2024

Automatic Hemorrhage Segmentation In Brain CT Scans Using Curriculum-based Semi-Supervised Learning

Solayman H. Emon
The University of Texas at El Paso

Tzu-Liang (Bill) Tseng
The University of Texas at El Paso

Michael Pokojovy
Old Dominion University, mpokojovy@odu.edu

Peter McCaffrey
University of Texas

Scott Moen
University of Texas

See next page for additional authors

Original Publication Citation

Emon, S. H., Tseng, T.-L., Pokojovy, M., McCaffrey, P., Moen, S., & Rahman, M. F. (2024). Automatic hemorrhage segmentation in brain CT scans using curriculum-based semi-supervised learning. In O. Colliot & J. Mitra (Eds.), *Medical Imaging 2024: Image Processing, Proc. of SPIE 12926* (129262M). SPIE. https://doi.org/10.1117/12.3006596

Automatic Hemorrhage Segmentation in Brain CT Scans Using Curriculum-Based Semi-Supervised Learning

Solayman Hossain Emon[1], Tzu-Liang (Bill) Tseng[2], Michael Pokojovy[3], Peter McCaffrey[4], Scott Moen[4], and Md Fashiar Rahman[2]
[1]Data Science Program, The University of Texas at El Paso, TX 79968, USA
[2]Department of Industrial, Manufacturing and Systems Engineering, The University of Texas at El Paso, TX 79968, USA
[3]Department of Mathematics and Statistics, Old Dominion University, VA 23529, USA
[4]University of Texas Medical Branch, Galveston, TX 77550, USA

ABSTRACT

One of the major neuropathological consequences of traumatic brain injury (TBI) is intracranial hemorrhage (ICH), which requires swift diagnosis to avert perilous outcomes. We present a new automatic hemorrhage segmentation technique via curriculum-based semi-supervised learning. It employs a pre-trained lightweight encoder-decoder framework (MobileNetV2) on labeled and unlabeled data. The model integrates consistency regularization for improved generalization, offering steady predictions from original and augmented versions of unlabeled data. The training procedure employs curriculum learning to progressively train the model at diverse complexity levels. We utilize the PhysioNet dataset to train and evaluate the proposed approach. The performance results surpass those of supervised model with an average Dice coefficient and Jaccard index of 0.573 and 0.428, respectively. Additionally, the method achieves 87.86% accuracy in hemorrhage classification and Cohen's Kappa value of 0.81, indicating substantial agreement with ground truth.

Keywords: Traumatic Brain Injury, Intracranial Hemorrhage, Brain CT Scans, Hemorrhage Segmentation, Semi-Supervised Learning, Curriculum Learning

1. INTRODUCTION

Traumatic brain injury (TBI) is a serious neurological emergency with high morbidity and mortality rates worldwide. In the United States, around 1.7 million people suffer annually from this silent endemic with expected total medical expenses of approximately $76.5 billion [1]. Primary brain injuries may result in temporary or permanent neuropathological deficits such as intracranial hemorrhage (ICH). Timely diagnosis of brain hemorrhage is extremely critical for mitigating possible complications due to delayed diagnosis. Manual segmentation of brain hemorrhages is a time-consuming and labor-intensive task for radiologists and medical professionals. It takes a lot of effort to manually visualize and inspect hemorrhage, especially in clinical emergencies. Automating this process with predictive modeling can save valuable time and resources, while allowing for timely intervention and treatment particularly in geographic areas with limited access to professional medical facilities.

Traditional machine learning approaches heavily rely on handcrafted features and rule-based methods, often struggling to adapt to the inherent complexity and variability of intracranial hemorrhages. Deep-learning-oriented [2,3] algorithms have proven to be highly useful due to their ability to automatically learn and extract intricate patterns and features from complex medical images. In particular, convolutional neural networks (CNNs) have been a paradigm shift towards data-driven methodologies that demonstrate reasonable performance in capturing intrinsic patterns and features associated with brain hemorrhages.

