

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

**Application of the U-Net architecture for
neuroimaging analysis**

Iván Emilio Guerrón Calero

Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero en Ciencias de la Computación

Quito, 20 de diciembre de 2022

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e ingenierías

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

**Application of the U-Net architecture for
neuroimaging analysis**

Iván Emilio Guerrón Calero

Nombre del profesor, Título académico

Noel Pérez, Ph. D

Quito, 20 de diciembre de 2022

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Iván Emilio Guerrón Calero

Código: 00206938

Cédula de identidad: 1723269526

Lugar y fecha: Quito, 20 de diciembre de 2022

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>.

RESUMEN

El ictus cerebral es la segunda causa de muerte en todo el mundo después de las cardiopatías, y uno de los tipos más preocupantes son las hemorragias intracraneales. Este tipo de hemorragia, causada por la rotura de vasos sanguíneos dentro del cerebro, afecta a éste, impide la oxigenación celular y causa daños en los nervios. Aunque los recientes avances en medicina han sido la salvación para muchos pacientes, los médicos siguen estando sujetos a errores humanos a la hora de detectar y segmentar las hemorragias intracraneales debido a las largas jornadas de trabajo. Por este motivo, se han introducido modelos de aprendizaje profundo para ayudar a reducir los errores en este campo médico. En este sentido, propusimos un nuevo método de aprendizaje profundo llamado D-Unet basado en la arquitectura estándar U-net para detectar y segmentar con éxito lesiones de hemorragia intracraneal en un conjunto de datos de imágenes de tomografía computarizada pertenecientes a 82 pacientes. Tanto la D-Unet como la U-net se entrenaron en las mismas condiciones experimentales utilizando un esquema de validación cruzada estratificada de diez pliegues, y las medias obtenidas de las puntuaciones del coeficiente IoU y DICE de 0,72 (0,05 de desviación estándar) y 0,84 (0,03 de desviación estándar) para la D-Unet, y 0,65 (0,04 de desviación estándar) y 0,79 (0,03 de desviación estándar) para la U-net, demostraron que la D-Unet tiende a funcionar mejor que su método de referencia. Además, el modelo mejor D-Unet seleccionado en la fase de entrenamiento se validó en un conjunto de pruebas externo, alcanzando puntuaciones de 0,86 para IoU y 0,89 para DICE. Esta evaluación del rendimiento en el conjunto de datos de prueba confirmó la calidad y la capacidad de generalización del modelo, lo que le permitió detectar y segmentar con éxito HIC de diferentes tipos, formas, tamaños y localizaciones.

Palabras clave: Unet, validación cruzada k-fold, Deep-learning, Hemorragias Intracraneales, Tomografía Computerizada, segmentación.

ABSTRACT

Brain stroke is the second-leading cause of death worldwide after heart disease, and one of the most concerning types are the intracranial hemorrhages. This type of bleeding, caused by ruptures of blood vessels within the brain, affects the brain, prevents cell oxygenation, and causes damage to the nerves. Although recent advances in medicine have been the salvation for many patients, doctors are still subject to human errors when detecting and segmenting intracranial hemorrhages due to long working hours. For this reason, deep-learning models have been introduced to help reduce errors in these medical field. In this regard, we proposed a new deep-learning method called D-Unet based on the standard U-net architecture to successfully detect and segment intracranial hemorrhage lesions on a data set of computerized tomography scan images belonging to 82 patients. Both the D-Unet and U-net were trained under the same experimental conditions using a stratified ten-fold cross-validation schema, and the obtained means of IoU and DICE coefficient scores of 0.72 (0.05 of standard deviation) and 0.84 (0.03 of standard deviation) for the D-Unet, and 0.65 (0.04 of standard deviation) and 0.79 (0.03 of standard deviation) for the U-net, demonstrated that D-Unet tends to perform better than its baseline method. Also, the best-selected D-Unet model in the training stage was validated in an external test set, reaching scores of 0.86 for IoU and 0.89 for DICE. This performance evaluation on the test data set confirmed the model's quality and generalization capacity, making it successful in detecting and segmenting ICH of different types, shapes, sizes, and locations.

