Deep Learning Approaches for Intracranial Hemorrhage Sub-Type Classification ⋆

Alexander William Wong[0000−0002−9773−7484]

University of Alberta, Edmonton AB, Canada
alex.wong@ualberta.ca

Abstract. This work applies deep learning approaches for predicting sub-categories of intracranial hemorrhages from brain computed topography scans. An intracranial hemorrhage, defined as internal bleeding within the skull, is a severe medical emergency requiring immediate and often invasive treatment. We provide a comprehensive overview for pre-processing cranial CT scans and explore multiple architectures for this classification experiment. Finally, we summarize the different approaches used by the top Kaggle RSNA Intracranial Hemorrhage Detection competition submissions. Our source code, models, and submission files are available at github.com/awwong1/rsna-hemorrhage-classification.

Keywords: Medical Image Analysis · Intracranial Hemorrhage

1 Introduction

Intracranial hemorrhage (ICH), when bleeding occurs in the cranium, is a severe medical emergency caused by physical trauma (head injury), non-traumatic cases (stroke, aneurysm), or disorders with blood clotting. Blood pools within the skull resulting in increased intracranial pressure which can irreversibly cause brain damage. In severe cases brain herniation, where parts of the brain are squeezed past the structures in the skull, may occur. To maximize survival probability, it is critical that physicians quickly recognize and treat the various types of ICH.

Diagnosing ICH is an often time consuming procedure involving coordination between multiple specialists [17]. First, a radiology technologist performs a computed topography (CT) scan to visualize the contents of the skull. Neurologists then review the scans of the cranium to find the position and type of ICH before proceeding with further therapy. Our work proposes an algorithm which takes as input a 2D slice of a CT scan and detects five different sub-categories of ICH, as specified by the RSNA Intracranial Hemorrhage Detection dataset [12].

2 Background

2.1 Task Overview

Our task is to detect the probability of five different variants of ICH existing within a CT scan of a human head: intraparenchymal, intraventricular, sub-

---

⋆ Project for CMPUT 617: Medical Image Analysis. Fall 2019. University of Alberta.
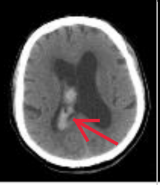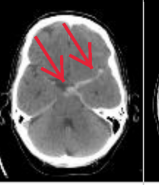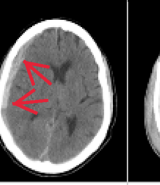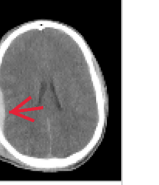
	Intraparenchymal	Intraventricular	Subarachnoid	Subdural	Epidural
Location	Inside of the brain	Inside of the ventricle	Between the arachnoid and the pia mater	Between the Dura and the arachnoid	Between the dura and the skull
Imaging					
Mechanism	High blood pressure, trauma, arteriovenous malformation, tumor, etc	Can be associated with both intraparenchymal and subarachnoid hemorrhages	Rupture of aneurysms or arteriovenous malformations or trauma	Trauma	Trauma or after surgery
Source	Arterial or venous	Arterial or venous	Predominantly arterial	Venous (bridging veins)	Arterial
Shape	Typically rounded	Conforms to ventricular shape	Tracks along the sulci and fissures	Crescent	Lentiform
Presentation	Acute (sudden onset of headache, nausea, vomiting)	Acute (sudden onset of headache, nausea, vomiting)	Acute (worst headache of life)	May be insidious (worsening headache)	Acute (skull fracture and altered mental status)

Fig. 1. Classification samples: five subtypes of intracranial hemorrhages with corresponding characteristics including mechanism, source, shape, and patient symptoms.

arachnoid, subdural, and epidural. A summary of the five ICH subtypes can be found in Fig. 1 with corresponding CT scan examples[1].

Intraparenchymal hemorrhage is characterized by bleeding within the brain parenchyma, or the functional substance composing the brain itself [4]. These hemorrhages are typically round in shape.

Intraventricular hemorrhage is where the bleeding occurs in the brain's ventricular system where the cerebrospinal fluid is produced [6]. Visually, this hemorrhage appears in the dark center of the brain.

Subarachnoid hemorrhage is when bleeding is released into the space surrounding the brain [1]. The primary visual characteristic separating subarachnoid and intraparenchymal hemorrhages are the tendril like tracks occurring along the fissures of the brain.

Subdural hematoma is the bleeding between the inner layer of the of the dura mater and the arachnoid mater [3]. This hemorrhage occurs in the intermediary membranes that surround the brain. The pooled blood is typically not in direct contact with the skull or brain parenchyma.

