Intracranial Hemorrhage Detection with Deep Convolutional Neural Networks

Brian Murphy BEM6982@NYU.EDU

Nivedha Satyamoorthi SATYAN01@NYU.EDU

Abstract

Intracranial hemorrhage (ICH) accounts for 10% of strokes in the U.S. Deep learning methods such as convolutional neural networks have been used to diagnose medical conditions from images including detecting ICH from CT scans. We compare and evaluate the performance of ResNet18 and DenseNet121 in detecting the presence of ICH and its subtypes from a dataset of cranial CT scans compiled by Radiological Society of North America (RSNA).

Keywords: Deep Learning, Convolutional Neural Network, Intracranial Hemorrhage, Computed Tomography Scans

1. Motivation

Intracranial hemorrhage (ICH), bleeding that occurs inside the cranium caused by vascular rupture, accounts for 10% of strokes in U.S. (Bhatia, 2019). Stroke is the number 5 cause of death (Heart.org) and a leading cause of disability in the U.S. Early identification of the location and type of any hemorrhage is crucial for prognosis as this condition can worsen within the first few hours after occurrence. Availability of trained radiologists is also a major concern. Automatic identification and classification of ICH using computer vision algorithms can help in timely diagnosis (Sage and Badura, 2020). We aim to use deep convolutional neural networks (CNN) to detect the presence of ICH and its subtypes from cranial computed tomography (CT) scans.

2. Introduction

ICH is a critical condition leading to death or disability. ICH can be connected to several causes such as vascular abnormalities, venous infarction, and tumors. Irrespective of the cause, ICH is a major threat to patients lives. Based on the location of hemorrhage, ICH is classified into epidural, subdural, subarachnoid, intraparenchymal, and intraventricular hemorrhage.

When a patient presents with ICH symptoms, CT scan is almost always the first imaging modality used to assess patients. CT scans are 3D medical images that comprise several 2D slices or images stacked upon each other. Highly trained radiologists typically analyze CT scans of the patient's brain to identify and determine the type of hemorrhage.

CNN have been used to diagnose several illnesses including ICH using medical images such as CT scans. Some of the recent and related works include Chang et al. (2018),

Napoletano et al. (2018), Tao et al. (2020), Tong et al. (2020), Sage and Badura (2020), Bhatia (2019), Serte and Demirel (2021), Wang et al. (2019), Ye et al. (2019).

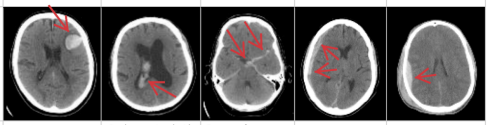
3. Hypothesis

The CNN is flexible to be trained with medical images and here, we propose two deep CNNs, ResNet and DenseNet, to detect the presence of ICH and its subtypes from 2D slices from head CT scans and compare their performance.

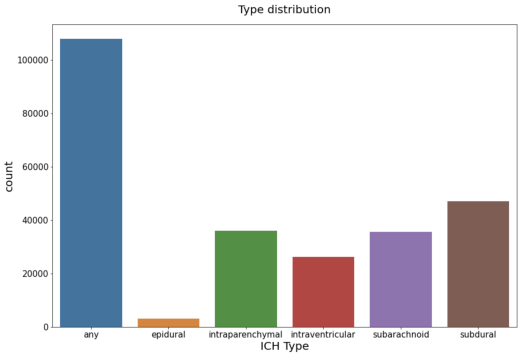
4. Data

The data used in this study were provided by the Radiological Society of North America (RSNA) through Kaggle. (RSNA, 2019). We used a subset of these data comprising 18,938 patient CT scans and 753,803 total 2D images in DICOM file format. Each DICOM file contains pixel data for a 512x512 pixel, 2D slice of a cranial CT scan and metadata for the image. The images each have six associated binary ground truth labels that indicate the presence of five ICH subtypes and an 'any' class, which indicates the presence of one or more of the five subtypes. (Kaggle, 2019).

Each hemorrhage subtype may require different image windowing to properly assess the the presence of pathology. (Figure 1(a), Sage and Badura (2020), Murphy, Stepwards (2019)). Figure 1(a) summarizes the characteristics of the ICH subtypes. Figure 1(b) shows the distribution of positive labels for the ICH subclasses.

	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)

(a) Summary of ICH subtypes



(b) Frequency of ICH diagnoses by subtype

Figure 1: Summary of ICH Subtypes in RSNA data. Figure 1(a) taken from Kaggle (2019)

5. Materials and Methods

This section details the methods used to generate a prediction probability for each subclass for each image (six binary prediction probabilities per image) in the test dataset.

