Comparing Augmented U-Net Models For Segmentation of Hematomas In Patients with Intracerebral Hemorrhaging

Sarah Liu
Department of Biomedical Engineering, Columbia, NY

***Abstract*—Formation of hematomas in the brain can be incredibly deadly, and are associated with a staggering mortality rate of about 40% within one month of initial observation. If a hematoma has been identified as dangerous, or if expansion seems to be likely, doctors often have to remove them surgically. This procedure, especially in a delicate environment such as the brain, is very invasive and can cause adverse affects such as infection, damage to surrounding structures, and recurrence of the hematoma. As such, to ensure proper diagnosis and to provide surgeons with ample information, it is imperative that hematomas can be effectively identified via biomedical imaging prior to the procedure. To this end, deep learning models–specifically U-Net models–have been shown to identify hematomas with accuracy that is on-par to that of human experts. This paper adapts and compares three different models based on U-Net architecture (Dimension Reduction U-Net, Dense U-Net, and vanilla U-Net) for segmentation of hematomas using brain MRI images of patients with intracerebral hemorrhaging.**

I. Introduction

INTRACEREBRAL hemorrhaging (ICH) occurs when a blood vessel in the brain bursts and is a serious condition that can be caused by a variety of events, such as physical trauma from a car accident or the spontaneous bursting due to hypertension. Furthermore, for patients regularly taking blood thinners, ICH can occur from even minors bumps to the head– and with greater severity [1]. ICH is the second most common cause of stroke, and also the most deadly. Unfortantely, it can also lead to even deadlier conditions. Resulting from the initial ICH, pools of blood clot and harden outside of blood vessels, forming dangerous hematomas that can be potentially life-threatening, and usually require immediate attention and treatment. To provide perspective on the issue, hematomas are associated with a mortality rate of about 40% within one month of initial observation [2].

In order to effectively remove hematomas from the brain, surgeons must be able to efficiently pinpoint their location and accurately approximate risk of future expansion. For this, several imaging modalities have been leveraged, namely computed tomography (CT) scans and magnetic resonance imaging (MRI). Due to its availability and ease of use, current standard practice in the field is to use a CT scan of the head to initially assess the extent of damage to the brain. Especially for specific cases such as traumatic subarachnoid hemorrhaging and epidural hematoma, CT is the preferred method of choice [3]. However, MRIs are increasingly becoming the modality of choice in emergency departments, due to them yielding greater sensitivity to smaller instances of hemorrhaging and excelling at detection of acute ICH [3]. As such, and since medical imaging acts as a primary method of diagnosis, being able to accurately identify intracerebral hematomas from both CT and MRI images is vital in providing the most immediate and appropriate possible treatment.

In identifying various artifacts from MRI scans, several publications have demonstrated that machine learning models can achieve results that are on par, if not better than human annotations [4]. One model that is commonly leveraged is the U-Net, a popular convolutional neural network (CNN) architecture designed for biomedical image segmentation tasks [5]. U-Nets are comprised of encoder and decoder paths, and utilize skip connections to maintain detailed features that may be lost in the compression process. The left encoder side of the U-Net consists of a series of convolutional and pooling layers that progressively downsample the input image, capturing high-level features while reducing spatial dimensions. The right decoder side consists of upsampling layers, which gradually increase the spatial dimensions of the feature maps. This path aims to recover spatial information lost during the downsampling phase and are concatenated with high-resolution feature maps from the encoder path via skip connections. Finally, in binary segmentation tasks, a last convolutional layer with a sigmoid activation function is applied, producing a pixel-wise segmentation mask where each pixel is assigned either a 0 or 1.

There have been several published works using various altered U-Nets or CNNs for the segmentation of hematomas using either CT or MRI images. One work proposed an adapted U-Net called **Dimension Reduction U-Net (DR -Net)** to model spontaneous intracerebral hemorrhage, achieving segmentation Dice scores of over 0.86 [6]. Instead of the traditional convolution layers found in a regular U-Net, DR U-Net used special reduced dimension convolution units, which each contain different series of normalizing, convolutional and activation layers. Another group likewise achieved high Dice scores in segmentation of hematomas using their model which they called a **Dense U-Net** [7]. This model also uses a general U-Net architectures, but replaces the convolutional layers in the encoder with interconnected dense blocks [8].