Epidural hematoma is when bleeding occurs between the outer membrane covering the brain (dura mater) and the skull [13]. One distinction between subdural and epidural hematomas are that epidural hematomas are typically oval/lens shaped, while the subdural hematomas are crescent shaped.

---

[1] Image source: kaggle.com/marcovasquez/basic-eda-data-visualization

2.2   Kaggle Challenge Data Overview

The CT scans provided by Kaggle are stored as Digital Imaging and Communications in Medicine (DICOM) files, an international standard for the representation, exchange, and communication of digital medical images [11]. A summary of the Kaggle `stage_2` challenge data is provided in Tab. 1.

Table 1. RSNA Intracranial Hemorrhage Detection Stage 2 file summary.

Dataset	# DICOM Files	Size (GiB)
Train + Validation	752,803	370.48
Test	121,232	59.66

All DICOM files in the training data are labeled with a `1` or `0` for each of the five ICH subtypes, with an additional `any` label indicating the presence of one or more ICH classifications. As shown in Tab. 2, a single DICOM file may have more than one ICH classification. The test dataset is used for the Kaggle competition submission and therefore the ground truth labels are not provided.

Table 2. ICH Stage 2 training data summary by sub-type and by ICH count per DICOM file. The `pos_weight` is the count of negative divided by positive samples.

Sub-Type Label	# Files	pos_weight		# ICH	# Files
Epidural	3,145	238.37		1	75,859
Intraparenchymal	36,118	19.84		2	24,826
Intraventricular	26,205	27.73		3	6,217
Subarachnoid	35,675	20.10		4	1,008
Subdural	47,166	14.96		5	23
Any	107,933	5.97		0	644,870

Each DICOM file contained a set of related meta tags storing additional information about the patient and study associated with the CT scan. For our experiment, we utilized the raw pixel data, the rescale intercept, rescale slope, window center, and window width. The raw pixel data was represented in a 512×512 matrix with values spanning the entire 16 bit signed integer space, ranging from -32,768 to 32,767.

2.3   Literature Survey

Brain CT scan analysis is an emerging domain of research in the medical computer vision community [14,7,10,9]. One popular architecture for biomedical image segmentation is U-Net, originally applied to the problem of brain tumor segmentation from 2D CT scans [14]. In [7], a deep learning approach for ICH

segmentation in 3D CT scan volumes is presented with dice accuracy comparable to radiologists. An explainable deep-learning algorithm to detect ICH in small datasets is provided in [10], accelerating physician confidence for adopting deep-learning systems into practice. Work from [9] showcased a deep learning window setting optimization module for determining appropriate CT scan windowing parameters to determine ICH and urinary stones, rather than relying on expert human knowledge.

3   Methodology

3.1   Data Pre-processing

The raw pixel data provided in the DICOM scans required additional processing to enhance the different characteristics of the human head. Windowing, or grey-level mapping, is a contrast enhancing technique used by radiologists to change the appearance of a CT scan to emphasize different structures [19]. A conversion into the Hounsfield scale, which reduced the maximum and minimum range of the raw pixel values, is required before the windowing is applied [20].
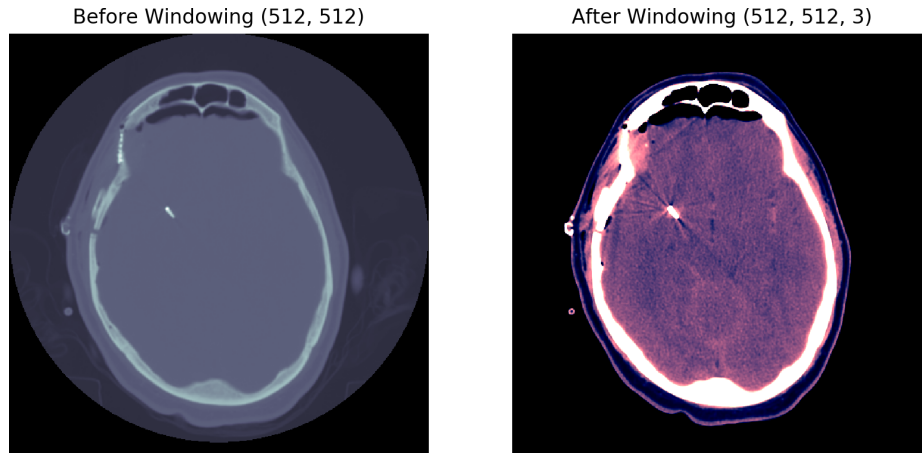

Fig. 2. ID_051688a0d: Left, grey scale raw pixel values; right, three channel brain, subdural, and soft tissue windowing displayed with `matplotlib.cm.bone` color map. The scan was labeled as having a subarachnoid and subdural intracranial hemorrhage.