Key words: Unet, k-fold cross validation, Deep-learning, Intracranial Hemorrhages, Computerized Tomography, segmentation

TABLE OF CONTENTS

Introduction.....	10
Materials and Methods.....	13
A. Database.....	13
B. Deep learning models	13
C. Proposed method.....	14
D. Experimental setup.....	16
Results and Discussion	18
A. Performance evaluation in the training set.....	18
B. Performance evaluation in the test set	20
Conclusions and Future Work	22
Acknowledgement	23
References.....	24

INDEX OF TABLES

Table 1. Performance results of deep learning models	18
--	----

INDEX OF FIGURES

Figure 1. Samples of images of our dataset from Physionet. The first row contains the original CT images and the second row contains the same CT images with their respective ground truth	13
Figure 2. Workflow of the proposed D-Unet.....	15
Figure 3. Performance of the best chosen model regarding our Dice loss function.....	20
Figure 4. Images from our test set. The first row contains the original CT images, the second row contains the same CT images with their respective ground truth, and the third row contains the images with our predicted masks.....	21

INTRODUCTION

An intracranial hemorrhage (ICH) is a type of bleeding or stroke that affects the brain, causing damage to the nerves, being, in most cases, fatal for the patient. Hemorrhages can result from various medical conditions such as hypertension, ruptured cerebral aneurysm, arteriovenous malformations, hemorrhagic conversion of ischemic infarction, and different types of traumas. Furthermore, intracranial hemorrhages are usually caused by ruptures of blood vessels within the brain, preventing cell oxygenation, increasing the brain pressure, and thus, ending with several brain cells and their functionalities [1]. Thanks to the studies carried out in this area, the symptoms, and effects of ICHs, such as tingling, facial paralysis, headaches, motor difficulties (swallowing or seeing), and even loss of balance and coordination, have become more evident, increasing the efficiency in the detection process [1].

Some conventional ICH detection methods include physical exams, CT scans, and MRIs [1]. Other tests such as blood glucose levels, electrocardiograms, urine analysis, vascular studies, and spinal fluid examinations are usually recommended. In recent years, efforts have been made to improve the quantity and quality of information that a single image can provide, such as adding structural information to the functional data of CT scans [2]. However, the analysis of these images is subject to errors, and many come from the human perception of the image [3]. Thus, the interest in new and more accurate techniques has grown to the point of being the subject of long investigations and collaborations.

Several ICH detection and classification solutions have been proposed based on machine learning and data mining. These have contributed to new possibilities for analyzing MRIs and CT scans, among others, quickly and effectively to provide a second opinion to medical specialists.

Several deep learning models have proven to be effective for various uses, such as SAE (Stacked Auto-Encoders) in [4], DBN (Deep belief Networks) in [5], DBM (Deep Boltzman Machine) in [6], RNN (recurrent neural networks) in [7], among others. Models based on a CNN (convolutional neural networks) architecture are considered structurally more practical to use in image analysis problems since they directly use the images (2D or 3D) as inputs to find relevant features, facilitating the pattern learning [8]. For example, in [9], an advanced CNN architecture was developed integrating 2D and 3D CNN networks, obtaining accuracy results of 85.3%, 80%, and 95.3% in the classification of Alzheimer's, lesions, and healthy brains. Likewise, in [10], an automated data channel and 2D and 3D deep neural networks were used to effectively estimate the volumes of intracranial hemorrhages. They reached a mean Dice coefficient of 0.914 and a volume correlation of 0.979, with a mean volume difference of 1.7ml concerning the ground truth. Also, in [11], CNN-based methods were proposed to classify CT images of brains into normal, ischemic, and hemorrhagic, obtaining 93.33% of accuracy.

