Multi-task Supervised Deep Learning Model for Classification and Segmentation of Intraparenchymal and Intraventricular Hemorrhage

Yuliang Xiao
Student ID: 1010073901
Department of Medical Biophysics
University of Toronto
yl.xiao@mail.utoronto.ca

Christian Chun Him Lai
Student ID: 1002168842
Department of Medical Biophysics
University of Toronto
christian.lai@mail.utoronto.ca

Liangbing Charlotte Liang
Student ID: 1002451768
Institute of Medical Science
University of Toronto
charlotte.liang@mail.utoronto.ca

Abstract

Timely and accurate identification and characterization of intracerebral hemorrhage (ICH) subtypes is critical in reducing mortality and disability in acute stroke. Deep learning models have been used to assist with these two tasks but often in isolation. In this work, we developed and evaluated a multi-task deep learning approach for the classification and segmentation of intraparenchymal (IPH) and intraventricular (IVH) hemorrhages using 222 3D CT scans. We assessed whether a multi-task strategy with shared feature representations could outperform conventional single-task pipelines. For classification, our proposed model significantly outperformed a baseline ResNet model, improving accuracy from 63.6% to 77.3%, recall from 63.6% to 81.8%, and area under the ROC curve (AUC) from 0.636 to 0.773. In the segmentation task, when compared against nnU-Net, our multi-task model maintained a comparable Dice Similarity Coefficient (DSC) while substantially reducing Hausdorff Distance (HD) from 45.02 mm to 15.42 mm, indicating more precise boundary delineation. These results demonstrate that combining classification and segmentation objectives within a single network yields more robust performance, providing both reliable hemorrhage detection and anatomically accurate lesion localization. Our findings suggest that incorporating auxiliary tasks enhances both classification efficacy and segmentation precision, potentially offering more reliable diagnostic support for clinicians managing intracranial hemorrhages. The code can be found here: https://github.com/mikami520/MultiHem.

The following members of the team confirm that we have provided feedback for the course to the department via the following link: https://www.surveymonkey.com/r/LPHD7YF

Full names: Charlotte Liang, Yuliang Xiao, Christian Chun Him Lai

Preprint. Under review.

1 Literature Review

1.1 Clinical Background on Hemorrhagic Stroke

Stroke is the second leading cause of death worldwide [1], and more than 10 million new stroke cases are reported every year [2]. It is caused by an interruption of blood supply to the brain, either due to a blockage (ischemic stroke) or a rupture of blood vessels (hemorrhagic stroke). This can cause permanent damage to the affected brain areas, resulting in long-term disabilities such as paralysis, memory loss, and impaired speech [3].

Among stroke cases, hemorrhagic stroke is less common compared to ischemic stroke, comprising approximately 15-20% of cases. However, it carries a higher mortality rate, and survivors often experience more severe disabilities [3]. Treatment approaches differ significantly as well; patients with hemorrhagic stroke typically require more complex interventions such as neurosurgery for hematoma removal, while those with ischemic stroke are often managed with anti-platelets and anti-coagulants [3]. Hemorrhagic stroke can be classified into five subtypes based on location. Intra-parenchymal hemorrhage (IPH) and intraventricular hemorrhage (IVH) occur within brain tissues, while subarachnoid hemorrhage, subdural hemorrhage, and epidural hemorrhage occur outside brain tissues [4]. Of these, IPH and IVH are considered the most severe. IPH results from the rupture of a cerebral vessel and subsequent bleeding into the brain parenchyma. It has the highest mortality rate among intracranial hemorrhage (ICH) subtypes, with a 1-year survival rate of approximately 50% and a 5-year survival rate below 30% [5].

IVH, by comparison, involves bleeding within the ventricular system—a network of fluid-filled spaces responsible for producing cerebrospinal fluid (CSF) that protects the brain. The presence of IVH significantly worsens clinical outcomes of hemorrhagic stroke by increasing intracranial pressure, impairing CSF circulation, and potentially leading to acute hydrocephalus. This condition is associated with higher mortality rates and poorer functional outcomes compared to hemorrhages without ventricular involvement. Survivors frequently experience severe long-term disabilities, including motor deficits, cognitive dysfunction, and reduced independence in daily living.

