Brain CT image hemorrhage classification

Sarah Deitch Erik Wilder Muskan Yadav Bowen Ma
Department of Mathematics, Northeastern University

May 1, 2022

1 Abstract

This paper, develops and evaluates the detection of intracranial hemorrhage in computed tomography (CT) images using deep learning algorithms and convolution neural networks (CNN) for segmentation and classification of the CT images. The motivation of this work is to democratize the access to accurate, safe and fast image-guidance, to unlock the use of image guidance directly at the point-of-care, and to enable new treatments in cases such as emergencies and bedside procedures.

2 Introduction

Hemorrhagic stroke refers to the loss of brain function due to the accumulation of blood inside the brain arising from compromised cerebral vasculature.[6] Intracranial hemorrhage (ICH) is defined as the bleeding occurring within the intracranial vault. The most common reasons include, vascular abnormalities, venous infarction, tumor, trauma effects, therapeutic anti coagulation, and cerebral aneurysm. The major symptoms that can be noticed are nausea and vomiting, confusion, dizziness, loss of vision or difficulty seeing, and loss of balance or coordination.[1-4].
Regardless of the actual cause, a hemorrhage constitutes a major threat. Therefore, an accurate diagnosis is very important for the treatment process and its success. ICH diagnosis is usually made based on the results of: an evaluation of your physical symptoms, patient medical history, computed tomography (CT) scan, magnetic resonance imaging (MRI) of your brain. These imaging tests are important to locate, categorize the extent and determine the cause of the bleed.[5].
Even though there were a number of diagnostic techniques available, several challenges remain with due to the urgency of the procedure, a complicated and time-consuming decision-making process, an inadequate level of experience. Hence, there is a significant requirement for a computer-aided diagnosis tool to assist the specialist.

Based on the brain's site of bleeding, several ICH subtypes can be classified (Figure 1). A hematoma is a collection of blood, in a clot or ball, outside of a blood vessel. **Subdural** hemorrhage (SDH) refers to bleeding between the dura and the arachnoid. It is a collection of blood on the surface of the brain. It is more frequent in older people and people with history of heavy alcohol use.
Whereas the **epidural subtype** (EDH) involves bleeding between the dura and the bone. Both frequently result from traumatic injuries. An epidural hematoma occurs when blood accumulates between the skull and the outermost covering of the brain.
Intraparenchymal hemorrhage (IPH) is bleeding within the area of brain parenchyma.
A hemorrhage inside the ventricular system is known as **intraventricular** (IVH). This is the most common type of ICH that occurs with a stroke.
Finally, blood within the subarachnoid space indicates **subarachnoid hemorrhage** (SAH) [1-5]. It is when there is bleeding between the brain and the thin tissues that cover the brain. These tissues are called meninges. The most common cause is trauma, but it can also be caused by rupture of a major blood vessel in the brain, such as from an intracerebral aneurysm. [3-4].
Hemorrhage detection and classification are particularly difficult due to similarities between the various ICH sub types and minute differences between healthy and bleeding tissues. Figure 1 shows the physiological

brain appearance and ICH sub types.

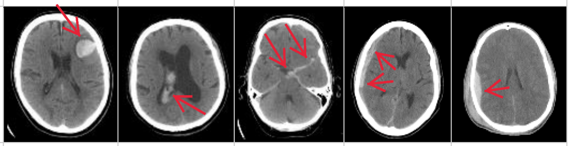
	Intraparenchymal	Intraventricular	Subarachnoid	Subdural	Epidural
Location	Inside of the brain	Inside of the ventricle	Between the arachnoid and the pia mater	Between the Dura and the arachnoid	Between the dura and the skull
Imaging					
Mechanism	High blood pressure, trauma, arteriovenous malformation, 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)

Figure 1. The data set consists of various brain CT scan slices, each of which has a hemorrhage (bleeding) within it. These hemorrhages are divided into different types, as according to the table below: (some of the slices contain multiple hemorrhages in them, and there is a 6th category for images with multiple sources of bleeding.) .

The presence of acute hemorrhagic stroke is identified clinically using computed tomography (CT) imaging, which clearly differentiates pathologic blood from normal brain tissue. The paper establishes application of machine learning and artificial neural network techniques, and programming using Python and TensorFlow to investigate the labeled data set. The goal of the project is to complete the machine learning model with python scripts for classification, and more importantly the segmentation of the brain hemorrhage in CT images.

3 Data Analysis

The data set consists of images presenting various types of bleeding: subdural, epidural, intraventricular, intraparenchymal, and subarachnoid. It is observed that many slices are not assigned to any of the types; while some contain multiple hemorrhages.The data set consists of 700,000 images, which provides a large data set for the study of ICH detection and sub type classification and segmentation.