5.1 Data Preprocessing

To prepare the data for analysis, we paired the pixel data from each DICOM file and with its associated label based on image IDs. We then converted the single-channel images into

three-channel images where each color channel used a different windowing setting. For each channel, we adjusted the images' pixel values by different window widths and levels to focus on a different part of the brain. The window width and level refer Hounsfield units (HUs), which measure the radiodensity in an image. (Bell and Greenway). The width refers to the HUs controlling the shade of gray, and the level refers the the HUs controlling the brightness of the image. (Stepwards, 2019). Using a function applied to the numerical pixel values generally based on input from Tang (2019) and Fink (2019), we created three channels for each image corresponding to the bone, subdural, and soft tissue windowing settings (Stepwards (2019), Murphy), and concatenated these into a three-channel image. The functions used can be found in our code base and the specific parameters used were determined empirically. Prior research has used a similar approach of incorporating multiple window settings in identifying ICH from CT scans. (Sage and Badura, 2020)

We split the data into a train, validation, and test set, using a 65%/15%/20% split ($n_{train} = 490,702; n_{val} = 112,779; n_{test} = 149,321$). We did this split randomly by patient ID because we did not want the models to train and test on images from the same patient. Figure 2 shows the percentage of positive samples for each subclass in each dataset. We randomly applied several data augmentation operations to images in the training set, including horizontal and vertical flipping, and 10-degree rotations. We also normalized the pixels arrays in all datasets and resized images to fit model input specifications.

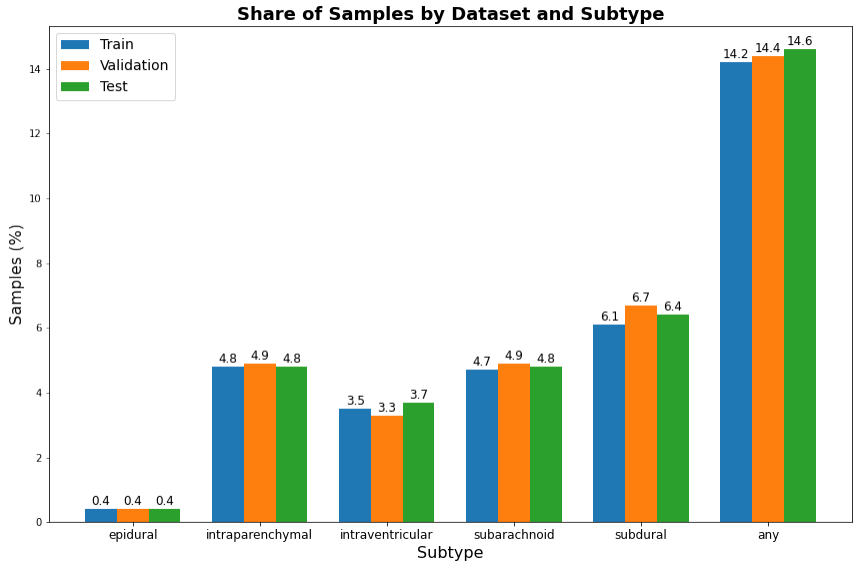


Figure 2: Percentage of positive ICH diagnoses in train, validation, and test sets by subtype

5.2 Model Selection

We wanted to compare the performance of two commonly used and well-performing deep CNN models in detecting ICH. We chose to evaluate the performance of ResNet18 and DenseNet121. (He et al. (2015), Huang et al. (2016)).

5.2.1 ResNet18

The residual neural network (ResNet) was developed in 2015 (He et al., 2015). The ResNet architecture introduced skip connections, also known as residual connections, to avoid information loss during training of deep networks. Skip connections allow for training very deep networks and can boost the performance of the model. The ResNet architecture is mainly composed of residual blocks. The residual connections preserve the gained knowledge during training and speed up the training time of the model by increasing the capacity of the network. (He et al., 2015)

The model was implemented using Pytorch. We used the smallest version of the architecture, ResNet18, to speed up the training process. We initialized the model with weights pretrained on the Imagenet database, and fine-tuned all weights. We modified the final fully-connected layer of the network to output six values for the six subclasses and added a sigmoid activation layer to produce probabilities between zero and one. (Inkawhich, 2017)

5.2.2 DenseNet121

The DenseNet architecture was first proposed in 2016. (Huang et al., 2016). DenseNet differs from previous CNN architectures because it uses the feature maps of each layer as an input to all subsequent layers. This structure helps avoid vanishing gradients, it improves feature propagation, and it incorporates feature reuse. Further, it does not increase computing time or memory usage because it does not require an increased number of parameters. (Huang et al., 2016). We chose to evaluate DenseNet because of its strong performance in image recognition, its efficiency, and because it has previously been used in image recognition tasks with CT scans. (Huang et al. (2016), Tong et al. (2020)).

