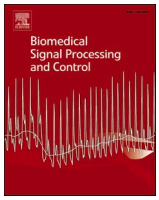
Biomedical Signal Processing and Control 102 (2025) 107386

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Uncertainty-Guided Semi-Supervised (UGSS) mean teacher framework for brain hemorrhage segmentation and volume quantification

Solayman Hossain Emon [a], Tzu-Liang (Bill) Tseng [b,*], Michael Pokojovy [c], Scott Moen [d], Peter McCaffrey [d], Eric Walser [d], Alexander Vo [d], Md Fashiar Rahman [b,*]

[a] Computational Science Program, The University of Texas at El Paso, TX 79968, USA
[b] Department of Industrial, Manufacturing and Systems Engineering, The University of Texas at El Paso, TX 79968, USA
[c] Department of Mathematics and Statistics, Old Dominion University, VA 23529, USA
[d] University of Texas Medical Branch, Galveston, TX 77550, USA

ARTICLE INFO

Keywords:
Deep learning
Brain hemorrhage
Semi-supervised learning
Uncertainty quantification

ABSTRACT

Traumatic brain injury (TBI) is considered a critical neurological emergency with substantial morbidity and mortality rates across the world. Among significant neuropathological consequences of brain injuries, intracranial hemorrhage (ICH) stands out as a particularly urgent condition necessitating prompt diagnosis to avert life-threatening complications. However, the traditional manual approach to detecting and segmenting brain hemorrhages in CT scans is time-consuming and labor-intensive. This study proposes a fully automated uncertainty-guided framework for intracranial hemorrhage segmentation in brain CT scans. The framework is trained on a semi-supervised scheme that leverages both labeled and unlabeled data. Notably, when trained on 80% of labeled data, the semi-supervised framework yields an average Dice coefficient of 0.613 ± 0.01 and a Jaccard index of 0.441 ± 0.02. These metrics significantly exceed their supervised counterparts, which demonstrates the efficacy of the proposed methodology. Moreover, the proposed approach exhibits an overall accuracy of 89.03% in brain hemorrhage classification with a Cohen's Kappa value of 0.835, which indicates substantial agreement between the model's predictions and the ground truth labels. In addition to its capabilities in intracranial hemorrhage detection and localization, the proposed framework offers a robust estimation of hemorrhage volume and provides a comprehensive 3D volumetric view. The accuracy and reliability of the volume quantification approach are justified through a comprehensive qualitative and quantitative assessment, utilizing visualization techniques and a goodness-of-fit test ($R^2 = 0.837$). In both instances, the method shows a notable alignment between the predicted hemorrhage volume and the actual hemorrhage volume. Thus, the proposed schemes of uncertainty-guided semi-supervised (UGSS) hemorrhage segmentation and volume quantification enhance model's applicability in clinical practice and research.

1. Introduction

The occurrence of intracranial hemorrhages (ICH) due to traumatic brain injuries (TBI) is one of the most critical medical emergencies in the world. Intracranial hemorrhage (ICH) is characterized by the accumulation of blood within the cranial cavity caused by an accident or any other trauma. As blood accumulates, it increases pressure within the cranial cavity, which can compress vital brain structures. This pressure can lead to brain herniation, where brain tissue is forced through openings within the skull, causing severe neurological deficits or even death. In such emergencies, timely diagnosis of brain hemorrhages becomes paramount. Healthcare professionals must act immediately, determining whether immediate surgical intervention is needed. To decide on the appropriate medical and surgical intervention, it is essential to detect and localize intracranial hemorrhage (ICH).

Traditionally, intracranial hemorrhage (ICH) can be diagnosed via inspecting a patient's imaging reports by a medical specialist. Several imaging modalities are used to diagnose intracranial hemorrhage (ICH) and assess its extent in the brain. Computed tomography (CT) [1] is an established non-invasive imaging standard for analyzing intracranial hemorrhage (ICH). When examining a brain CT scan for hemorrhages, radiologists look for areas of abnormal density that signify the potential

* Corresponding authors.
E-mail addresses: btseng@utep.edu (T.-L.(B. Tseng), mrahman13@utep.edu (M.F. Rahman).

https://doi.org/10.1016/j.bspc.2024.107386
Received 24 May 2024; Received in revised form 12 November 2024; Accepted 16 December 2024
Available online 25 December 2024

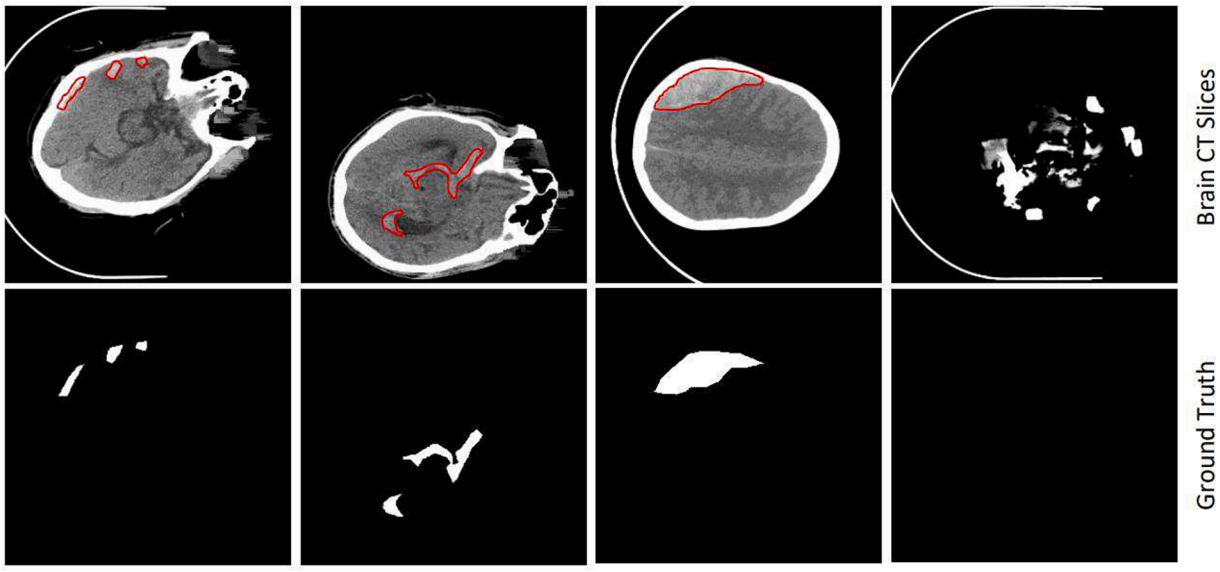


Fig. 1. Example of different challenging cases in hemorrhage images (pre-processed from the PhysioNet [5] dataset). The top row displays brain CT slices, while the bottom row represents their corresponding ground truth mask of those brain slices. Note that the dynamic shapes, various sizes, and positions make the segmentation task challenging. The red strokes pointing towards specific areas indicate the locations of hemorrhage.

presence of blood. Therefore, proficient clinical expertise and intensive experience are imperative for manual detection and segmentation of hemorrhages in brain CT scans. Additionally, it is a time-consuming and labor-intensive task, which may hinder rapid assessment and decision-making. Moreover, the extended assessment duration and unexpected delays, particularly in instances of high patient volumes or emergency scenarios, could have significant consequences. Radiologists can save time by automating the ICH segmentation task so that it can be quickly used to analyze and interpret the clinical consequences of the hemorrhage. Automated segmentation algorithms can rapidly analyze CT scan images and identify hemorrhagic regions within seconds or minutes. This speed is particularly crucial in emergency situations, enabling prompt diagnosis and timely intervention. Therefore, automated segmentation can expedite the diagnosis process, leading to faster interventions and better patient outcomes.

However, automatic hemorrhage segmentation is challenging due to several complex factors. One of the primary challenges arises from the diverse and dynamic nature of intracranial hemorrhages. Hemorrhages can manifest themselves with varying degrees of severity, resulting in a wide range of sizes, shapes, and distributions within the brain tissue. The subtle density differences between hemorrhages and adjacent structures can result in ambiguous boundaries that are hard to discern even for experienced radiologists. Moreover, CT images are characterized by limited soft tissue contrast, which can hinder the clear distinction between hemorrhagic areas and surrounding brain tissue. Fig. 1 shows examples of brain CT scan images (slices), which demonstrate the difficulties of accurate segmentation. These variabilities make it difficult to rely solely on statistical approaches for accurate segmentation. In complex images, statistical methods like hidden Markov models (HMMs) [2], gaussian mixture models (GMMs) [3], and *k*-means clustering [4] have trouble handling ambiguous boundaries and variations in object appearance due to their reliance on local context. To address these challenges, there is a requirement for innovative approaches that harness the power of sophisticated machine learning algorithms, including deep neural networks. These methods possibly learn intricate features and patterns from large datasets, enabling them to adapt to the varying sizes, shapes, and appearances of hemorrhages. Deep learning techniques are especially capable of automatically learning hierarchical features from raw pixel data, making them more adaptive to variations in object appearance and background clutter.

In recent years, the domain of medical image segmentation has witnessed remarkable advancements due to the emergence of deep learning-based techniques [6–9]. The efficacy of deep learning-based methods relies on their ability to fine-tune millions of parameters. The overall learning process heavily depends on huge datasets with precise annotations. However, accurate data level annotations can only be provided by experts with domain knowledge. It is very challenging to acquire reliable annotations for large-scale data (e.g., volumetric CT or MRI scans), especially for tasks like semantic segmentation which require pixel/voxel-level annotation for each instance. Consequently, the researchers and practitioners have been exploring strategies to mitigate the scarcity of meticulously annotated data concerning the difficulties. To alleviate annotation scarcity, a feasible approach is to take semi-supervised learning (SSL) strategies which leverage both labeled and unlabeled data to effectively train the deep neural networks [10]. Semi-supervised learning (SSL) for medical image segmentation presents an exciting avenue for addressing the aforementioned challenges. It offers a pathway to reducing the exhaustive manual effort and costs associated with labeling vast amounts of data. Moreover, the integration of unlabeled data into the learning process not only enhances the model's learning capabilities but also potentially alleviates the requirement for an exorbitant amount of fully labeled data. This is achieved by exploiting the inherent structure and relationships present in unlabeled data, thereby assisting the model in capturing underlying patterns and representations of data. In most existing approaches, medical image segmentation can be considered as a pixel-level classification task to generate segmentation probabilistic maps and assign each pixel a certain class.

Recently, several semi-supervised methods have been presented for medical image segmentation. Existing semi-supervised methods can be broadly divided into three categories: (i) Consistency regularization (ii) Pseudo-labeling (iii) Auxiliary task learning. The technique of pseudo-labeling involves iteratively repeating this procedure in a self-training cycle while utilizing a model trained on the combination of the labeled samples and pseudo-labeled samples. For instance, Bai et al. [11] employed conditional random field (CRF) to improve pseudo labels and iteratively updated the network parameters and pseudo segmentation labels. Kim et al. [12] employ a semi-supervised learning approach to establish dense correspondences among semantically similar images by combining a limited number of ground-truth correspondences alongside

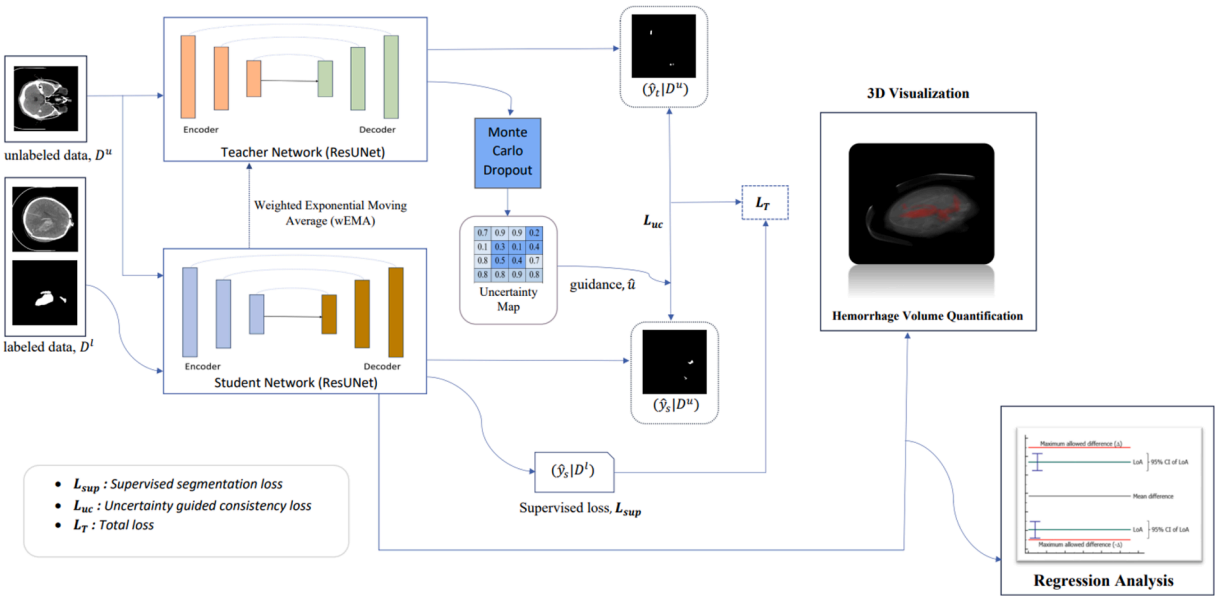


Fig. 2. Schematic illustration of the proposed framework for semi-supervised hemorrhage segmentation.