Address all correspondence to:

Md Fashiar Rahman, mrahman13@utep.edu

Medical Imaging 2024: Image Processing, edited by Olivier Colliot, Jhimli Mitra, Proc. of SPIE Vol. 12926, 129262M · © 2024 SPIE
1605-7422 · doi: 10.1117/12.3006596

In the context of brain hemorrhage segmentation, the U-Net [4] architecture has emerged as a widely employed framework that has proven effective in accurate image segmentation. However, training deep learning models for image segmentation typically requires large and statistically diverse labeled datasets. Endowing large datasets with pixel level annotation can be expensive, especially when the task requires expert knowledge and/or manual verification. To address this issue, semi-supervised machine learning emerges as a valuable approach by combining both labeled and unlabeled data. Besides labeled data, a portion of unlabeled data can additionally provide valuable information about the underlying data distribution and statistical relationships between different data points. For instance, Yap et al. [5] proposed a semi-supervised learning approach for lesion segmentation, leveraging cut-paste augmentation and consistency regularization. For semi-supervised gross target volume segmentation, Luo et al. [6] introduced a technique focused on learning from multi-scale consistency between outputs from different scales for semi-supervised gross target volume segmentation. Furthermore, Wang et al. [7] proposed a modified U-Net and curriculum learning strategy, integrated into a multi-task semi-supervised attention-based model for intracranial hemorrhage (ICH) segmentation.

In response to these challenges, the purpose of this work is to develop and present an automatic brain hemorrhage segmentation approach by utilizing curriculum-based semi-supervised learning. We incorporate consistency regularization which helps our model leverage the information contained in unlabeled data effectively, leading to better generalization and performance. Curriculum learning [8] can help the model learn from simpler examples first and then gradually tackle more challenging ones, potentially leading to improved convergence and generalization. To assess the efficacy of the training strategy, we evaluate the developed model with pixel-level segmentation accuracy as well as slice-level classification accuracy.

2. METHOD

The overall methodology is based on the semi-supervised learning paradigm, where both labeled (D_l) and unlabeled (D_u) data are utilized to improve the model performance. An augmented version of unlabeled ($D_{u'}$) data is created for incorporating the consistency regularization. As a base model, we employ an encoder-decoder framework (MobUNet) where the encoder part captures context and reduces the spatial dimensions of the input image, while the decoder part recovers the spatial resolution and generates the segmentation mask. The consistency regularization technique encourages the model to produce consistent predictions for both original and augmented versions of unlabeled data, which enhances model's generalization capability. We computed consistency loss (L_c) by computing the difference between the predictions on unlabeled and augmented version of the unlabeled data. The total loss (L_T) is utilized to train the framework by combining the supervised loss (L_s) and the consistency loss (L_c). Instead of randomly sampling training examples, we use curriculum learning to introduce the training sample to the model in a systematic way. The overall framework is shown in Figure 1.

2.1 Data Preparation

In this study, the data preparation process incorporates both labeled and unlabeled samples. A fraction of the dataset (80%) is used for labeled data which consists of pairs of brain images and their corresponding masks. The remaining portion (20%) of the dataset is used for unlabeled data. We use the PhysioNet [9] dataset which has 82 CT brain scans (46 male, 36 female), 36 of which are diagnosed with intracranial hemorrhage (ICH). For every CT scan, there exists a corresponding mask that contains annotations of a specific region of the hemorrhage. The image values for the raw CT scans are defined over a broad range of Housefield Units (HU) [10]. Hence, we used a windowing [11] technique to pre-process raw CT scans. This approach takes a CT scan as input and applies windowing to adjust the contrast in the range between 0 to 255 and transforms the images to grayscale. The dimension of the resulting images, which were further used for training the model, is 256 × 256.

2.2 Model Architecture

We chose a symmetric encoder-decoder architecture (displayed in Figure 1) named MobUNet. The model architecture is basically a U-Net [4] variant, commonly used for semantic segmentation tasks. The encoder part of the model is based on a pre-trained MobileNetV2 [12], a lightweight convolutional neural network (CNN), to extract features from input images. To minimize the domain shift problem, the encoder MobileNetV2 has been trained on the RSNA dataset [13] before being utilized for the present image segmentation task. This helps the encoder capture more meaningful features specific to hemorrhage segmentation tasks. We trained the decoder part starting with the bottleneck layer's output from the encoder and gradually upsampled the feature maps while concatenating them with corresponding feature maps from the skip connections. This process helps retain high-resolution details lost during downsampling.

Skip connections are established between specific layers of the encoder and corresponding decoder layers. These connections allow information from different scales to be combined during the decoding process, resulting in more accurate segmentation. Residual connections are introduced between the upsampling layers to enable the model to learn residual representations and mitigate the vanishing gradient problem effectively. The usual l_2 regularization is applied to the convolutional layers, which helps prevent overfitting during the training stage.