In this work, I adapt the two models for MRI images instead of CT. In later data analysis, I then compare the performance of both models, as well as that of a vanilla U-Net, in segmenting hematomas in MRI scans of patients with ICH. Performance

evaluation demonstrates that all three models achieved effectively similar predication accuracy, despite different having architectures, runtimes and computational requirements.

II. Methods

A. Data Set

The dataset used to test all three models was comprised of MRI brain scans from 345 subjects, which resulted in 11,052 total two-dimensional slices. Each slice was 256x256 pixels and paired with expert annotations that outlined hematomas via a pixel-wise binary mask, if they were present. From this aggregated dataset, 2,842 two-dimensional MRI slices were found to contain visible hematomas. From the slices with visible hematomas, 2,345 were used for training and the remaining 497 were used for testing.

B. Models

Vanilla U-Net. The model was constructed using Pytorch, and included five basic convolution blocks for both the encoder and the decoder. Dimension reduction and expansion was done with maxpooling layers and upsampling. The model included skip connections from the encoding to the decoding blocks in order to re-introduce higher-resolution features that may have been lost in the downsampling process. Finally, a sigmoid activation function was applied to generate a binary mask.

Dimension Reduction U-Net. Designed by Yu, et. al. to process and segment CT brain scans, DR U-Net functionally acts as a U-Net, but with altered data processing [6]. Instead of the traditional convolution layers found in a regular U-Net, DR U-Net was implemented with special reduced-dimension convolution units, which each contain different series of normalizing, convolutional and activation layers. For the encoding path of the model, this means that traditional convolution layers are replaced with Blocks 1 and 3 before applying Maxpooling operations. For the decoding path, the model applies Blocks 2 and then 3 to recover information and improve segmentation. Each block is comprised of two branches which process input continuously. The side branch is incorporated to avoid deep learning gradient problems, and added back to the main branch at the end of the block. Functionally, the main branch of each block works by compressing the given data to a quarter of the channels specified for that layer of the UNet before then re-encoding them back to their original dimensionality. Additionally, instead of the typical rectified linear unit, an exponential linear unit was used for all activation layers.

Dense U-Net. This model was developed by Xu, et. al. and was originally intended for the segmentation and quantification of intracranial hemorrhage from the analysis of CT scans [7]. Dense U-Net contains 4 dense blocks that make up the encoding layer, each made up of 3, 6, 12 and 8 layers with a growth factor of $k = 8$. These dense blocks were first introduced by Huang, et. al. as Dense Convolutional Networks (DenseNets), and whereas normal convolutional layers would only receive input from the directly preceding feature map, each layer in the dense block is connected with every layer preceding it [8]. The decoding layer, in contrast, follows that of a regular U-Net, including skip connections which concatenate

Fig. 1: Dimension Reduction U-Net Architecture. DR U-Net uses special dimension reduction "blocks" instead of simple convolution layers. Horizontal colored arrows indicate when the blocks are being applied: Block 1 (yellow), Block 2 (red) or Block 3 (green). Vertical arrows represent 2x2 Maxpooling (blue) and 2x2 deconvolution (purple) operations. C is the number of channels for each layer. [6]

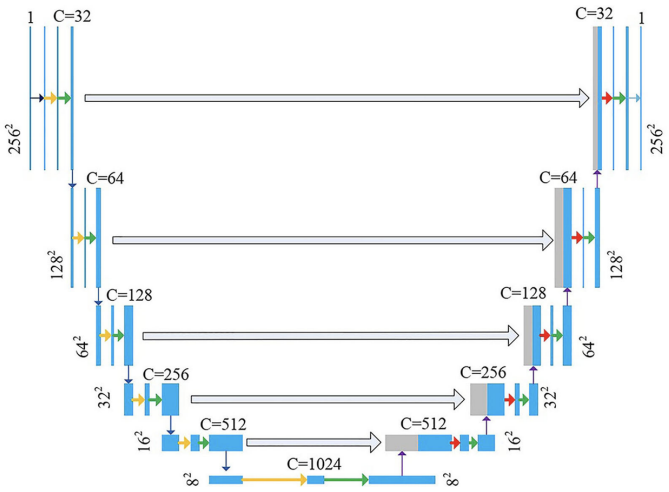
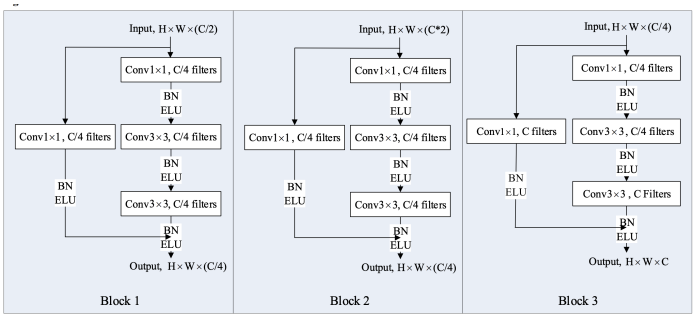


