An ensemble approach to intracranial hemorrhage segmentation*

Emmanuel Montagnon[1 [0000−0003−3091−9651]] and Laurent Letourneau-Guillon[1 [0000−0002−6325−7538]] as part of the 2I-mtl team

Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Montréal, Québec, Canada laurent.letourneau-guillon.1@umontreal.ca, emmanuel.montagnon.chum@ssss.gouv.qc.ca

Abstract. Herein we describe the semantic segmentation approach used for our submission to the MICCAI INSTANCE 2022 challenge. In an attempt to regularize predictions, we opted for an ensemble approach including the Attention U-Net and SegResNet (with or without variational autoencoder) architectures combined with different loss functions. The predicted output was further filtered to remove potential false positives.

Keywords: Intracranial hemorrhage · Segmentation · Deep learning.

1 Introduction

1.1 Background

Intracranial hemorrhage (ICH) is a critical neurological condition characterized by bleeding in one of five compartments: intracerebral, intraventricular, subarachnoid, subdural or epidural, see Fig. 1. There are many potential underlying causes that can be classified as either traumatic or spontaneous (non-traumatic). Etiologies also depend on the hemorrhage compartment. For example, common causes of spontaneous intracerebral hemorrhage are hypertension and amyloid angiopathy whereas the most common cause of spontaneous subarachnoid hemorrhage is a ruptured aneurysm.

Regardless of the cause, baseline hemorrhage volume is an important prognostic factor [22, 8]. In the case of intracerebral hematoma, an ellipsoid approximation using the maximal 3 plan diameters is often used in routine clinical practice to estimate this volume, the so-called ABC/2 formula [27]. This formula has also been applied to subdural and epidural hematomas [28, 4] However, there are limitations associated with this approach, especially in large and irregular hematomas [27, 26, 5, 4]. In addition, this formula cannot be used to

* Supported by Grants from 1. the Fonds de Recherche en Sante du Quebec - Association des Radiologistes du Quebec / Radiology Research and 2. the Radiological Society of North America -Seed Grant. Laurent Letourneau-Guillon is also supported by the Fonds de Recherche en Sante du Quebec - Clinical Research Scholars Junior 1.

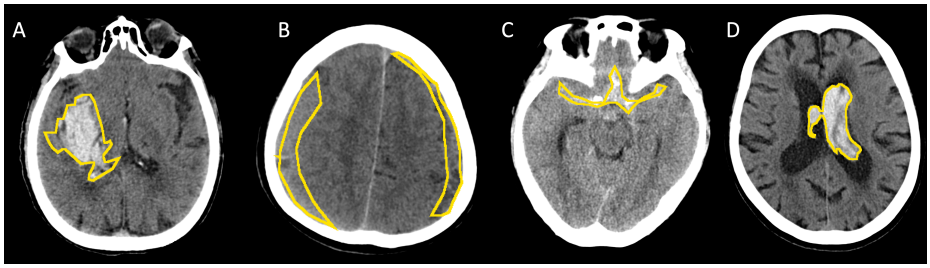

Fig. 1. Examples of intracranial hemorrhage subtypes included in the INSTANCE 2022 dataset [11], including intracerebral (A, case 13), subdural (B, case 52), subarachnoid (C, case 15) and intraventricular (D, case 26) hemorrhages. Subdural and subarachnoid hemorrhages were underrepresented in the dataset.

estimate subarachnoid and intraventricular hemorrhage volumes given their different imaging patterns. In such situations, clinicians resort to use subjective semi-quantitative scales (such as the Fisher and Graeb score. respectively [9, 2]).

The gold standard for extracting intracranial hemorrhage volume remains manual segmentation performed by expert annotators. However, the time consuming nature of such task coupled with the limited availability of expert annotators limit the use of this approach in routine clinical practice. In the last decade, significant progresses in the field of biomedical image segmentation was brought by iterative refinements of fully convolutional neural network architectures. The introduction in 2015 of the U-Net architecture [20] marked a turning point. Since then, several improvements were proposed, including the addition of various paradigms to the U-Net architecture such as DenseNet (e.g. UNet++) [32] and transformer networks (e.g. Swin-UNet [3]) as well as adaptation to 3D volumes (e.g. V-Net) [15] and to various tasks (e.g. nn-Unet) [6], among many others.

