Intracerebral Hemorrhage Segmentation on Noncontrast Computed Tomography Using a Masked Loss Function U-Net Approach

Nadine A. Coorens, MSc,† Kevin Groot Lipman, MSc,*† Sanjith P. Krishnam, MBBS,* Can Ozan Tan, PhD,*‡ Lejla Alic, PhD,† and Rajiv Gupta, MD, PhD**

Objective: Intracerebral hemorrhage (ICH) volume is a strong predictor of outcome in patients presenting with acute hemorrhagic stroke. It is necessary to segment the hematoma for ICH volume estimation and for computerized extraction of features, such as spot sign, texture parameters, or extravasated iodine content at dual-energy computed tomography. Manual and semiautomatic segmentation methods to delineate the hematoma are tedious, user dependent, and require trained personnel. This article presents a convolutional neural network to automatically delineate ICH from noncontrast computed tomography scans of the head.

Methods: A model combining a U-Net architecture with a masked loss function was trained on standard noncontrast computed tomography images that were down sampled to 256 × 256 size. Data augmentation was applied to prevent overfitting, and the loss score was calculated using the soft Dice loss function. The Dice coefficient and the Hausdorff distance were computed to quantitatively evaluate the segmentation performance of the model, together with the sensitivity and specificity to determine the ICH detection accuracy.

Results: The results demonstrate a median Dice coefficient of 75.9% and Hausdorff distance of 2.65 pixels in segmentation performance, with a detection sensitivity of 77.0% and specificity of 96.2%.

Conclusions: The proposed masked loss U-Net is accurate in the automatic segmentation of ICH. Future research should focus on increasing the detection sensitivity of the model and comparing its performance with other model architectures.

Key Words: hemorrhagic stroke, convolutional neural model, dice coefficient, brain extraction mask

(*J Comput Assist Tomogr* 2023;47: 93–101)

Intracerebral hemorrhage (ICH) is a second most common subtype of stroke that accounts for 10% to 20% of all strokes[1,2] and has a 1-month mortality rate of approximately 40%.[3,4] Furthermore, a long-term functional disability in most patients surviving the acute stage causes a substantial burden on healthcare systems worldwide.[5] Known prognostic factors associated with a poor outcome include initial ICH volume at presentation, subsequent hematoma expansion, intraventricular hemorrhage, and contrast extravasation (which may be detected as a spot sign) in computed tomography (CT) images.[4–6] The initial volume is the strongest predictive factor for mortality and poor outcome.[4,7] Therefore, it is commonly used in clinical decision making. However, hemorrhage volume is typically estimated by manual or semiautomatic delineation of the hemorrhage on noncontrast CT scans. This makes it time consuming and prone to interrater and intrarater variability. Another common method for volume estimation of ICH is to assume an elliptical shape and approximate the volume using the dimensions of largest cross-sectional dimensions in 3 directions, referred to as the ABC/2 method.[8] While it is acceptable for small and nicely shaped lesions, this approximation is inaccurate for larger and complex shaped hemorrhages.[9] Because current standard-of-care methods are limited by unmanageable efforts and inaccuracies, there is a need for an automated, accurate, and fast segmentation method estimating ICH volume in CT images.

Recently, deep networks based on the convolutional neural networks have emerged as a powerful, automatic tool in solving a variety of image segmentation and interpretation tasks.[10–12] Furthermore, some of these models, notably U-Net architecture, are very fast and perform well even with small training data sets.[13] Previous work in automated hemorrhagic stroke workflow has mainly consisted of detection or classification of ICH,[14–17] while only few attempts have been made to automatically segment ICH based on U-Net architecture.[18–20] The work of Abramova et al[19] (2021) introduced squeeze-and-excitation blocks to the network, whereas Arab et al[20] (2020) combined a modified U-Net with deep supervision. Both approaches achieved promising results (Dice coefficients of 0.86 and 0.84, respectively) but were limited by the availability of clinical data (76 and 55 cases, respectively). Larger data sets may possibly reflect a broader range of irregular hematoma shapes and sizes. In addition, segmentation performance may be increased through the implementation of a masked loss function.[21] Given the legitimate necessity in generating accurate ICH segmentations, the goal of this article is to develop a fully automatic deep learning framework to segment ICH from noncontrast CT images.

MATERIALS AND METHODS

Study Subjects and Data Acquisition

Patients presenting with an acute ICH to the emergency department at the Massachusetts General Hospital (Boston, Mass) from January 2015 to September 2018 were retrospectively included in this study (IRB approval 2015P000607). Patients were considered eligible based on the availability of virtual noncontrast images from a dual-energy CT scan with 1-mm slice thickness. Cases with external hardware such as intracranial pressure monitors and drainage catheters were excluded from the study. Based on these eligibility criteria, a total of 116 subjects were accepted for the study. Our processing algorithm uses skull stripping from CT scans as an intermediate step. Therefore, the cases where skull stripping or the semiautomatic segmentation method failed manual quality check on boundaries were also excluded, resulting in a total number of 107 included subjects.

From the *Department of Radiology, Massachusetts General Hospital, Boston, MA; and †Magnetic Detection and Imaging Group, Technical Medical Centre, and ‡Robotics and Mechatronics, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, the Netherlands.
Received for publication April 17, 2022; accepted June 21, 2022.
Correspondence to: Rajiv Gupta, MD, PhD, Mass General Imaging: Neuroradiology, 55 Fruit St, Boston, MA 02114 (e-mail: RGUPTA1@mgh.harvard.edu).
The authors declare no conflict of interest.
Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jcat.org).

DOI: 10.1097/RCT.0000000000001380

Virtual noncontrast images were acquired from dual-energy CT scans (SIEMENS SOMATOM Force) with slice thickness 1 mm and exposure time 1000 ms. A standard 120 kilovoltage peak (kVp) data set was generated by blending low-energy (80 kVp) and high-energy (140 kVp) images (software version Syngo CT VA50A).