Fig. 2: Reduced-Dimension Convolution Units. Each block contains two branches and changes the total channels of the inputs. Side branches are reintroduced to the main branches at the end of each block. [6]



the dense layers with the upsampled layers and a sigmoid activation function. Additionally, Maxpooling and deconvolution transitional layers were used for dimensionality reduction and expansion, also following typical U-Net architecture.

C. Implementation and Data Analysis

Model Implementations. DR U-Net and Dense U-Net model architectures were both intended for CT scans and were adapted for 256x256 pixel MRI inputs. DR U-Net was originally developed in Tensorflow, so the dimension reduction blocks as well as overall model architecture were adapted to PyTorch [6]. The original code not having been published publicly, Dense U-Net was implemented in Pytorch using available model outlines (see Fig. 4.) and the available code for DenseNet [8]. Because the models were implemented and run through Kaggle, GPU storage was a limitation, particularly

Fig. 3: Dense U-Net Architecture. The encoding layer of the U-Net is made up of 4 dense blocks and downsampled with 1x1 convolution and 2x2 Maxpooling (red arrows). Each layer of the dense blocks is connected with preceding layers. The decoding layer is made up of regular convolutional blocks and upsampled with 2x2 deconvolution (green arrows). Batch normalization, ReLU activation and 3x3 convolution was applied between layers (black arrows). As the input images were 256x256 instead of 512x512 pixels, the dimensions for each feature map, descending, is instead as follows: 256x256, 128x128, 64x64, 32x32, 16x16. [7]

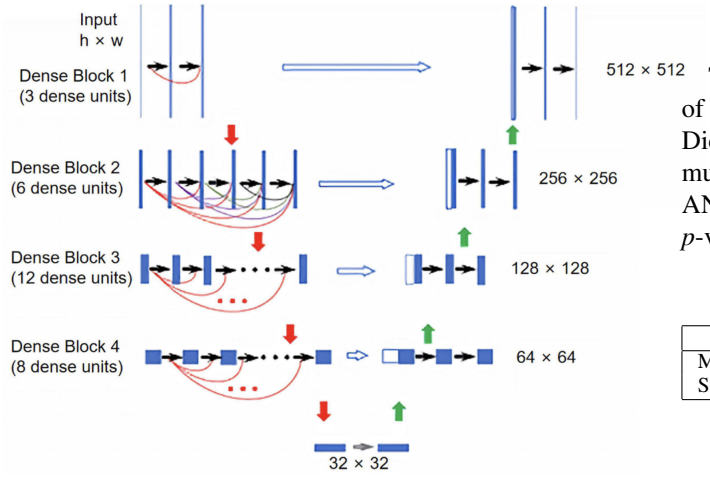
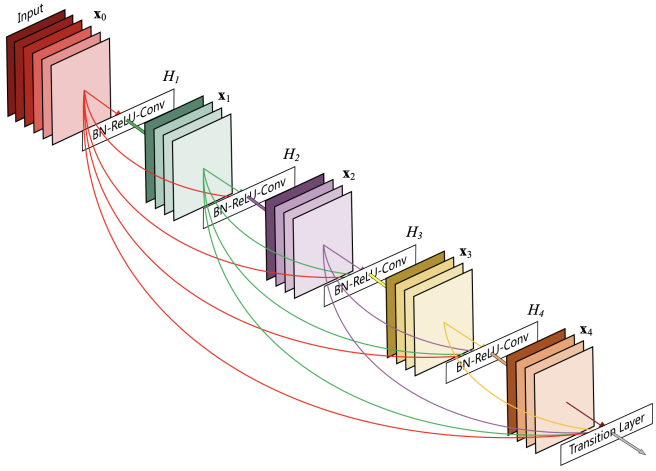


Fig. 4: DenseNet Block. This example block contains five dense layers, with a growth rate of $k = 4$. Each layer receives input from all layers before it. [8]



for Dense U-Net. In order to successfully train the model, a smaller growth rate was used ($k = 8$, instead of $k = 32$ from the model by Xu, et. al.) and bottlenecking layers were introduced to the dense layers to minimize GPU usage [7][9]. In the training phase, batch sizes were 8 and the learning rate was set to 0.001, decreasing with a factor of approximately 0.32 when loss has plateaued. All models were trained using the Dice score as a loss function and the Adam optimizer was used to optimize model parameters.

