Detecting Intracranial Hemorrhage with Deep Learning

Arjun Majumdar, *Member, IEEE,* Laura Brattain, *Member, IEEE*, Brian Telfer, *Senior Member, IEEE,* Chad Farris, Jonathan Scalera

***Abstract*— Initial results are reported on automated detection of intracranial hemorrhage from CT, which would be valuable in a computer-aided diagnosis system to help the radiologist detect subtle hemorrhages. Previous work has taken a classic approach involving multiple steps of alignment, image processing, image corrections, handcrafted feature extraction, and classification. Our current work instead uses a deep convolutional neural network to simultaneously learn features and classification, eliminating the multiple hand-tuned steps. Performance is improved by computing the mean output for rotations of the input image. Postprocessing is additionally applied to the CNN output to significantly improve specificity. The database consists of 134 CT cases (4,300 images), divided into 60, 5, and 69 cases for training, validation, and test. Each case typically includes multiple hemorrhages. Performance on the test set was 81% sensitivity per lesion (34/42 lesions) and 98% specificity per case (45/46 cases). The sensitivity is comparable to previous results (on different datasets), but with a significantly higher specificity. In addition, insights are shared to improve performance as the database is expanded.**

I. Introduction

Intracranial hemorrhage refers to bleeding within the skull. Intracranial hemorrhage is an important cause of death and disability and is a cause of stroke [1]. Intracranial hemorrhage can occur spontaneously or in the setting of trauma [2-4]. Spontaneous intracranial hemorrhage can be related to variety of disease processes including, but not limited to arteriovenous malformations, ruptured aneurysms, anticoagulation, tumors, venous sinus thrombosis, hypertension, cerebral amyloid angiopathy, and hemorrhagic conversion of strokes [1,3,5,6]. Traumatic intracranial hemorrhage can occur in anyone in the setting of trauma, but patients on anticoagulation are at significantly increased risk for intracranial hemorrhage [3]. Intracranial hemorrhage is an emergency with rapid diagnosis of paramount importance for improving patient outcomes due to fast patient decline within the first few hours after onset of symptoms [2].

There are four major types of intracranial hemorrhage: epidural, subdural, subarachnoid, and intraparenchymal, which refer to the location of bleeding. Unenhanced computed tomography (CT) scans of the brain are commonly used to evaluate for intracranial hemorrhage [2]. Differences in x-ray attenuation and location of intracranial hemorrhage on unenhanced CT scans of the brain make them detectable and allow the different types of intracranial hemorrhage to be differentiated [7]. Hemorrhages can be subtle for several reasons: small size; blood pooling along the edge of normal intracranial structures with similar attenuation, in particular, the falx cerebri; or hemorrhage age of days or weeks, in which case the attenuation on CT becomes more similar to that of the brain tissue. Although intracranial hemorrhages are typically hyperattenuating compared to gray matter, they are sometimes hypoattenuating. Other features in the brain, such as calcifications of the pineal glands and choroid plexus, can also have attenuation similar to intracranial hemorrhages.

At a busy Level 1 trauma center, thousands of CT scans are generated each year to determine whether head trauma has caused intracranial hemorrhage. Although radiologists can accurately detect hematomas, a "second set of eyes" to aid in detecting subtle intracranial hemorrhages would be valuable. In addition to cueing a clinician to subtle hemorrhages, a computer-aided diagnosis (CAD) system could prioritize cases in the radiologist's queue. Flagging scans with possible significant hemorrhages could help triage, which imaging studies the radiologist should review first, and this could significantly decrease the time to detection of a potential life-threating intracranial hemorrhage. Then, once detected, intracranial hemorrhages are commonly followed with serial CT scans to evaluate the stability of the size of the hemorrhage to help make the clinical decision of whether surgery is necessary or not and automated detection and size measurement of known hemorrhages would allow for rapid comparison across multiple scans.