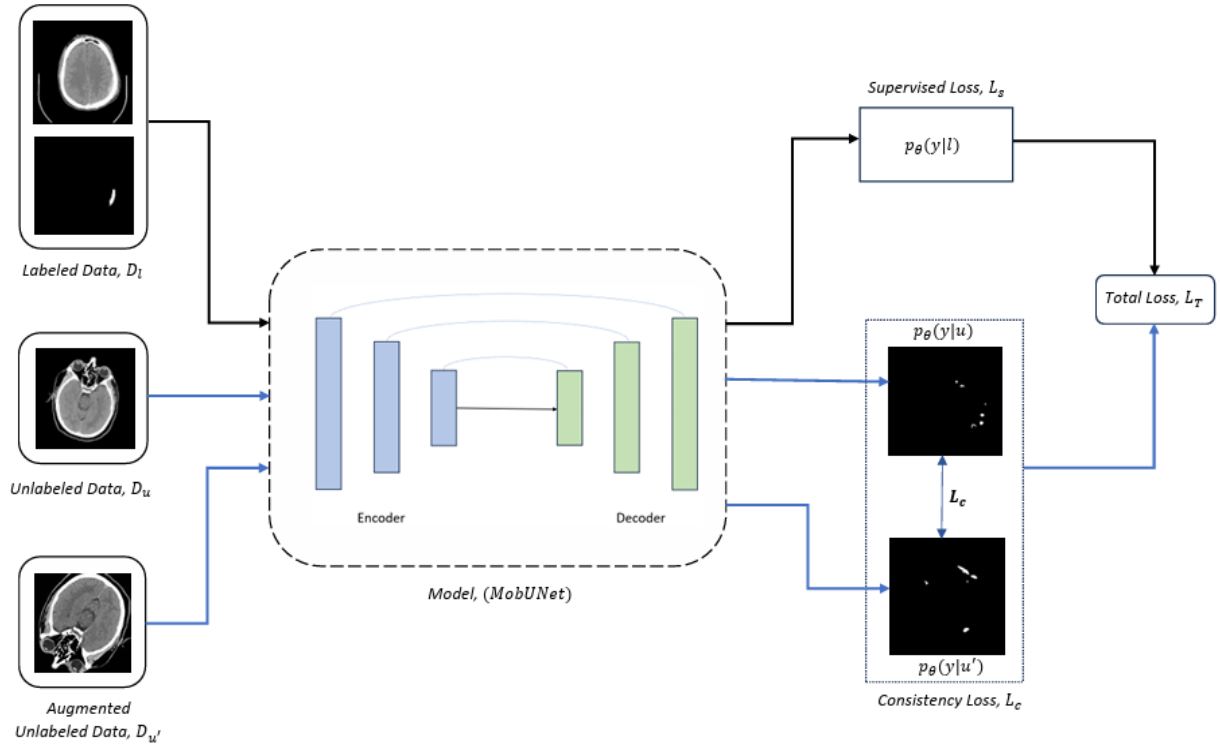

Figure 1. Schematic illustration of the proposed method, where the supervised branch generates segmentation probabilities, while the unsupervised branch computes the consistency utilizing the unlabeled data.

2.3 Consistency Regularization

Consistency regularization is a technique used to leverage unlabeled data by encouraging the model to produce consistent predictions under small perturbations of the input data. The data augmentation is performed using random horizontal flips, rotations, and zooms. We compute the consistency loss as the difference between the predictions on the original and augmented data using the (empirical) mean squared error (MSE):

$$L_c = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{p}_{u_i} - \hat{p}_{u_i'}\right)^2 \tag{1}$$

where $\hat{p}_u$, $\hat{p}_{u'}$ denote the predictions for original and augmented unlabeled data, respectively, and n stands for the number of data points. The expression L_c quantifies the discrepancy between model's predictions on the original and augmented data, encouraging the model to maintain consistency in predictions despite the perturbations introduced.

2.4 Curriculum Learning

We employ curriculum learning to present training examples in a meaningful order, gradually increasing the complexity or difficulty of the examples as the model learns. In the curriculum learning setup, we calculate the difficulty of each image based on its average pixel intensity:

$$\tau = \frac{1}{M \times N}\sum I(i,j) \tag{2}$$

where, *I(i, j)* represents the pixel intensity value at location *(i, j)* in the image, while *M* and *N* stand for the number of rows and columns in the image, respectively.