Evaluating Model Performance. To evaluate the performance of each model against the annotated ground truth masks, a Dice score was calculate for each predicted mask:

$$DICE = 2 * \frac{predicted \cap ground\,truth}{predicted + ground\,truth} \quad (1)$$

These scores were taken for each test each batch and averaged to one number, where a higher Dice score indicates a greater degree of similarity. The significance of the differences between models' Dice scores was then calculated using an Analysis of Variance (ANOVA).

III. Results

Table 1 demonstrates the averages and standard deviations of the Dice scores for each model. DR U-Net has the highest Dice score, but it is apparent that the accuracy does not vary much from model to model. Both the visual segmentations and ANOVA test corroborates this, with a notably large calculated p-value of 0.9455 ($p < 0.05$ indicates signifcance).

TABLE I: Dice Scores Across Models

	DR U-Net	Dense U-Net	U-Net
Mean	0.8361	0.8281	0.8359
Standard Deviation	0.156	0.1727	0.131

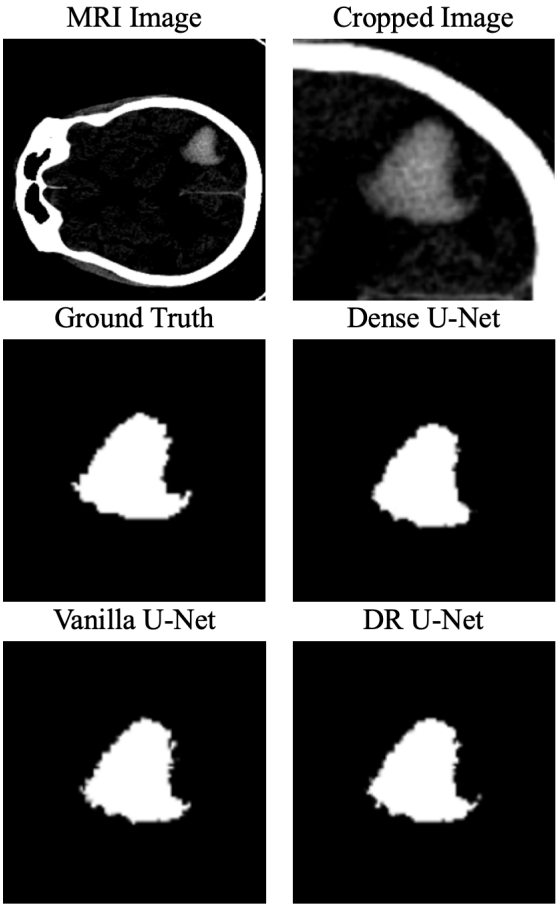
IV. Discussion

Overall, the results indicate that there is not a significant difference in model segmentation performance between regular U-Net, Dimension Reduction U-Net and Dense U-Net. This suggests that DR U-Net, the model that requires the least memory storage and least trainable parameters, can achieve results that are on par with the other models (Dice score of 0.836), even with less computational resources. DR U-Net uses 10,953,249 trainable parameters, while the vanilla U-Net equivalent uses 31,114,305 [6]. This means that the model can produce results with fewer computations, supposedly resulting in faster training times and less overall resources [6]. This phenomenon is supported by Figure 7, which shows how DR U-Net is faster if all models are trained for the same number of epochs.

However, DR U-Net takes more epochs to fully optimize its weights as compared to U-Net and Dense U-Net, as illustrated by Figure 6. This means that the model will require several more rounds of training and validation, effectively doubling or tripling its training time. Nevertheless, even though DR U-Net may take longer to effectively tune its weights, its lower number of computations due to having many fewer parameters also allows it to predict segmentation masks faster. For a commercial user, who is only looking to apply the pre-trained model, the faster prediction times can lead to faster diagnoses and more immediate administration of treatment.