a substantial collection of confident pseudo-labels generated from model predictions. Consistency regularization is one of the most widely used techniques for semi-supervised learning (SSL), the objective typically being the training of models that maintain invariance to a range of data augmentations. For example, Yap et al. [13] presented a semi-supervised learning approach for lesion segmentation tasks based on the ideas of cut-paste augmentation and consistency regularization. However, inconsistent data distributions between labeled and unlabeled data can lead to overfitting due to the poor tuned of hyperparameters. The model may become too tailored to the labeled data's distribution and unable to generalize to unlabeled data's distribution. To alleviate the over-fitting, Xie et al. [14] exploited the attention mechanism to learn the pair-wise link between labeled and unlabeled data. Wang et al. [15] introduced a curriculum learning approach that employs a multi-task semi-supervised attention-based model for the segmentation of intracerebral hemorrhage (ICH) from CT images. Samuli et al.'s [16] temporal ensembling technique recommended utilizing predictions based on exponential moving average (EMA) for unlabeled data as targets for consistency. However, maintaining the EMA predictions during the training process is a heavy burden. In order to solve the issue, Tarvainen et al. [17] suggested a mean teacher model by averaging model weights instead of predictions. Yu et al. [18] further extended the mean teacher paradigm with an uncertainty quantification strategy to improve the performance of the consistency-based approach so as to learn from more meaningful and reliable targets during training. Meyer et al. [19] proposed an uncertainty-aware temporal self-learning for semi-supervised segmentation of prostate zones and beyond. Instead of perturbing networks or data for consistency learning, another line of research focuses on building task-level regularization by adding auxiliary tasks to leverage geometric information with distance maps. The concept of auxiliary task learning encompasses training of a model on not only the primary task (task of interest, e.g., semantic segmentation), but also on one or more auxiliary tasks concurrently. The auxiliary task (e.g., distance map prediction) can help the model learn spatial relationships and geometric properties within the data. By jointly learning the primary task and auxiliary tasks, the model can benefit from the shared knowledge and representations across tasks. Li et al. [20] created a multi-task network to construct adversarial regularized shape-aware restrictions. Luo et al. [21] introduced a dual-task consistency learning approach by combining regression and segmentation tasks. In order to further improve performance, Zhang et al. [22] combined the mutual learning of dual-task networks with the learning of cross-task consistency. Although these models have reported good results for semi-supervised medical image segmentation, they still fail to adequately address the necessity for further volumetric analysis of intracranial hemorrhage.

In this study, we utilized an uncertainty-guided semi-supervised (UGSS) learning framework for fully automatic intracranial hemorrhage segmentation use case. The model is trained using a semi-supervised scheme that leverages both labeled and unlabeled data. It consists of two different branches, namely teacher and student networks, which typically have similar architectures. The role of the student model is to learn from the teacher model's predictions and gradually improve its own performance. The purpose of the teacher network is to provide guidance and act as a source of knowledge for the student network. The student model is the target model to be trained, and it assigns the weighted exponential moving average (wEMA) of its weights to the teacher model at each step of training. On the other hand, prediction of the teacher model offers an additional channel of supervision for the student model to learn by the utilization of uncertainty quantification. For the final inference, we utilized the trained student model to perform the hemorrhage segmentation task. In addition to segmentation capabilities, we further incorporate a comprehensive 3D volumetric view of brain hemorrhage as well as quantify the volume of that hemorrhage. Our main contributions are four-fold: (i) We propose a comprehensive framework that combines two essential components: uncertainty quantification and consistency loss. These components work synergistically to enhance the robustness and generalization capabilities of segmentation models, particularly in scenarios with limited labeled data. (ii) We dynamically adjust the influence of the consistency loss using an exponential schedule, enabling the model to gradually focus on unlabeled data as training progresses. (iii) We adopt a weighted exponential moving average (wEMA) that adapts to the influence of different sources of information during training. It provides an additional layer of control and customization over conventional exponential moving average. (iv) We integrate the hemorrhage volume (e.g., cubic centimeters) quantification and comprehensive 3D volumetric view into the framework that further elevates diagnostic and visualization abilities. These features offer a comprehensive and accurate assessment of intracranial hemorrhages, aiding in treatment planning, patient monitoring, and benefiting overall decision-making.

Table 1
Pseudo code for the proposed method.

Algorithm 1. The core learning algorithm of the proposed framework

Input: A batch of $\{x_l, y_l\}$ from labeled dataset D^l and $\{x_u\}$ from unlabeled dataset D^u
Output: The trained teacher model and student model
1. Represent the output predictions of teacher and student network $\hat{y}_t$ and $\hat{y}_s$ respectively.
2. while stopping criteria not met
3. $(x_l^i, y_l^i), (x_u^i) \leftarrow$ sampled form D^l and D^u
4. Calculate uncertainty guided consistency loss, L_{uc} as Eq. (5)
5. Calculate supervised segmentation loss, L_{sup} as Eq. (8)
6. Calculate overall loss, L_T as Eq. (9)
7. Update the student model's weights as Eq. (8)
8. Update the teacher model's weights with weighted exponential moving average (wEMA) as Eq. (11)
9. end while

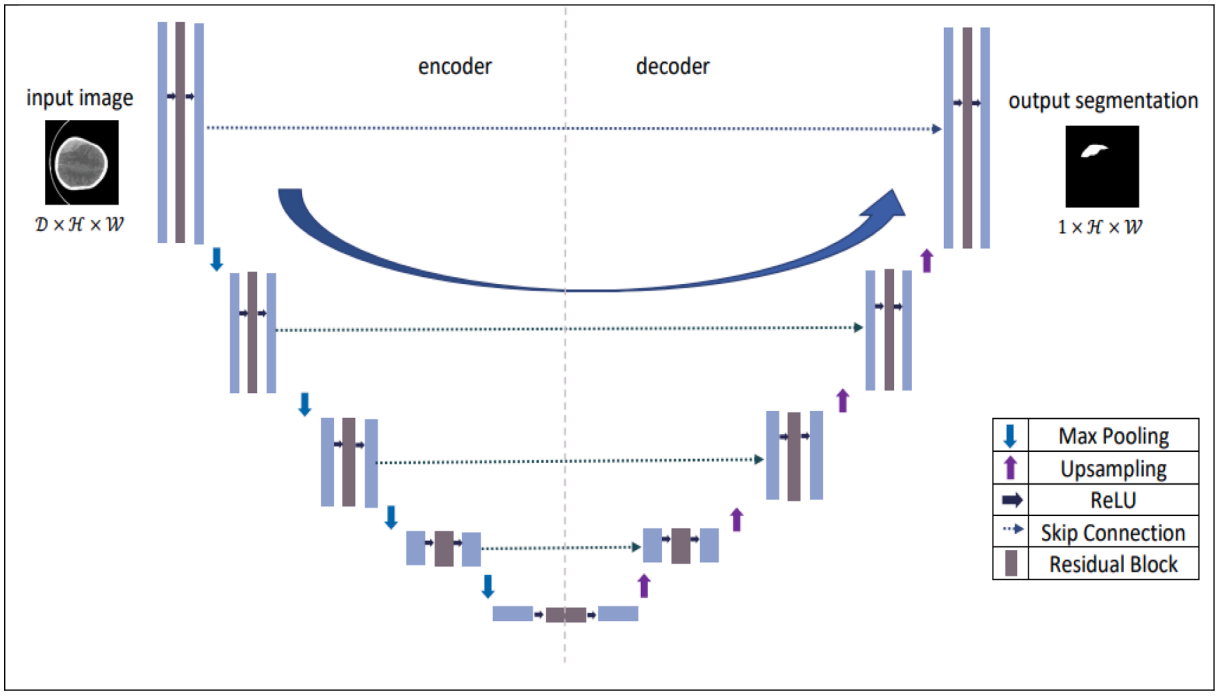


Fig. 3. ResUNet architecture. The rectangular boxes represent multi-channel feature maps and arrows of various colors represent different mathematical operations for the binary image segmentation task.

2. Proposed methodology

In this section, we first explain the backbone architecture of the teacher and student models and consistency regularization technique. Next, we describe the uncertainty-guided semi-supervised learning scheme and loss functions that are utilized for training the backbone network. An overview of our proposed framework is illustrated in Fig. 2, consisting of a teacher model and a student model, following the mean teacher [17] idea. Before introducing our model, we define the semi-supervised segmentation problem with a set of notations. In our problem setting, we are given a training set containing N labeled samples and M unlabeled sample images with spatial dimension $H \times W$. The labeled set is defined as, $D^l = \{x_l^i, y_l^i\}_{i=1}^{N}$ and unlabeled set as $D^u = \{x_u^i\}_{i=N+1}^{N+M}$ where $x^i \in R^{H \times W}$ is the input image and $y^i \in \{0,1\}^{H \times W}$ is the corresponding pixel-wise segmentation label mask. Here, (x_l^i, y_l^i) represent sample from labeled dataset, while (x_u^i) denote sample from unlabeled dataset. The symbols $\hat{y}_t$ and $\hat{y}_s$ represent the output predictions of the teacher and student models, respectively. In summary, the student model is trained by minimizing (i) the supervised segmentation loss (L_{sup}) on labeled data, D^l (ii) uncertainty-guided consistency loss (L_{uc}) between the student model and teacher model on both labeled, (D^l) and unlabeled, (D^u). The pseudo code of the proposed uncertainty-guided mutual consistency learning framework is given in Table 1.

As shown in Fig. 2, the mean teacher framework typically consists of two branches, teacher and student who utilize the same ResUNet (shown in Fig. 3) architecture as a backbone. Following the design of the mean teacher framework, the student model is trained both on labeled (D^l) and unlabeled (D^u) data. The exponential moving average (EMA) of the student model's weights is assigned to the teacher model. The estimated uncertainty from the teacher model is used to guide the learning of the student model to learn more reliable information from unlabeled data for semi-supervised learning.

2.1. Backbone architecture

For both teacher and student networks, we utilize a similar architecture called UNet [23] which has demonstrated remarkable performance in various segmentation tasks. It is typically a convolutional neural network (CNN) architecture, which has a symmetric encoder-decoder scheme. In this use case, we utilize a tailored variant of the standard UNet architecture, integrating residual blocks to create what we refer to as ResUNet. The residual block component helps mitigate the vanishing gradient problem and promotes better gradient flow. Fig. 3

depicts the structure of the customized ResUNet.

The architecture generates pixel-wise classification by using a combination of encoding, decoding, and skip connections to capture and fuse information at multiple scales and ultimately producing a pixel-wise classification mask that delineates the target hemorrhage regions within an image. The pixel-wise classification indicates whether the pixel belongs to the object or region of interest (foreground) or the background in the case of binary segmentation. The encoder block is responsible for extracting features at different levels from the input. Each level of the encoder is linked to a corresponding level in the decoder, allowing the model to capture both local and global context. More specifically, the encoder consists of multiple convolutions followed by Rectified Linear Unit (ReLU) [24] and 2 × 2 max pooling operation, as shown in Fig. 3. Specifically, the encoder consists of several consecutive residual blocks, each followed by a 2 × 2 max pooling operation. The residual blocks solve the risk of vanishing gradients problem for deeper networks. In this case, the encoder has four residual blocks, each increasing the number of feature channels from 3 to 16, 16 to 32, 32 to 64, and 64 to 128, respectively. Like the encoder, the decoder also consists of multiple residual blocks. The dashed arrow lines show skip connections that directly connect encoder layers to corresponding decoder layers. These connections help in preserving fine-grained details from the encoder. As the encoder progresses, it down-samples the spatial dimensions of the feature maps, effectively compressing the information while preserving important features. The decoder portion of the architecture begins to up-sample the feature maps, gradually recovering the original image's spatial resolution. It uses transposed convolutional layers (also known as deconvolution or up-sampling layers) to increase the spatial dimensions of the feature maps. At the final stage of the decoder, a sigmoid activation is used to produce a mask where each pixel is assigned a value between 0 and 1, representing "probabilistic" predictions for hemorrhage regions.

2.2. Consistency learning

By incorporating consistency regularization into semi-supervised framework, the neural network is encouraged to produce consistent predictions across different perturbations of the same input. This study leverages a consistency-based approach which encourages similar predictions between student and teacher networks for the unlabeled (D^u) data. Both student and teacher networks use the softmax function to generate discrete probability distribution for each class (segmentation classes). We incorporate Kullback-Leibler (KL) divergence [25] to measure the difference between the probability distributions of both model's predictions. Subsequently, a measure of consistency loss (denoted as L_{con}) is computed by minimizing the divergence between the student and teacher model's predictions. For each class i at a specific pixel/region, let's denote student and teacher model's probability densities as S and$T$ respectively.

$$L_{con} = D_{KL}(S\|T) = \sum_i S(i) log\left(\frac{S(i)}{T(i)}\right) \tag{1}$$

In the Equation (1), initially we calculate the element-wise ratio of student probabilities to teacher probabilities for each class i at a specific pixel/region, then multiply each element of the logarithm ratio with the corresponding student probability $S(i)$.