1.2 Importance of Rapid Radiological Assessments

The rapid and accurate detection of IPH and IVH is crucial for reducing mortality and mitigating long-term complications in affected patients. Both conditions typically present with sudden neurological symptoms such as altered consciousness, headache, seizures, and focal deficits, but their clinical presentations alone are often non-specific [6, 7]. Therefore, precise radiological differentiation is important to inform prognosis and treatment approaches.

In clinical emergency settings, computed tomography (CT) is the most commonly used and first-line imaging modality to confirm the presence, type, location, and extent of bleeding. CT is a computerized X-ray imaging procedure which employs a narrow beam of X-rays that rotates quickly around the patient's body. This generates signals that are processed to create tomographic images. These images can be stacked to produce three-dimensional (3D) volumes, improving visualization and characterization of hemorrhages.

One of the most useful properties in this context is attenuation, a measure of radiodensity expressed in Hounsfield units (HU), which exhibits a linear relationship with hematocrit values (percentage by volume of red cells in blood). Whole blood typically has an attenuation value around 56 HU, while normal white and gray matter attenuate at 30-34 HU and 37-41 HU, respectively [8]. This contrast in attenuation values allows easy and rapid identification of hemorrhage regions and provides quantitative information regarding hematoma volume, shape, and proximity to critical brain structures such as the brain stem or ventricular system. These metrics are crucial for treatment planning, for surgical decision-making, and for prognostication.

1.3 Deep Learning for Hemorrhage Classification and Segmentation

The recent development of deep learning approaches has significantly improved the automatic classification and segmentation of intracranial hemorrhage, though these technologies are not yet readily available to assist with clinical decision-making in emergency settings. Current methods can be broadly categorized into classification approaches (identification of hemorrhage subtypes) and seg-

mentation approaches (outlining the boundaries and providing spatial information of hemorrhagic lesions).

In terms of hemorrhage subtype classification, Ye *et al.* applied a jointed 3-dimensional (3D) Convolutional Neural Network and Recurrent Neural Network (CNN-RNN) for distinguishing the five subtypes of hemorrhage. The 3D CNN was used to extract features from brain CT 3D volumes, while the RNN extracted features from sequential changes in different brain slices across volumes. Their group reported an accuracy exceeding 0.8 for all five subtypes of hemorrhage and required only approximately 30 seconds to perform the classification task for one 3D CT brain volume [9]. In a similar approach, Burduja et al. combined CNN with a bidirectional Long Short-Term Memory (LSTM) network to perform subtype classification and achieved accuracies ranging between 0.8 and 1.0, depending on different subtypes [10].

Another aspect that has drawn significant attention is hemorrhage segmentation. The most commonly employed methods for segmentation tasks are deep learning approaches with U-Net architecture. Using a small dataset containing only 82 CT scans, Hssayeni *et al.* employed a standard U-Net model with 5-fold cross-validation [11]. Dice similarity coefficient (DSC) is a commonly used metric for assessing segmentation performance based on overall overlap between predicted and actual masks. A DSC of 0 indicates no overlap, while a DSC of 1 indicates complete overlap. Hssayeni *et al.* reported DSCs ranging from 0.23 to 0.52 for the five different hemorrhage subtypes, specifically 0.28 for IPH and 0.30 for IVH [11]. To address the problem of class imbalance caused by small hemorrhagic regions against large non-hemorrhagic backgrounds, Patel *et al.* applied a weighted map in their U-Net segmentation for training [12]. Their trained model was subsequently applied to two small datasets of 25 and 50 CT scans, respectively. They reported DSCs of approximately 0.9 solely for IPH, while other hemorrhage subtypes were not included in the study. More recently, Khan *et al.* built multiple models by integrating different pre-trained encoders, such as DenseNet, Inception, and ResNet, into U-Net [13]. They also performed skull-region removal, image contrast adjustment, and inversion to improve the segmentation of hemorrhagic regions. They reported that the DenseNet201-U-Net model achieved the best performance among all tested models when applied to a dataset of 82 CT scans containing a mix of the five different hemorrhage subtypes. The DSC of the DenseNet201-U-Net model was 0.86 in this dataset. Khan *et al.* also demonstrated that their model could provide additional information for 3D visualization and hemorrhage volume measurement with approximately 20% error margin [13].