Data Preprocessing

Three-dimensional (3D) CT images were preprocessed slice-wise (summarized in Fig. 1) to produce 3 sets of 2D images: standardized CT images and label images representing ICH and binary brain masks. The standardized CT images were used as input to the U-Net and the ICH label images were used to validate and test segmentations produced by the model. Both the standardized CT images and ICH labels were masked using the brain mask, thereby setting all values outside the brain to zero. In addition, the brain masks were multiplied with the output of the model to configure a masked loss function.

Data Scaling

Neural networks tend to achieve a lower performance using data with large differences in scales, because the estimated weights of the model will be updated at different rates rather than in a similar manner. Hence, to achieve standard Gaussian distribution of the data, content of the brain mask in CT images were standardized to achieve zero-mean-at-unit-variance by subtracting the mean and dividing the result by the standard deviation of the original image:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

where z denotes the standardized gray value for each pixel, x is the original CT gray value, μ is the mean pixel value of the original CT image, and σ the standard deviation of the original CT data.

Skull Stripping

The brain mask was generated by skull stripping (FSL Toolbox, FMRIB Software Library, Release 6.0(c) 2018, The University of Oxford) of the nonstandardized CT scans. This method is originally intended for magnetic resonance imaging data as a brain extraction tool (BET),[22] which is a deformable surface method. Nevertheless, applying 2 threshold steps and a 3D Gaussian filter before BET allows FSL Toolbox to be used for skull stripping of CT images,[23] in which 2 threshold parameters can be adjusted: the fractional intensity threshold and the vertical gradient threshold. The former threshold is in the range of 0 to 1, where a smaller value leads to a larger brain outline estimate. The latter threshold is in the range of −1 to 1, where positive values provide a larger brain outline at the bottom and a smaller outline at the top axial slices. Here, the fractional intensity threshold and the vertical gradient threshold were set to a value of 0.01 and 0, respectively. The final step in constructing the brain mask was a 2D erosion on each binary image slice with a square kernel of 5 × 5 pixels.

Intracerebral Hemorrhage Annotation

Intracerebral hemorrhage labels were generated through semiautomatic segmentation of ICH in the virtual noncontrast CT data. First, the CT scans were thresholded to include pixels between 75 and 100 HU, followed by an erosion and dilation procedure to reduce noise. Consecutively, region growing was performed with the ICH center of mass as a seed point and holes inside the ICH segmentations were filled. The threshold values and seed point location were adjusted manually for each CT scan. The resulting semiautomatic ICH segmentations were complemented by a manual revision to optimize the results. Finally, the ICH label images were multiplied with the corresponding brain mask images to ensure that all annotations are within the brain area.

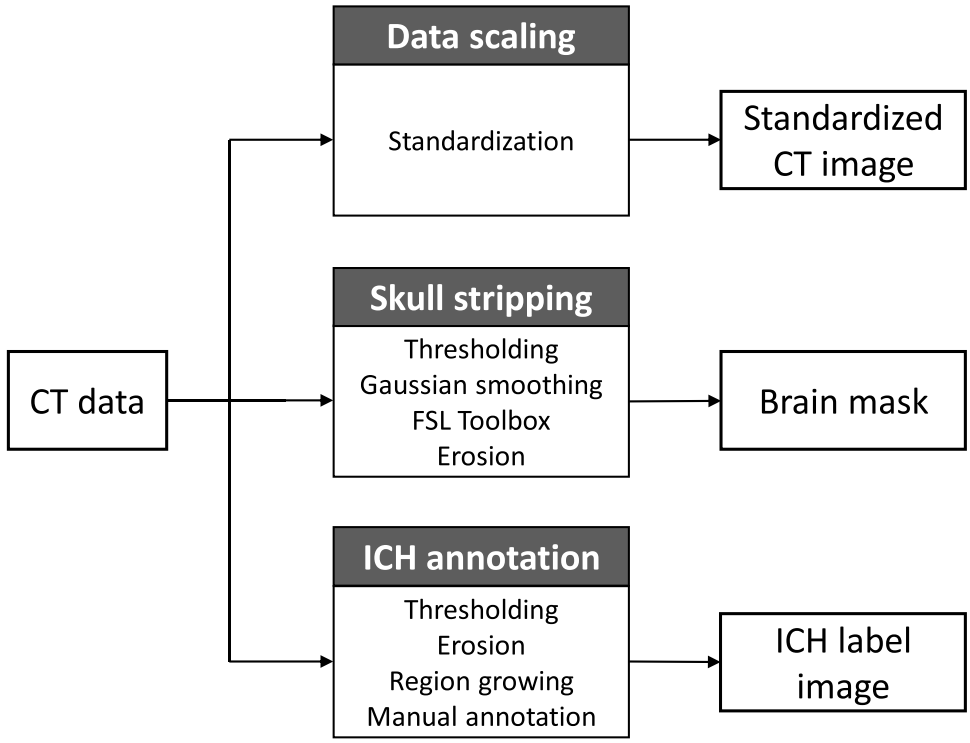

FIGURE 1. The data preprocessing steps that where undertaken to acquire each of the 3 input images. From the original CT data, a standardized CT image, brain mask, and ICH label image are acquired.

Image Centering

To finalize the preprocessing of the data, all 3 image types were centered to the image space by padding the bounding box of the brain area. Then, each slice in the 3D image volumes was resampled to a size of 256 × 256 pixels to decrease training time. A bilinear interpolation method was used to resize the standardized CT images, whereas a nearest-neighbor interpolation was used for downscaling the ICH annotation and brain mask images. All slices outside the brain area (slices with corresponding mask image that had only zero values) were removed.