2.3. Uncertainty quantification

Quantification of uncertainty can take on various forms, depending on the context. For semi-supervised learning, the uncertainty can be used to judge whether the model provides accurate and confident prediction, which can be leveraged to further exploit the unlabeled data. Uncertainty, in this context, is a measure of how confident or reliable the teacher model is in segmenting each data point (pixel). As the teacher model's predictions exhibit higher reliability, it helps the student model in a more effective assimilation of knowledge from the teacher model. Consequently, it allows the student model to be more confident and consistent when dealing with data points where the teacher model is confident and vice versa. The uncertainty is estimated using dropout-based techniques and entropy calculations, which is a measure of the randomness or uncertainty of model's predictions. Higher entropy values generally indicate more uncertainty in the predictions. First, we calculate the entropy [26] for each set of predictions (sampled predictions), and then takes the mean entropy across all the samples. Afterwards, we utilize Monte Carlo (MC) dropout [27] during the inference to observe the variability in predictions. Essentially, we run the multiple forward passes with the dropout enabled and observe how the prediction changes each time. with dropout (a randomness-inducing technique) enabled, and we observe how the predictions change each time. The uncertainty ($\widehat{u}$) is estimated using Equations (2) and (3) by performing T stochastic forward passes on each input sample image from the teacher model with random dropout:

0.7	0.9	0.9	0.2
0.1	0.3	0.1	0.4
0.8	0.5	0.4	0.7
0.8	0.8	0.9	0.8
Uncertainty Map			

0.3	0.1	0.1	0.8
0.9	0.7	0.9	0.6
0.2	0.5	0.6	0.3
0.2	0.2	0.1	0.2
Confidence Map			

Fig. 4. Example of estimated uncertainty and corresponding confidence map.

$$p_k = \frac{1}{T}\sum_{t=1}^{T} p_k^t, \tag{2}$$

$$\widehat{u} = -\sum_{k=1}^{C} p_k \log_2 p_k, \tag{3}$$

where p_k^t is the prediction logits of class k at the t^{th} time in the forward pass, C is the number of classes in the segmentation tasks,p_k is the average softmax probability of T stochastic passes from teacher model and $\widehat{u}$ is the estimated segmentation uncertainty. This measure of uncertainty is then further utilized in the uncertainty-guided consistency loss (L_{uc}). The uncertainty feature map shows the low-confidence area of the teacher model (as displayed in Fig. 4), which helps guide the student model to enhance its segmentation ability by learning from only the most meaningful and reliable predictions from the teacher model. This adaptability helps the model become more robust and reduces its sensitivity to potentially noisy or mislabeled data.

We incorporate the concept of pseudo-label confidence through uncertainty quantification. The pseudo-label confidence is defined as the inverse of the entropy, normalized within the range $[0, 1]$ ensuring that low-entropy (high-confidence) predictions receive higher weights. The pseudo-label confidence $\left(p_{conf}\right)$ can be defined as equation (4).

$$p_{conf}(i) = 1 - \frac{\widehat{u}_i}{\widehat{u}_{max}}. \tag{4}$$

Where, $\widehat{u}_i$ is the entropy at pixel i and $\widehat{u}_{max}$ is the maximum possible entropy for a pixel (i.e., Logarithm of number of classes; $\log(C)$, where C is the number of classes).

2.4. Uncertainty-guided consistency loss

We utilized uncertainty-guided consistency loss (L_{uc}) to ensure the student network's predictions more aligned with the teacher's knowledge, especially in situations where the teacher's predictions are more reliable or confident. It is obtained by multiplying the consistency loss (L_{con}) by the segmentation uncertainty estimate ($\widehat{u}$) from the teacher network. The idea is to assign higher weights to data points with lower

uncertainty (i.e., more reliable) and lower weights to data points with higher uncertainty (i.e., less reliable). It gives more weight to consistency loss where the teacher model is more confident and less weight for uncertain or ambiguous predictions. The student network would pay more attention to data points where the teacher's predictions are more reliable (low uncertainty) and less attention where the teacher's predictions are uncertain. The uncertainty aware consistency loss (L_{uc}) function can be formally expressed as:

$$L_{uc} = \sum_{x_u \in D^u} L_{con} \times \widehat{u} \tag{5}$$

Here, $\widehat{u}$ is the estimated segmentation uncertainty by the teacher network and L_{uc} represent the consistency loss. This uncertainty guidance helps in reducing the sensitivity of the student model when dealing with noisy or uncertain data and adaptive to the varying levels of reliability in the teacher's predictions.

2.5. Overall loss function

As the framework is trained under semi-supervised settings, we employed both supervised (L_{sup}) and unsupervised loss (L_{uc}) functions in the model. The student model is trained by minimizing the supervised loss (L_{sup}) on labeled data and consistency loss (L_{con}) on unlabeled data. Notably, the unlabeled part takes unlabeled samples as inputs, in contrast supervised part only takes labeled sample for learning purposes. In the context of semantic segmentation, the goal is to classify each pixel as either the object of interest (positive class) or the background (negative class). Thus, there is often a class imbalance between the background (negative class) and the object of interest (positive class). As the hemorrhage segmentation task is a heavily imbalanced task with respect to pixel-level instances, we incorporate weighted binary cross-entropy loss (L_{wbc}) to address that class imbalance problem. It tends to assign a higher weight to the class that is less frequent, effectively giving it more importance during the training process. This can help the model pay more attention to the class with fewer instances and potentially improve the segmentation performance. The weighted binary cross-entropy loss is formalized as Equation (6).

$$L_{wbc} = -\frac{1}{N}\sum_{i=1}^{N}\left(w_{pos} y_i \log(p_i) + w_{neg}(1-y_i)\log(1-p_i)\right), \tag{6}$$

where, N is the total number of samples in the dataset. The true label (ground truth) is y_i for the i^{th} sample. Here, p_i is the predicted probability of the i^{th} sample given by the model's output. The positional weight (w_{pos}) in the weighted binary cross-entropy loss function can be defined as

$$w_{pos} = \log \frac{p_T}{p_{pos}}. \tag{7}$$

The calculated positional weight (w_{pos}) reflects the relative occurrence of positive and negative classes across all the mask images. This weight helps adjust the impact of the positive class in the loss calculation, addressing class imbalance and putting more emphasis on the positive class during training. In Equation (7), the logarithm compresses the ratio of all the pixel count (p_T) to positive pixel count (p_{pos}). The logarithmic transformation provides a smooth transition in weight values as the ratio of positive to all pixel's changes. This ensures that the weight adjustments are gradual rather than sudden, which can be beneficial for training stability. Then, we estimate both supervised segmentation loss function (L_{sup}) and overall loss function (L_T) with Equations (8) and (9), respectively. The overall loss function would be the combination of supervised and unsupervised loss:

$$L_{sup} = \sum_{(x_l, y_l) \in D^l} L_{wbc}, \tag{8}$$

$$L_T = L_{sup} + \lambda_u L_{uc}, \tag{9}$$

Here, λ_u is a hyperparameter controlling the contribution of the unlabeled loss (L_{uc}) term. We use an exponential scheduler to adjust the contribution of the hyperparameter based on the progression of training. The goal of this scheduling is to gradually increase the impact of the consistency loss on the overall training process as the training progresses. The formula for the exponential scheduler can be expressed as Equation (10).

$$\lambda_u = \alpha + \left(\frac{\delta_c}{\delta_T}\right)^e \times (m - \alpha). \tag{10}$$

where, δ_c is the current global training step and δ_T is the is the total number of training steps in an epoch or the entire training process. The hyper-parameter α is the initial weight assigned to the consistency loss. The maximum assigned weight to the consistency loss is denoted by m and the exponent e controls the rate of increase. A higher exponent value leads to a faster ramp-up of the weight.

2.6. Weighted exponential moving average (wEMA)

In the realm of knowledge distillation, the exchange of insights between teacher and student models plays a pivotal role in enhancing the learning process. One popular technique, known as exponential moving average (EMA) [28], facilitates transferring knowledge between the teacher network and student network. The student model is the desired model to be trained, and at each training step, the exponential moving average (EMA) of its model weights is assigned to the teacher model. In this context, we expand upon this concept by introducing the weighted exponential moving average (wEMA). Unlike conventional exponential moving average (EMA), the weighted exponential moving average (wEMA) provides an additional layer of control and customization by assigning different weights to different sources of information during the update process. It provides the flexibility to adaptively adjust the importance of different sources of information as training progresses. The weight of the teacher network (θ') and the weight of the student network (θ) determine the balance between the teacher's and student's contributions. The weighted moving average (wEMA) process is defined as:

$$\theta'_t = \alpha' \theta'_{t-1} + (1-\alpha')\left(w_s \times \theta_t + w_t \times \theta'_{t-1}\right) \tag{11}$$

At a training step t, we update the weights of the teacher network (θ') in Equation (11). Here α' denotes the wEMA decay that regulates the update rate. The α' value in the provided training framework is determined using a cosine annealing schedule, which smoothly adjusts the value over the course of training. The adjustment is achieved through a cosine function that oscillates between -1 and 1 over the interval $[0, \pi]$, where the current training step is normalized by the total number of training steps. The cosine annealing schedule promote a balanced and effective update of the teacher network parameters.

2.7. Hemorrhage volume quantification

Hemorrhage volume quantification plays a vital role in clinical practice, and patient care by providing valuable information for diagnosis, and monitoring. It allows medical professionals to assess whether the condition is improving, stable, or worsening and adjust the treatment accordingly. Healthcare providers can make more informed decisions regarding surgery, medication, or other interventions based on the size and location of the hemorrhage. The hemorrhage volume is typically measured in cubic centimeters (cm^3). In this context, the intracranial hemorrhage volume is calculated by counting the number of voxels (V_i) within the hemorrhage region and multiplying it by the size of each voxel. Voxel size (S_z) represents the volume of each element in

Table 2
Subject demographics of PhysioNet dataset.

Imaging modality	Computerized Tomography (CT)
Total number of subjects	82
Number of subjects with hemorrhage	36
Total number of extracted slices from CT scans	2814 (No hemorrhage: 2496)
Raw file format	NIfTI
CT scan's unit measurement	Hounsfield Unit (HU)
Slice thickness	5 mm
Annotation Pattern	Slice level, Pixel level
Patient age range	27.8 ± 19.5
Gender	46-Male, 36-Female

the three-dimensional grid of the medical image. We calculate the voxel size by taking the product of the three voxel dimensions (v_x, v_y, and v_z) obtained from the CT scan header information. Then we calculate the count of pixels (voxels) in the hemorrhage region by summing the values in the hemorrhage ground truth or prediction. This count represents the number of voxels that make up the hemorrhage region. The volume is calculated by counting the number of voxels within the hemorrhage region and multiplying it by the size of each voxel. The formula for calculating the hemorrhage volume, denoted as V_H, is given by Equation (12).

$$V_H = \sum_i V_i \times S_z \tag{12}$$

Here, V_i is i^{th} voxel within the hemorrhage region and S_z is the size of voxel.

3. Experimental setup

This section discusses the details of the dataset and empirical results obtained for intracranial hemorrhage segmentation task. Additionally, ablation studies are conducted to study the impact of each component in the proposed method.

3.1. Dataset

In this study, we analyze PhysioNet [5] dataset which contains a collection of CT studies regarding intracranial hemorrhage. The Siemens/ SOMATOM CT scanner was used to acquire the data which had an isotropic resolution of 0.33 mm, 100 kV. The dataset contains 82 subjects (CT scans), including 36 scans for patients diagnosed with intracranial hemorrhage. Each patient's CT scan typically includes 34 slices (cross-sectional images) on average with 5 mm slice-thickness. The raw CT scans are in NIfTI (Neuroimaging Informatics Technology Initiative) format. The unit of measurement of the CT scans is the Hounsfield Unit (HU) which is a measure of radiodensity. More demographic details regarding the dataset are shown in Table 2.

The raw CT scans are originally the collection of 2D slices in different 3D orientations (axial, sagittal, and coronal). This 3D rendering provides a comprehensive view of anatomical structures. In Fig. 5, we display a representative section of a raw 3D CT scan extracted from a NIfTI (Neuroimaging Informatics Technology Initiative) file.

The PhysioNet dataset is highly imbalanced at slice-level as well as pixel-level (positive hemorrhage instances). A total of 2,814 slices were extracted from the CT scan data. Most of the slices in the dataset do not contain positive cases of intracranial hemorrhage. Among those slices, only 318 of them have the presence of intracranial hemorrhages. This makes it a heavily imbalanced dataset thus extra challenging for automatic hemorrhage detection and segmentation tasks. Fig. 6 demonstrate the distribution of hemorrhage both in slice and pixel level instances. In managing a limited set of observations (2814 images from 82 CT scans), we employed augmentation techniques to enhance the diversity and variability within the dataset. We utilized transformations such as rotation, flipping, and cropping in the augmentation process.