Development of CAD systems in general has focused on other pathologies, such as cancer. No commercial CT systems offer CAD for intracranial hemorrhage. Academic research for intracranial hemorrhage has focused on standard

A. Majumdar, L. Brattain, and B. Telfer are with the MIT Lincoln Laboratory, Lexington MA 02421 USA (telfer@ll.mit.edu).

C. Farris and J. Scalera are with the Boston Medical Center, Boston MA 02118 USA (jonathan.scalera@bmc.org).

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

This material is based upon work supported by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Assistant Secretary of Defense for Research and Engineering.

multi-step image processing approaches. Our work is novel in that it applies deep learning to this application to perform the detection in a single classification step, optimized based on data, as well as developing postprocessing to significantly improve specificity from CNN outputs.

II. Prior Work

Intracranial hemorrhage image attenuation significantly overlap with those of gray matter, meaning that simple thresholding is ineffective [8]. A variety of methods have been developed to detect intracranial hemorrhages or to measure hemorrhage volume using standard image processing approaches [7, 9-17]. These approaches typically take a sequential approach of detecting the head within the image, aligning the head, removing the skull, compensating for CT cupping artifacts, extracting handcrafted features from the imaged brain tissue, and classifying intracranial hemorrhage voxels based on the features. Most have used small datasets of 11-30 cases. In some approaches, manual intervention is sometimes needed, such as for aligning the head [7]. Comparing performance of these methods is difficult because of the different types of intracranial hemorrhages considered and because of the different performance metrics used. Some papers only consider large intracranial hemorrhages, which are the easiest to detect. Previous work has not reported on the ability to detect subtle hemorrhages, particularly along the falx cerebri. Often detection is assessed as a percentage of intracranial hemorrhage voxels, rather than considering the detection of each intracranial hemorrhage, which is more informative to a radiologist.

One of the more sophisticated approaches [7] specifically focuses on detecting small (<1cm) acute intracranial hemorrhages. In addition to executing the processing steps listed above, this approach also registers the head to an anatomic model and adjusts detections based on anatomic regions. A follow-on study [17] determined that this computer-aided diagnosis system improved detection of these hemorrhages by emergency physicians and radiology residents, although not by radiology specialists.

III. Methods

A. Data

This study retrospectively identified unenhanced head CT examinations performed from 7/1/2014-7/1/2016 at Boston University Medical Center that demonstrated different types of intracranial hemorrhage including epidural, subdural, subarachnoid, and intraparenchymal hemorrhage, as well as unenhanced CT scans of the brain of normal controls that demonstrated no evidence of intracranial hemorrhage according the inclusion and exclusion criteria. Inclusion criteria included patients over the age of 18 with unenhanced head CTs that demonstrated intracranial hemorrhage or were considered normal. Scans that were not deemed to be of diagnostic quality were excluded from the study. We identified cases from the radiology information system database by searching for keywords: "epidural", "subdural", "subarachnoid", "intraparenchymal", "unremarkable head CT", and/or "normal head CT". All the selected cases of different types of intracranial hemorrhage and controls were then verified for hemorrhage presence by checking the associated radiology report and verified cases were de-identified. The de-identified scans were provided in DICOM format to MIT Lincoln Laboratory for algorithm development and testing. This retrospective HIPAA-compliant study was approved by the Institutional Review Board at both Boston University Medical Center and MIT; the requirement for informed consent was waived.

These cases vary considerably in terms of: hemorrhage age and CT attenuation (hypo-, iso-, and hyperattenuating), hemorrhage size and type, and patient interventions, in that some patients had craniotomies, craniectomies, external ventricular drains, and/or subdural drains inserted. Head orientation also varied.

Each scan consisted of 3D images with 5mm and 1.25mm slices. Since the 5mm slices have higher signal-to-noise ratio, radiologists typically view these first to detect hemorrhages, and then use the 1.25mm data to examine in additional detail. Following this practice, the 5mm data were used for machine learning. For the 5mm data, each 3D image consists of $512 \times 512 \times 28$-40 (32 typical) voxels. The x,y image resolution is typically about 1.4mm.