We converted the single channel, grey scale pixel data into a three channel image where channels represented the brain window, the subdural (blood) window, and the soft tissues window. The reference window values used were taken from [8], where brain was centered 40 with a width of 80, subdural was centered 80 with a width of 200, and soft tissues was centered 40 with a width of 380. A sample CT scan, with the multiple window transformations and concatenated three channel image, can be found in Fig. 2.

The windowed characteristics of the cranium are emphasized in Fig. 3. The subdural channel revealed a mass near the top left of the image, indicating the pooled blood/subdural hematoma. Without the image processing, the presence of the pooled blood would be difficult to distinguish from the rest of the image.
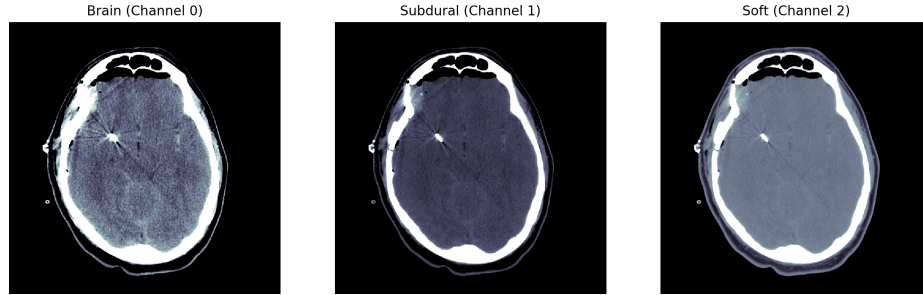

Fig. 3. ID_051688a0d: Isolated brain, subdural, and soft tissue windowing results.

3.2 Model Architecture and Training

We approach this task as a typical image classification problem using the popular ResNet family of model architectures [5]. We first initialized our models with pre-trained weights from the ILSVRC 2012 image classification challenge [15]. To adhere to our classification task, the final linear layer was modified to output 6 values. We used the Binary Cross Entropy with Logits Loss (`BCEWithLogitsLoss`). For our training optimizer, we used Adam with a learning rate of $2 * 10^{-5}$ and trained our models for 5 epochs. We used the `NVIDIA/apex`[2] PyTorch extension with optimization level `"O1"` for automatic mixed precision support and GPU acceleration during training.

For both training and evaluation, we resized the images to be 224 by 224 pixels. In addition, we applied a random shift scale rotate transformation (rotation limit of 14 degrees), and a random horizontal flip ($p = 0.5$) to the training dataset. For local model parameter/architecture search, we split the provided labelled data into a 95% train and 5% validation dataset using the `random_split` method. To evaluate the effect of weighting the positive label losses relative to its sparse representation within the dataset we reran the experiments, setting our loss `pos_weight` parameter to $\frac{\#Neg}{\#Pos}$ for each of the 6 labels.

For each architecture ResNet-34, ResNet-50 and ResNet-101, we repeated the training process three times to determine the final Kaggle submission model parameters. To generate the Kaggle submission files, we resumed training using the parameters from the model with the lowest validation loss. We fine tuned over the entire labelled dataset for an additional 1 epochs with a learning rate of $1 * 10^{-5}$ before performing inference over the unlabelled test failes.

---

[2] PyTorch mixed precision and acceleration extension: github.com/NVIDIA/apex
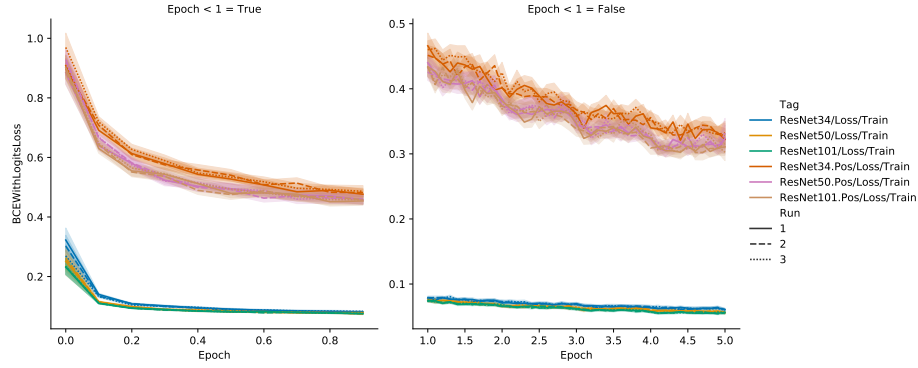
4   Results