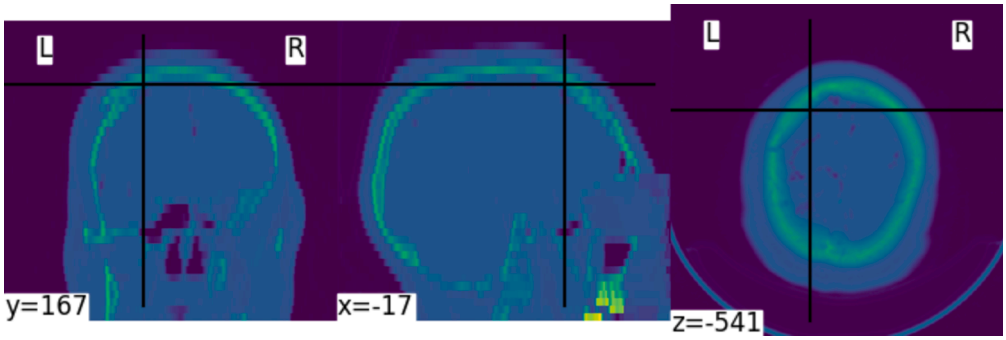


Fig. 5. Illustration of a raw CT sample from the dataset.

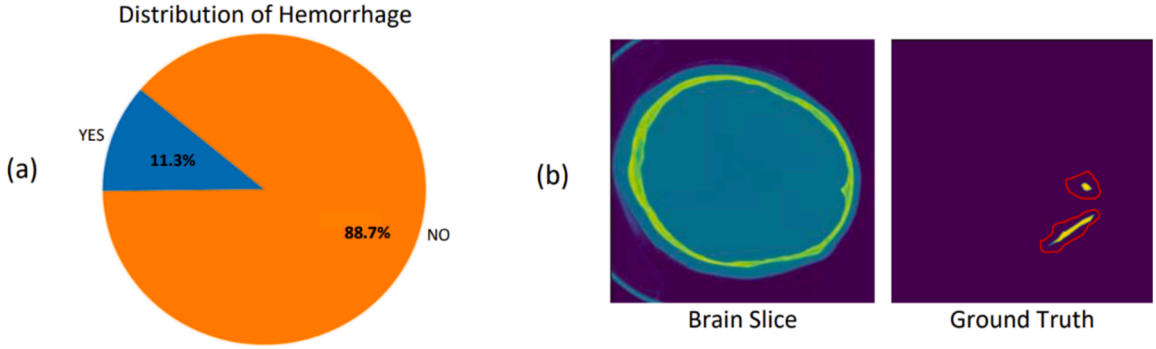


Fig. 6. The left panel (a) represents the slice level distribution of the intracranial hemorrhage. The blue portion indicates that those slices have the presence of hemorrhage, while orange portion indicates the opposite. On the right panel (b), the two images illustrate pixel-level data imbalance (marked in red) within the brain CT slices.

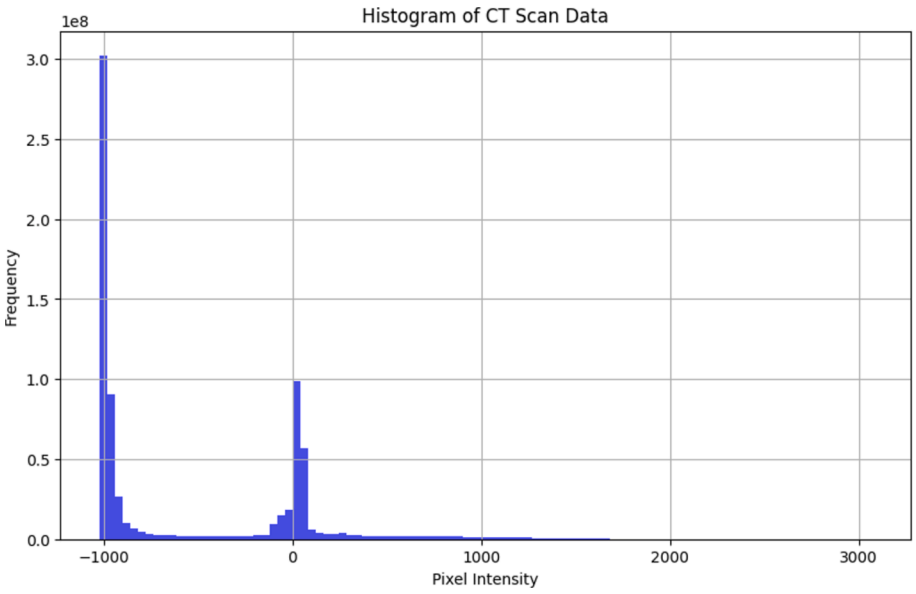


Fig. 7. Histogram showing pixel intensities and counts for raw CT scans.

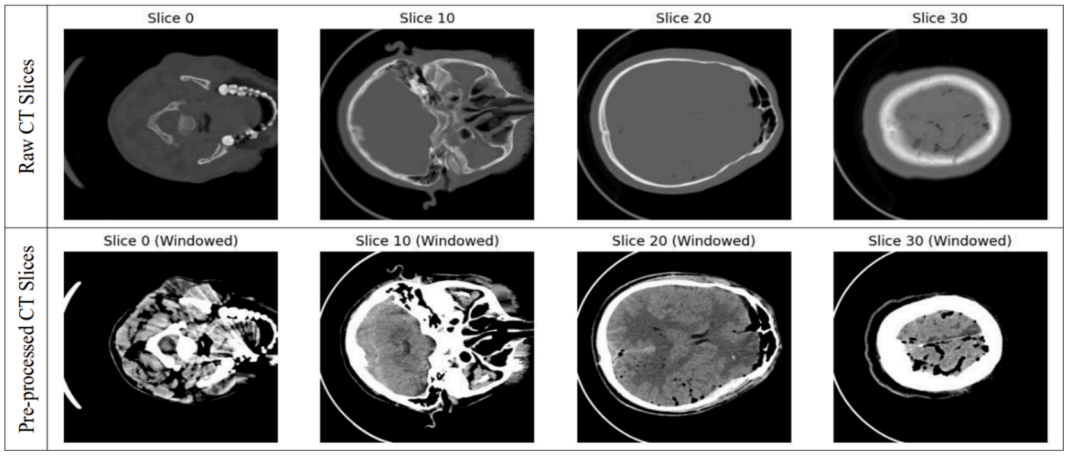


Fig. 8. The first row represents a snapshot of every 10^{th} raw slice from CT scan and the second row illustrates the pre-processed version after applying the window technique to those raw CT slices.

3.2. Preprocessing

The CT scan modality uses Hounsfield Units (HU) [29] to represent pixel intensities that quantify the radiodensity of tissues in the body. They often have a wide range of intensity values due to variations in tissue density. The intensity values cover a wide dynamic range, typically from −1000 HU to ∼+2000 HU (referred to in Fig. 7).

The variability in contrast levels within raw CT scans makes it difficult to distinguish between different anatomical structures accurately. Hence it creates a challenge for segmentation algorithms to define the appropriate regions of interest. To tackle this issue, we use windowing technique [30] to preprocess the raw CT scans, as depicted in Fig. 8. Windowing is a common image processing technique used in medical imaging, particularly in Computed Tomography (CT) to adjust the display of pixel intensity values within a specified range to enhance the visibility of specific structures or tissues in the image. It adjusts the range of pixel values in the image to emphasize certain structures or pathologies while maintaining the overall contrast and dynamic range. The technique involves two key parameters: window level (wl) and window width (ww). The window level (wl) defines the center of the window and corresponds to the Hounsfield Unit (HU) value that will be mapped to the middle of the grayscale display. Mathematically, it is the midpoint of the range of pixel values after windowing. The relation can be expressed as Equations (13) and (14).

$$wl = \frac{min + max}{2} \tag{13}$$

$$ww = max - min, \tag{14}$$

where, min and max, are the windowed range's minimum and maximum Hounsfield Unit (HU) values respectively and window width (ww) defines the range of Hounsfield Unit (HU) values that will be mapped to the grayscale display. The windowing operation then maps pixel intensity values within the range between $(wl - ww/2)$ and $(wl + ww/2)$ to the full grayscale display range (usually 0 to 255 for 8-bit images). Mathematically, we have:

$$P_{new} = \left(\frac{\left(P - \left(wl - \frac{ww}{2} \right) \right)}{ww} \right) \times 255. \tag{15}$$

Here, P denotes the pixel intensity value in the original CT scan before

Table 3
Quantitative comparison between different paradigms and the proposed method using varying proportion of labeled data on PhysioNet dataset.

Labeled Data		50 %		70 %		80 %		100 %		
Paradigm	**Method**	***JCD***	***DSC***	***JCD***	***DSC***	***JCD***	***DSC***	***JCD***	***DSC***	**PA**
sup	UNet	0.248	0.394	0.281	0.420	0.321	0.457	0.367	0.460	0.798
sup	ResUNet	0.270	0.412	0.298	0.458	0.353	0.466	0.380	0.497	0.823
semi-sup	ResUNet-CutMix	0.312	0.426	0.370	0.545	0.3936	0.564	0.402	0.571	0.887
semi-sup	ResUNet-CutPaste	0.309	0.432	0.381	0.549	0.404	0.576	0.390	0.586	0.941
semi-sup **[proposed]**	**UGSS**	**0.326**	**0.454**	**0.398**	**0.573**	**0.421**	**0.598**	**0.441**	**0.613**	**0.952**

applying the windowing operation and P_{new} represents the new gray scale pixel intensity value after the windowing operation has been applied.

3.3. Evaluation metrics

To gauge the segmentation performance of our model, we used both the Dice coefficient (*DSC*) and Jaccard index (*JCD*), given in Equations (16) and (17) respectively**.** The Dice coefficient, also known as the Sørensen–Dice coefficient [31], is a commonly used metric for measuring the similarity or overlap between two sets. In the context of our semantic image segmentation task, we used both Dice coefficient and Jaccard index [32] to evaluate the performance of the framework by comparing similarity between the predicted binary mask and ground truth binary mask.

$$DSC = 2 \times \frac{|y \cap \hat{y}|}{|y| + |\hat{y}| + \varepsilon}, \tag{16}$$

$$JCD = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}| + \varepsilon}. \tag{17}$$

Here, y is the set of pixels or voxels that are positive according to the ground truth and $\hat{y}$ is the set of pixels or voxels that are predicted as positive (belonging to the target class). The set cardinality is denoted by symbol $|\bullet|$. To handle the cases where both the predicted and ground truth masks are empty (i.e., they contain no positive elements), we add a small epsilon (ε) value to ensure that the metrics are well-defined. To evaluate the reliability of the proposed method, standard Cohen's kappa coefficient [33] is computed by measuring the agreement rate between the ground truth and the proposed automated method. It assesses the level of agreement between the actual and predicted instances. In addition, we consider another evaluation metric called Pixel Accuracy (PA) to assess how well the model is performing across the entire image, considering all pixels, regardless of whether they belong to the object of interest or the background. This metric measures the ratio of correctly classified pixels (both foreground and background) to the total number of pixels in the image and simply can be expressed by Equation (18).

$$PixelAccuracy(PA) = \frac{NumberofCorrectlyClassifiedPixels}{TotalNumberofPixelsintheImage} \tag{18}$$

We utilize linear regression and Bland-Altman plot to assess the accuracy of the estimated volumes compared to the reference standard (ground truth volumes). Bland-Altman plot [34] is generated to visualize the agreement between two variables, which are assumed to represent some form of measurement or observation. First, we calculate the difference (Δ_v) between the estimated volume $(\hat{V})$ and the reference/true volume (V) for each data point via:

$$\Delta_v = \hat{V} - V \tag{19}$$

The plot visually shows the spread of the measurement differences and their relationship to the mean value. The x-axis of that plot represents the mean of the estimated and reference volumes, $((\hat{V} + V)/2)$ and the y-axis represents the differences. Each data point in the scatter plot corresponds to a pair of measurements. Additionally, two horizontal dashed lines are plotted at the limits of agreement, which are calculated as 1.96 times the standard deviation of the differences (95 % two-sided confidence interval), both above and below the mean. These lines help assess the precision and agreement between the two measurement methods. Points outside these limits may indicate outliers or poor agreement. Additionally, we fit a linear regression model to the data, where the estimated volumes are the independent variable, and the reference volumes are the dependent variable. We utilize the metric R^2 to evaluate the model's goodness of fit. A significant linear relationship between the estimated and reference volumes is indicated by a higher R^2 value (close to 1).

3.4. Model configurations

The proposed uncertainty guided framework (UGSS) is benchmarked against three settings, including a fully supervised (i.e., training on labeled data only) baseline and several semi-supervised learning baselines such as CutMix [35] based regularization (ResUNet-CutMix) and CutPaste [36] based regularization (ResUNet-CutPaste), respectively. To standardize the benchmarking process and ensure fair comparisons, all methods are implemented in a shared codebase using the PyTorch [37] framework. The framework is optimized using the AdamW [38] optimizer for up to 200 epochs. We utilized a workstation equipped with Intel® Core i7- 12700F, 2100 Mhz, 12 Core(s); NVIDIA GeForce RTX™ 3050, 4 GB GDDR6 for training the model. During training, the learning rate is linearly warmup in the first 10 epochs and gradually decayed after the 10^{th} epoch using a cosine scheduler. Early stopping is used depending on the performance on the validation dataset and 10 % of the training samples are randomly chosen as the validation dataset.