A total of 134 CT scans were included in the database. Of these, 88 included at least one hemorrhage and 46 were normal controls. The cases were divided into training, validation, and test sets, as summarized in Table I. The normal cases were reserved entirely for the test set. The number of hemorrhages of each type, according to the radiologist reports, is also included in Table I. (No attempt was made to count the number of lesions of a particular type within a particular case report.) The total database consists of 4,300 images containing approximately 1.1 billion voxels.

An annotation tool was developed in Matlab to allow the hemorrhage boundaries to be marked. Annotations were performed by an attending radiologist with 9 years of experience, a radiology resident with 3 years of experience, and supplemented by an engineer with guidance from the radiologists. Radiology reports were reviewed and used to guide annotations.

Examples of intracranial hemorrhages and annotations are shown in Figure 1. In the upper image, the annotation in green of an intraventricular hemorrhage in the left occipital horn was considered by the radiologists to be a challenge case for the detection algorithm.

TABLE I. NUMBER OF CASES.

		Datasets			
		Train	Validation	Test	Total
Number Cases	Total	60	5	69	134
	Normal	0	0	46	46
Hemorrhage Type	Epidural	6	2	1	9
	Subdural	17	1	9	27
	Subarachnoid	30	3	9	42
	Intraparenchymal	48	2	18	63
	Intraventricular	21	1	3	26
	Total	122	9	40	

In the lower image, both annotations were considered as challenging for a detection algorithm.

B. Machine Learning

Recently, convolutional neural networks (CNNs) have become the prevailing method for automated medical image analysis (see [18] for a recent survey). For medical image segmentation, methods often derive from U-Net [19], which extended fully convolutional networks (FCNs) [20] to the medical image domain. In this study, we utilize a modified version of the U-Net model together with data augmentation techniques specific to CT.

The network architecture is composed of 9 convolutional blocks and operates on 2D slices. Each block contains k convolutional layers, where a layer is defined as a 3x3 convolution followed by batch normalization and a rectified linear unit (ReLU). We experiment with $k \in \{1,2\}$ and find that both models perform similarly. Following the first 4 blocks, features are downsampled using 2x2 max-pooling. After the next 4 blocks, features are upsampled using 2x2 nearest neighbor expansion followed by a 3x3 convolution, batch normalization, and ReLU. Following [19], features from blocks 1, 2, 3, and 4 are concatenated with the upsampled outputs from blocks 8, 7, 6, and 5, respectively. Finally, the output of the 9th convolutional block is processed by a 1x1 convolutional layer to produce pixel-level detections.

Two forms of data augmentation are used to improve the performance of the model: random left-right flipping and random rotation of ±10 degrees. In this study we found that data augmentation was beneficial during both training and testing. Accordingly, outputs for evaluation are produced by averaging the outputs from multiple random augmentations of the same slice.

IV. RESULTS

The trained classifier was evaluated on the 69 cases (23 with 42 intracranial hemorrhages, 46 normal) in the test set. A variety of metrics have been assessed in past papers. For a system that is intended to cue radiologists, primary metrics considered are sensitivity per lesion and specificity per

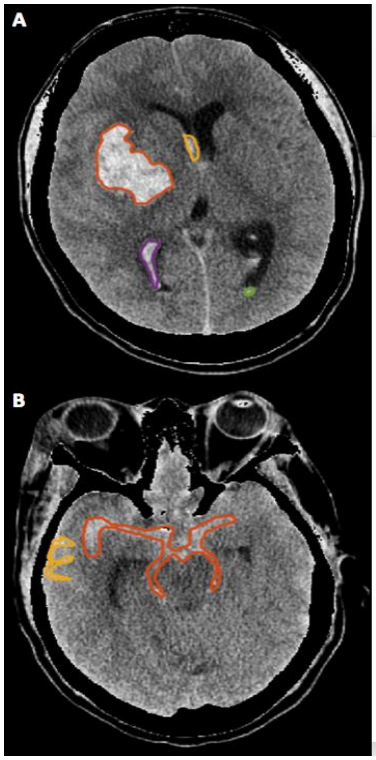

Figure 1. Examples of intracranial hemorrhages and annotations: (A) intraparenchymal (orange), intraventricular (yellow, purple, green). (B) subarachnoid (orange, yellow). CT attenuations of <0 or >100 HU are displayed as black (bones appear as black).