Despite these advances in both classification and segmentation approaches, multi-task models that can comprehensively perform both hemorrhage subtype classification and segmentation simultaneously remain lacking. Such integrated models would be favorable for more efficient and accurate detection and diagnosis in clinical settings.

2 Project Aim/Hypothesis

We aimed to develop a multi-task deep learning model that simultaneously performs two critical functions: (1) accurately classifying IPH and IVH hemorrhage subtypes with improved accuracy and recall over baseline ResNet models, and (2) precisely segmenting the spatial boundaries of hemorrhagic regions with reduced Hausdorff distance compared to conventional nnU-Net approaches. Using a dataset of 222 3D CT scans, our model employs a shared encoder with attention mechanisms to leverage the complementary information between classification and segmentation tasks, potentially improving clinical decision support for hemorrhage assessment and treatment planning.

3 Methods

3.1 Dataset

In 2019, the Radiology Society of North America (RSNA) held the Intracranial Hemorrhage Detection Challenge, searching for more efficient and informative methodologies in identifying the presence, types and locations of intracranial hemorrhage. The dataset for this challenge contains more than 25000 CT scans contributed by 3 different institutes and data annotation performed by neuroradiologists across the world [9]. More than 1000 teams around the world participated in this challenge and contributed a large variety of methods mainly in detecting the presence of hemorrhage, classifying different types of hemorrhage and segmenting hemorrhagic areas.

Song *et al.* specifically focused on performing voxel-level segmentation of IPH and IVH in a dataset comprised of 222 CT scan volumes selected from the RSNA Intracranial Hemorrhage Detection Challenge dataset [14]. Since this dataset contains both hemorrhage subtype classification labels and segmentation masks, we used the same dataset for training and testing in our multi-task model. The IPH and IVH classification labels provided in the RSNA Intracranial Hemorrhage Detection Challenge dataset were used as ground truth labels for the classification part of our model, and segmentation masks were generated and used as ground truth masks for the segmentation part of our model. The 222 CT image stacks were registered as 3D volumes. 22 of the registered CT volumes were kept as test set, and the remaining 200 were split into 80% training set and 20% validation set.

3.2 Model Network and Losses

For our multi-task model training, registered CT volumes underwent a preprocessing pipeline that included intensity normalization and spatial resampling. We applied isotropic spacing at 0.5mm for both X-axis and Y-axis dimensions, while the Z-axis was resampled at 5mm. The X-axis and Y-axis dimensions were resized to 128×128 pixels, followed by padding the Z-axis to a uniform depth of 128 slices.

These transformed volumes were then fed into a common encoder—a 3D convolutional neural network with 5 layers—for hierarchical feature extraction. Features extracted from different encoder scales were subsequently fused with spatial and channel attention mechanisms applied. These attention mechanisms produced cleaner feature maps by focusing the model's processing capabilities on the hemorrhage regions and on features relevant to hemorrhage subtype classification. The fused feature maps were passed to the classification decoder, which generated hemorrhage subtype predictions and computed the Cross Entropy Loss. For the segmentation component, our model extracted features directly from the encoder without fusion and passed them to the segmentation decoder, which generated predicted hemorrhage masks and computed the Dice Loss. The weighted Cross Entropy Loss from the classification component and the Dice Loss from the segmentation component were combined into a total loss for joint optimization in this multi-task model.

The model was trained using the Adam optimizer with a weight decay of 1×10^{-4}, learning rate of 5×10^{-4}, a batch size of 16, for 300 epochs. We employed a cosine annealing learning rate schedule with warm restarts [15] to improve convergence. To prevent overfitting, we used early stopping based on validation loss with a patience of 50 epochs.

We compared our multi-task model's performance against two baselines: a classification-specific model (ResNet3D) and a segmentation-specific model (3D nnU-Net). The ResNet3D is a modified version of the ResNet-18 architecture adapted to handle 3D spatial features. The nnU-Net, in contrast, is an automatic self-configuring deep learning model that optimizes preprocessing, network architecture, training, and post-processing for unseen segmentation tasks [16].