The 2022 INSTANCE challenge allows the application of these segmentation algorithms to the task of intracranial hematoma segmentation. Such segmentation can be used to automatically calculate hematoma volumes. In the case of spontaneous intracerebral hematoma, such segmentation can also be used for the purpose of hematoma progression prediction, an active area of research [12]. Indeed, recent papers suggest improved predictive performance of radiomics [29, 23, 10] or deep learning models [31, 25] compared to subjective expert interpretation of hematoma expansion markers [17]. Both the radiomics and deep learning approach could benefit from accurate hematoma segmentation either as input in the predictive model and/or to evaluate the volumetric progression over time.

2 Methods

Exploratory Data Analysis In this semantic segmentation challenge, 100 CT scans of brain hemorrhages and associated binary labels were provided for the training phase. Spatial sampling was (0.45±0.04, 0.45±0.04, 4.97±0.14) along

x, y, z axes respectively. In order to assess hemorrhage properties, we used DBSCAN[19], a density based clustering algorithm, in order to extract connected pixels corresponding to hemorrhagic areas in each exam. A total of 204 islands or clusters (*i.e* groups of connected positive labels) were found.

Table 1. Global properties of hemorrhages in the training dataset.

Property	**Mean ± Standard deviation**
Volume [mm^3]	12436.29 ± 25531.29
Elevation (z axis) [mm]	25.52 ± 23.90
Cluster minimum [HU]	19.53 ± 12.15
Cluster maximum [HU]	117.64 ± 157.13
Cluster mean [HU]	51.65 ± 8.04

We noticed that 23 of the 204 clusters exhibited maximum densities above 140 HU, with a maximum value at 1308 HU for case 52. Those values are above typical ranges associated with intracranial hemorrhages. We thus assume that some labels provided may contain pixels of tissues/structures other than hemorrhage (e.g. bone or calcified structure adjacent to the hemorrhage). For the remainder of this study, we clipped images in the range [-10; 140] HU, as we aimed to predict only hemorrhagic regions.

Data augmentation Taking into account the intracranial hemorrhage subtype distribution in the training dataset, we specifically augmented images from cases 52, 77, 95, 98, representing underrepresented subarachnoid and subdural hemorrhage subtypes, using Euler transforms consisting in rotations of either $-\frac{\pi}{2}$ or $-\frac{\pi}{2}$ around z-axis and translations ranging from -30 to 30 pixels, 10 pixels stepwise, using the SimpleITK framework [30]. The final train dataset then consisted in 6,285 2D images. Additionally, considering the limited size of the dataset, we used online data augmentation during training aiming improved generalization : random orthogonal rotations (probability .4) and cropping (probability .1) were applied to images in the training phase, using built-in methods available from the MONAI library. In order to limit class imbalance issues, models were trained only on images containing at least one pixel of positive class. All models were trained using original images size (*i.e.* 512 x 512), clipped within $[-10; 140]$ and divided by the range of considered densities, which is 150 in our configuration.

2.1 Neural Networks Architectures

In our experiments, we considered only 2D models due to the limited amount of available cases and the irregular shapes exhibited by hemorrhages. We used the MONAI framework to conduct our experiments [1]. Among available architectures, we opted for Attention U-Net [18] as well SegResNet with or without variational autoencoder [16], based on a preliminary evaluation of the different

segmentation architectures provided by the MONAI framework. Briefly, Attention U-Net uses a soft attention mechanism to highlight salient features useful for specific tasks. This is achieved by the addition of attention gates in the standard U-Net architecture [18, 20]. Such refinement produced improved prediction performances compared to U-Net in the original description and is a less computationnaly expensive alternative to a cascaded framework [18]. SegResNet, the 2018 BraTS segmentation challenge winner, uses a encoder-decoder based CNN architecture using ResNet-like blocks [16]. This architecture also incorporates a variational autoencoder branch during training, which provides an additional constraints on its layers [16]. The latter would theoretically be helpful in the current setting with a relatively small training dataset by regularizing the model. Using these two architectures, we tested different loss functions for training including combinations of Dice (Equation 1) with either Cross-Entropy loss (Equation 2) or Focal loss [13] (Equation 3), Tversky loss [21] (Equation 5), and Generalised Dice loss [24]. In both *Dice + Focal* and *Dice + Cross-Entropy*, weights were 1 for *Dice* and 0.75 for *Focal* or *Crossentropy* respectively.