We implemented the model using Pytorch, and for maximum efficiency, we chose to use DenseNet121, the smallest version of the architecture. We initialized the model weights and modified the output layer in the same way as with ResNet.

5.3 Training and Hyperparameters

For each model, we selected hyperparameters empirically by randomly subsetting the train and validation datasets and performing a grid search on the learning rate and batch size. For both models, we found that a learning rate of 0.0001 was optimal in terms of validation subset area under the receiver operating characteristic curve (ROC AUC). We also decayed the learning rate using the function 0.5^n where n is the current epoch, which improved results relative to a static learning rate. A batch size of 32 gave us the best results in both models.

We used Pytorch's version of the Adam optimizer (Kingma and Ba (2015), Loshchilov and Hutter (2019)) and binary cross entropy (BCE) loss to optimize both models. We chose to use BCE loss rather than a multi-class loss metric because our model was designed to separately predict the probability of the presence of each subclass, and because each image could have positive labels for multiple subclasses. This strategy has been used by other researchers for ICH detection with deep CNNs and has often been shown to result in higher precision than multi-class approaches. (Sage and Badura, 2020). We set our training to run for 20 epochs for each model, but stopped DenseNet training after ten epochs because loss plateaued.

6. Results

We evaluated the performance of our models on predicting subtype labels from test data samples based on accuracy, average log loss, ROC AUC, and precision-recall (PR) AUC. The results for both models are shown in Figures 3, 4, and 5.

Both models achieved high overall accuracy in the test set. The average log loss was 0.06 for DenseNet121 and .09 for ResNet18. ResNet18 and DenseNet121 both performed best at predicting the 'intraventricular' subclass in terms of ROC AUC (0.975 and 0.992 ROC AUC, respectively), and achieved 0.949 and 0.977 ROC AUC, respectively, for the 'any' subclass. Both ResNet18 and DenseNet121 performed best at predicting the 'any' subclass in terms of PR AUC, achieving 0.842 and 0.917 PR AUC, respectively. ResNet18 and DenseNet121 achieved only 0.066 and 0.159 PR AUC, respectively, for the epidural subclass.

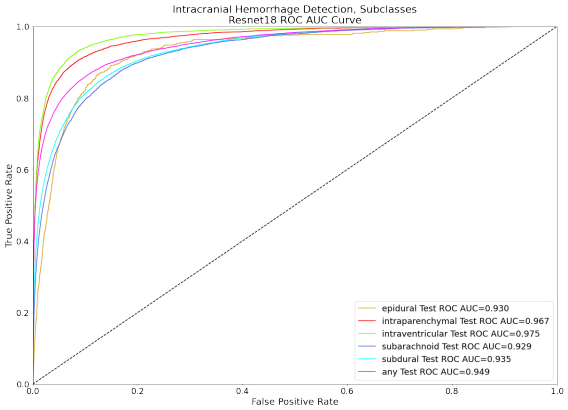
Resnet Accuracy and Average Log Loss		
Class	Accuracy	Avg Log Loss
All classes	94.8%	0.090
Epidural	93.0%	0.010
Intraparenchymal	96.7%	0.077
Intraventricular	97.5%	0.062
Subarachnoid	92.9%	0.111
Subdural	93.5%	0.120
Any	94.9%	0.176

(a) ResNet18

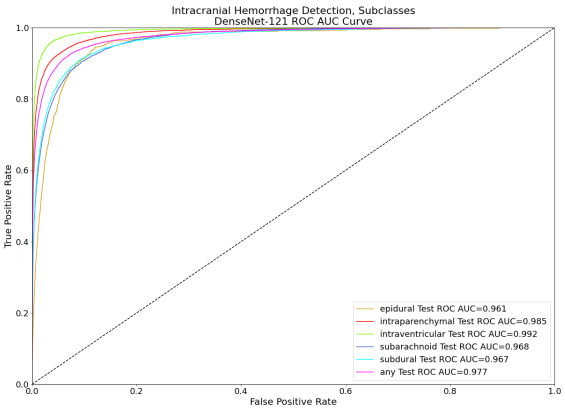
DenseNet Accuracy and Average Log Loss		
Class	Accuracy	Avg Log Loss
All classes	97.5%	0.067
Epidural	99.6%	0.016
Intraparenchymal	98.1%	0.054
Intraventricular	98.8%	0.034
Subarachnoid	96.9%	0.081
Subdural	96.4%	0.096
Any	95.4%	0.123