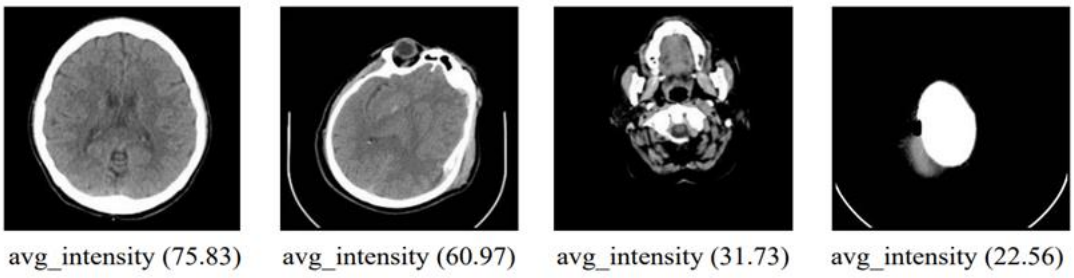

Figure 2. Easier samples to relatively complex samples (left to right) based on the average intensity value

During the training process, we progressively present the model with these curriculum datasets, starting from the easiest subsets and moving toward more challenging ones (depicted in Figure 2). Larger average pixel intensity values imply that the image is generally brighter or has higher overall pixel values so that we consider them an "easier" subset. This progressive learning paradigm prevents the model from becoming overwhelmed by a barrage of difficult cases right from the start.

2.5 Evaluation Metrics

To assess the efficacy of the proposed framework, we employed both Dice coefficient [14] and Jaccard index [15]. These metrics are widely used metrics for quantifying the similarity or overlap between two sets, with the segmentation task specifically compares the predicted binary mask and the ground truth binary mask. We define the Dice coefficient and Jaccard index in Equations (3) and (4), respectively:

$$\text{Dice} = 2 \times \frac{|p \cap \hat{p}|}{|p| + |\hat{p}| + \varepsilon}, \tag{3}$$

$$\text{Jaccard} = \frac{|p \cap \hat{p}|}{|p \cup \hat{p}| + \varepsilon}, \tag{4}$$

where p denotes the ground truth mask and $\hat{p}$ represents the predicted mask. In scenarios where both predicted and ground truth masks are void of positive elements, we introduce a small positive value, ε, to ensure the metrics are always well-defined.

2.6 Model Training

The overall training approach involves optimizing a segmentation model with a combination of supervised and consistency losses using a curriculum learning mechanism. For each batch of labeled data, the supervised loss between the model prediction and the ground truth mask is calculated based on the Dice loss. The model leverages unlabeled data with consistency regularization. The consistency loss is computed based on the mean squared error (MSE) between the model predictions on original and augmented data. The combined loss (L_T), which is the sum of the supervised and consistency losses, is used to compute the gradients during backpropagation and update model weights via the Adam optimizer. We utilized a workstation equipped with Intel® Core i7-12700F, 2100 MHz, 12 Cores; NVIDIA GeForce RTX™ 3050, 4 GB GDDR6 for training the model.

3. RESULTS

We used the 5-fold cross-validation, where three folds were used for training, one for the validation and the remaining hold-out set to test the method. Once trained, the model was evaluated empirically using the Dice coefficient and Jaccard index, which are widely employed for quantitative evaluation in medical imaging. In the fully supervised approach within the 100% labeled data setting (see Table 1), the MobUNet-based method outperforms the UNet-based approach, achieving higher values in terms of both the Jaccard index and the Dice coefficient. Proceeding to semi-supervised methods, we introduce the proportion of unlabeled data to evaluate the potential benefits of using mixed (80% labeled and 20% unlabeled) data.

Table 1. Comparison of Dice coefficients (%) and Jaccard indices (%) amongst the methods on the PhysioNet dataset across different ratios of labeled examples

Method		**Data Proportion**		**Metrics**	
Baseline	Architecture	Labeled Data [%]	Unlabeled Data [%]	Jaccard Index	Dice Coefficient
Supervised	UNet	100%	-	0.275 ± 0.02	0.398 ± 0.04
Supervised	MobUNet	100%	-	0.292 ± 0.01	0.410 ± 0.03
Semi-Supervised	UNet	80%	20%	0.368 ± 0.02	0.496 ± 0.05
Semi-Supervised	MobUNet	80%	20%	0.410 ± 0.04	0.542 ± 0.02
Semi-Supervised **[Proposed]**	MobUNet + Curriculum	80%	20%	**0.428 ± 0.02**	**0.573 ± 0.01**

The results listed in Table 1 indicate that incorporating unlabeled data improves the performance of both models significantly, as reflected in the increased indices values compared to their fully supervised counterparts. The proposed semi-supervised approach combining MobUNet with curriculum-based training attains the highest Jaccard index and Dice coefficient values among all competing methods. Our proposed method achieved an average Dice coefficient and Jaccard index values of 57.3% and 42.8%, respectively, for hemorrhage segmentation. In addition, the proposed approach is able to classify brain hemorrhage with an overall accuracy of 87.86% (see Table 2).