Design of the Masked Loss U-Net

Model Architecture

The model (illustrated in Fig. 2) uses the U-Net architecture developed by Ronneberger et al[13] (2015). It consists of a contracting path and an expansive path with a bottleneck in between. To improve the generalization of the model, dropout (with rate = 0.5) was performed before and after the bottleneck ensuring 50% probability to output a 0-valued activation of the corresponding neurons. A sigmoid activation function was used in the final layer to ensure that the predicted output values are in the range of 0 to 1.

Masked Loss Function

A feedback signal, referred to as a loss score, to adjust the model parameters (eg, weights and biases) is provided by the loss function based on the Dice coefficient, specifically designed for image segmentation tasks. Because of its overlap-based characteristics, the Dice coefficient is particularly useful in evaluating class imbalanced data.[24–26] The Dice coefficient ranges from 0 to 1, where a score of 1 denotes a perfect overlap.[27]

To define a loss function that can be minimized, the soft Dice loss (1 – Dice) was used. In addition, a smooth factor $s = 1e{-}7$ was added to both the nominator and denominator to avoid mathematical inconsistencies, but still allowing the perception of small gradients in the loss score. This leads to the following loss function:

$$\text{Loss function} = 1 - \frac{2|A \cap B| + s}{|A| + |B| + s} \quad (2)$$

A possible way to increase segmentation performance is through the implementation of a masked loss function. Previously, a masked loss function has been used to bypass missing data in the input image during the calculation of the loss score.[21] Similarly, a masked loss function may be used for ICH segmentation to restrict the calculation of the loss score to the area within the brain and inhibit false predictions outside the brain area. In addition, the masked output potentially updates the model parameters with greater accuracy. The masked loss function in this study was achieved by multiplying the brain mask image with the predicted output of the U-Net, causing the predictions outside of the brain to be zero. The implementation of the masked loss function in combination with the U-Net is visualized in Figure 2 and illustrated in more detail in Figure 3.

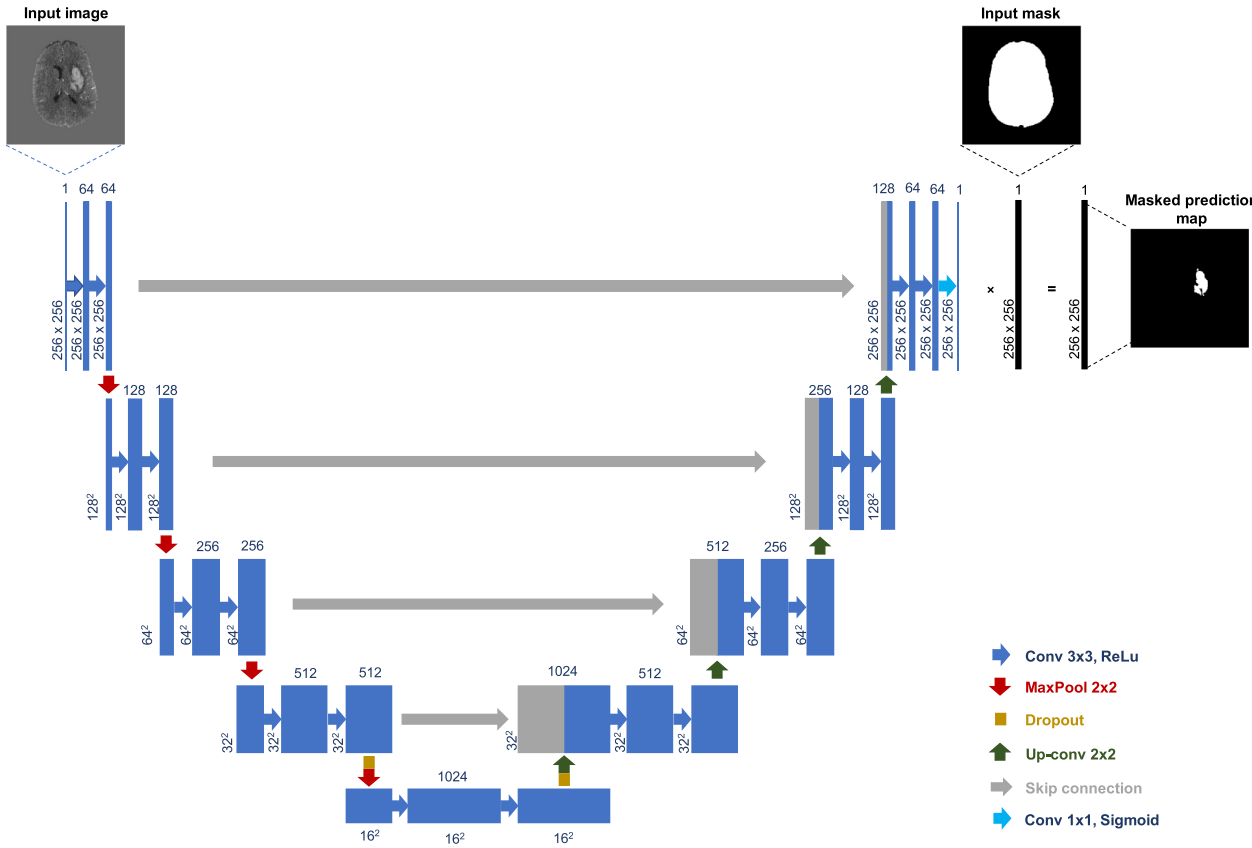

FIGURE 2. Architecture of the proposed model to automatically delineate ICH on noncontrast CT images. A U-Net[13] is combined with a masked loss function. A standardized and masked CT image of size 256 × 256 serves as the input image of the model and is processed by the convolutional blocks in the contracting and expanding paths of the U-Net. After the final convolutional layer, the predicted output is multiplied by the corresponding brain mask image (input mask) before the calculation of the loss score. Figure 2 can be viewed online in color at www.jcat.org.