normal case. However, before assessing those metrics, the sensitivity at a voxel level is considered. Figure 2 compares receiver operating characteristics (ROCs) of the two CNN variations described in Section III (k = 1, 2) vs. simple image thresholding to detect voxels that fall between x and 80 Hounsfield Units (HU), where x is varied. When x = 40 HU (green operating point in Figure 2), this can be considered as a baseline rule-of-thumb approach (i.e., most hemorrhage attenuations are nominally considered to fall between 40 and 80 HU), with a per-voxel specificity of 0.935, resulting in false alarms in every test case. At a nominal operating point from applying a threshold of 0.5, the CNN approaches reduce the voxel-wise false alarms by three orders of magnitude at the cost of reducing the voxel-wise detections by a factor of 2. (Recall that our goal is to detect hemorrhages, not to detect every voxel in each hemorrhage, so a reduction in the voxel-wise detections is not necessarily a concern.) The k = 1, 2 CNNs

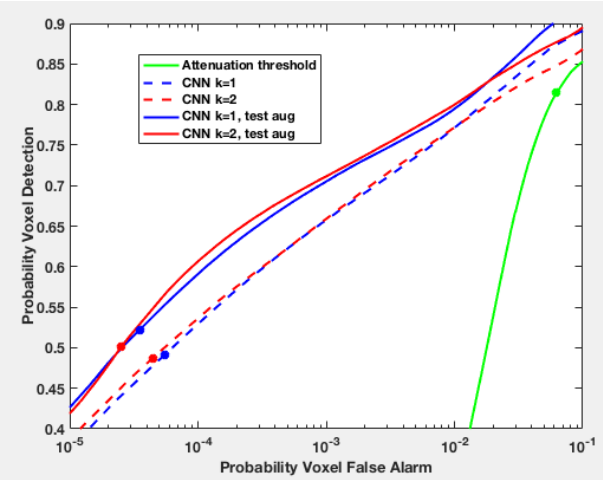

Figure 2. Voxel-wise ROC. Solid circles plot operating points for attentuation thresholding at 40 and 80 HU and for CNN output threshold at 0.5.

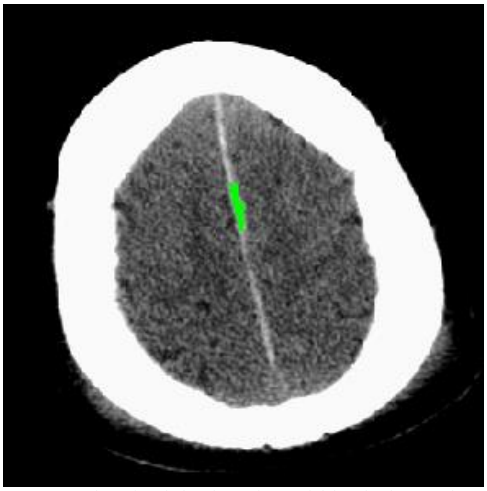

Figure 3. Example of subtle detection of blood along edge of falx cerebri.

TABLE II. CNN SENSITIVITY WITH POSTPROCESSING LOGIC.

		Number Detected	Total
Hematoma Type	Epidural	1	1
	Subdural	7	8
	Subarachnoid	7	10
	Intraparenchymal	15	17
	Intraventricular	4	6
Total		34	42

perform nearly identically. Averaging outputs from augmented test images improves performance. Considering the $k = 2$ CNN with augmented inputs, 20 of 46 normal test cases still contain false alarms.