Fig. 4. Training loss over epoch for ResNet-[34,50,101], three runs each. Shaded region represents 95% confidence interval over 0.1 epoch bins.
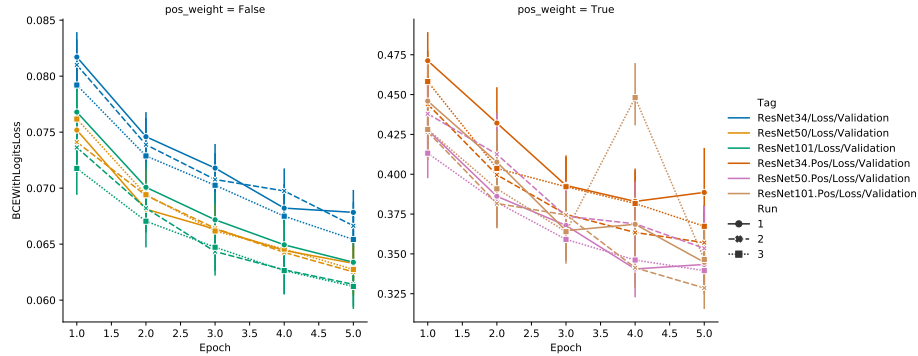

Fig. 5. Validation loss after training epoch for ResNet-[34,50,101], three runs each. Bars represent 95% confidence interval over the entire validation dataset.

The training and validation losses can be found in Fig. 4 and Fig. 5. A best-of summary for our models with comparisons to the Kaggle Top 3 leaderboard solutions can be found in Tab. 3. Our submissions were submitted after the challenge deadline, therefore estimated leaderboard positions were derived from submissions with scores nearest to ours.

The "Kaggle Score" evaluation was performed using a *weighted multi-label logarithmic* loss applied to the test dataset submission. In an effort to prevent competitors from taking advantage of the competition, the weights per label

Table 3. Experiment result summary and comparison with scores obtained by top three Kaggle competition winners.

Model	Pos	Run	Validation Loss	Kaggle Score	Est. Leaderboard
ResNet-34	N	3	0.001245	0.80829	421
	Y	3	0.079299	0.64680	412
ResNet-50	N	1	0.019656	0.97402	424
	Y	1	0.024243	0.71733	421
ResNet-101	N	3	0.000912	0.78709	421
	Y	2	0.030089	0.77097	421
Kaggle Team Name				Kaggle Score	Leaderboard
SeuTao [16]				0.04383	1
NoBrainer [2]				0.04484	2
takuoko [18]				0.04510	3

were not provided to contestants. As the test dataset labels were not released, submitting our predictions to Kaggle was the only way to derive this metric.

Our model with the best score was **ResNet-34** with `pos_weight` accounting for the dataset label sparsity. This result suggests that ResNet-50 and ResNet-101 architectures were overfitting to the labelled data set, and did not generalize well to the unlabelled test set. Additionally, all submissions that used `pos_weight` outperformed submissions using the same model architecture but did not account for the label sparsity in the dataset.

Our best approach, compared to the competition submissions, would have achieved an approximate rank of 412 on the competition leaderboard. However, the top 347 positions attained a score < 0.1, suggesting a large attainable improvement in test set performance.

5   Discussion & Future Work

The top 12 teams from the leaderboard were bound by the competition guidelines to submit a summary of their neural network architecture, including training method, and feature selection/engineering methods. Additionally, all their source code and configurations for executing and generating their submission file were made open source. All 12 gold medalists used a windowing approach to convert the raw CT scan into a multi-channel image before extracting image features using a convolutional neural network.

Our approach evaluated DICOM CT scans one at a time, using the raw pixel data and windowing pre-processing only. However, the high ranking Kaggle submissions exploited additional ID tags provided in the DICOM meta. Using various configurations of the meta `Patient`, `Series`, and `Study` ID tags, competitors were able to create sequences of DICOM files. The representation of the DICOM scans as sequences of data, rather than individual DICOM scans as isolated samples, was the key missing insight for attaining competitiveness in this challenge. Our original intention behind not taking advantage of these

tags was for our algorithm to maximize generalizeability using pixel data only. For future competitions, our approaches should incorporate all meaningful meta data to better facilitate the challenge task.

Another unexplored approach is to train a CT scan encoder network and classify on the bottleneck learned feature vectors. No ranked submissions used an encoder network as part of their deep learning pipeline, opting instead for convolutional feature extraction to recurrent neural network classification only.

