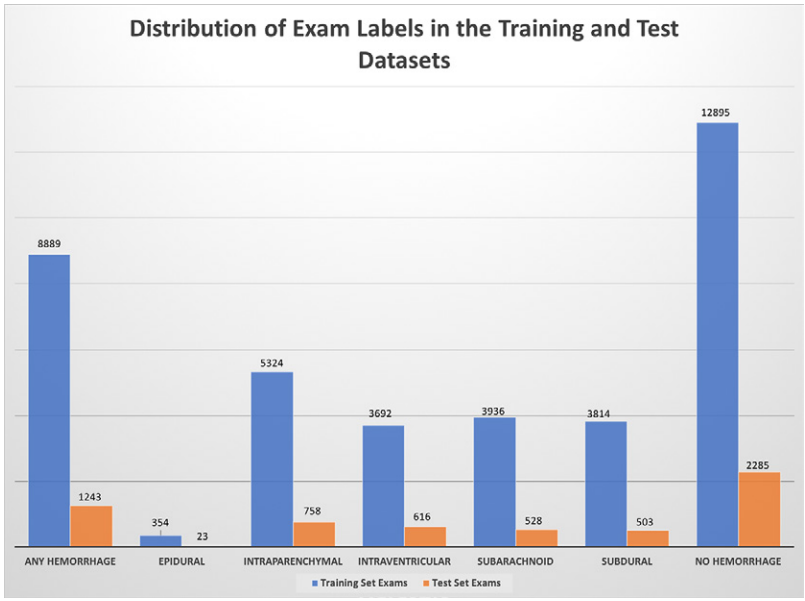

Figure 4: Distribution of examination labels in the final training (blue) and test (orange) datasets. The "any hemorrhage" designation represents when one or more of the hemorrhage subclasses were present in the entire examination.

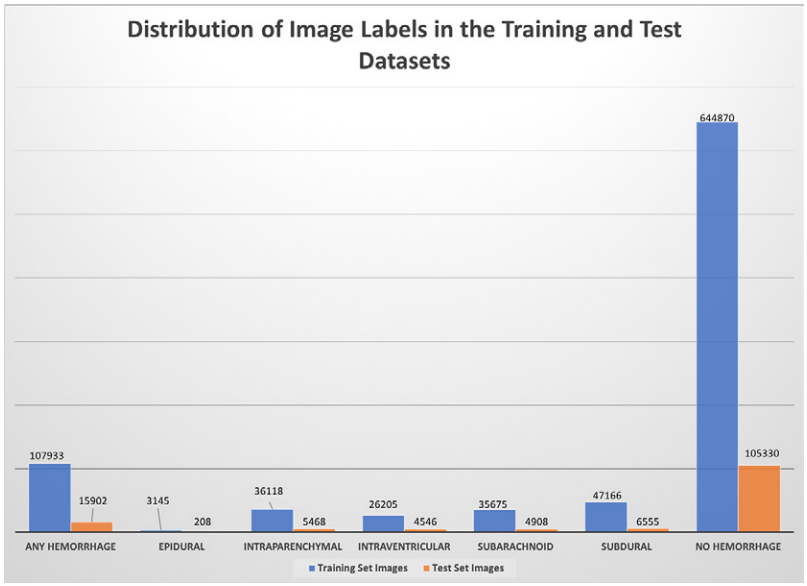

Figure 5: Distribution of image-based labels in the final training (blue) and test (orange) datasets. The "any hemorrhage" designation represents when one or more of the hemorrhage subclasses were present on an image.

Summary

The RSNA Brain Hemorrhage CT Dataset (*https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection*) is the largest public dataset of its kind containing a very large and heterogeneous collection of brain CT studies from multiple scanner manufacturers, institutions, and countries. It is also a "real-world" dataset containing complex examples of cerebral hemorrhage in both the inpatient and emergency setting. The dataset, released under a noncommercial license, has representation of a large variety of cerebral pathologic states for use in future machine learning applications. The objective of engaging with a subspe-

cialty society to leverage their unique expertise in developing a clinical use case and high-quality dataset is an effective and useful model to follow for future collaborations.