From examining the nature of the voxel-level false alarms, two common characteristics were noted: 1) false alarms tended not to occur at the same position in adjacent slices, whereas true detections often did, 2) false alarms often occurred adjacent to cortical bones or bone marrow. Post-processing logic was applied to mitigate these causes. Voxel detections were retained only if there was: 1) a voxel detection at the same position in an adjacent slice (within ±4 voxels in x and y), 2) no adjacent voxels with an intensity of 100 HU or more. Applying this postprocessing logic significantly improved the specificity, resulting in only 1 case with false alarms, a specificity of 98% per case.

Applying the same logic to assess sensitivity, a detection was declared for a particular type of hemorrhage when at least one voxel detection fell within the annotation bounds for each intracranial hemorrhage. No attempt was made to account for detections of separate clusters of a particular type of hemorrhage in a case; these were counted as one hemorrhage and one detection or missed detection. The results are provided in Table II, yielding 34 detections out of 42 for a sensitivity of 81%.

Considering the subtle case of blood pooling along the edge of the falx cerebri, Figure 3 shows an example of the algorithm detecting this correctly (assessed by radiologist annotations) in the area where the falx appears slightly thicker.

V. DISCUSSION

Although the number of hemorrhages in the test set, especially when broken down by type, is small, the results provide important insights. The intraparenchymal hemorrhages were detected with the highest probability. These were typically hyperattenuating and surrounded by normal tissue. Similarly, the epidural hemorrhage was straightforward to detect. The subdural hemorrhage that was missed was primarily hypoattenuating with respect to normal tissue. Hypoattenuating examples were not well represented in the training set. The four intraventricular hemorrhages that were detected were larger and hyperattenuating, whereas the two that were missed were small, at the posterior of the occipital horn. The subarachnoid hemorrhages were relatively difficult to detect. These are typically narrow, with blood filling the sulci (grooves or fissures in the cortex) and sometimes isoattenuating.

The overall sensitivity of 81% is not as meaningful as the insights into the results, since the overall rate depends on the frequency and subtlety of each hematoma. However, the 98% specificity is expected to apply generally to normal cases, subject to statistical confidence arising from testing on 46 normal cases. These results cannot be directly compared to results from prior work, since the characteristics of the hematomas in these works differ, as do the performance metrics. The most relevant prior work [7] yielded a similar 82.6% sensitivity per lesion (≤1 cm) but much lower 71% specificity per case, based on essentially the same number of test cases (69 in this work vs. 70). The hemorrhages in this

prior work were more stressing to detect because they were all small, but easier to detect because they were all acute and likely hyperattenuating. The key difference is that the approach in [7] was designed carefully by hand, even including an anatomic mapping, and thus would be challenging to further optimize with a larger training set. In addition, even with careful design, manual intervention was sometimes required. In comparison, the CNN approach in this work is completely automated and is expected to further improve as the training set is refined and increased.

Future work will improve accuracy through expanding the size of the database, and will also consider including 1.25 mm slices (in addition to the 5 mm) and classifying in 3D or in additional views (sagittal and coronal planes in addition to axial). The additional annotations will be driven by active learning based on CNN performance and the particular data examples that are most needed, particularly for hypoattenuating and subarachnoid hemorrhages. It was observed that the CNN detections sometimes were improved over the radiologist annotations (partly based on limitations of the annotation tool as well as the radiologist's time), so these improved detections will be incorporated back into the training set as part of the active learning approach.

Of course, the goal of a CAD system is to improve the radiologist's accuracy, by detecting subtle hemorrhages that may be missed. The effectiveness of a CAD system needs to be assessed in conjunction with a radiologist, not only in terms of sensitivity and specificity. For example, the CAD system evaluated in [17] was found to improve radiologists' diagnoses, even with similar sensitivity and lower specificity than the results reported herein. This evaluation will be addressed in future work.