Nevertheless, several factors may have impacted these results. For both DR U-Net and Dense U-Net, the model architectures had to be modified to fit within RAM and GPU storage memory constraints. For example, if Dense U-Net were to be run with a growth factor of 32 instead of 8, as

Fig. 5: Model Results. Three models, Regular U-Net, Dimension Reduction U-Net and Dense U-Net, perform segmentation on the same MRI image. No significant difference was found between the performance of the models.



originally intended, the greater depth and network complexity may possibly allow it to outperform the DR U-Net and regular U-Net models. Additionally, the data set is comprised of relatively low-resolution black and white images. If the same models were applied to images with greater resolution and larger feature maps, the higher complexity Dense U-Net model may again outperform the models with less trainable parameters.

V. Concluding Remarks

In this work, segmentation performance of hematomas in MRI images of patients with ICH was compared across three U-Net models: Dimension Reduction U-Net, Dense U-Net and a vanilla U-Net. Across all three models, performance was found to be about the same and there was no significant difference between the outputs (p = 0.95). These findings corroborated the claim that DR U-Net, when provided with comparatively longer training times, can achieve segmentation accuracy to the same level of U-Net models containing more trainable parameters. Furthermore, this work shows that Dense U-Net and DR U-Net can achieve good segmentation accuracy even when adapted from processing CT scans to MRI images.

Fig. 6: Graph of Validation Loss.

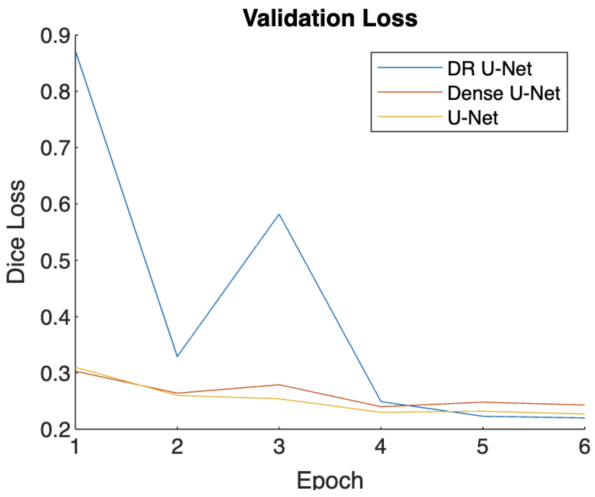
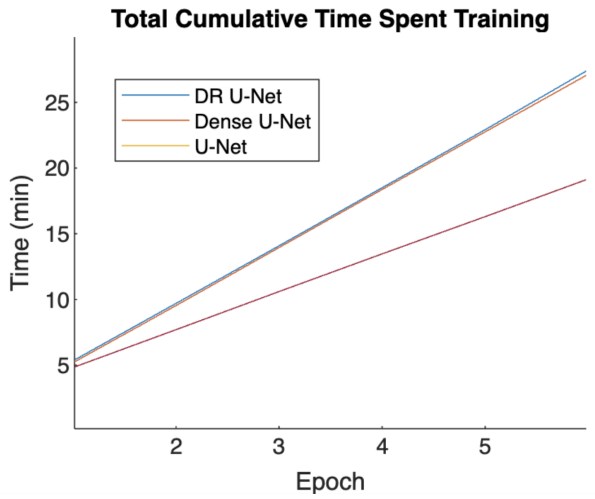


Fig. 7: Graph of Total Cumulative Time Spent Training. DR U-Net has shorter training times per epoch, requiring about only 19 minutes to train over 6 epochs, as compared to 27 minutes for both regular U-Net and Dense U-Net. Despite shorter computation times, DR U-Net achieves similar performance to the other models.



In future work, these models should be tested with higher-resolution images and with greater computational resources so that their full network capabilities may be measured against each other.

Acknowledgments

I would like to thank Dr. Jia Guo and Professor Andrew Laine with their mentorship over the course of this project, as well as Yanting Yang, Sneha Naik and Ye Tian for their continued help and support throughout the semester.