3.3 Evaluation Metrics

To assess the classification task's performance, we employed standard metrics (accuracy, precision, recall, specificity, F1-score, MCC, Cohen's kappa, and AUC), and compared these metrics between our model and the baseline classification model. We also plotted both confusion matrices and ROC curves for direct visual comparison. Accuracy, the most commonly used metric for assessing classification performance, is calculated simply as the proportion of correct predictions. An accuracy of 0 indicates no correct predictions, while 1 represents perfect predictions. Although accuracy provides a general assessment of classification performance, it does not account for class imbalance in the dataset. If a model achieves high accuracy by correctly predicting the majority class while misclassifying the minority class, relying solely on accuracy can be misleading. Since correct predictions of both IPH and IVH are equally important, we also employed the F1-score, which provides better evaluation on imbalanced datasets. An F1-score of 1 suggests perfect predictions for all classes, while 0 indicates that one or all classes have no correct predictions. Additionally, we plotted Receiver Operating Characteristic (ROC) curves and quantified Areas Under the Curve (AUCs). ROC-AUCs assess model performance based on true detection rates versus false detection rates, with an AUC of 1 representing perfect classification and 0.5 representing random guessing.

We used Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) to assess the performance of the segmentation component of our multi-task model. DSC measures the general overlap between

regions—specifically between the predicted hemorrhage segmentation mask and the ground truth mask. A DSC of 0 indicates no overlap between predicted and ground truth masks, while 1 indicates complete overlap. Complementing this, HD measures the maximum distance between points in the boundaries of the predicted and actual masks. Because HD is sensitive to small misalignments, it is particularly useful in assessing boundary accuracy. A smaller HD value indicates a more precise predicted segmentation mask compared to the ground truth mask.

4 Results

4.1 Comparisons of Classification Performance

The performance evaluation of the baseline ResNet3D model and our multi-task deep learning model on a binary classification task showed that our multi-task model achieved a strong recall of 0.818 and moderate specificity (0.727), with an overall accuracy of 0.773, and outperformed the ResNet baseline on all reported metrics (Table 1).

Figure 1 shows the confusion matrices for the ResNet model and the multi-task model. Darker squares along the diagonal represent correct classifications. It reveals fewer false positives and false negatives in our multi-task model. Figure 2 shows the ROC curves of both models, and likewise, our multi-task model shows a higher AUC, indicating a stronger overall discriminative ability.

Metric	ResNet	Multi-Task	% Gain
Accuracy (↑)	0.636	**0.773**	21.54%
Precision (↑)	0.636	**0.750**	17.92%
Recall (Sensitivity) (↑)	0.636	**0.818**	28.62%
Specificity (↑)	0.636	**0.727**	14.33%
F1-Score (↑)	0.636	**0.780**	22.64%
MCC (↑)	0.273	**0.547**	100.37%
Cohen's Kappa (↑)	0.273	**0.546**	100.00%
AUC (↑)	0.636	**0.773**	21.54%

Table 1: Comparison of Baseline ResNet and Multi-Task Model Classification Metrics with Percent Gain Over ResNet.

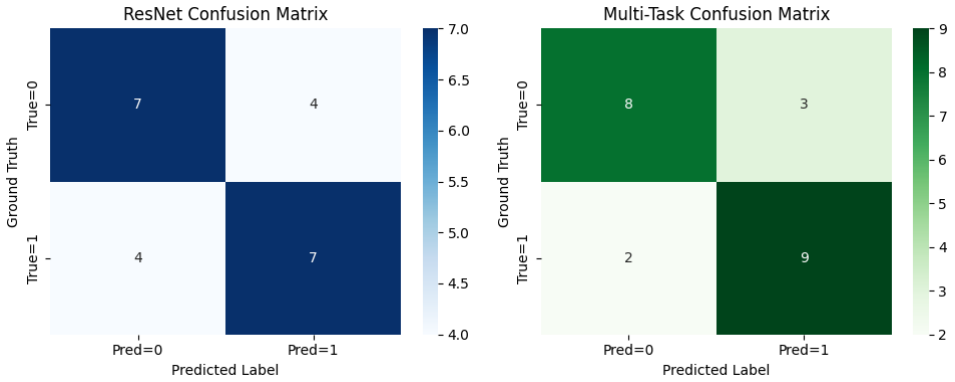


Figure 1: Confusion matrices for ResNet (left) and multi-task (right)