At first the data file provided by ”Zeta Surgical” was downloaded. The labels file was imported, and each image was given an additional category label, which indicated the type of hemorrhage present using the numeric labels 0 through 6. A label of 0 meant that the image was of a normal brain, while labels 1 through 5 represented epidural, intraparenchymal, intraventricular, subarachnoid, and subdural hemorrhages respectively. The label 6 was reserved for images with multiple types of hemorrhages present.

The image paths were then added to the dataframe, so that the images could be easily imported later.

	image_path	Image	any	epidural	intraparenchymal	intraventricular	subarachnoid	subdural	category
0	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_0002d9086.jpg	0	0	0	0	0	0	0
1	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_00227fb3a.jpg	0	0	0	0	0	0	0
2	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_00a421e28.jpg	0	0	0	0	0	0	0
3	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_013bff525.jpg	1	0	1	0	1	0	6
4	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_0aec3f605.jpg	1	0	0	0	0	1	5
...	...	...	...	...	...	...	...	...	...
95	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_fa1c52d1f.jpg	1	0	0	0	1	1	6
96	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_fbe8892cc.jpg	1	0	0	0	1	0	4
97	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_fc464f1b3.jpg	1	0	0	0	1	1	6
98	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_fda990c87.jpg	1	0	0	0	1	1	6
99	C:/Users/ewild/Documents/School/2022 Spring/72...	ID_ff62735bd.jpg	1	0	0	1	0	0	3

Figure 2. The data set consists of sub types that have been normalized between 0 and 1

On observing the data set shown in Figure 2 the major difficulty faced were speculated to be as the following:

The large size of the data set provided. The data set available had about 700,000 rows of data when viewed on the labels file. From the given 400,000 images a 100,000 images matched the data in the labels file. To process the given data set a lot of computer memory was required. Additionally, along with a need of enough memory there was still a requirement of a lot of time to run the given simulations.

The proposed framework to troubleshoot these problems was as follows: As a preliminary step, the code would first be run on a small random sample of images to verify its efficacy. Then, once the code was sufficiently debugged, it would be re-run on a larger random selection of images, this time chosen so that there would be an approximately equal number of images from each hemorrhage type.

To decrease memory usage with the CNN model in particular, the images were down sampled to 128 x 128 x 3 immediately upon opening, and incorrectly sized images were filtered out as necessary. To increase classification accuracy, the code would be repeated several times with different sets of randomly selected images, which would take advantage of the breadth of the data set without overloading the computer's memory at any given time.

4 Data Modeling

Various models were implemented on this data set in order to find the most accurate classification model.

Logistic Model In statistics, the logistic model is a statistical model that models the probability of one event taking place by having the log-odds (the logarithm of the odds) for the event be a linear combination of one or more independent variables. In regression analysis, logistic regression estimates the parameters of a logistic model (the coefficients in the linear combination). Formally, in logistic regression there is a single binary dependent variable, coded by a indicator variable, where the two values are labeled "0" and "1", while the independent variables can each be a binary variable or a continuous variable . The corresponding probability of the value labeled "1" can vary between 0 and 1, hence the labeling; the function that converts log-odds to probability is the logistic function, hence the name.

For the given data set, since logistic regression could only be used for binary classification, it was decided that it would only be used to classify normal and irregular images. If the model showed a high accuracy rate, it could aid the company in avoiding redundant work and help them focus on irregular brain CT images, that is their ultimate goal.

Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant, a method used in statistics and other fields, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.[7]

Quadratic Discriminant Analysis (QDA) is a generative model. QDA assumes that each class follow a Gaussian distribution. The class-specific prior is simply the proportion of data points that belong to the class. The class-specific mean vector is the average of the input variables that belong to the class.[8]

In this paper, LDA and QDA is used firstly to classify normal and irregular images, since it gives a higher accuracy rate. In addition, it is also of great value to find irregular brain CT images in practical applications. Secondly, LDA and QDA is used to classify images according to their category labels. Unfortunately, the accuracy rate may not have been too high, because the images were complex and the models were comparatively simpler methods. Regardless, it is believed that this method had an advantage of completing the classification without taking a lot of time.

Residual neural network (ResNet) A residual neural network (ResNet) is an artificial neural network (ANN). Residual neural networks utilize skip connections, or shortcuts to jump over some layers. Typical ResNet models are implemented with double- or triple- layer skips that contain nonlinearities (ReLU) and batch normalization in between. An additional weight matrix may be used to learn the skip weights; these models are known as HighwayNets. Models with several parallel skips are referred to as DenseNets. In the context of residual neural networks, a non-residual network may be described as a plain network.[9]

ResNet is known to be a pre-trained model. The pre-trained model has been previously trained on a dataset and contains the weights and biases that represent the features of whichever dataset it was trained on. Learned features are often transferable to different data.. Hence, it was attempted to achieve a higher accuracy in image classification.