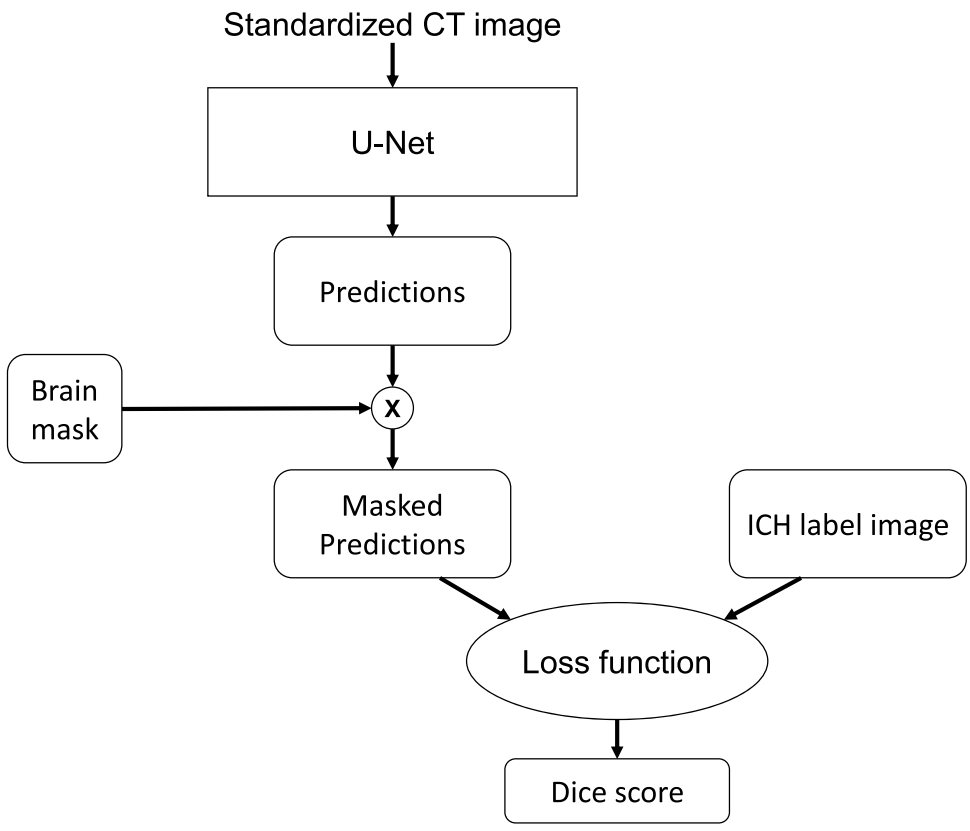

FIGURE 3. Illustration of how the output prediction map is multiplied by the mask image, before the calculation of the loss.

Training

A total of 107 CT brain scans were randomly split into training, validation, and test sets with approximately 60%, 20%, and 20% of cases, respectively. To prevent overfitting of the model, data augmentation was used to generate more training samples from existing training data. Random transformations were applied considering that invariance of head CT images to translation and rotation as well as robustness to gray value variations are of primary importance in ICH segmentation. An overview of the data augmentation techniques is shown in Table 1. In case of the rotation of the CT image, a bilinear interpolation was performed, while a nearest neighbor interpolation was used for the corresponding label and mask images to preserve their binary property.

To prevent disproportionate activations in some layers of the model with consequent lack of contribution from remaining layers, the weights in this study were initialized using normalized initialization of the weights.[28] This type of initialization draws random samples from a uniform distribution with limits based on the number of input and output units in the concerning layer.

TABLE 1. Overview of the Data Augmentation Techniques

Augmentation	Variable	Range
Horizontal flip	Probability p	Yes or no, $P = 0.50$
Rotation	Angle α	$\alpha = -20$ to $20°$
Horizontal translation	Distance x	$x = -10$ to 10 pixels
Vertical translation	Distance y	$y = -10$ to 10 pixels
Gray value change	Multiplier m	$m = 0.9$ to 1.1

The acquisition of the ICH probability map using the U-Net was implemented with Keras in Python. The model was trained for 600 epochs, where each epoch included 1261 iterations with a mini batch size of 10. The Adam training algorithm was used as an optimizer, with a learning rate of 1e−5. Adam optimization is appropriate for sparse gradients on noisy problems.[29] The time taken to train the model was approximately 70 hours on a single Nvidia GTX 1080 Ti GPU (Nvidia Corporation, Santa Clara, Calif).

Evaluation of the Test Results

The performance of the proposed masked loss U-Net to segment ICH on noncontrast CT images was evaluated by comparing the results to the ICH labels using Dice score and Hausdorff distance (HD). The Dice coefficient is an overlap-based measure, whereas the HD is a distance-based metric, which takes the spatial position of pixels into consideration and quantifies the mismatch between 2 sets. Distance-based metrics are recommended when the overall accuracy of the segmentation, such as the boundary contour, is of importance.[25] In this study, all distances were calculated in pixel length, which means that the pixel sizes were not considered. In addition, the detection of ICH on each CT slice was assessed through a confusion matrix and the Pearson correlation coefficient (r) was computed to evaluate the correlation between automatic and annotated ICH volume.

RESULTS

A total of 107 subjects were included (62 male, 72 female) with an average age of 66 ± 17 (mean ± SD) years. The result of the random data partitioning into the training, validation, and test

TABLE 2. Characteristics of the Training, Validation and Test Sets

	Training	Validation	Test
Subjects (male, female)	64 (40 M, 24 F)	21 (15 M, 6 F)	22 (10 M, 12 F)
No.	12614	3734	4095
Slices with ICH	4708	1404	1094
Slices without ICH	7906	2330	3001
Mean ICH size, pixels	13,531 ± 6809	15,178 ± 6180	13,819 ± 6699

set is presented in Table 2, showing the number of slices with and without ICH and the mean ICH size (in pixels) per set.