References

[1] Flaherty, M. L., Tao, H., Haverbusch, M., Sekar, P., Kleindorfer, D., Kissela, B., Khatri, P., Stettler, B., Adeoye, O., Moomaw, C. J., Broderick, J. P., & Woo, D. (2008). Warfarin use leads to larger intracerebral hematomas. Neurology, 71(14), 1084–1089. https://doi.org/10.1212/01.wnl.0000326895.58992.27

[2] Van Asch, C. J., Luitse, M. J., Rinkel, G. J., van der Tweel, I., Algra, A., & Klijn, C. J. (2010). Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. The Lancet. Neurology, 9(2), 167–176. https://doi.org/10.1016/S1474-4422(09)70340-0

[3] Heit, J. J., Iv, M., & Wintermark, M. (2017). Imaging of Intracranial Hemorrhage. Journal of stroke, 19(1), 11–27. https://doi.org/10.5853/jos.2016.00563

[4] Isaksson, L.J., Pepa, M., Summers, P. et al. (2023). Comparison of automated segmentation techniques for magnetic resonance images of the prostate. BMC Med Imaging 23, 32. https://doi.org/10.1186/s12880-023-00974-y

[5] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597. Retrieved from https://arxiv.org/abs/1505.04597

[6] Yu, N., Yu, H., Li, H., Ma, N., Hu, C., & Wang, J. (2021). A robust deep learning segmentation method for hematoma volumetric detection in intracerebral hemorrhage. Stroke, 53, 167–176. https://doi.org/10.1161/STROKEAHA.120.032243

[7] Xu, J., Zhang, R., Zhou, Z., Wu, C., Gong, Q., Zhang, H., Wu, S., Wu, G., Deng, Y., Xia, C., & Ma, J. (2021). Deep Network for the Automatic Segmentation and Quantification of Intracranial Hemorrhage on CT. Frontiers in Neuroscience, 14. https://doi.org/10.3389/fnins.2020.541817

[8] Huang, G., Liu, Z., and Weinberger, K. Q. (2016). Densely Connected Convolutional Networks. CoRR, abs/1608.06993. http://arxiv.org/abs/1608.06993

[9] Veit, A. (2018). Densenet-pytorch. GitHub Repository. Retrieved from https://github.com/andreasveit/densenet-pytorch

[10] Li, Z., You, M., Long, C., Bi, R., Xu, H., He, Q., & Hu, B. (2020). Hematoma Expansion in Intracerebral Hemorrhage: An Update on Prediction and Treatment. Frontiers in neurology, 11, 702. https://doi.org/10.3389/fneur.2020.00702.

[11] Yao, H., Williamson, C., Gryak, J., & Najarian, K. (2020). Automated hematoma segmentation and outcome prediction for patients with traumatic brain injury, Artificial intelligence in medicine, 107, https://doi.org/10.1016/j.artmed.2020.101910.

[12] Tanioka, S., Yago, T., Tanaka, K., Ishida, F., Kishimoto, T., Tsuda, K., Ikezawa, M., Araki, T., Miura, Y., & Suzuki, H. (2022). Machine learning prediction of hematoma expansion in acute intracerebral hemorrhage. Scientific reports, 12, https://doi.org/10.1038/s41598-022-15400-6.

[13] Ma, C., Wang, L., Gao, C., Liu, D., Yang, K., Meng, Z., Liang, S., Zhang, Y., & Wang, G. (2022). Automatic and Efficient Prediction of Hematoma Expansion in Patients with Hypertensive Intracerebral Hemorrhage Using Deep Learning Based on CT Images. Journal of personalized medicine, 12(5), 779. https://doi.org/10.3390/jpm12050779.

[14] Lee, H., Lee, J., Jang, J., Hwang, I., Choi, K., Park, J., Chung, J., & Choi, S. (2024). Predicting hematoma expansion in acute spontaneous intracerebral hemorrhage: integrating clinical factors with a multitask deep learning model for non-contrast head CT. Neuroradiology, 66, 577–587. https://doi.org/10.1007/s00234-024-03298-y.

[15] Brouwers, H. B., Chang, Y., Falcone, G. J., Cai, X., Ayres, A. M., Battey, T. W., Vashkevich, A., McNamara, K. A., Valant, V., Schwab, K., Orzell, S. C., Bresette, L. M., Feske, S. K., Rost, N. S., Romero, J. M., Viswanathan, A., Chou, S. H., Greenberg, S. M., Rosand, J., & Goldstein, J. N. (2014). Predicting hematoma expansion after primary intracerebral hemorrhage. JAMA neurology, 71(2), 158–164. https://doi.org/10.1001/jamaneurol.2013.5433.