Acknowledgments: The authors would like to thank the ASNR Board of Directors, Mary Beth Hepp, Executive Director/CEO, and Rahul Bhala, MBA, MPH, of the ASNR for their support and coordination of this project. The authors also acknowledge the contributions of Paras Lakhani, MD, Richard J. T. Gorniak, MD, Katherine Andriole, PhD, Bradley Erickson, MD, PhD, Daniel Rubin, MD, MS, Tara Retson, MD, Carol Wu, MD, Phil Culliton, Julia Elliott, Chris Carr, and Jamie Dulkowski.

Contributors: Amit Agarwal, MD, Fernando Aparici, MD, Ahmad Atassi, MD, Yulia Bronstein, MD, Karen Buch, MD, Rafael Burbano, MD, Harsha Chadaga, MD, Christopher Chan, MD, Amy Chen, MD, Arthur Cortez, MD, Bradley Delman, MD, Cliff Eskey, MD, Luciano Farage, MD, Wende Gibbs, MD, Gloria Guzman Perez-Carrillo, MD, Bronwyn Hamilton, MD, Christopher Hancock, MD, Alvand Hassankhani, MD, Mark Herbst, MD, Tarek Hijaz, MD, Michael Hollander, MD, Samuel Hou, MD, Rajan Jain, MD, David Joyner, MD, Jinsuh Kim, MD, Xiaojie Luo, MD, Ben McGuinness, MD, Anthony Miller, MD, Todd Mulderink, MD, Momin Muzaffar, MD, Thomas O'Neill, MD, Mark Oswood, MD, David Pastel, MD, Boyan Petrovic, MD, Vaishali Phalke, MD, Douglas Phillips, MD, Sanjay Prabhu, MD, Gary Press, MD, Rafael Rojas, MD, Zoran Rumboldt, MD, Lubdha Shah, MD, Dinesh Sharma, MD, Achint Singh, MD, Shira Slasky, MD, James Smirniotopoulos, MD, Joel Stein, MD, Mary Tenenbaum, MD, Vinson Uytana, MD.

Author contributions: Guarantors of integrity of entire study, A.E.F., S.S.H., F.C.K.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, A.E.F., L.M.P., G.S., S.S.H., M.P.L., A.A., J.N.; clinical studies, S.S.H., F.C.K., M.P.L., G.C., L. Cala, M.M., F.M., I.I., C.L.,. L.S., A.A., J.N.; experimental studies, S.S.H., F.C.K., I.I., C.L., D.J.; statistical analysis, J.K., R.B., A.S., I.I.; and manuscript editing, A.E.F., L.M.P., G.S., S.S.H., J.K., R.B., J.T.M., A.S., M.P.L.

Disclosures of Conflicts of Interest: A.E.F. disclosed no relevant relationships. **L.M.P.** Activities related to the present article: institution receives time-limited loaner access by OSU AI Lab to NVIDIA GPUs via Master Research Agreement between The Ohio State University and NVIDIA, no money transfer; time-limited access by OSU AI Lab to WIP postprocessing software via Master Research Agreement between OSU and Siemens Healthineers. No money transfer; unrestrictive support of OSU AI Lab by DeBartolo Family Funds. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **G.S.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: board member and shareholder of MD.ai but no money paid to author or MD.ai. Other relationships: disclosed no relevant relationships. **S.S.H.** disclosed no relevant relationships. **J.K.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: author is consultant for INFOTECH, Soft; institution receives grants from Genentech and GE. Other relationships: disclosed no relevant relationships. **R.B.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: previously employed by Roam Analytics (employment ended prior to working on this study); stockholder in Roam Analytics. Other relationships: disclosed no relevant relationships. **J.T.M.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: consultant for Siemens (PAMA clinical decision support related); institution receives grants from GE (past and pending grants related to PTX detection); author may receive future royalties from GE related to PTX AI detector licensed to GE. Other relationships: disclosed no relevant relationships. **A.S.** Activities related to the present article: employee and shareholder in MD.ai. Activities not related to the present article: employee and shareholder in MD.ai. Other relationships: disclosed no relevant relationships. **F.C.K.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: consultant for MD.ai; employed by Diagnósticos da América (DASA). Other relationships: disclosed no relevant relationships. **M.P.L.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: consultant for and stockholder in Nines, SegMed, and Bunker Hill. Other relationships: disclosed no relevant relationships. **G.C.** disclosed no relevant relationships. **L. Cala** disclosed no relevant relationships. **L. Coelho** disclosed no relevant relationships. **M.M.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: consultant for Cerebrotech Medical Systems. Other relationships: disclosed no relevant relationships. **F.M.** disclosed no relevant relationships. **E.M.** disclosed no relevant relationships. **I.I.** disclosed no relevant relationships. **V.Z.** disclosed no relevant relationships. **O.M.** disclosed no relevant relationships. **C.L.** disclosed no relevant relationships. **L.S.** disclosed no relevant relationships. **D.J.** disclosed no relevant relationships. **A.A.** disclosed no relevant relationships. **R.K.L.** disclosed no relevant relationships. **J.N.** disclosed no relevant relationships.