Because the number of slices with ICH versus the number of slices without ICH is highly imbalanced, the performance measures were computed for each class separately. The Dice coefficients and HDs are presented in Table 3. The mean Dice score calculated on the total test set is 86.0% ± 30.5% and the mean HD is 0.88 ± 1.56. The histograms of the Dice coefficients and the HDs of the slices with ICH are shown in Figures 4 and 5, respectively. In Supplemental Digital Content Figure 1, http://links.lww.com/RCT/A145, the evaluation loss during training is presented. Figure 6 shows examples of segmentation results with a poor prediction (Dice = 75.3%, HD = 8.31), a medium prediction (Dice = 76.3%, HD = 2.65), and a good prediction (Dice = 97.6%, HD = 2.00).

The sensitivity and specificity, which evaluate the ICH detection accuracy of the model, are 77.0% and 96.2%, respectively. A complete confusion matrix of the performance is presented in Table 4. The correlation between automatic and annotated ICH volume was assessed by a scatterplot (Fig. 7). Pearson correlation coefficient (r) was found to be 0.97, suggesting a good significant linear correlation between these 2 methods. In addition, agreement between these 2 methods was assessed by the Bland-Altman plot (Supplemental Digital Content Figure 2, http://links.lww.com/RCT/A146). The bias, upper limit, and lower limit of agreement were found to be 0.89, 6.93, and −4.75 mL, respectively.

DISCUSSION

This study presents a novel deep learning method to segment ICH from noncontrast CT images using a masked loss U-Net. The masked loss function uses a binary brain mask to indicate the specific region of interest (ie, the brain) within the CT image. By multiplying the brain mask with the prediction map, all prediction values outside the area of interest are set to zero. In this way, any outlying predictions are not considered when updating the weights of the model via backpropagation. This offers the model a head start and more freedom in optimizing the set of weights.

The overall performance of the model can be described by a combination of the ICH segmentation accuracy and the ICH detection accuracy. Examples of the segmentation results are visualized in Figure 6. The segmentation performance of the model was evaluated by the Dice coefficients and HDs. Table 3 shows the mean and median values of these measures, calculated on the sets of CT images with and without ICH. As can be derived from this table, the mean Dice score is much higher in the group of images without ICH (96.2%) compared with the images with ICH (58.0%). The fact that the mean Dice score in images without ICH is not perfect indicates that the model incorrectly detects hemorrhage in some cases. The low mean Dice coefficient for the ICH group could be caused by either a low ICH detection accuracy of the model, an impaired segmentation performance or a combination of the two.

To further investigate this, the histograms of both the Dice coefficients and the HDs were plotted for the set of images with ICH. From Figure 4, it can be deduced that there is a significant number of images with a Dice score of 0%, indicating that the model does not detect ICH in these cases. Thus, the segmentation performance in the image set with ICH is biased by false-negative detections. Likewise, false-positive detections affect the performance on the images without ICH. If the false-negative images are disregarded, the histogram in Figure 4 would be left skewed with most of the images having a Dice score greater than 65%. Therefore, the median Dice coefficient of the segmentations in the set of images with ICH (75.9%) is more relevant than the mean Dice coefficient in evaluating the segmentation performance. Moreover, for the subgroup with Dice scores greater than 50%, the median Dice coefficient was 85.6%.

The right-skewed histogram of the HDs displayed in Figure 5 demonstrates that the model predicts segmentations of great quality in most situations. The majority (90%) of the predicted images have a HD less than 4 pixels, whereas there are only a few outliers with a larger HD. Figure 7 shows a correlation coefficient of 0.97 demonstrating efficient performance of this model on segmentation tasks.

To define the ICH detection accuracy of the model, the sensitivity and specificity were calculated. The sensitivity of 77.0% and specificity of 96.2% support the previous statements: the low mean Dice scores are most likely a consequence of a lower ICH detection rate of the model. Although the model is good in ruling out hemorrhage, the model does not recognize the ICH in a considerable number of images. Furthermore, the Bland-Altman plot in Supplemental Digital Content Figure 2, http://links.lww.com/RCT/A146, shows a bias of 0.89 mL, upper and lower limits of agreement of 6.93 and −4.75 mL, respectively. Additional training may potentially further narrow the limits of agreement.

Notably, qualitative assessment of the subset of false negative images (ie, ICH not detected) indicates that the majority (78%) are

TABLE 3. Mean and Median Dice Coefficients and Hausdorff Distances (in Pixels) for the Masked Loss U-Net on CT Images With and Without ICH

Data	Dice Coefficient		Hausdorff Distance	
Images	Mean ± SD	Median (Range)	Mean ± SD	Median (Range)
With ICH	58.0% ± 37.7%	75.9% (98.0%)	3.01 ± 1.45	2.65 (7.31)
Without ICH	96.2% ± 19.0%	100% (100%)	0.10 ± 0.56	0 (5.57)