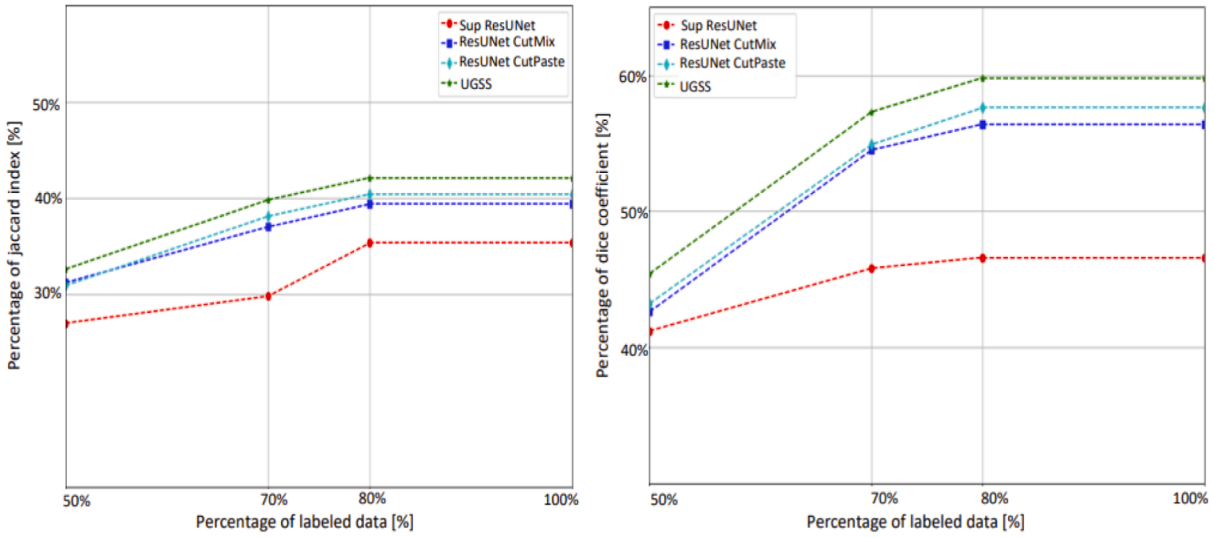
4. Results and discussion

This section provides an overview of the data split, evaluation metrics, and performance comparison with respect to different baselines. Additionally, this section provides some graphical representation of the quantitative results for the sake of concise interpretation. For a more comprehensive insight, we also conduct a qualitative analysis of the model outcomes.

4.1. Overall performance

We benchmarked the results of the proposed uncertainty guided semi-supervised (UGSS) framework with the supervised (sup) UNet, supervised (sup) Residual UNet (ResUNet) and semi-supervised (semi-sup) counterparts: ResUNet-CutMix, ResUNet-CutPaste. Table 3 presents a comprehensive evaluation of various hemorrhage segmentation methods under different scenarios, considering the percentage of labeled data during training. First, we explore the impact of the data mix (labeled and unlabeled) on the performance of these methods.

The labeled data split ranges from 50 % to 100 %, representing the proportion of data with ground truth labels used for training. When using 50 % labeled data, the supervised (sup) UNet achieves a Jaccard index of 0.248 and a Dice coefficient of 0.394. As the amount of labeled data increases, going up to 100 % labeled data; Jaccard, Dice coefficient and PA show improvements peaking at 0.367, 0.460, and 0.798,



(a) Comparison of Jaccard indices of different methods

(b) Comparison of Dice coefficients of different methods

Fig. 9. Comparison of segmentation performance of the proposed method (in green) compared to other paradigms using different percentages of labeled data (50%, 70%, 80% and 100%) on PhysioNet dataset.

respectively. Another supervised model ResUNet achieves improvements of approximately 3.54 % in Jaccard index, 8.04 % in Dice coefficient, compared to supervised (sup) UNet. This demonstrates the effectiveness of integrating residual blocks in ResUNet for enhancing segmentation performance. The supervised approach typically benefits from additional supervision as the proportion of labeled data rises, leading to better accuracy and agreement with ground truth segmentation. Semi-supervised models exhibit the ability to achieve comparable results, akin to the supervised approach, even with a lower proportion of labeled data compared to the fully supervised model. For instance, when using 80 % labeled data, the supervised method ResUNet yields a Jaccard score of 0.353 and a Dice coefficient of 0.466. In contrast, the semi-supervised baseline models (ResUNet-CutMix, ResUNet-CutPaste) achieve better performance than supervised model, even when utilizing only 70 % of labeled data. These results indicate the effectiveness of semi-supervised learning in improving hemorrhage segmentation. Across all data split scenarios, the semi-supervised models (including proposed method) outperform the supervised model. The comparison suggests that applying semi-supervised techniques with data augmentation such as CutMix [35] and CutPaste [36], can lead to improved model performance with less labeled data. This observation underscores the value of leveraging unlabeled data in enhancing model generalization and segmentation accuracy. The performance of the semi-supervised models improved with inclusion of proportion of labeled data (50 % to 100 %). This trend is expected as more labeled data generally leads to better understanding of the context. Delving into the semi-supervised settings, it can be observed that the proposed UGSS outperforms both ResUNet-CutMix and ResUNet-CutPaste in terms of Jaccard coefficient and Dice coefficient at different proportions of labeled data. At 100 % labeled data, UGSS surpasses ResUNet-CutMix by 12.23 % in Jaccard index, 7.35 % in Dice coefficient. When compared to ResUNet-CutPaste, UGSS achieves improvements of 13.08 % in Jaccard index, 4.60 % in Dice coefficient. We further visualize the segmentation results with a comparison plot, shown in Fig. 9. It can be observed that all semi-supervised models perform better than the supervised baseline in different labeled data settings. Besides, the proposed method outperforms other semi-supervised segmentation methods across different proportions of labeled data. This suggests that the proposed approach effectively harnesses the benefits of semi-supervised learning, achieving significant improvements in segmentation accuracy.

The evaluation results for hemorrhage detection based on sensitivity, specificity, overall accuracy, and Cohen's kappa coefficient are shown in Table 4. Upon careful examination of the results, it is evident that the proposed UGSS method outperforms the other methodologies across the board. The "semi-supervised" variations, namely "ResUNet-CutMix" and "ResUNet-CutPaste," also demonstrate good performance, showing increased sensitivity and specificity compared to the fully supervised method. However, they fall slightly short of the performance achieved by the proposed UGSS. The fully supervised method (denoted as "sup") shows the lowest performance in terms of sensitivity, specificity, accuracy, and cohen's kappa. These findings highlight the importance of exploring with unlabeled data, particularly in challenging scenarios. The proposed method exhibits the highest sensitivity, specificity, accuracy, and Cohen's kappa values which indicates its efficacy in accurately segmenting images, making it a promising instrument for image segmentation tasks. It achieves an accuracy of 89.03 %, indicating a high level of overall correctness in hemorrhage detection. The reasonable sensitivity value indicates its reliability for identifying hemorrhages, which is crucial in clinical applications to avoid false negatives. Besides, the computed specificity suggests that the method can effectively differentiate between hemorrhage regions, reducing the likelihood of

Table 4
Evaluation of intracranial hemorrhage detection performance on different paradigms and the proposed method.

Paradigm	Method	Sensitivity	Specificity	Accuracy (%)	Cohen's kappa
sup	ResUNet	0.657	0.818	85.21 %	0.792
semi-sup	ResUNet-CutMix	0.705	0.851	87.10 %	0.817
semi-sup	ResUNet-CutPaste	0.723	0.839	85.89 %	0.811
semi-sup[proposed]	UGSS	0.751	0.872	89.03 %	0.835

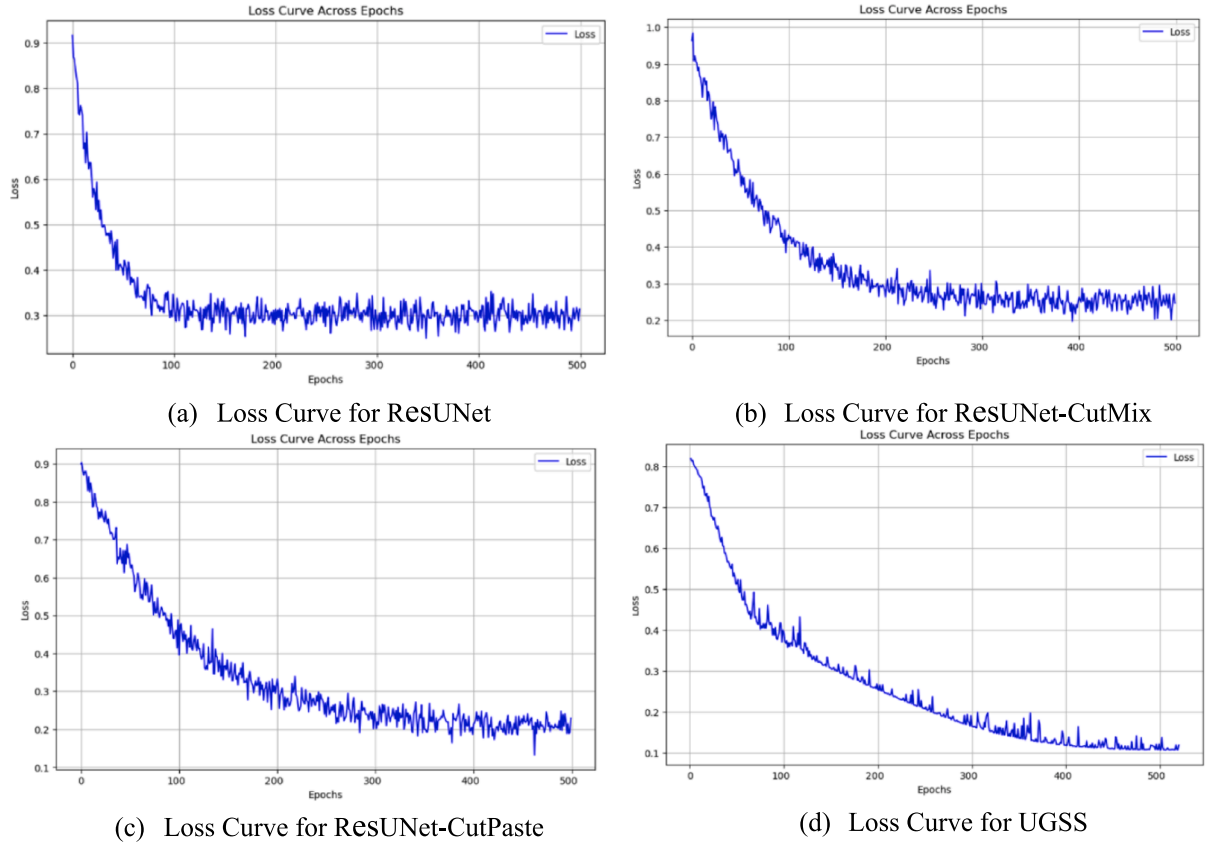


Fig. 10. Loss curves for various model configurations: (a) ResUNet, (b) ResUNet-CutMix, (c) ResUNet-CutPaste, and (d) UGSS across training epochs.

false positives. Cohen's kappa of 0.835 signifies substantial agreement between the model's predictions and the ground truth, indicating robustness and consistency in hemorrhage detection.

To better understand the training dynamics, we performed the loss curve analysis across various frameworks. The Fig. 10 presents the loss curves across epochs for different model configurations. Each subplot displays the evolution of the loss value over 500 training epochs, providing insight into the convergence behavior of each model. The semi-supervised settings (ResUNet-CutMix & ResUNet-CutPaste) exhibits a similar convergence trend, stabilizing at a slightly lower loss value compared to supervised training scheme. The proposed UGSS achieves steady-state loss among the models with lower loss metric values.

Additionally, we create a Receiver Operating Characteristic (ROC) curve (referred in Fig. 11). The ROC curve [39] depicts the trade-off between sensitivity and specificity, offering a visual representation of our proposed framework's discrimination ability. The results reveal a promising performance for the UGSS method, as evidenced by an AUC value of 0.93. This suggests that the proposed method is effective at detecting hemorrhages as well, compared to random label assignment.

We assessed the accuracy of estimated hemorrhage volume $(\widehat{V})$ compared to true hemorrhage volume (V) through the utilization of a Bland-Altman plot and regression analysis. This analysis provides valuable insights into the agreement between two measurements, and the overall performance of the estimation technique. Fig. 12 presents the Bland-Altman plot for our proposed method, where the y-axis represents the differences between estimated and true hemorrhage volumes, and the x-axis denotes the average of the two measurements. The central horizontal line in the plot represents the mean difference between those two measurements. The upper and lower horizontal lines (red dashed lines) delineate the limits of agreement, which are expected to encompass 95 % of data points. The plot showed that, on average, the estimated hemorrhage volumes were in good agreement with the true hemorrhage volumes. The center line of the plot, representing the mean difference between the two methods, was close to zero. This suggests that, on average, the estimation method did not exhibit a significant bias in its measurements. To further explore the relationship between estimated and true hemorrhage volumes, a linear regression analysis was conducted. The R^2 value of 0.837 from the regression analysis indicates that approximately 83.7 % of the variability in the estimated hemorrhage volumes can be explained by the variability in the true hemorrhage volumes. In other words, there is a strong linear relationship between the estimated and true volumes, and this relationship accounts for a significant proportion of the total variability observed in the data. This high value suggests that the estimation method is reliable and is capturing a substantial portion of the true hemorrhage volume variation.