6   Conclusion

To improve the speed for diagnosing ICH, we presented an algorithm to detect five different ICH sub-categories from DICOM files containing 2D CT scans of brain slices. We presented an overview of the windowing approaches for brain ICH detection and a method using `BCEWithLogitsLoss` with `pos_weight` for handling dataset label imbalance in a multi-label classification task. We compared our results using three different ResNet architectures with the Kaggle competition winners and acknowledged the key missing insight of utilizing additional tags in the DICOM files to create sequences of CT scans.

References

1. Abraham, M.K., Chang, W.T.W.: Subarachnoid hemorrhage. Emergency Medicine Clinics of North America **34**(4), 901 – 916 (2016). https://doi.org/10.1016/j.emc.2016.06.011, https://doi.org/10.1016/j.emc.2016.06.011
2. Darragh: 2nd place solution - sequential model (November 2019), https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/discussion/117228
3. Edlmann, E., Giorgi-Coll, S., Whitfield, P.C., Carpenter, K., Hutchinson, P.J.: Pathophysiology of chronic subdural haematoma: inflammation, angiogenesis and implications for pharmacotherapy. Journal of Neuroinflammation **14**(1), 108 (05 2017). https://doi.org/10.1186/s12974-017-0881-y, https://dx.doi.org/10.1186/s12974-017-0881-y
4. Gross, B.A., Jankowitz, B.T., Friedlander, R.M.: Cerebral Intraparenchymal Hemorrhage: A Review. JAMA **321**(13), 1295–1303 (04 2019). https://doi.org/10.1001/jama.2019.2413, https://doi.org/10.1001/jama.2019.2413
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
6. Hinson, H. E., H.D.F., Ziai, W.C.: Management of intraventricular hemorrhage. Current Neurology and Neuroscience Reports **10**(2), 73–82 (03 2010). https://doi.org/10.1007/s11910-010-0086-6, https://doi.org/10.1007/s11910-010-0086-6
7. Islam, M., Sanghani, P., See, A.A.Q., James, M.L., King, N.K.K., Ren, H.: Ichnet: Intracerebral hemorrhage (ich) segmentation using deep learning. In: Crimi,

A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 456–463. Springer International Publishing, Cham (2019), `https://link.springer.com/chapter/10.1007/978-3-030-11723-8_46`

8. Knipe, H., Murphy, A.: Windowing (ct) (2019), `https://radiopaedia.org/articles/windowing-ct`

9. Lee, H., Kim, M., Do, S.: Practical window setting optimization for medical image deep learning (2018), machine Learning for Health (ML4H) Workshop at NeurIPS 2018

10. Lee, H., Yune, S., Mansouri, M., Kim, M., Tajmir, S.H., Guerrier, C.E., Ebert, S.A., Pomerantz, S.R., Romero, J.M., Kamalian, S., Gonzalez, R.G., Lev, M.H., Do, S.: An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. Nature Biomedical Engineering **3**(3), 173–182 (March 2019), `https://doi.org/10.1038/s41551-018-0324-9`

11. Mildenberger, P., Eichelberg, M., Martin, E.: Introduction to the dicom standard. European radiology **12**(4), 920–927 (2002)

12. of North America, R.S.: Kaggle: Rsna intracranial hemorrhage detection (2019), `https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection`

13. Pryse-Phillips, W.: Companion to Clinical Neurology. Oxford University Press, 3 edn. (2009), p. 335

14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015), `https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28`

15. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y

16. SeuTao: 1st place solution. sequence model wins (November 2019), `https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/discussion/117210`

17. Stieg, P., Kase, C.: Intracranial hemorrhage: diagnosis and emergency management. Neurologic clinics **16**(2), 373—390 (May 1998). https://doi.org/10.1016/s0733-8619(05)70069-4, `https://doi.org/10.1016/s0733-8619(05)70069-4`

18. takuoko: 3rd place solution & become gm!! (updated with code) (November 2019), `https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/discussion/117223`

19. Tidwell, A.S.: Advanced imaging concepts: A pictorial glossary of ct and mri technology. Clinical Techniques in Small Animal Practice **14**(2), 65 – 111 (1999). https://doi.org/https://doi.org/10.1016/S1096-2867(99)80008-5, `http://www.sciencedirect.com/science/article/pii/S1096286799800085`, updates in CT and MRI, Part 1

20. Vos, W.D., Casselman, J., Swennen, G.: Cone-beam computerized tomography (cbct) imaging of the oral and maxillofacial region: A systematic review of the literature. International Journal of Oral and Maxillofacial Surgery **38**(6), 609 – 625 (2009). https://doi.org/https://doi.org/10.1016/j.ijom.2009.02.028, `http://www.sciencedirect.com/science/article/pii/S0901502709008649`