In order to leverage as much cases as possible in the training dataset, we considered the first 96 cases in the training step, and the remaining for validation and test. Splitting by patient avoids data leakage, where slices from the same patient may appear in both the training and validation sets, thus introducing a bias in metrics evaluation. We chose the *AdamW* [14] for optimization, with a constant learning rate of 0.001, a batch size of 16 samples and an early stopping callback relying on validation metrics behaviour. Models were trained on two computers with NVIDIA RTX 2060 Super and NVIDIA RTX 2080Ti graphical processor units, respectively.

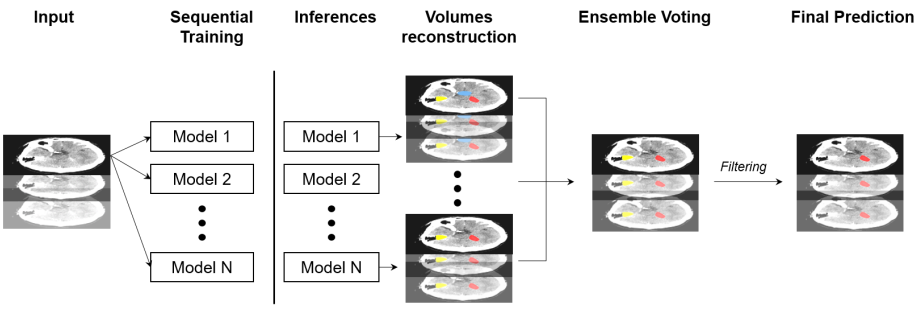

Fig. 2. Global workflow applied in this study. 2-D networks such as Attention U-Net or SegResNet (with ou without a variational auto-encoder) are first independently trained. Inference phase consists in aggregating volume predictions from each model, followed by ensemble voting and islands filtering, based on basic hemorrhages properties observed in training dataset such as volume or density.

Volume prediction and regularization The prediction process consisted in three steps: first, for each case, all selected models inferred on each volume slices. Secondly, leveraging all predictions, an ensemble voting approach allowed prediction of a final volume. This approach was inspired by the winning solution of the 2017 BraTS challenge, which potentially mitigates the risk of overfitting by combining the predictions of different models thus averaging away potential erroneous predictions from a specific model [7]. Finally, to further remove potential false positive predictions, predicted clusters were filtered by preserving ones with a volume larger than 36 pixels, an elevation above or equal to 3 slices and a mean density within $[40; 80]$ HU range. Those parameters were determined empirically, leveraging statistical properties of the training dataset hemorrhages as presented in Table 1.

3 Results

Metrics obtained on our validation dataset (cases 96 to 100 inclusively) are provided in 3.

Table 2. Metrics on cases 96 to 100 for various models and loss functions.

Model	Loss function	Mean Dice	IoU	Hausdorff distance
AttentionUNET	Dice Cross-entropy	0.220	**0.919**	108.913
AttentionUNET	Generalized Dice	0.181	0.841	123.763
AttentionUNET	Tversky	0.167	0.910	121.067
SegResNetVAE	Generalized Dice	**0.284**	0.845	**104.173**
SegResNetVAE	Dice Cross-entropy	0.191	0.908	111.074
SegResNet	Generalized Dice	0.151	0.834	148.273

4 Discussion

As shown in Table 3, no model exhibited optimal values for all three metrics simultaneously. Such observation enforces the proposed ensemble approach in order to leverage the performances of multiple models. A voting strategy basically consists in finding a trade off between predictions. For instance, looking at Figure 3, one can see that SegResNet with Generalized Dice Loss (bottom center) does not predict the small intraventricular hemorrhage region in the left atrium (red circle). Still, the voted solution includes this region thanks to other models predictions. At the opposite, if only one model provides the right prediction, its contribution to the final prediction will be lowered. Ensemble approach combined with a dedicated filtering approach thus aim to regularize predictions, diminishing false positives, but possibly at the cost of increased false negatives. As ensemble approach offers numerous possible configurations, future works may