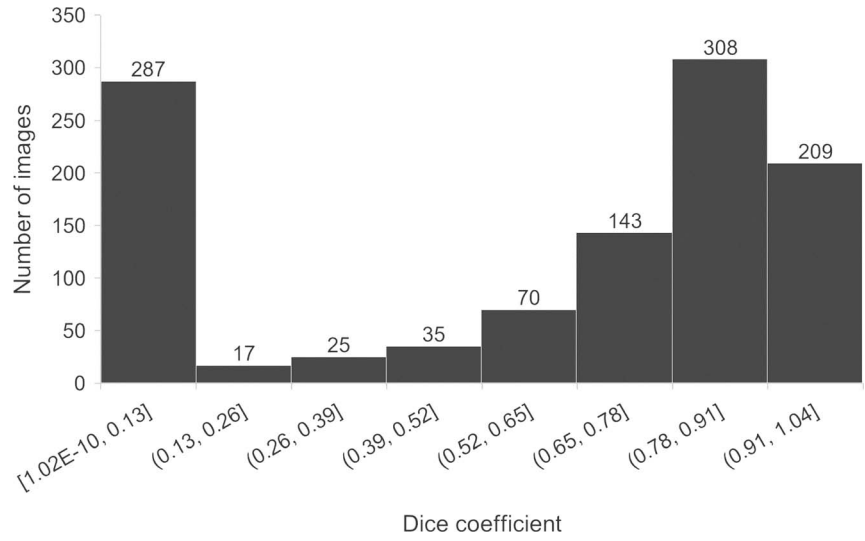

FIGURE 4. Histogram of the Dice coefficients for ICH segmentation using a masked loss U-Net. The Dice coefficients were calculated only on CT images containing ICH. Because of the bimodal distribution, the mean (58.0%) and median (75.9%) Dice coefficients are skewed. A total of 364 images had a Dice score less than 50%, but the median Dice coefficient is 85.6% in the subgroup with Dice scores greater than 50%.

boundary images of the hemorrhages. In these cases, the hemorrhage diminishes and transitions into the surrounding brain tissue. The model failed to detect hemorrhage in all images with ICH of 2 subjects, accounting for 15% of false-negative images. These hemorrhages were small compared with the hemorrhages of other subjects. Thus, the inaccurate detections are mainly a consequence of small (portions of) hemorrhages that transition into the surrounding tissue.

Although a fair number of CT images are misclassified as images without ICH, the combined results of the median Dice coefficient and HD and the sensitivity and specificity suggest that the masked loss U-Net performs well in delineating ICH when the model detects the hemorrhage on the image.

Limitations and Recommendations

The impaired performance of the model to detect ICH could be attributed to, among others, the imbalanced data set; there are twice as many images without ICH as there are images including ICH. Because the model converged at approximately 91% accuracy during training, a possible approach to address this problem is to improve the training phase of the model. This could be achieved by implementing adaptive learning rate methods to gradually reduce the learning rate during training. As the learning rate decreases, the model will be allowed to optimize the set of weights.

Another method to improve the detection accuracy is by stratifying the data set. In this study, the data set was randomly split into a training, validation, and test set. Thus, a uniform distribution of subject characteristics such as ICH size, ICH subtype, sex, and age was not ensured when generating the 3 sets, see Table 2. Data stratification before partitioning can add to the robustness of the model during training. Moreover, the random distribution of the data resulted in an imbalance in the ratio of images with and without ICH between the 3 sets. For the training and

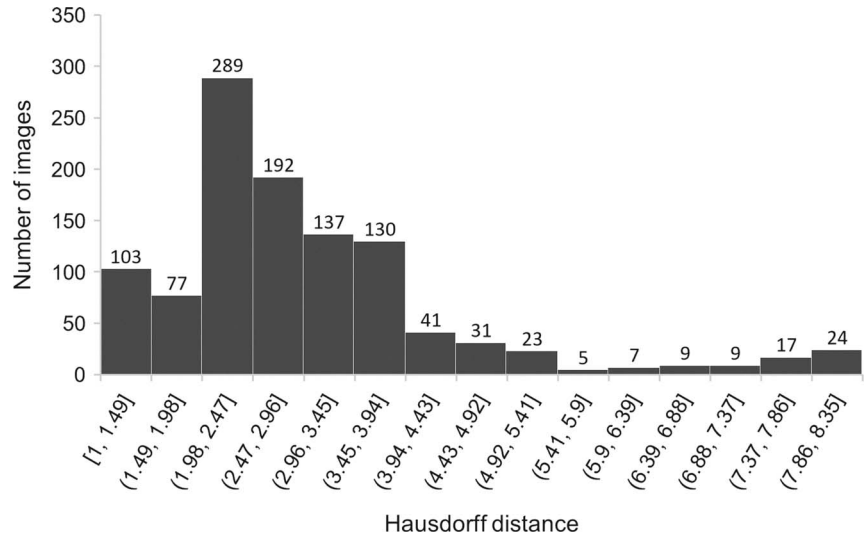

FIGURE 5. Histogram of the HDs (in pixels) for ICH segmentation using a masked loss U-Net. The HDs were calculated only on CT images containing ICH. Mean, 3.01; median, 2.65.