Table 2. Slice-level accuracy on the PhysioNet dataset with 80% labeled data

Baseline	**Architecture**	**Sensitivity**	**Specificity**	**Accuracy [%]**	**Cohen's Kappa**
Semi-supervised **[Proposed]**	MobUNet + Curriculum	0.7186	0.8860	87.86%	0.8106

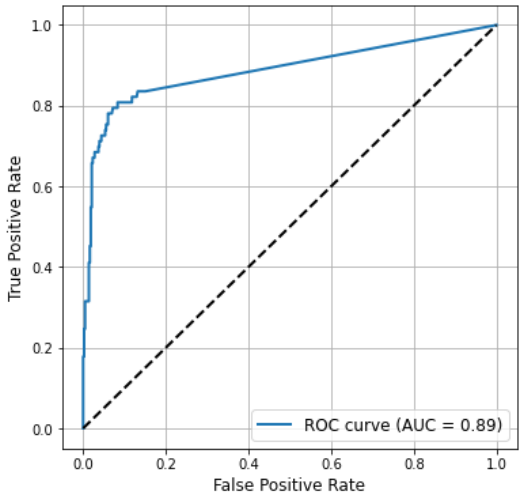

Figure 3. Receiver operating characteristic (ROC) curve

The Cohen's Kappa [16] value of 0.81 indicates that there is a substantial agreement between the model's predictions and ground truth labels. Moreover, the assessment of model concordance is measured through the Area Under Curve, commonly referred to as AUC. The AUC curve provides a holistic representation of the trade-off between true positive rates and false positive rates. The Area Under the Curve (AUC) value of 0.89 suggests that the model has a good discriminatory capacity, efficiently distinguishing between positive and negative instances (displayed in Figure 3). At

the same time, as depicted in Figure 4, there is a reasonable visual agreement between predicted and the ground truth segmentation masks.

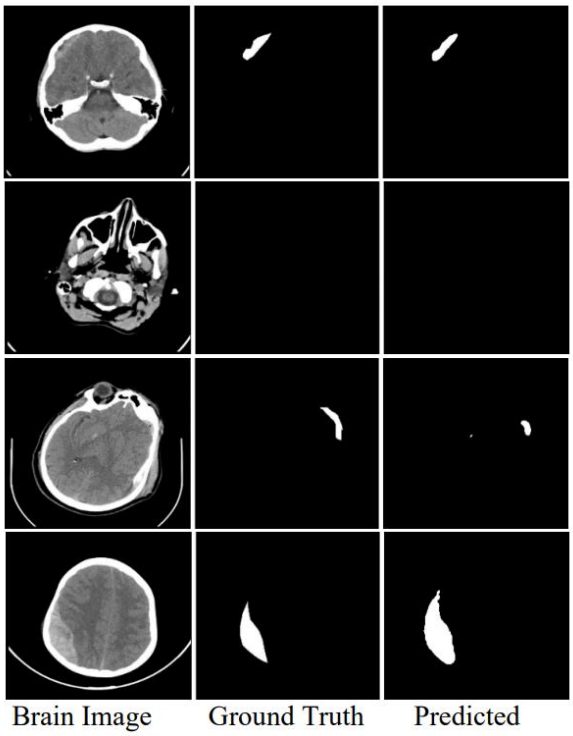

Figure 4. Comparison between predicted masks under our proposed model and ground truth

4. DISCUSSION

Our findings highlight the efficacy of the proposed approach, demonstrating its ability to surpass supervised counterparts in terms of both the Dice coefficient and the Jaccard index. In Figure 4, we present four randomly chosen segmentation outcomes from the proposed model which showcase a spectrum of accuracy levels. Notably, the first case demonstrates high segmentation accuracy, reflecting model's proficiency in precisely delineating object boundaries and capturing detailed features. Furthermore, the second case presents correct identification of negative instances in segmentation. By effectively discerning between positive and negative instances, the proposed model becomes more dependable in distinguishing relevant features. However, it is imperative to acknowledge the cases where the segmentation accuracy was moderate. In the third case, the model exhibited a relatively ambiguous performance, encountering challenges in accurately segmenting certain regions. This underscores the need for further research to enhance model's adaptability to diverse image characteristics. Lastly, the fourth case exhibits reasonable performance in segmenting the hemorrhage area. While the proposed model encountered challenges with segmenting some of the bone regions, somewhat better segmentation accuracy in tissue regions was achieved.