4.2 Comparisons of Segmentation Performance

We compared the segmentation performance of the baseline nnU-Net and the proposed multi-task model across 22 test cases. Performance was evaluated using the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD). Table 2 summarizes the performance comparison between the baseline nnU-Net and the proposed multi-task model. While the baseline model achieved a lower mean DSC (0.779 ± 0.157) compared to the multi-task model (0.822 ± 0.174), this difference was not statistically significant ($t(21) = -1.74$, $p = 0.097$). Only 36.4% of cases demonstrated improved

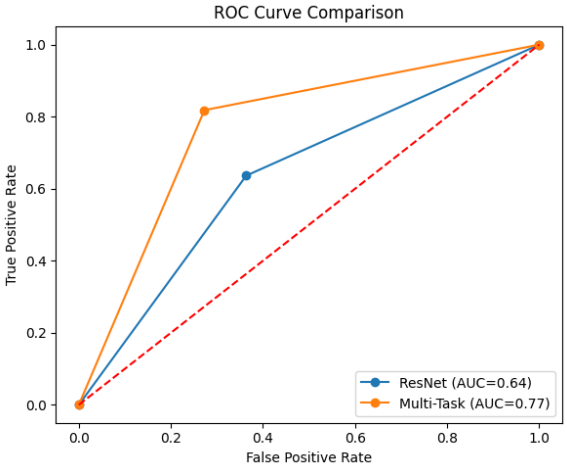


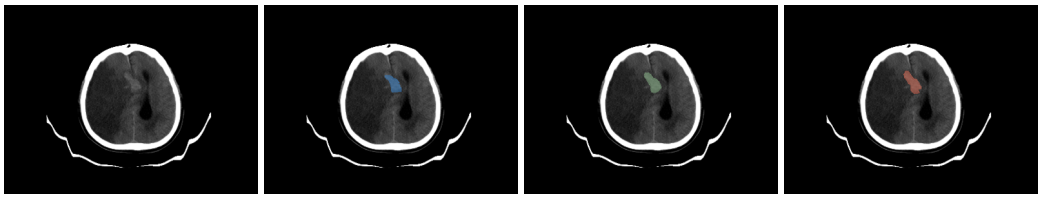
Figure 2: ROC curves for baseline ResNet and multi-task model

DSC with the multi-task model (Table 2). In contrast, the multi-task model significantly outperformed the baseline in HD, reducing the mean HD from 45.02 ± 31.12 mm to 15.42 ± 14.24 mm ($t(21) = -5.43$, $p < 0.001$), with 81.8% of cases showing improvement (Table 2). These results suggest that while the multi-task model may not consistently enhance segmentation overlap (measured by DSC), it substantially reduces boundary error (measured by HD), reflecting more anatomically precise segmentation.

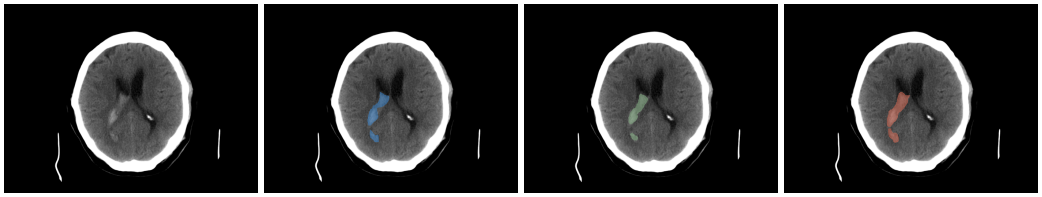
Metric	Baseline	Multi-task
DSC (↑)	0.779 ± 0.157	**0.822 ± 0.174**
HD (mm) (↓)	45.02 ± 31.12	**15.42 ± 14.24**

Table 2: Comparison of model performance (mean ± standard deviation) using DSC and HD.

To illustrate our model's performance in comparison to the ground truth segmentation mask and baseline nnU-Net model, we present two axial slices of IPH and IVH, respectively, in Figure 3. We selected slices where the hemorrhages were identified but with variations in boundary definitions.



(a) Intraparenchymal Hemorrhage (IPH)



(b) Intraventricular Hemorrhage (IVH)

Figure 3: Visual examples of hemorrhage segmentation (a) for IPH and (b) for IVH. In each row, the images from left to right show: (1) the original CT slice, (2) the ground truth overlay (blue), (3) the baseline (nnU-Net) predicted segmentation overlay (green), and (4) the multi-task model predicted segmentation overlay (red).