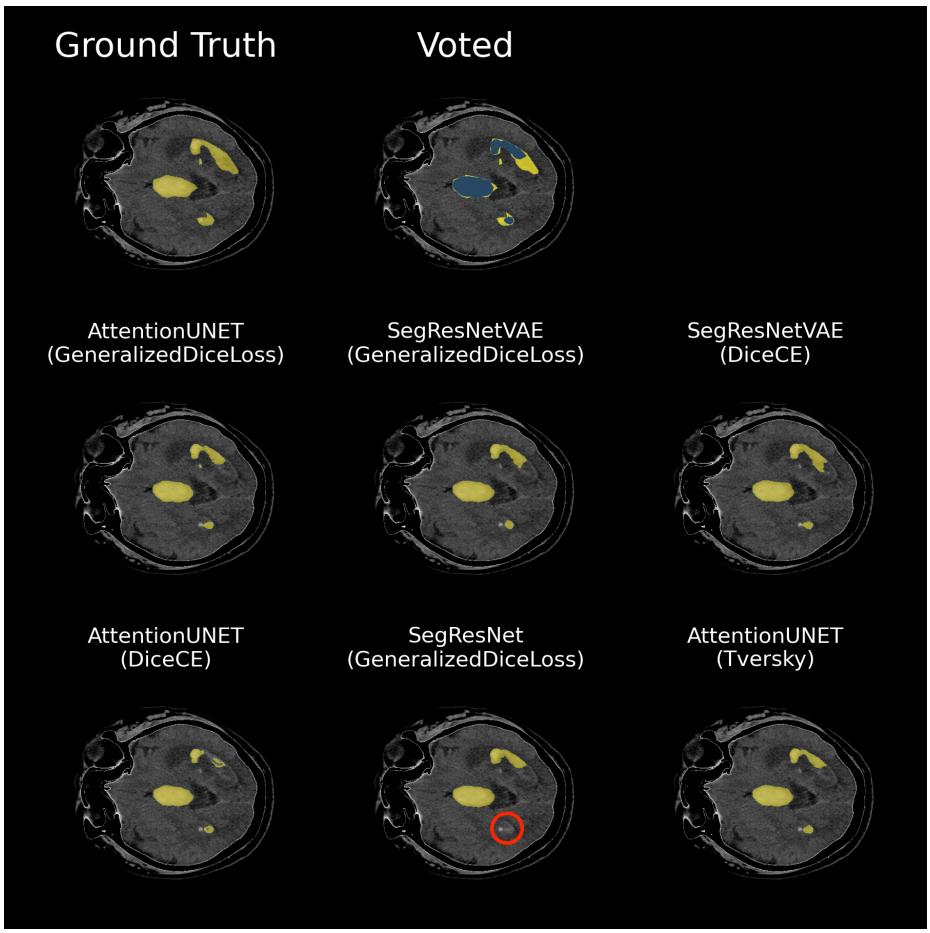

Fig. 3. Example of models predictions. All individual predictions (two last rows) are considered to create the final voted result (center top, blue mask).

address a thorough investigation of model architectures and loss functions in order to accurately select the best elements for the ensemble approach, while maintaining consistent generalization performances.

5 Appendix

We provide here the mathematical expressions of the various loss functions used in this study:

Dice As provided in [21], considering P as positive predictions and G as ground truth, Dice parameter is defined as:

$$D(P, G) = \frac{2|PG|}{|P| + |G|} \tag{1}$$

Cross Entropy

$$\mathrm{CE}(p, y) = \begin{cases} -log(p) & \text{if } y = 1 \\ -log(1-p) & \text{otherwise} \end{cases} \tag{2}$$

with y the ground truth value and p the probability provided by the model.