5. CONCLUSION

The proposed methodology adopts the semi-supervised learning paradigm, where both labeled and unlabeled data are utilized to improve model performance. As a base model, we employ an encoder-decoder framework where the encoder part captures context and reduces spatial dimensions of the input image, while the decoder part recovers the spatial resolution and generates a segmentation mask. The consistency regularization technique encourages the model to produce consistent predictions for both original and augmented versions of unlabeled data, which enhances model’s generalization capability. Our findings offer valuable insights into advantages and limitations in the realm of automatic hemorrhage detection. Our future efforts will include more sophisticated deep learning architectures and more comprehensive datasets to improve segmentation accuracy and robustness for clinical applications. Other aspects such as intracranial pressure (ICP) and midline shift (MLS) will also be considered.

ACKNOWLEDGEMENTS

This work was partially supported by the National Science Foundation (IUSE 2216396) and the Department of Education (FIPSE P116S210004, MSEIP P120A220044). 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, and acknowledge their valuable support and feedback.

REFERENCES

[1] A. Gudigar, U. Raghavendra, A. Hegde, G. R. Menon, F. Molinari, E. J. Ciaccio, and U. R. Acharya, "Automated detection and screening of traumatic brain injury (TBI) using computed tomography images: a comprehensive review and future perspectives," *International Journal of Environmental Research and Public Health*, vol. 18, no. 12, pp. 6499, 2021.

[2] S. Ahmmed, P. Podder, M. R. H. Mondal, S. A. Rahman, S. Kannan, M. J. Hasan, A. Rohan, and A. E. Prosvirin, "Enhancing Brain Tumor Classification with Transfer Learning across Multiple Classes: An In-Depth Analysis," *BioMedInformatics*, vol. 3, no. 4, pp. 1124-1144, 2023.

[3] M. F. Rahman, T.-L. B. Tseng, M. Pokojovy, W. Qian, B. Totada, and H. Xu, "An automatic approach to lung region segmentation in chest X-ray images using adapted U-Net architecture." pp. 894-901.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation." pp. 234-241.

[5] B. P. Yap, and B. K. Ng, "Cut-Paste Consistency Learning for Semi-Supervised Lesion Segmentation." pp. 6160-6169.

[6] X. Luo, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, N. Chen, G. Wang, and S. Zhang, "Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency." pp. 318-329.

[7] J. L. Wang, H. Farooq, H. Zhuang, and A. K. Ibrahim, "Segmentation of intracranial hemorrhage using semi-supervised multi-task attention-based U-net," *Applied Sciences*, vol. 10, no. 9, pp. 3297, 2020.

[8] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum Learning." pp. 41-48.

[9] M. D. Hssayeni, M. S. Croock, A. D. Salman, H. F. Al-khafaji, Z. A. Yahya, and B. Ghoraani, "Intracranial hemorrhage segmentation using a deep convolutional model," *Data*, vol. 5, no. 1, pp. 14, 2020.

[10] T. D. DenOtter, and J. Schubert, "Hounsfield Unit," 2019.

[11] H. Lee, M. Kim, and S. Do, "Practical window setting optimization for medical image deep learning," *arXiv preprint arXiv:1812.00572*, 2018.

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks." pp. 4510-4520.

[13] A. E. Flanders, L. M. Prevedello, G. Shih, S. S. Halabi, J. Kalpathy-Cramer, R. Ball, J. T. Mongan, A. Stein, F. C. Kitamura, and M. P. Lungren, "Construction of a machine learning dataset through collaboration: The RSNA 2019 brain CT hemorrhage challenge," *Radiology: Artificial Intelligence*, vol. 2, no. 3, pp. e190211, 2020.

[14] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells III, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index1: Dcientific reports," *Academic Radiology*, vol. 11, no. 2, pp. 178-189, 2004.

[15] I. E. Vorontsov, I. V. Kulakovskiy, and V. J. Makeev, "Jaccard index based similarity measure to compare transcription factor binding site models," *Algorithms for Molecular Biology*, vol. 8, no. 1, pp. 1-11, 2013.

[16] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276-282, 2012.