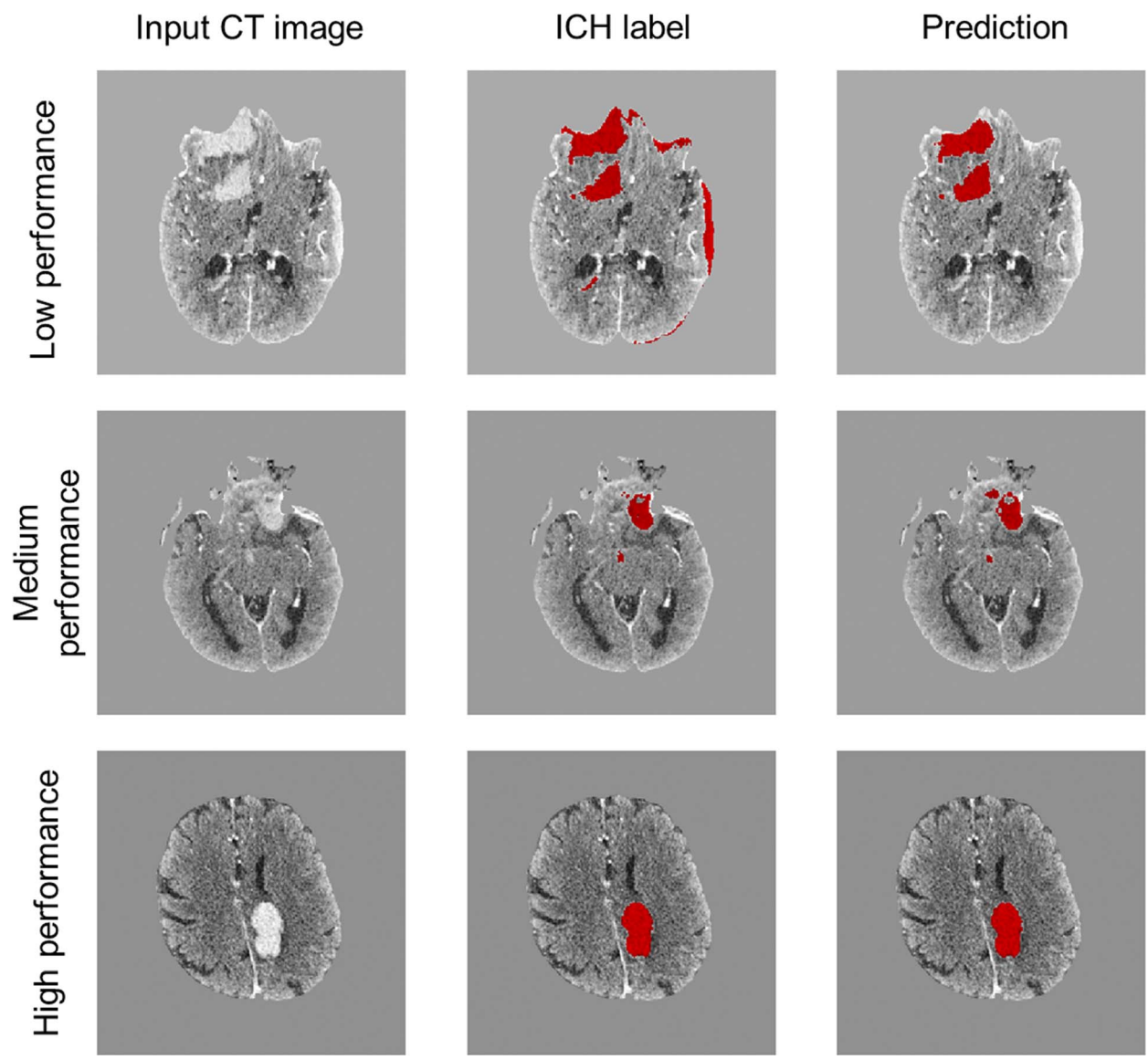

FIGURE 6. Intracerebral hemorrhage segmentation results of the masked loss U-Net. The left column depicts the CT images as presented to the U-Net. The corresponding ICH label images are shown in the middle column, and the predicted segmentation results are displayed on the right. The ICH labels and segmentation results are visualized in red. From top to bottom, a low performing prediction (Dice = 75.3%, HD = 8.31), a medium prediction (Dice = 76.3%, HD = 2.65), and a good prediction (Dice = 97.6%, HD = 2.00) are shown. Figure 6 can be viewed online in color at www.jcat.org.

TABLE 4. Confusion Matrix of Performance Measures

		Predicted		
		Positive	**Negative**	
Actual	Positive	True positive 842	False negative 252	Sensitivity 77.0%
	Negative	False positive 113	True negative 2888	Specificity 96.2%
		Precision 88.2%	Negative predictive value 92.0%	Accuracy 91.1%

validation set, approximately 37% of the images contain ICH, whereas in the test set, only 27% includes ICH. For a fairer evaluation, the ratio should be the same in every set.

In addition to these limitations of the training and evaluation phase, shortcomings in the data preprocessing method might contribute to an impaired performance of the model. First, the semiautomatic method to generate the ICH label images may have its flaws as the results were not verified by expert radiologists. When the predictions of the model are compared with poor quality labels, the model might be unable to pick up important features. In particular, the quality of the labels is less reliable at the boundaries of the hemorrhages, thereby contributing to the impaired detection accuracy in these areas. An inevitable step to improve the model performance is therefore to refine the label quality.

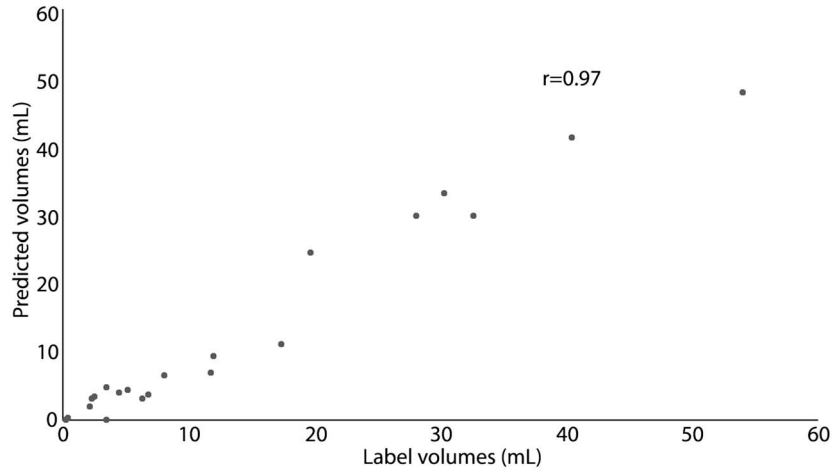

FIGURE 7. Scatter plot of the annotated (label) ICH volumes versus predicted ICH volumes.