5 Discussion

This study evaluated a multi-task deep learning model for simultaneous segmentation and classification of IPH and IVH, while comparing its performance against a baseline nnU-Net model. The classification results obtained in this study highlight a clear advantage of the proposed multi-task deep learning model over the baseline ResNet approach. Specifically, the multi-task model achieved higher accuracy (77.3% vs. 63.6%), recall (81.8% vs. 63.6%), and an improved AUC (0.773 vs. 0.636). One plausible explanation for the performance improvement is the enhanced feature representation learned through multi-tasking [17]. By sharing information across additional tasks, such as segmentation, the model likely gains a broader contextual understanding of hemorrhage characteristics. This context may help in distinguishing subtle signal differences between hemorrhagic and non-hemorrhagic tissue, ultimately reducing both false positives and false negatives. Moreover, the higher recall of the multi-task model indicates that fewer true hemorrhage cases are missed, which is critical in clinical practice, because missing a hemorrhage can have serious consequences.

Additionally, while the multi-task model exhibited a similar performance in DSC to the baseline model, it achieved a substantial improvement in HD which indicates reduced boundary error. These findings highlight a trade-off between segmentation overlap and boundary precision when classification gradients are introduced into the learning process in our dual-task model. Our results of low DSC score may be because our model is mistakenly labeling areas outside the hemorrhage. This was discussed in the work of Sharrock *et al.*, who developed a 3D deep neural network (DeepBleed) for segmenting ICH using non-contrast CT [18]. Their model, trained and validated on over 600 scans from phase II/III clinical trials, achieved a high mean DSC of 0.914 but required substantial labeled data and careful design of anatomical context windows to avoid the scenario where the model mistakenly labels other brain structures or areas as hemorrhage.

Our model's advantage in boundary accuracy (indicated by lower HD) suggests that a concurrent classification task may have helped the encoder learn higher-level semantic features, such as the global presence, type, and morphology of hemorrhage, which refined boundary prediction, even if at the cost of pixel-level mask coherence. The classification task plays an important role in shaping the shared feature representations learned by the encoder. It encourages the model to learn features that are informative for both tasks, which may be especially helpful in cases where hemorrhage boundaries are unclear. However, this setup introduces the risk of gradient interference, where the learning signals from the classification and segmentation branches compete during training [19]. This could explain why the segmentation accuracy (as measured by DSC) decreases in some cases, even though the model achieved better boundary delineation. The improvements in the performance of both classification and segmentation may indicate that multi-task learning serves as a "regularizer" and allows both tasks to benefit from each other [17].

Our findings of better performance using multi-task learning in medical imaging align with a recent study conducted by Xu *et al.* [20]. They also reported that combining segmentation with classification tasks in nasopharyngeal carcinoma improved both performance and model interpretability, achieving a DSC of 0.74. Another study by Kordnoori *et al.* [21] also demonstrated a significantly better classification (97% accuracy) in a multi-task learning framework for three primary brain tumor types on MR images. They also demonstrated that a shared encoder with separate branches for each task outperformed single-task baselines. Their findings, combined with our results, further demonstrate the values of multi-task learning in neuroimaging applications. In addition to enhanced performance, a unified model could streamline workflows by providing multiple clinically informative outputs. To achieve this goal, rather than deep learning methods, Liao *et al.* [22] proposed a multi-resolution binary-level set algorithm to segment intracranial hematomas using a coarse-to-fine image representation. Their model effectively captured both global structure and local boundaries without requiring large-scale training data. Compared to these methods, our multi-task model combines the advantages of automatic feature learning from deep models with the flexibility of dual objectives.

The strength of our approach lies in its ability to refine anatomical boundaries, which may be more important in clinical settings, including surgical planning or longitudinal tracking of hemorrhage evolution. However, despite notable improvements in classification performance, caution is warranted when generalizing to larger or more heterogeneous datasets. Another limitation remains in the instability of DSC in some cases, which may be due to competing gradient flows between segmentation and classification branches. Other limitations include reliance on dual-labeled data and

low segmentation consistency across cases. These findings suggest that although multi-task learning can enhance semantic awareness and improve certain performance metrics, careful design and tuning are needed to prevent adverse effects on other metrics.