The Table 5 summarizes the experimental results of 3D volume quantification, specifically focusing on the Goodness-of-Fit (R^2) and Average Volume Difference measured in cubic centimeters (cm^3). Among the other architectures, supervised learning model ResUNet shows the lowest performance, with a Goodness-of-Fit of 0.698 and an Average Volume Difference of 6.85 cm^3. In contrast, the semi-supervised models exhibit better 3D hemorrhage volume quantification performance than the supervised model. The improved performance trend of semi-supervised paradigms over the supervised was also followed while comparing the 3D experimental performance analysis. The proposed semi-supervised UGSS method demonstrates the best performance compared to the others model, achieving the highest Goodness-of-Fit ($R^2 = 0.837$) and the lowest Average Volume Difference (2.03 cm^3).

Table 6 presents a comparative analysis of the computational complexities associated with various model architectures utilized in both supervised and semi-supervised learning paradigms. The semi-supervised models ResUNet-CutMix, ResUNet-CutPaste exhibit

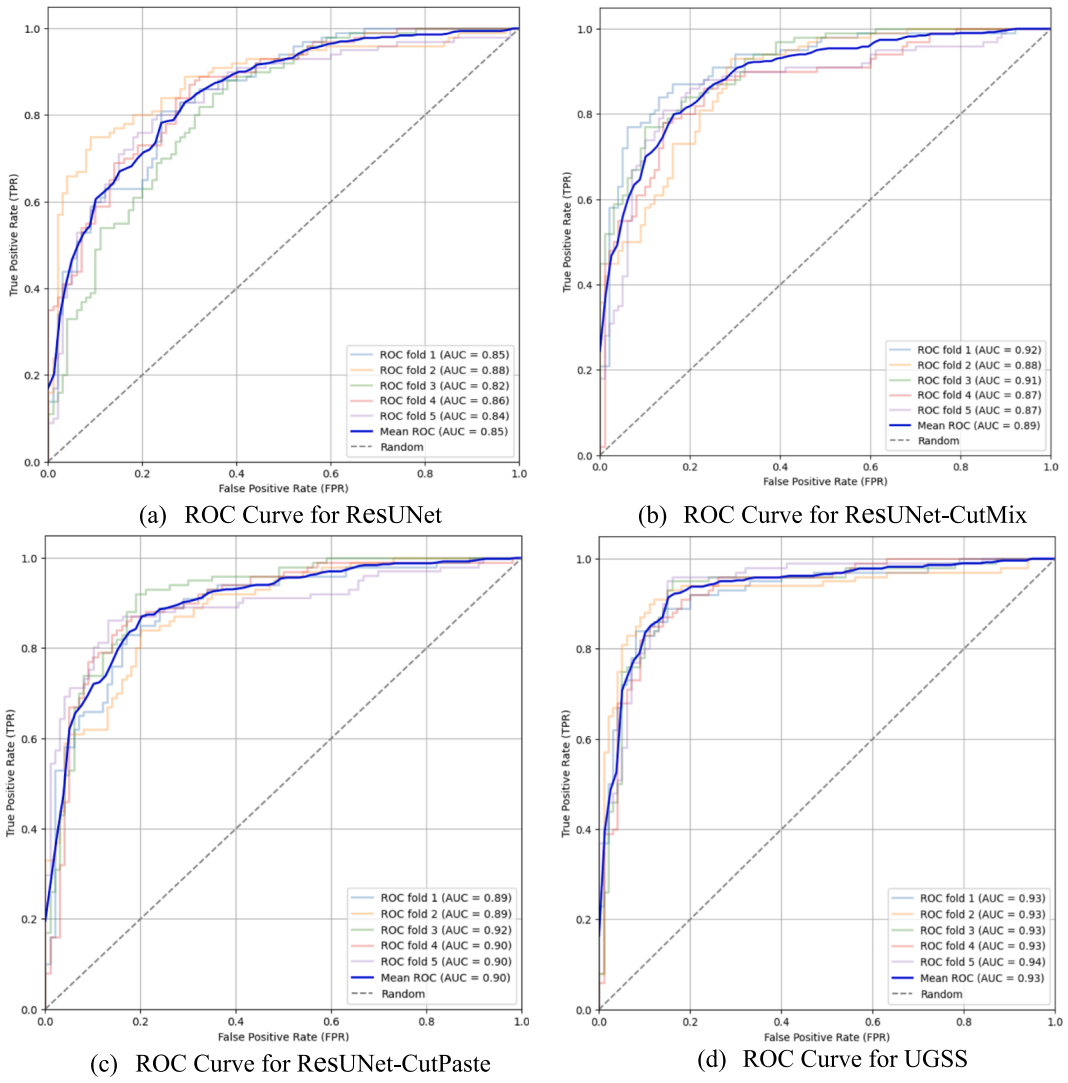


Fig. 11. Comparison of Receiver Operating Characteristic (ROC) plots across various models. The mean ROC value utilizes 5-fold cross validation results.

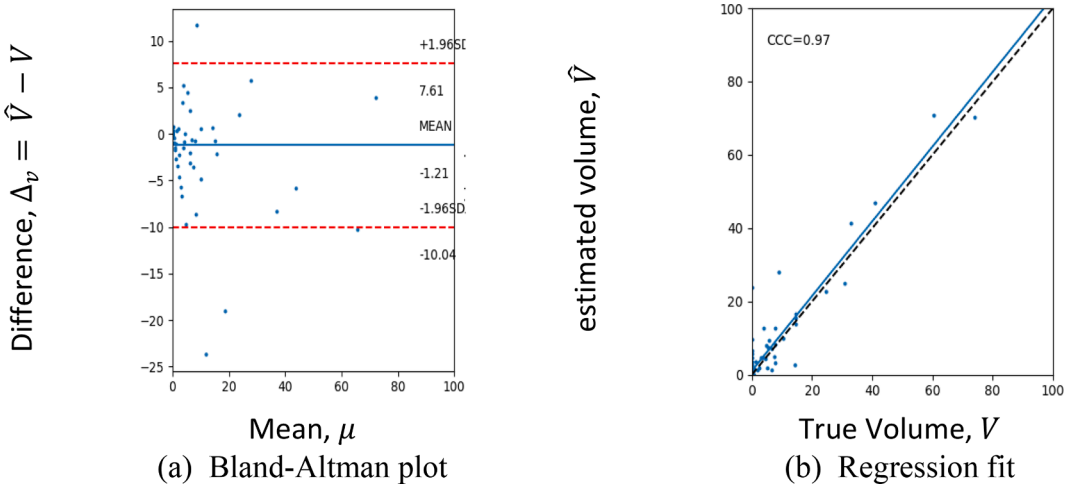


Fig. 12. Comparison of estimated and true intracranial hemorrhage volume (cc). The left panel represents the Bland-Altman plot, and the right panel shows the linear regression fit between estimated hemorrhage volumes (y-axis) and true hemorrhage volumes (x-axis).

Table 5
Comparative performance of 3D volume quantification across different paradigms.

Paradigm	Method	Goodness-of-Fit (R^2)	Average Volume Difference (c m^3)
sup	ResUNet	0.698	6.85
semi-sup	ResUNet-CutMix	0.765	4.04
semi-sup	ResUNet-CutPaste	0.781	4.16
semi-sup [proposed]	UGSS	0.837	2.03

quadratic training complexity of $O(n^2)$, indicating a significant increase of training time as the dataset size expands. The proposed UGSS model improves upon this by achieving a training time complexity of $O(n\log n)$, which balances performance and scalability. The supervised model ResUNet operate a liner time complexity of $O(n)$ during the training period. While the supervised model exhibits efficient computational advantages, its segmentation performance is not superior to that of the proposed UGSS model.

In the aspect of memory utilization, the supervised ResUNet model is categorized as low memory consumption, making it suitable for environments with limited resources. The semi-supervised ResUNet-CutMix and ResUNet-CutPaste models require moderate and high memory, respectively. Because those methods operate data augmentation and repeated training procedure with the data manipulation which requires additional memory and time. The proposed UGSS model strikes a balance with logarithm complexity $O(n\log n)$, classified as having moderate memory requirements. Along with the moderate time and memory requirements, the proposed UGSS can provide competitive detection and segmentation performance in data scarcity situations.

We additionally conducted a comparative analysis of several advanced automatic segmentation techniques applied to brain hemorrhage cases, highlighting their performance in terms of key evaluation metrics and additional features. The primary evaluation metrics considered are the Jaccard Index and the Dice Coefficient, which are standard measures for assessing segmentation accuracy. In addition to these metrics, we have also evaluated whether the methods incorporate 3D visualization and volume quantification features, which are crucial for comprehensive analysis in clinical applications.

Table 7 summarizes the performance of numerous automatic segmentation methods for brain hemorrhage. The table includes results from various recent methods such as UNet-based, GAN-based, Transformer-based, YOLOVx architecture for the brain hemorrhage use case. For better justification, we compared the proposed UGSS model with two different model settings: UNet series and non-UNet series. For UNet series architecture, DCAU-Net and D-UNet methods provide reasonable accuracy in segmentation tasks, however there is no inclusion of hemorrhage volumetric analysis for one of the methods. For the clinical utility, exclusion of volumetric analysis of hemorrhages would have a crucial effect. which could be crucial in clinical utility. The proposed method provides a feature of volume quantification, an aspect not addressed by certain other advanced methods. The UGSS framework is evaluated on the PhysioNet dataset and achieves a reasonable Jaccard Index (0.441) and Dice Coefficient (0.613) compared to previous works on the same dataset. In comparison with non-UNet series architecture, transformed-based and YOLOVx models show promising performance. However, some of the settings still lack the inclusion of volumetrics analysis and 3D visualization. This comparative analysis highlights the advantages of the proposed UGSS method in terms of both segmentation accuracy and the inclusion of advanced features like volume quantification and 3D visualization, which are essential for a holistic evaluation of brain hemorrhage.

4.2. Qualitative analysis

The results of various methods are also visualized for qualitative analysis, shown in Fig. 13. The rows represent four random samples, with the first column displaying the original CT slice and the rest of the columns showcasing the ground truth mask and hemorrhage prediction from different paradigms. Comparative analysis serves as a useful means of assessing the strengths and weaknesses of each approach. It can be observed that the proposed approach performs effectively for different hemorrhages. In addition, uncertainty awareness enhances the ability of the proposed scheme to select the optimal output among all the inferences. This consistency highlights the robustness and effectiveness of our approach in handling varying hemorrhage patterns and sizes. Fig. 13 (a) and (b) exemplify instances where the method demonstrates proficient segmentation, accurately delineating hemorrhage regions with a high degree of precision. Fig. 13 (c) illustrates a more complicated case of complex hemorrhage where the discrepancies between the 'ground truth' and the predicted segmentation. However, the proposed method efficiently handles this challenge by generating appropriate prediction (no presence of hemorrhage).

By rendering the hemorrhages in a three-dimensional space, we construct a 3D volumetric view (shown in Fig. 14) which can be utilized to visualize brain hemorrhage size, shape, and location. The hemorrhage

Table 6
Comparative analysis of algorithmic complexities on different paradigms.

Paradigm	Method	Time Complexity (Train)	Time Complexity (Inference)	Space Complexity
sup	ResUNet	$O(n)$	$O(n)$	Low
semi-sup	ResUNet-CutMix	$O(n^2)$	$O(n)$	Moderate
semi-sup	ResUNet-CutPaste	$O(n^2)$	$O(n)$	High
semi-sup[proposed]	UGSS	$O(n\log n)$	$O(n)$	Moderate

*n refers to the number of training samples

Table 7
Comparative analysis of advanced automatic segmentation methods.