Focal Loss As provided in [13], focal loss is defined as:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} log(p_t) \tag{3}$$

with $\gamma \geq 0$ a parameter which balances contribution of easy *vs.* hard cases, α controls the importance of positive or negative examples, and

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases} \tag{4}$$

Tversky loss As provided in [21], considering P as positive preditions and G as ground truth, Tversky loss is defined as:

$$S(P, G, \alpha, \beta) = \frac{|PG|}{|PG| + \alpha |P \backslash G| + \beta |P \backslash G|} \tag{5}$$

References

1. Project monai. `https://docs.monai.io/en/stable/index.html`, accessed: 2022-08-01
2. Bisson, D.A., Flaherty, M.L., Shatil, A.S., Gladstone, D., Dowlatshahi, D., Carrozzella, J., Zhang, L., Hill, M.D., Demchuck, A., Aviv, R.I., et al.: Original and modified graeb score correlation with intraventricular hemorrhage and clinical outcome prediction in hyperacute intracranial hemorrhage. Stroke **51**(6), 1696–1702 (2020)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. arXiv preprint arXiv:2105.05537 (2021)
4. Hu, T.T., Yan, L., Yan, P.F., Wang, X., Yue, G.F.: Assessment of the abc/2 method of epidural hematoma volume measurement as compared to computer-assisted planimetric analysis. Biological Research for Nursing **18**(1), 5–11 (2016)
5. Huttner, H.B., Steiner, T., Hartmann, M., Köhrmann, M., Juettler, E., Mueller, S., Wikner, J., Meyding-Lamade, U., Schramm, P., Schwab, S., et al.: Comparison of abc/2 estimation technique to computer-assisted planimetric analysis in warfarin-related intracerebral parenchymal hemorrhage. Stroke **37**(2), 404–408 (2006)
6. Isensee, F., Petersen, J., Kohl, S.A., Jäger, P.F., Maier-Hein, K.H.: nnu-net: Breaking the spell on successful medical image segmentation. arXiv preprint arXiv:1904.08128 **1**(1-8), 2 (2019)
7. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: International MICCAI brainlesion workshop. pp. 450–462. Springer (2017)
8. Ko, S.B., Choi, H.A., Carpenter, A.M., Helbok, R., Schmidt, J.M., Badjatia, N., Claassen, J., Connolly, E.S., Mayer, S.A., Lee, K.: Quantitative analysis of hemorrhage volume for predicting delayed cerebral ischemia after subarachnoid hemorrhage. Stroke **42**(3), 669–674 (2011)
9. Kramer, A.H., Hehir, M., Nathan, B., Gress, D., Dumont, A.S., Kassell, N.F., Bleck, T.P.: A comparison of 3 radiographic scales for the prediction of delayed ischemia and prognosis following subarachnoid hemorrhage. Journal of neurosurgery **109**(2), 199–207 (2008)
10. Li, H., Xie, Y., Wang, X., Chen, F., Sun, J., Jiang, X.: Radiomics features on non-contrast computed tomography predict early enlargement of spontaneous intracerebral hemorrhage. Clinical neurology and neurosurgery **185**, 105491 (2019)
11. Li, X., Luo, G., Wang, W., Wang, K., Gao, Y., Li, S.: Hematoma expansion context guided intracranial hemorrhage segmentation and uncertainty estimation. IEEE Journal of Biomedical and Health Informatics **26**(3), 1140–1151 (2022). https://doi.org/10.1109/JBHI.2021.3103850
12. Li, Z., You, M., Long, C., Bi, R., Xu, H., He, Q., Hu, B.: Hematoma expansion in intracerebral hemorrhage: an update on prediction and treatment. Frontiers in Neurology **11**, 702 (2020)
13. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
14. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)

15. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. IEEE (2016)
16. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization (Nov 2018). https://doi.org/10.48550/arXiv.1810.11654, http://arxiv.org/abs/1810.11654, arXiv:1810.11654 [cs, q-bio]
17. Nehme, A., Ducroux, C., Panzini, M.A., Bard, C., Bereznyakova, O., Boisseau, W., Deschaintre, Y., Diestro, J.D.B., Guilbert, F., Jacquin, G., et al.: Non-contrast ct markers of intracerebral hematoma expansion: a reliability study. European Radiology pp. 1–10 (2022)
18. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas (May 2018). https://doi.org/10.48550/arXiv.1804.03999, http://arxiv.org/abs/1804.03999, arXiv:1804.03999 [cs]
19. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
21. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks (Jun 2017). https://doi.org/10.48550/arXiv.1706.05721, http://arxiv.org/abs/1706.05721, arXiv:1706.05721 [cs]
22. Salman, R.A.S., Frantzias, J., Lee, R.J., Lyden, P.D., Battey, T.W., Ayres, A.M., Goldstein, J.N., Mayer, S.A., Steiner, T., Wang, X., et al.: Absolute risk and predictors of the growth of acute spontaneous intracerebral haemorrhage: a systematic review and meta-analysis of individual patient data. The Lancet Neurology **17**(10), 885–894 (2018)
23. Shen, Q., Shan, Y., Hu, Z., Chen, W., Yang, B., Han, J., Huang, Y., Xu, W., Feng, Z.: Quantitative parameters of ct texture analysis as potential markers for early prediction of spontaneous intracranial hemorrhage enlargement. European radiology **28**(10), 4389–4396 (2018)
24. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 240–248. Springer (2017)
25. Teng, L., Ren, Q., Zhang, P., Wu, Z., Guo, W., Ren, T.: Artificial intelligence can effectively predict early hematoma expansion of intracerebral hemorrhage analyzing noncontrast computed tomography image. Frontiers in Aging Neuroscience p. 212 (2021)
26. Wang, C.W., Juan, C.J., Liu, Y.J., Hsu, H.H., Liu, H.S., Chen, C.Y., Hsueh, C.J., Lo, C.P., Kao, H.W., Huang, G.S.: Volume-dependent overestimation of spontaneous intracerebral hematoma volume by the abc/2 formula. Acta Radiologica **50**(3), 306–311 (2009)
27. Webb, A.J., Ullman, N.L., Morgan, T.C., Muschelli, J., Kornbluth, J., Awad, I.A., Mayo, S., Rosenblum, M., Ziai, W., Zuccarrello, M., Aldrich, F., John, S., Harnof, S., Lopez, G., Broaddus, W.C., Wijman, C., Vespa, P., Bullock, R., Haines,

S.J., Cruz-Flores, S., Tuhrim, S., Hill, M.D., Narayan, R., Hanley, D.F., Tayal, A.H., Diaz, A.T., Hoffer, A., Moheet, A., Agarwal, S., Markowski, M., Rosenthal, G., Kureshi, I., Varelas, P., Pinto, P.T.C., Marti-Fabregas, J., Camarata, P., Seder, D.B., Freeman, W., Helms, A., Lazaridis, C., Shahi, K., Sinclair, D., Gizzi, M., LeDoux, D., Torbey, M., Cockroft, K., Antezana, D., Latorre, J.G., Margalit, N., Crabtree, H.M., Lovick, D., Venkatasubramanian, C., Weaver, M., Jallo, J., Reinert, M., Butcher, K., Leonardo, J., Ardelt, A., Adeoye, O., Csiba, L., Wartenberg, K., Bosel, J., Testai, F., Adams, H., Hobohm, C., Kerz, T., Wechsler, L., Gardner, P., Decker, D., Bulters, D., Gonzales, N., Caron, J.L., Hall, C., Ansari, S., Luft, A., Arikan, F., Graham, R.S., Tucker, K.: Accuracy of the abc/2 score for intracerebral hemorrhage. Stroke **46**(9), 2470–2476 (2015). https://doi.org/10.1161/STROKEAHA.114.007343

28. Won, S.Y., Zagorcic, A., Dubinski, D., Quick-Weller, J., Herrmann, E., Seifert, V., Konczalla, J.: Excellent accuracy of abc/2 volume formula compared to computer-assisted volumetric analysis of subdural hematomas. PLoS One **13**(6), e0199809 (2018)
29. Xie, H., Ma, S., Wang, X., Zhang, X.: Noncontrast computer tomography–based radiomics model for predicting intracerebral hemorrhage expansion: preliminary findings and comparison with conventional radiological model. European radiology **30**(1), 87–98 (2020)
30. Yaniv, Z., Lowekamp, B.C., Johnson, H.J., Beare, R.: Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research. Journal of digital imaging **31**(3), 290–303 (2018)
31. Zhong, J.w., Jin, Y.j., Song, Z.j., Lin, B., Lu, X.h., Chen, F., Tong, L.s.: Deep learning for automatically predicting early haematoma expansion in chinese patients. Stroke and vascular neurology **6**(4) (2021)
32. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)