References

1. Cordonnier C, Demchuk A, Ziai W, Anderson CS. Intracerebral haemorrhage: current approaches to acute management. Lancet 2018;392(10154):1257–1268.
2. Majumdar A, Brattain L, Telfer B, Farris C, Scalera J. Detecting Intracranial Hemorrhage with Deep Learning. Conf Proc IEEE Eng Med Biol Soc 2018;2018:583–587.
3. Kuo W, Häne C, Mukherjee P, Malik J, Yuh EL. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. Proc Natl Acad Sci U S A 2019;116(45):22737–22745.
4. Dawud AM, Yurtkan K, Oztoprak H. Application of Deep Learning in Neuroradiology: Brain Haemorrhage Classification Using Transfer Learning. Comput Intell Neurosci 2019;2019:4629859.
5. Ginat DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. Neuroradiology 2020;62(3):335–340.
6. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018;392(10162):2388–2396.
7. Prevedello LM, Halabi SS, Shih G, et al. Challenges Related to Artificial Intelligence Research in Medical Imaging and the Importance of Image Analysis Competitions. Radiol Artif Intell 2019;1(1):e180031.
8. Shih G, Wu CC, Halabi SS, et al. Augmenting the National Institutes of Health Chest Radiograph Dataset with Expert Annotations of Possible Pneumonia. Radiol Artif Intell 2019;1(1):e180041.

Erratum: Construction of a Machine Learning Dataset through Collaboration

Originally published online:
Radiology: Artificial Intelligence 2020;2(3)
https://doi.org/10.1148/ryai.2020190211
Construction of a Machine Learning Dataset through Collaboration: The RSNA 2019 Brain CT Hemorrhage Challenge

Adam E. Flanders, Luciano M. Prevedello, George Shih, Safwan S. Halabi, Jayashree Kalpathy-Cramer, Robyn Ball, John T. Mongan, Anouk Stein, Felipe C. Kitamura, Matthew P. Lungren, Gagandeep Choudhary, Lesley Cala, Luiz Coelho, Monique Mogensen, Fanny Morón, Elka Miller, Ichiro Ikuta, Vahe Zohrabian, Olivia McDonnell, Christie Lincoln, Lubdha Shah, David Joyner, Amit Agarwal, Ryan K. Lee, Jaya Nath, For the RSNA-ASNR 2019 Brain Hemorrhage CT Annotators

Erratum in:
Radiology: Artificial Intelligence 2020;2(4)
DOI:10.1148/ryai.2020209002

This erratum corrects the inadvertent omission of names to the Acknowledgments section. The Acknowledgments should be listed as follows:

The authors would like to thank the ASNR Board of Directors, Mary Beth Hepp, Executive Director/CEO, and Rahul Bhala, MBA, MPH, of the ASNR for their support and coordination of this project. The authors also acknowledge the contributions of Paras Lakhani, MD, Richard J. T. Gorniak, MD, **Sarah Hooper, Jared Dunnmon, Henrique Carrete Jr., MD, PhD, Nitamar Abdala, MD PhD, Ernandez Santos, BIT, Raquel Nascimento Lopes, MD, Thiago Yoshio Kobayashi, MD,** Katherine Andriole, PhD, Bradley Erickson, MD, PhD, Daniel Rubin, MD, MS, Tara Retson, MD, Carol Wu, MD, Phil Culliton, Julia Elliott, Chris Carr, and Jamie Dulkowski.

This copy is for personal use only. To order printed copies, contact reprints@rsna.org