5.1 Future Directions

To improve upon the current limitations and extend the utility of our model to facilitate faster triage, surgical decision-making and monitoring ICH progression, we propose the following future work:

- **Gradient flow balancing:** Fine-tuning the conflicting tasks between classification and segmentation losses during training—potentially through uncertainty weighting or gradient normalization—may help mitigate performance drops in segmentation metrics like DSC.
- **Multi-label and multi-class expansion:** Incorporating datasets with more granular labels (e.g., IPH+IVH, subdural/epidural hemorrhages, and hemorrhage internal architecture) will improve generalization and enable the model to learn more clinically nuanced features. Future work might also investigate the utility of multi-tasking on substantially larger cohorts or employ additional auxiliary objectives. Such follow-up studies could further clarify how best to harness multi-task learning to reduce misclassifications while retaining clinical interpretability and ease of deployment.
- **Semi-supervised learning:** To address the scarcity of pixel-level annotations, we plan to implement semi-supervised methods such as pseudo-labeling or consistency regularization, leveraging unlabeled CT data for robust segmentation under limited supervision.

6 Conclusion

This study demonstrated the effectiveness of a multi-task deep learning approach for simultaneous classification and segmentation of intraparenchymal (IPH) and intraventricular (IVH) hemorrhages from CT scans. Our proposed model significantly outperformed baseline models in key metrics, showing a 21.54% improvement in classification accuracy (from 63.6% to 77.3%) and a substantial 65.7% reduction in Hausdorff Distance (from 45.02 mm to 15.42 mm), indicating much more precise boundary delineation. These improvements highlight the advantages of employing a shared feature representation across tasks. By leveraging information from both classification and segmentation simultaneously, the model develops a more comprehensive understanding of hemorrhage characteristics. This shared learning appears particularly beneficial for boundary precision, which is crucial for clinical applications such as surgical planning, volume assessment, and monitoring hemorrhage evolution.

While our multi-task model showed moderate improvements in DSC (0.779 to 0.822), the most significant gains were observed in classification recall (28.62% improvement) and boundary precision. This suggests that the approach is particularly effective at reducing false negatives – a critical factor in clinical settings where missing a hemorrhage could have serious consequences for patient care. The results support our hypothesis that a multi-task architecture can enhance both hemorrhage detection and accurate localization. However, we also observed potential gradient interference between tasks that may affect segmentation consistency across cases, indicating that careful balancing of competing objectives is necessary for optimal performance.

In summary, our multi-task model represents a promising step toward more efficient and accurate hemorrhage assessment, demonstrating that joint learning of classification and segmentation tasks can enhance overall performance compared to single-task approaches. The model's improved boundary delineation and higher classification recall make it particularly valuable in clinical settings where both detection accuracy and spatial precision are essential for proper diagnosis and treatment planning of intracranial hemorrhages.

References

[1] Centers for Disease Control and Prevention. Stroke facts: Data and statistics, 2024. Accessed: 2025-04-09.

[2] Liyuan Pu, Li Wang, Ruijie Zhang, Tian Zhao, Yannan Jiang, and Liyuan Han. Projected global trends in ischemic stroke incidence, deaths and disability-adjusted life years from 2020 to 2030. *Stroke*, 54(5):1330–1339, 2023.

[3] Scott D Smith and Clifford J Eskey. Hemorrhagic stroke. *Radiologic Clinics*, 49(1):27–45, 2011.

[4] Manikandan Rajagopal, Suvarna Buradagunta, Meshari Almeshari, Yasser Alzamil, Rajakumar Ramalingam, and Vinayakumar Ravi. An efficient framework to detect intracranial hemorrhage using hybrid deep neural networks. *Brain Sciences*, 13(3):400, 2023.

[5] Laurent Puy, Adrian R Parry-Jones, Else Charlotte Sandset, Dar Dowlatshahi, Wendy Ziai, and Charlotte Cordonnier. Intracerebral haemorrhage. *Nature Reviews Disease Primers*, 9(1):14, 2023.

[6] Adnan I Qureshi, A David Mendelow, and Daniel F Hanley. Intracerebral haemorrhage. *The Lancet*, 373(9675):1632–1644, 2009.

[7] J Claude Hemphill III, Steven M Greenberg, Craig S Anderson, Kyra Becker, Bernard R Bendok, Mary Cushman, Gordon L Fung, Joshua N Goldstein, R Loch Macdonald, Pamela H Mitchell, et al. Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the american heart association/american stroke association. *Stroke*, 46(7):2032–2060, 2015.