ACKNOWLEDGMENT

The contributions of Mihnea Bulugioiu to early development of annotation and visualization tools is gratefully acknowledged.

REFERENCES

[1] N.J. Fischbein, C.A.C. Wijma, "Nontraumatic intracranial hemorrhage," *Neuroimaging Clin N Am*, 2010;20(4):469–492.

[2] J.C. Hemphill, S.M. Greenberg, C.S. Anderson, et al., "Guidelines for the management of spontaneous intracerebral hemorrhage," *Stroke*, 2015;46(7):2032–2060.

[3] D.K. Nishijima, S.R. Offerman, D.W. Ballard, et al., "Risk of traumatic intracranial hemorrhage in patients with head injury and preinjury Warfarin or Clopidogrel use," S Zehtabchi, editor, *Acad Emerg Med*, 2013;20(2):140–145.

[4] J.C. Purrucker, K. Haas, T. Rizos, et al., "Early clinical and radiological course, management, and outcome of intracerebral hemorrhage related to new oral anticoagulants," *JAMA Neurol,* 2016;73(2):169.

[5] S. Sacco, C. Marini, D. Toni, L. Olivieri, A. Carolei, "Incidence and 10-year survival of intracerebral hemorrhage in a population-based registry," *Stroke*, 2009;40(2):394–399.

[6] D.B. Zahuranec, N.R. Gonzales, D.L. Brown, et al., "Presentation of intracerebral haemorrhage in a community," *J Neurol Neurosurg Psychiatry*, 2005;77(3):340–344.

[7] T. Chan, "Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain," *Comput Med Imaging Graph*, 2007;31(4–5):285–298.

[8] W.L Nowinski et al., "Characterization of interventricular and intracerebral intracranial hemorrhages in non-contrast CT," *Neuroradiology J.*, 27: 299-315, 2014.

[9] P. Maduskar, M. Acharyya, "Automatic identification of intracranial hemorrhage in non-contrast CT with large slice thickness for trauma cases," *Proc SPIE*, vol. 7260, 2009.

[10] C. Liao et al., "Computer-aided diagnosis of intracranial intracranial hemorrhage with brain deformation on computed tomography," *Computerized Medical Imaging and Graphics*, 34 (2010) 563-571.

[11] J. Kosior, "Quantomo: validation of a computer-assisted methodology for the volumetric analysis of intracerebral haemorrhage," *Int J Stroke*, vol. 6, 2011, 302-305.

[12] H.-L. Tong et al., "Automated hemorrhage slices detection for CT brain images," *Proc IVIC*, 268-279, 2011.

[13] D. Dowlashahi, "Planimetric intracranial hemorrhage measurement in patients with intraventricular hemorrhage," *Stroke*, 43, 2012, 1961-1963.

[14] K. Prakesh et al., "Segmentation and quantification of intra-ventricular/cerebral hemorrhage in CT scans by modified distance regularized level set evolution technique," *Int J Computer Assisted Radiological Surgery*, 7(5), 2012, 785-798.

[15] S. Soroushmehr, "CT image segmentation in traumatic brain injury," *Proc IEEE EMBC*, 2015, 2973-2976.

[16] M. Sun et al., "Intracranial hemorrhage detection by 3D voxel segmentation on brain CT images," *Proc 2015 IEEE Int Conf on Wireless Communications and Signal Proc*, 2015.

[17] T. Chan, H. Kuang, "Effect of a computer-aided diagnosis system on clinician's performance in detection of small acute intracranial hemorrhage on computed tomography," *Acad Radiol*, 2008, 15:290-299.

[18] G. Litjens, T. Kooi, B.E. Bejnordi, A.A. Setio, F. Ciompi, M. Ghafoorian, et al., "A survey on deep learning in medical image analysis," *Medical image Analysis*, 42, 2017, 60-88.

[19] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Proc. International Conference on Medical Image Computing and Computer-assisted Intervention,* Oct 2015, 234-241.

[20] J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, 3431-3440.