Paradigm	Method	Dataset	Jaccard Index	Dice Coeficient	Volume Quantification
UNet Series	(UNet) Hssayeni et al. [5]	PhysioNet	0.18	0.31	–
	(D-UNet) Kuang et al. [40]	NCCT	0.48	0.65	✓
	(DCAU-Net) Yuan et al. [41]	MICCAI Adam	–	0.745	–
	UGSS [proposed]	**PhysioNet**	**0.441**	**0.613**	✓
Non-UNet Series	(CycleGAN) Ganeshkumar et al. [42]	PhysioNet	0.22	0.32	–
	(3D EIS-Net) Hulin et al. [43]	NCCT	–	0.448	✓
	(YOLOV5 + TransDeeplab) Jagath et al. [44]	CQ500	–	0.76	–
	(Residual Transformer) Zhixiang et al. [45]	NCCT	–	0.586	✓

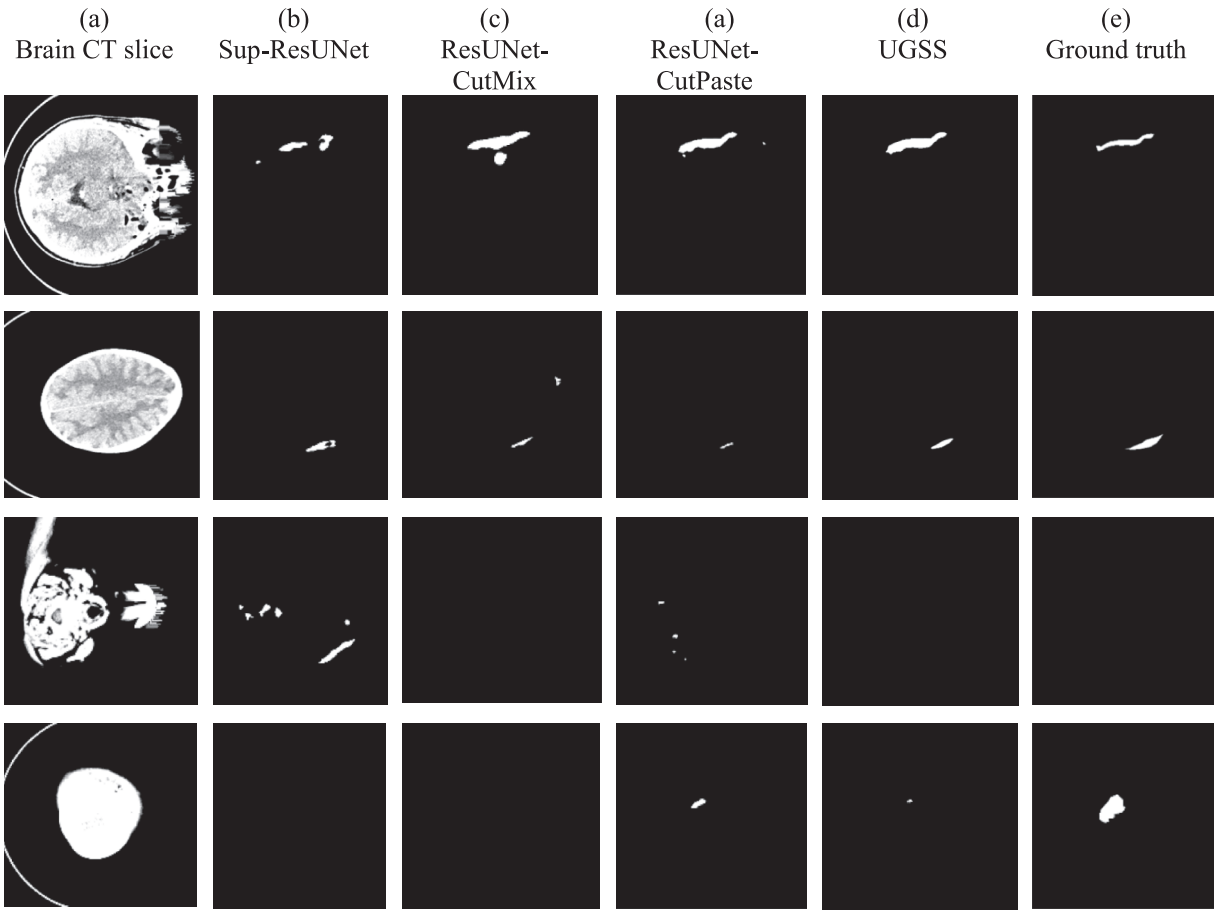


Fig. 13. Examples showing the segmentation outcomes from different methods. Each row represents a different sample, the column (a) represents the CT slice, (b) to (d) shows the segmentation results using Sup-ResUNet, ResUNet-CutMix, ResUNet-CutPaste, and UGSS methods, respectively, and column (e) shows the corresponding ground truth mask.

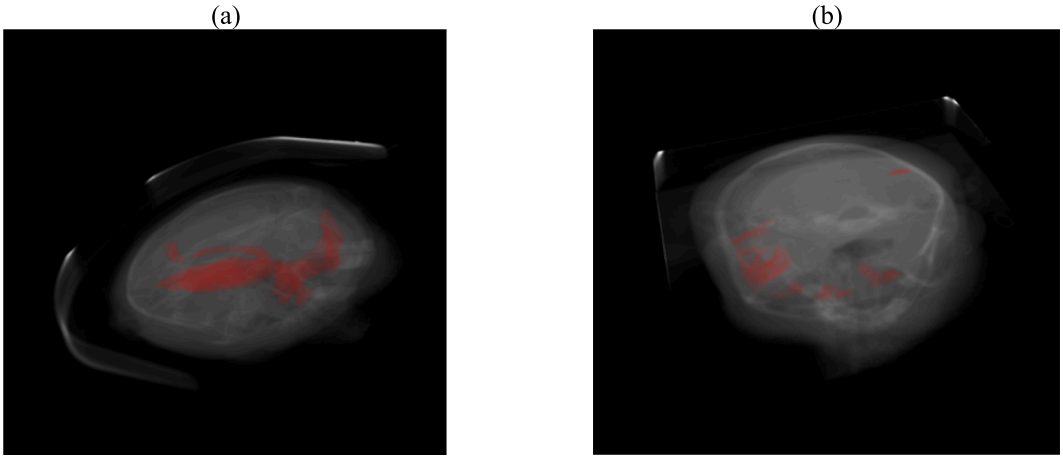


Fig. 14. 3D volumetric view of intracranial hemorrhage (ICH). The left and right panel display the same brain object from different orientations. The reddish masks represent the hemorrhage regions inside brain.

regions within the brain are visually highlighted in a distinctive reddish hue, allowing for clear differentiation from the surrounding anatomical structures. Incorporating a 3D volumetric view into the framework further elevates its diagnostic and visualization capabilities. It offers a comprehensive perspective that aids in treatment planning and surgical procedure, helping practitioners to identify the optimal approach and minimize damage to healthy brain tissue. Additionally, this visualization technique supports medical education and interdisciplinary communication, as it provides a clear and intuitive representation of the pathology.

4.3. Ablation study

The Ablation studies were conducted on the PhysioNet dataset. The

Table 8
Comparative analysis of weight update methods on model performance metrics.

Paradigm	Method	Weight Update	Metrics		
			JCD	*DSC*	PA
Semi-Supervised	UGSS	Without EMA	0.350	0.482	0.824
		EMA	0.403	0.596	0.880
		wEMA	0.441	0.613	0.952

Table 9
Comparative analysis of pseudo-label confidence on model performance metrics.

Paradigm	Method	Pseudo-label	Metrics		
			JCD	*DSC*	PA
Semi-Supervised	UGSS		0.374	0.489	0.798
		✓	0.441	0.613	0.952

different experimental settings were designed to thoroughly evaluate the effectiveness of the proposed model's components as well as the analysis of key process parameters.

4.3.1. Analysis of model components

To investigate the effectiveness of each component of the proposed architecture, we conducted a comprehensive analysis with a series of experiments in different model settings. First, we examine the influence of different weight update methods on the proposed model's performance. We assessed three variations: (i) The model without any exponential moving average (EMA), (ii) The model with a conventional EMA update, and (iii) The model with a weighted EMA (wEMA) update mechanism.

As seen in the Table 8, each weight update method progressively boosts the performance across all the evaluation metrics. The model with exponential moving average (EMA) performs better than without inclusion of weight update techniques. The standard exponential moving average (EMA) technique significantly boosts the performance across all metrics, with notable improvements in DSC (+23.6 %), JCD (+15.1 %), and PA (+6.7 %). It justifies the primary inclusion of weight update mechanisms for the proposed framework. Furthermore, we explore the inclusion of a more dynamic weight update method named weighted exponential moving average (wEMA). Incorporating the weighted exponential moving average (wEMA) improves the performance across all metrics, reinforcing the efficacy of the proposed UGSS method.

The Table 9 presents the comparative performance of the Uncertainty-Guided Semi-Supervised (UGSS) framework with and without the use of pseudo-label confidence. The inclusion of pseudo-label confidence significantly improves the model's performance across key metrics, including the Jaccard Index (JCD), Dice Similarity Coefficient (DSC), and Pixel Accuracy (PA). This demonstrates that pseudo-label confidence enables the model to focus more on reliable predictions, taking take of the uncertainty issues.

Table 10 presents a comprehensive analysis of loss functions, uncertainty guidance, and weight update strategies in the context of semi-supervised hemorrhage detection. The study compares different variants of the proposed UGSS model, each employing different loss functions and weight update mechanisms. To justify the choice of loss functions, uncertainty guidance and weight update mechanisms, we compared the performances metrics for variants of model. Notably, the UGSS with the proposed combined components consistently outperformed the others variant under various loss functions and weight update mechanisms. The introduction of weighted exponential moving average (wEMA) and consistency loss in UGSS significantly improved segmentation accuracy. Furthermore, we experiment with different loss functions, including *KL* divergence and Smooth L_1 loss [46], which reveal varying degrees of impact on the results. The approach combined with the *KL* divergence and weighted EMA (wEMA) consistently produces better results, with the highest JCD and DSC values of 0.441 and 0.613, respectively. Moreover, the incorporation of an uncertainty-guided weight update strategy, leveraging Kullback-Leibler (*KL*) divergence, further yields the highest Jaccard index and Dice coefficient values. These findings emphasize the importance of incorporating uncertainty-aware methodologies, which indicates the potential of our proposed "UGSS" architecture.

4.3.2. Effects of key process parameters

To thoroughly assess the impact of different key process parameters on the model's performance, we made experimental settings for parameters uncertainty threshold (δ), Consistency loss weight (λ_u) and wEMA decay factor (α'). Table 11 showcases the effects of varying uncertainty threshold (δ) values on segmentation performance metrics such as Jaccard Index (JCD) and Dice Similarity Coefficient (DSC).

The results indicate that optimal performance is achieved when the uncertainty threshold and Consistency loss weight are set to $\delta = 0.7$, $\lambda_u = 0.6$ resulting in the highest segmentation scores. Additionally, we examined the continuously varying wEMA decay factor and keep track the metrics value. For decay factor α' ≈ 0.95, we ended up with a better metrics performance. Fig. 15 visually demonstrates the necessity of fine-

Table 11
Impact of key process parameters on model performance metrics.

Paradigm	Method	Key Process Parameters	Scale	JCD	DSC
Semi-Supervised	UGSS	Uncertainty threshold (δ)	0.5	0.421	0.598
			0.6	0.384	0.601
			0.7	**0.441**	**0.613**
			0.8	0.426	0.552
			0.9	0.415	0.587
		Consistency loss weight (λ_u)	0.1	0.350	0.551
			0.2	0.354	0.587
			0.3	0.371	0.590
			0.4	0.362	0.603
			0.5	0.396	0.581
			0.6	**0.441**	**0.613**
			0.7	0.402	0.578
			0.8	0.387	0.600
			0.9	0.364	0.591
			1.0	0.409	0.604

Table 10
Ablation analysis of the proposed framework on PhysioNet dataset.

Paradigm	Method	Supervised Loss	Consistency Loss	Uncertainty Guided	Weight Update	Metrics	
						JCD	*DSC*
Semi-Supervised	UGSS	✓	Smooth L_1		EMA	0.350	0.482
			KL divergence			0.389	0.547
			Smooth L_1		wEMA	0.374	0.489
			KL divergence			0.391	0.562
			Smooth L_1	✓	EMA	0.371	0.505
			KL divergence			0.403	0.596
			Smooth L_1	✓	wEMA	0.396	0.549
			KL divergence			0.441	0.613

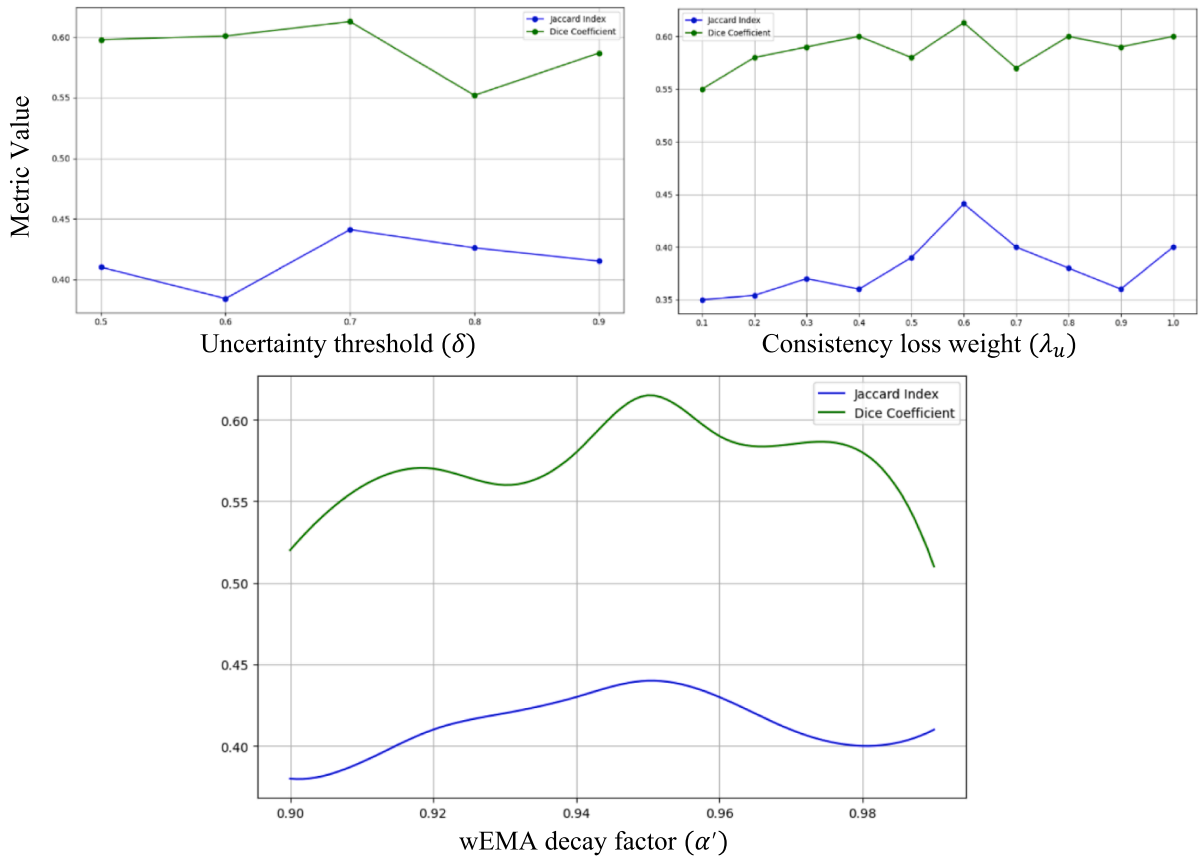


Fig. 15. Effects of key process parameters on model performance.

tuning the key process parameters to ensure robust and accurate brain hemorrhage segmentation in semi-supervised learning. The results highlight the importance of carefully tuning process parameters to maximize the proposed UGSS framework's efficacy.