[8] Go Shirota, Wataru Gonoi, Masanori Ishida, Hidemi Okuma, Yukako Shintani, Hiroyuki Abe, Yutaka Takazawa, Masako Ikemura, Masashi Fukayama, and Kuni Ohtomo. Brain swelling and loss of gray and white matter differentiation in human postmortem cases by computed tomography. *PLoS One*, 10(11):e0143848, 2015.

[9] Adam E Flanders, Luciano M Prevedello, George Shih, Safwan S Halabi, Jayashree Kalpathy-Cramer, Robyn Ball, John T Mongan, Anouk Stein, Felipe C Kitamura, Matthew P Lungren, et al. Construction of a machine learning dataset through collaboration: the rsna 2019 brain ct hemorrhage challenge. *Radiology: Artificial Intelligence*, 2(3):e190211, 2020.

[10] Jewel Sengupta, Robertas Alzbutas, Przemysław Falkowski-Gilski, and Bożena Falkowska-Gilska. Intracranial hemorrhage detection in 3d computed tomography images using a bi-directional long short-term memory network-based modified genetic algorithm. *Frontiers in neuroscience*, 17:1200630, 2023.

[11] Murtadha D Hssayeni, Muayad S Croock, Aymen D Salman, Hassan Falah Al-Khafaji, Zakaria A Yahya, and Behnaz Ghoraani. Intracranial hemorrhage segmentation using a deep convolutional model. *Data*, 5(1):14, 2020.

[12] Ajay Patel, Floris HBM Schreuder, Catharina JM Klijn, Mathias Prokop, Bram van Ginneken, Henk A Marquering, Yvo BWEM Roos, M Irem Baharoglu, Frederick JA Meijer, and Rashindra Manniesing. Intracerebral haemorrhage segmentation in non-contrast ct. *Scientific reports*, 9(1):17858, 2019.

[13] Muntakim Mahmud Khan, Muhammad EH Chowdhury, ASM Shamsul Arefin, Kanchon Kanti Podder, Md Sakib Abrar Hossain, Abdulrahman Alqahtani, M Murugappan, Amith Khandakar, Adam Mushtak, and Md Nahiduzzaman. A deep learning-based automatic segmentation and 3d visualization technique for intracranial hemorrhage detection using computed tomography images. *Diagnostics*, 13(15):2537, 2023.

[14] Changwei Song, Qing Zhao, Jianqiang Li, Xin Yue, Ruoyun Gao, Zhaoxuan Wang, An Gao, and Guanghui Fu. Hemseg-200: A voxel-annotated dataset for intracerebral hemorrhages segmentation in brain ct scans. In *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3376–3377. IEEE, 2024.

[15] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[16] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[17] Phuoc-Nguyen Bui, Duc-Tai Le, Junghyun Bum, and Hyunseung Choo. Multi-scale feature enhancement in multi-task learning for medical image analysis. *arXiv preprint arXiv:2412.00351*, 2024.

[18] Matthew F Sharrock, W Andrew Mould, Hasan Ali, Meghan Hildreth, Issam A Awad, Daniel F Hanley, and John Muschelli. 3d deep neural network segmentation of intracerebral hemorrhage: development and validation for clinical trials. *Neuroinformatics*, 19:403–415, 2021.

[19] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33:5824–5836, 2020.

[20] Kaifan Yang, Xiuyu Dong, Fan Tang, Feng Ye, Bei Chen, Shujun Liang, Yu Zhang, and Yikai Xu. A transformer-based multi-task deep learning model for simultaneous t-stage identification and segmentation of nasopharyngeal carcinoma. *Frontiers in Oncology*, 14:1377366, 2024.

[21] Shirin Kordnoori, Maliheh Sabeti, Mohammad Hossein Shakoor, and Ehsan Moradi. Deep multi-task learning structure for segmentation and classification of supratentorial brain tumors in mr images. *Interdisciplinary Neurosurgery*, 36:101931, 2024.

[22] Chun-Chih Liao, Furen Xiao, Jau-Min Wong, and I-Jen Chiang. A multiresolution binary level set method and its application to intracranial hematoma segmentation. *Computerized Medical Imaging and Graphics*, 33(6):423–430, 2009.