In deep learning, a **convolutional neural network (CNN)** is a class of deep neural networks, most commonly applied to analyze visual imagery. Convolutional neural networks are composed of multiple layers of artificial neurons. Artificial neurons, are mathematical functions that calculate the weighted sum of multiple inputs and outputs an activation value. When you input an image in a convolution neural network , each layer generates several activation functions that are passed on to the next layer. [10]

The goal was to import an approximately equal number of images from each category for use in training and testing the CNN. In order to maximize the number of images available for training and testing, all images were immediately down sampled to size (128, 128, 3) when they were imported. The incorrectly sized images were removed, and the remaining data was split into our test and training data. A separate "y" list was created with the category labels for each selected image; those that corresponded with an incorrectly sized images were deleted.

5 Simulations and Results

Due to computer limitations, it was decided to begin with the simplest analyses. On a small sample of 100 images, a logistic regression was run on the binary case, simply classifying each image as have a hemorrhage or not having a hemorrhage. While the sample size was small, it demonstrated remarkable good accuracy.

```
max_iter reached after 14 seconds
Logistic Regression Score: 0.950
```

Figure 3.

Figure 3 shows that logistic regression was 95 % accurate with an efficiency with respect to execution time as well. It was concluded that logistic regression works well for classifying normal and irregular images.

After observation, LDA performed similarly well in the binary case, though QDA was far less accurate. Both model took considerably less time to run.

```
## Linear Discriminant Analysis
lda = LinearDiscriminantAnalysis(store_covariance=False)
lda.fit(X_train, y_train)
print("LDA Score: %.3f"%lda.score(X_test,y_test))
```

```
LDA Score: 0.950
```

```
## Quadratic Discriminant Analysis
qda = QuadraticDiscriminantAnalysis(store_covariance=False)
qda.fit(X_train, y_train)
print("QDA Score: %.3f"%qda.score(X_test,y_test))
```

```
QDA Score: 0.450
```

Figure 4.

LDA and QDA were then applied to the multi-class case, with far worse results. The results of the analyses are in Figure 5.

```
## Linear Discriminant Analysis
lda = LinearDiscriminantAnalysis(store_covariance=False)
lda.fit(X_train, y_train)
print("LDA Score: %.3f"%lda.score(X_test,y_test))
```

```
LDA Score: 0.250
```

```
## Quadratic Discriminant Analysis
qda = QuadraticDiscriminantAnalysis(store_covariance=False)
qda.fit(X_train, y_train)
print("QDA Score: %.3f"%qda.score(X_test,y_test))
```

```
QDA Score: 0.050
```

Figure 5.

In this case, LDA and QDA performed with 25% and 5% accuracy, respectively. The analysis was then rerun on a data set with 10000 images, with somewhat improved results.

```
## Linear Discriminant Analysis
lda = LinearDiscriminantAnalysis(store_covariance=False)
lda.fit(X_train, y_train)
print("LDA Score: %.3f"%lda.score(X_test,y_test))
```

```
LDA Score: 0.402
```

```
## Quadratic Discriminant Analysis
qda = QuadraticDiscriminantAnalysis(store_covariance=False)
qda.fit(X_train, y_train)
print("QDA Score: %.3f"%qda.score(X_test,y_test))
```

```
QDA Score: 0.167
```

Figure 6.

Given the low accuracy, it was concluded that these method were likely insufficient for dealing with such a challenging classification problem.

Next for ResNet50 initially 500 images per class were used for a total of 3500 images. However, the test accuracy rate was only observed to be about 30%. Therefore, it was recommended to import more images and then observe the accuracy rate. Hence, 1000 images per class were imported for a total of 7000 images. Surprisingly the accuracy rate improved to about 40% as seen in Figure 7 below.

```
Epoch 17/20
175/175 [==============================] - 723s 4s/step - loss: 0.0042 - accuracy: 0.9996 - val_loss: 3.2610 - val_accuracy: 0.4057
```

Figure 7.

The training accuracy rate was found to be very high, but the test accuracy rate did not perform equally well. Regardless this method was found to be promising and it was concluded that there was still a lot of room for improvement. Following this further in the future it will be helpful to consider modifying some parameters or increasing the number of images used in order to improve the accuracy.

Finally it was decided to use CNN as the last classification method to achieve a better accuracy. The final attempt was to build and train a CNN from scratch. The CNN was developed with three iterations of convolution, pooling and batch normalization, along with a few dropout layers. The filter size was left relatively small, with a 5x5 matrix for the first layer and 3x3 matrices for the two subsequent layers. The activation function at each step was ReLu, with a softmax activation at the very end.