(b) DenseNet121

Figure 3: Test Set Accuracy and Average Log Loss per Prediction

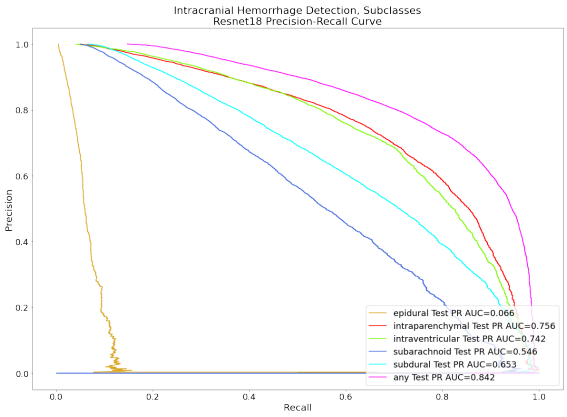


(a) ResNet18

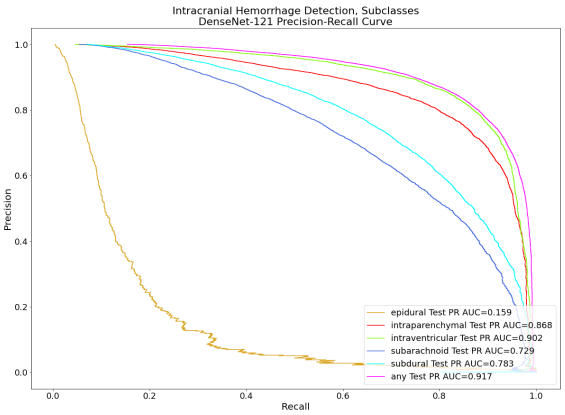


(b) DenseNet121

Figure 4: Test Set ROC AUC Curves



(a) ResNet18



(b) DenseNet121

Figure 5: Test Set Precision-Recall AUC Curves

7. Discussion

Because the percentage of positive ICH diagnoses is relatively low and an important aim of the task is to reduce false negatives, we felt that the PR AUC was the strongest indicator of model performance. For the same reason, we felt that accuracy and average log loss were not very informative. We also placed greater emphasis on diagnosing the 'any' subclass when evaluating the models because detecting the presence of ICH regardless of subclass would carry greater importance in a clinical setting than identifying subtypes of the illness.

Figure 6 shows a comparison of the models for several key performance metrics. DenseNet performed overall better than ResNet, particularly considering the PR AUC results. DenseNet was particularly successful at predicting the 'any' subclass, with 0.917 PR AUC. The models essentially failed to predict the 'epidural' subclass, with very low PR AUC scores. This was likely because there is a class imbalance and epidural does not have significant images to train the models.

One potential improvement could be to use over/under-sampling for subclasses. Additionally, training the images for the each subclass separately may improve subclass predictions. Different windowing settings could also improve performance for some subclasses. Use of 3D CNN or use a sequential model after the output of the current model could potentially improve performance.

Model Comparison		
Evaluation Metric	**Resnet18**	**DenseNet121**
Overall Acccuracy	94.8%	97.5%
Avg Log Loss	0.090	0.067
PR AUC ('Any' Class)	0.842	0.917
Best Subclass Prediction	'Any'	'Any'
Worst Subclass Prediction	'Epidural'	'Epidural'
Training Time	Slower	Faster

Figure 6: Comparison of ResNet18 and DenseNet121 Performance on ICH classification

References

Daniel J Bell and Kyle Greenway. Hounsfield unit. Radiopaedia. URL https://radiopaedia.org/articles/hounsfield-unit?lang=us. Accessed May 6, 2021.

Amit Bhatia. Application of computer vision in intracranial hemorrhage (ich) detection. *Stanford University, Computer Science Department*, 2019. URL http://cs230.stanford.edu/projects_fall_2019/reports/26256990.pdf.