Second, this model requires skull stripping of the CT scan to generate a brain mask. Although this step can be accomplished automatically, the method based on BET still introduces the possibility of failure. The failure rate to produce a brain mask was 3.4%. In addition, the quality of the brain mask differed at various locations in the head. A critical step in BET is the calculation of the center of gravity of the slices, which should be located within the brain area. However, in case the head scan includes much neck portion, the center of gravity of the volume will be outside of the brain, failing BET to extract the brain region in these slices. Thus, in some slices, the brain mask included tissue outside the brain. An additional erosion step was performed to remove any excess pixels located in the skull area. Because a fixed erosion kernel size did not suffice to the entire data set, high value pixels outlining the brain were still present in some scans. These distinct pixels might affect the training process of the model.

Furthermore, the model was trained on preprocessed CT images, which were down sampled from 512 × 512 to 256 × 256 pixels to speed up the training process. However, the rescaling of the images might have been accompanied by a certain amount of information loss, affecting the overall outcome of the model. In addition, the original image size is preferred to estimate the ICH volume more accurately as well as using the obtained segmentations for other research purposes. Therefore, next steps in future research should also focus on training the model on 512 × 512 sized images.

Finally, because most misclassified images are boundary images of the hemorrhage, detection accuracy in these areas may be improved by including spatial information from the axial direction in the model. However, this requires different architecture (eg, 3D U-Net).

CONCLUSIONS

This work presented a deep learning approach to automatically segment ICH on noncontrast CT images. The core strategy was to combine a masked loss function and a U-Net architecture. The median Dice coefficient of 75.9% and HD of 2.65 pixels showed that the model is accurate in delineating ICH. However, with a sensitivity of 77.0% and specificity of 96.2%, the model is still unable to detect ICH in a fair amount of CT images. Nevertheless, a masked loss function U-Net is a promising method to enable fast and accurate estimation of ICH volume in CT images, thereby supporting clinical management of ICH. Future research should focus on improving the quality of the labels, using 512 × 512 sized images, increasing the detection sensitivity of the model and comparing its performance to other model architectures.

REFERENCES

1. Feigin VL, Lawes CM, Bennett DA, et al. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol*. 2009;8:355–369.
2. Ikram MA, Wieberdink RG, Koudstaal PJ. International epidemiology of intracerebral hemorrhage. *Curr Atheroscler Rep*. 2012;14:300–306.
3. van Asch CJ, Luitse MJ, Rinkel GJ, et al. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. *Lancet Neurol*. 2010;9:167–176.
4. Brouwers HB, Battey TW, Musial HH, et al. Rate of contrast extravasation on computed tomographic angiography predicts hematoma expansion and mortality in primary intracerebral hemorrhage. *Stroke*. 2015;46: 2498–2503.
5. An SJ, Kim TJ, Yoon BW. Epidemiology, risk factors, and clinical features of intracerebral hemorrhage: an update. *J Stroke*. 2017;19:3–10.
6. Wada R, Aviv RI, Fox AJ, et al. CT angiography "spot sign" predicts hematoma expansion in acute intracerebral hemorrhage. *Stroke*. 2007;38: 1257–1262.
7. Broderick JP, Brott TG, Duldner JE, et al. Volume of intracerebral hemorrhage. A powerful and easy-to-use predictor of 30-day mortality. *Stroke*. 1993;24:987–993.
8. Kothari RU, Brott T, Broderick JP, et al. The abcs of measuring intracerebral hemorrhage volumes. *Stroke*. 1996;27:1304–1305.
9. Freeman WD, Barrett KM, Bestic JM, et al. Computer-assisted volumetric analysis compared with abc/2 method for assessing warfarin-related intracranial hemorrhage volumes. *Neurocrit Care*. 2008;9:307–312.
10. Ibtehaz N, Rahman MS. Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Netw*. 2020;121: 74–87.
11. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *CoRR*. 2015;abs/1409.1556.
12. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, June 27–30, 2016, Las Vegas. 2016;770–778.
13. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *MICCAI*. 2015;234–241.

14. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study. *Lancet*. 2018;392:2388–2396.
15. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digit Med*. 2018;1:9.
16. Ye H, Gao F, Yin Y, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur Radiol*. 2019;29:6191–6201.
17. Chang PD, Kuoy E, Grinband J, et al. Hybrid 3d/2d convolutional neural network for hemorrhage evaluation on head ct. *AJNR Am J Neuroradiol*. 2018;39:1609–1616.
18. Hssayeni MD, Croock MS, Salman AD, et al. Intracranial hemorrhage segmentation using a deep convolutional model. *Data*. 2020;5.
19. Abramova V, Clèrigues A, Quiles A, et al. Hemorrhagic stroke lesion segmentation using a 3d u-net with squeeze-and-excitation blocks. *Comput Med Imaging Graph*. 2021;90:101908.
20. Arab A, Chinda B, Medvedev G, et al. A fast and fully-automated deep-learning approach for accurate hemorrhage segmentation and volume quantification in non-contrast whole-head ct. *Sci Rep*. 2020;10:19389.
21. Xu J, Wu S, Zhu S, et al. Masked loss residual convolutional neural network for facial keypoint detection. *MOBIMEDIA*. 2017;234–239.
22. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp*. 2002; 17:143–155.
23. Muschelli J, Ullman NL, Mould WA, et al. Validated automatic brain extraction of head ct images. *Neuroimage*. 2015;114:379–385.
24. Milletari F, Navab N, Ahmadi S. V-net: fully convolutional neural networks for volumetric medical image segmentation. *3DV*. 2016;565–571.
25. Taha AA, Hanbury A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015; 15:29.
26. Islam M, Sanghani P, See AAQ, et al. Ichnet: intracerebral hemorrhage (ich) segmentation using deep learning. *BrainLes*. 2019;456–463.
27. Eelbode T, Bertels J, Berman M, et al. Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index. *IEEE Trans Med Imaging*. 2020;39:3679–3690.
28. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *AISTATS*. 2010;9:249–256.
29. Kingma DP, Ba J. *Adam: a method for stochastic optimization. CoRR*. 2015;abs/1412.6980.