The main challenge of the work was to get high enough accuracy. An initial attempt was made 500 images from each category, trained through 67 epochs (stopped early due to early stopping) yielded a validation accuracy of 45%.

To improve the accuracy, an iterative approach was taken. First, the maximum number of images were opened for use as training and test data (13000). The data was processed as described previously, and the CNN was allowed to run for 200 epochs with an early stopping criterion of 20. On the first iteration, it yielded approximately 50% validation accuracy, with 55% training accuracy.

The weights of the model were then saved, and a new set of images were imported. The saved CNN was then trained again for 100 epochs on this new set of training data, with a validation accuracy of about 52%. The process of training and then saving the weights was then repeated two more times with new sets of training data. Each iteration gave a 2 – 3% improvement in validation accuracy, with a final maximal validation accuracy of 58.93%. By the final iteration, the training accuracy was approximately 64%, which indicated that over fitting likely not occurring in this case.

The graph of the training vs. test accuracy and loss can be seen in Figure 8:

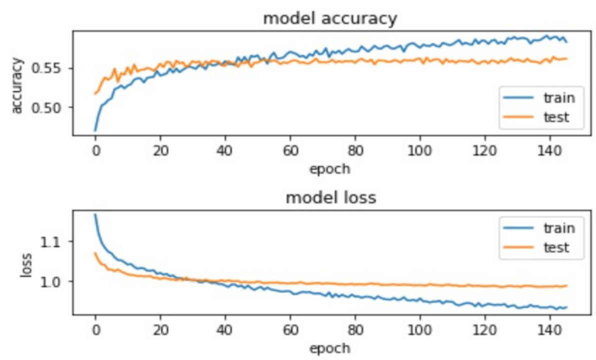


Figure 8.

6 Conclusion

CNN demonstrated more reliability compared to other established machine-learning approaches. As an improvement to the model, running the code on a computer that could process a larger portion of the 100,000 images given would be helpful. Additionally, based on the research done, this likely means that the CNN needs more training data.

7 Acknowledgement

We would like to thank Zeta Surgical for providing the data set.

We thank the company contact CTO: Mr Raahil Sha who provided insight and expertise that greatly assisted the project.

We would like to express our deep gratitude to Professor He Wang for his patient guidance, enthusiastic encouragement and useful critiques of this project work.

References

[1] Naidech, A.M. Intracranial Hemorrhage. Am. J. Respir. Crit. Care Med. 2011, 184, 998–1006.

[2] Tominari, S.; Morita, A.; Ishibashi, T.; Yamazaki, T.; Takao, H.; Murayama, Y.; Sonobe, M.; Yonekura, M.; Saito, N.; Shiokawa, Y.; et al. Prediction model for 3-year rupture risk of unruptured cerebral aneurysms in Japanese patients. Ann. Neurol. 2015, 77, 1050–1059.

[3] Yasugi, M.; Hossain, B.; Nii, M.; Kobashi, S. Relationship Between Cerebral Aneurysm Development and Cerebral Artery Shape. J. Adv. Comput. Intell. Intell. Inform. 2018, 22, 249–255.

[4] Freeman, W.D.; Aguilar, M.I.. Intracranial Hemorrhage: Diagnosis and Management. Neurol. Clin. 2012, 30, 211–240.

[5] Marcolini, E.; Stretz, C.; DeWitt, K.M. Intracranial Hemorrhage and Intracranial Hypertension. Emerg. Med. Clin. N. Am. 2019, 37, 529–544.

[6] Brain bleed/hemorrhage (intracranial hemorrhage): Causes, symptoms, treatment. Cleveland Clinic. (2020, May 4). Retrieved May 1, 2022, from https://my.clevelandclinic.org/health/diseases/14480-brain-bleed-hemorrhage-intracranial-hemorrhage/

[7] httpDash, S. K. (2021, August 18). Linear discriminant analysis: What is linear discriminant analysis. Analytics Vidhya. Retrieved May 1, 2022, from https://www.analyticsvidhya.com/blog/2021/08/a-brief-introduction-to-linear-discriminant-analysis/

[8] Hrouda-Rasmussen, S. (2021, April 7). Quadratic discriminant analysis. Towards Data Science. Retrieved May 1, 2022, from https://towardsdatascience.com/quadratic-discriminant-analysis-ae55d8a8148a/

[9] ”raghunandepu”. (2019, November 16). Understanding and implementation of residual networks(resnets). Medium. Retrieved May 1, 2022, from https://medium.com/analytics-vidhya/understanding-and-implementation-of-residual-networks-resnets-b80f9a507b9c/

[10] Mandal, M. (2021, July 23). CNN for deep learning: Convolutional Neural Networks. Analytics Vidhya. Retrieved May 1, 2022, from https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/