5. Conclusions

This study proposed an uncertainty aware mean teacher framework for fully automated intracranial hemorrhage segmentation in brain CT scans. We trained the proposed model using semi-supervised learning technique which enables the exploitation of unlabeled data in the supervised and unsupervised phases. To alleviate the label scarcity, we trained the proposed model using semi-supervised learning technique, which can effectively utilize the unlabeled data to improve segmentation performance. To eliminate the possible intricacy caused by the noisy unlabeled data, we employed uncertainty quantification to guide the overall training procedure. The experimental results have demonstrated detection as well as the segmentation capabilities of the proposed method. Additionally, we incorporate hemorrhage volume estimation and 3D volumetric representation of brain hemorrhage, which offers a comprehensive and accurate assessment of intracranial hemorrhages, benefiting both clinical practice and research endeavors. This multifaceted approach holds great promise for improving patient care, advancing medical knowledge, and ultimately mitigating the impact of TBI on individuals and healthcare systems globally. While the proposed framework demonstrates reasonable performance in brain hemorrhage segmentation, there are several limitations worth acknowledging. The framework was developed utilizing only a single imaging modality—CT scans. However, clinical diagnosis and treatment planning often involve multimodal imaging, such as MRI or PET scans, which provide complementary pictorial information. This information is crucial for clinical evaluations and decision making. Therefore, multi-modal imaging integration would add great potential to clinical settings. Another key limitation of the current study is the relatively small sized dataset (PhysioNet) used for training and evaluation. The limited dataset size can impact the model's ability to generalize the wide variety of clinical cases. Expanding the dataset to include more diverse data across different distributions, potentially would improve the model's generalization capabilities. The computational complexity of the proposed framework is substantial, primarily due to the need for multiple base model's forward passes with uncertainty estimation. This can result in significant resource consumption and longer training periods. Future work should explore strategies to reduce computational demands which will increase efficiency in clinical environments. Due to the cost associated with obtaining precise ground annotations, we aim to integrate weakly labeled techniques in our future studies. Moreover, we plan to employ a severity index regarding intracranial hemorrhage, which potentially offers clinicians a more comprehensive understanding the impact of brain hemorrhage, aiding in treatment planning, patient monitoring, and overall decision-making.

CRediT authorship contribution statement

Solayman Hossain Emon: Writing – original draft, Methodology, Formal analysis, Data curation. **Tzu-Liang (Bill) Tseng:** . **Michael Pokojovy:** Writing – review & editing, Investigation, Conceptualization. **Scott Moen:** Writing – review & editing, Validation. **Peter McCaffrey:** Writing – review & editing, Validation. **Eric Walser:** Writing – review & editing, Validation. **Alexander Vo:** Validation. **Md Fashiar Rahman:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially supported by the United States (U.S.) National Science Foundation (ERC-ASPIRE-1941524, IUSE 2216396) and the U.S. Department of Education (FIPSE P116S210004, MSEIP P120A220044, MSEIP P120A240054). The authors wish to express sincere gratitude for their financial support. Additionally, the authors would like to express appreciation for the collaboration with the University of Texas at Medical Branch (UTMB) in Galveston, TX, USA, and acknowledge their valuable support and feedback.

Data availability

Data will be made available on request.

References

[1] T.M. Buzug, Computed tomography," Springer handbook of medical technology, Springer, 2011, pp. 311–342.
[2] L. Rabiner, B. Juang, An introduction to hidden Markov models, IEEE ASSP Mag. 3 (1) (1986) 4–16.
[3] D. A. Reynolds, "Gaussian mixture models," Encyclopedia of Biometrics, vol. 741, no. 659-663, 2009.
[4] J.A. Hartigan, M.A. Wong, Algorithm AS 136: A k-means clustering algorithm, J. R. Stat. Soc. Ser. C (Appl. Stat.) 28 (1) (1979) 100–108.
[5] M.D. Hssayeni, M.S. Croock, A.D. Salman, H.F. Al-khafaji, Z.A. Yahya, B. Ghoraani, Intracranial hemorrhage segmentation using a deep convolutional model, Data 5 (1) (2020) 14.
[6] M.F. Rahman, Y. Zhuang, T.-L.-B. Tseng, M. Pokojovy, P. McCaffrey, E. Walser, S. Moen, A. Vo, Improving lung region segmentation accuracy in chest X-ray images using a two-model deep learning ensemble approach, J. Vis. Commun. Image Represent. 85 (2022) 103521.
[7] M.F. Rahman, T.-L.B. Tseng, M. Pokojovy, W. Qian, B. Totada, H. Xu, An automatic approach to lung region segmentation in chest X-ray images using adapted U-Net architecture. Medical Imaging 2021: Physics of Medical Imaging vol. 11595, SPIE, 2021, pp. 894–901.
[8] Y. Zhuang, M.F. Rahman, Y. Wen, M. Pokojovy, P. McCaffrey, A. Vo, E. Walser, S. Moen, H. Xu, T.-L.-B. Tseng, An interpretable multi-task system for clinically applicable COVID-19 diagnosis using CXR, J. Xray Sci. Technol. 30 (5) (2022) 847–862.
[9] W. Sun, T.-L.-B. Tseng, J. Zhang, W. Qian, Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data, Comput. Med. Imaging Graph. 57 (2017) 4–9.
[10] W. Sun, T.-L.-B. Tseng, J. Zhang, W. Qian, Computerized breast cancer analysis system using three stage semi-supervised learning method, Comput. Methods Programs Biomed. 135 (2016) 77–88.
[11] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, D. Rueckert, Semi-supervised learning for network-based cardiac MR image segmentation, in: Medical Image Computing and Computer-Assisted Intervention– MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20 (pp. 253–260). Springer International Publishing.
[12] J. Kim, K. Ryoo, J. Seo, G. Lee, D. Kim, H. Cho, and S. Kim, "Semi-supervised learning of semantic correspondence with pseudo-labels." pp. 19699-19709.
[13] B. P. Yap, and B. K. Ng, "Cut-Paste Consistency Learning for Semi-Supervised Lesion Segmentation." pp. 6160-6169.
[14] Y. Xie, J. Zhang, Z. Liao, J. Verjans, C. Shen, Y. Xia, Pairwise relation learning for semi-supervised gland segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23 (pp. 417–427). Springer International Publishing.
[15] J.L. Wang, H. Farooq, H. Zhuang, A.K. Ibrahim, Segmentation of intracranial hemorrhage using semi-supervised multi-task attention-based U-net, Appl. Sci. 10 (9) (2020) 3297.
[16] S. Laine, and T. Aila, "Temporal ensembling for semi-supervised learning," arXiv preprint arXiv:1610.02242, 2016.
[17] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, Adv. Neural Inf. Proces. Syst. 30 (2017).
[18] L. Yu, S. Wang, X. Li, C.-W. Fu, P.-A. Heng, Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation, in: Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II 22 (pp. 605–613). Springer International Publishing.
[19] A. Meyer, S. Ghosh, D. Schindele, M. Schostak, S. Stober, C. Hansen, M. Rak, Uncertainty-aware temporal self-learning (UATS): Semi-supervised learning for segmentation of prostate zones and beyond, Artif. Intell. Med. 116 (2021) 102073.
[20] S. Li, C. Zhang, X. He, Shape-aware semi-supervised 3D semantic segmentation for medical images, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23 (pp. 552–561). Springer International Publishing.
[21] X. Luo, J. Chen, T. Song, G. Wang, Semi-supervised medical image segmentation through dual-task consistency, in: Proceedings of the AAAI conference on artificial intelligence, vol. 35, No. 10, pp. 8801–8809.
[22] Y. Zhang, J. Zhang, Dual-task mutual learning for semi-supervised medical image segmentation, in: Pattern Recognition and Computer Vision: 4th Chinese Conference, PRCV 2021, Beijing, China, October 29–November 1, 2021, Proceedings, Part III 4 (pp. 548–559). Springer International Publishing.
[23] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18 (pp. 234–241). Springer International Publishing.
[24] A. F. Agarap, "Deep learning using rectified linear units (relu)," arXiv preprint arXiv:1803.08375, 2018.
[25] J.R. Hershey, P.A. Olsen, Approximating the Kullback Leibler divergence between Gaussian mixture models, in: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, vol. 4, IEEE, pp. IV-317.
[26] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. Inf. Theory 37 (1) (1991) 145–151.
[27] M. Hasan, A. Khosravi, I. Hossain, A. Rahman, and S. Nahavandi, "Controlled Dropout for Uncertainty Estimation," arXiv preprint arXiv:2205.03109, 2022.
[28] Z. Cai, A. Ravichandran, S. Maji, C. Fowlkes, Z. Tu, S. Soatto, Exponential moving average normalization for self-supervised and semi-supervised learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 194–203.
[29] U. Schneider, E. Pedroni, A. Lomax, The calibration of CT Hounsfield units for radiotherapy treatment planning, Phys. Med. Biol. 41 (1) (1996) 111.
[30] H. Lee, M. Kim, and S. Do, "Practical window setting optimization for medical image deep learning," arXiv preprint arXiv:1812.00572, 2018.
[31] K.H. Zou, S.K. Warfield, A. Bharatha, C.M. Tempany, M.R. Kaus, S.J. Haker, W. M. Wells III, F.A. Jolesz, R. Kikinis, Statistical Validation of Image segmentation quality based on a spatial overlap index1: scientific reports, Acad. Radiol. 11 (2) (2004) 178–189.
[32] I.E. Vorontsov, I.V. Kulakovskiy, V.J. Makeev, Jaccard index based similarity measure to compare transcription factor binding site models, Algorithms Mol. Biol. 8 (1) (2013) 1–11.
[33] M.L. McHugh, Interrater reliability: the kappa statistic, Biochemia Medica 22 (3) (2012) 276–282.
[34] D. Giavarina, Understanding bland altman analysis, Biochemia Medica 25 (2) (2015) 141–151.
[35] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, Y. Yoo, Cutmix: Regularization strategy to train strong classifiers with localizable features, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6023–6032.
[36] C.-L. Li, K. Sohn, J. Yoon, T. Pfister, Cutpaste: Self-supervised learning for anomaly detection and localization, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 9664–9674.
[37] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, Pytorch: An imperative style, high-performance deep learning library, Adv. Neural Inf. Proces. Syst. 32 (2019).
[38] I. Loshchilov, and F. Hutter, "Decoupled weight decay regularization," arXiv preprint arXiv:1711.05101, 2017.
[39] S.H. Park, J.M. Goo, C.-H. Jo, Receiver operating characteristic (ROC) curve: practical review for radiologists, Korean J. Radiol. 5 (1) (2004) 11–18.
[40] H. Kuang, B.K. Menon, W. Qiu, Segmenting hemorrhagic and ischemic infarct simultaneously from follow-up non-contrast CT images in patients with acute ischemic stroke, IEEE Access 7 (2019) 39842–39851.
[41] W. Yuan, Y. Peng, Y. Guo, Y. Ren, Q. Xue, DCAU-Net: dense convolutional attention U-Net for segmentation of intracranial aneurysm images, Visual Comput. Ind., Biomed., Art 5 (1) (2022) 9.
[42] M. Ganeshkumar, V. Ravi, V. Sowmya, E. Gopalakrishnan, K. Soman, C. Chakraborty, Identification of intracranial haemorrhage (ICH) using ResNet with data augmentation using CycleGAN and ICH segmentation using SegAN, Multimed. Tools Appl. 81 (25) (2022) 36257–36273.
[43] H. Kuang, B.K. Menon, S.I. Sohn, W. Qiu, EIS-Net: Segmenting early infarct and scoring ASPECTS simultaneously on non-contrast CT of patients with acute ischemic stroke, Med. Image Anal. 70 (2021) 101984.
[44] J.C. Rajapakse, C.H. How, Y.H. Chan, L.C.P. Hao, A. Padhi, V. Adrakatti, I. Khan, T. Lim, Two-stage approach to intracranial hemorrhage segmentation from head CT images, IEEE Access (2024).
[45] Z. Xu, C. Ding, Combining convolutional attention mechanism and residual deformable Transformer for infarct segmentation from CT scans of acute ischemic stroke patients, Front. Neurol. 14 (2023) 1178637.
[46] Q. Wang, Y. Ma, K. Zhao, Y. Tian, A comprehensive survey of loss functions in machine learning, Ann. Data Sci. (2020) 1–26.