P. D. Chang, E. Kuoy, J. Grinband, B. D. Weinberg, M. Thompson, R. Homo, J. Chen, H. Abcede, M. Shafie, L. Sugrue, C. G. Filippi, M. Y. Su, W. Yu, C. Hess, and D. Chow. Hybrid 3D/2D Convolutional Neural Network for Hemorrhage Evaluation on Head CT. *AJNR Am J Neuroradiol*, 39(9):1609–1616, 09 2018.

Laura Fink. Rsna ih detection - eda. Kaggle, 2019. URL https://www.kaggle.com/allunia/rsna-ih-detection-eda. Accessed May 6, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. URL http://arxiv.org/abs/1512.03385.

Heart.org. Facts, causes and risks of stroke. URL https://www.heart.org/en/about-heart-disease-in-women/facts/facts-causes-risks-and-prevention-of-stroke. Accessed May 5, 2021.

Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016. URL http://arxiv.org/abs/1608.06993.

Nathan Inkawhich. Finetuning torchvision models. Pytorch, 2017. URL https://pytorch.org/tutorials/beginner/finetuning_torchvision_models_tutorial.html. Accessed May 5, 2021.

Kaggle. RSNA intracranial hemorrhage detection: Identify acute intracranial hemorrhage and its subtypes, 2019. URL https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/. Accessed May 5, 2021.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Andrew Murphy. Windowing (CT). Radiopaedia. URL https://radiopaedia.org/articles/windowing-ct?lang=us. Accessed May 5, 2021.

P. Napoletano, F. Piccoli, and R. Schettini. Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity. *Sensors (Basel)*, 18(1), Jan 2018.

RSNA. RSNA intracranial hemorrhage detection challenge (2019), 2019. URL https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-intracranial-hemorrhage-detection-challenge-2019. Accessed May 5, 2021.

Agata Sage and Pawel Badura. Intracranial hemorrhage detection in head ct using double-branch convolutional neural network, support vector machine, and random forest. *Applied Sciences*, 10(21), 2020. ISSN 2076-3417. URL https://www.mdpi.com/2076-3417/10/21/7577.

S. Serte and H. Demirel. Deep learning for diagnosis of COVID-19 using 3D CT scans. *Comput Biol Med*, 132:104306, Mar 2021.

Stepwards. Fundamentals of computed tomography studies: Windowing, 2019. URL https://www.stepwards.com/?page_id=21646. Accessed May 5, 2021.

David Tang. See like a radiologist with systematic windowing. Kaggle, 2019. URL https://www.kaggle.com/dcstang/see-like-a-radiologist-with-systematic-windowing. Accessed May 6, 2021.

Z. Tao, H. Bingqiang, L. Huiling, Y. Zaoli, and S. Hongbin. NSCR-Based DenseNet for Lung Tumor Recognition Using Chest CT Image. *Biomed Res Int*, 2020:6636321, 2020.

Nuo Tong, Shuiping Gou, Tianye Niu, Shuyuan Yang, and Ke Sheng. Self-paced DenseNet with boundary constraint for automated multi-organ segmentation on abdominal CT images. *Physics in Medicine & Biology*, 65(13):135011, July 2020. doi: 10.1088/1361-6560/ab9b57. URL https://doi.org/10.1088/1361-6560/ab9b57.

S. Wang, C. Tang, J. Sun, and Y. Zhang. Cerebral Micro-Bleeding Detection Based on Densely Connected Neural Network. *Front Neurosci*, 13:422, 2019.

H. Ye, F. Gao, Y. Yin, D. Guo, P. Zhao, Y. Lu, X. Wang, J. Bai, K. Cao, Q. Song, H. Zhang, W. Chen, X. Guo, and J. Xia. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur Radiol*, 29(11):6191–6201, Nov 2019.

Team Responsibilities

Brian Murphy: DenseNet121 model selection, implementation, and evaluation, drafted manuscript sections 4, 5.1, 5.2.2, and 5.3.
Nivedha Satyamoorthi: ResNet18 model selection, implementation, and evaluation, drafted manuscript sections 1, 2, 3, 5.2.1, and 7.
Both: Drafted all other sections, reviewed and edited entire manuscript, background research, data preprocessing, interpretation of results.

Data and Code Availability

The data used in this project are available at https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/. Code used for data preprocessing and analysis can be accessed here: https://github.com/bmurphy1993/ICH_detection. We set up a Singularity container (see, https://sylabs.io/docs/) for the python environment and squash file for data on NYU's Greene high performance computing cluster. It may be necessary to set up a similar data file to run code.