Other deep learning models have been developed from the CNN architecture, such as the U-Net presented in [12]. The O-Net presented in [13] obtained an average Dice coefficient of 97.26% and 98.29% for brain segmentation in magnetic resonance images on two public data sets. In addition, the Ψ -Net is designed to suppress unwanted information and segment the areas of hemorrhage more precisely. It was validated on two data sets (spontaneous ICH and traumatic ICH), with a total of 4989 slices (1309 of which showed ICH), obtaining dice coefficient scores of 95% and 89.47%, respectively [14]. Another variant of the U-Net architecture was presented in [15]. It was validated on a data set of 134 CT images (60 for training, 5 for validation, and 69 for testing), achieving sensitivity and specificity scores of 81% and 98% per lesion, respectively.

Despite the developments in deep learning regarding the models, architectures, and training techniques, several factors still threaten overall successful performance, and further research in this field is required. In this context, we propose a new deep learning model derived from the U-Net architecture for the identification and segmentation of intracranial hemorrhages in computed tomography scans. The main contributions behind the objective are related to modifications in the number of filters in each layer and inclusions of dropout layers to seek a high-performance ICH detector. These modifications followed a heuristic search, looking for competitive results based on judgement and knowledge, applying the ISO/IEC TR 24372:2021 standard [16] and the development of this research considered the principles of design engineering [17].

MATERIALS AND METHODS

A. Database

We employed a brain CT images database from Physionet, which was revised and approved by the ethics committee of the Iraqi Ministry of Health. This database includes 82 patient studies in which 32 of them were diagnosed with ICH of different types (Subarachnoid, Intraparenchymal, Intraventricular, Subdural, and Epidural). Each study contains around 30 slices (5 mm slice thickness). For the case of patients with ICH diagnosed, it conserves the original CT scan image and the lesion's mask (ground truth) segmented by two radiologists. Moreover, all the images are compressed in NIfTI (Neuroimaging Informatics Technology Initiative) file format [18].

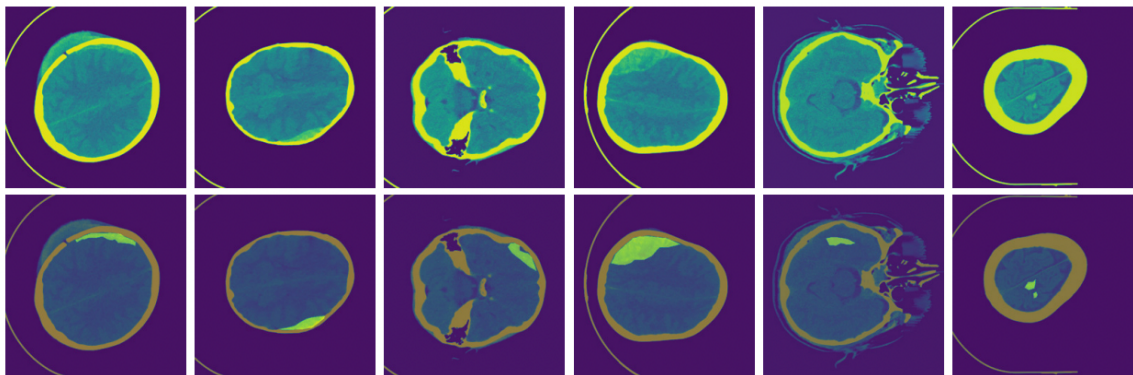


Figure 1. Samples of images of our dataset from Physionet. The first row contains the original CT images, and the second row contains the same CT images with their respective ground truth

B. Deep learning models

Deep learning is a machine learning technique that consists of several neuronal layers which will contribute with higher complexity to the corresponding architectures. The high complexity, in addition to an iterative task (epochs), and usually a large quantity of data, results in progressive learning [19]. Deep learning has led to many discoveries and improvements in different areas such as visual object detection [20], image segmentation [12], speech recognition

[21], image classification [22], genomics [23], time series predictions [24], and medical image analysis [25]. Regarding lesion analysis, the U-net model arises as one of the most convenient methods to use due to its singular architecture, which combines different convolutional operations, assembling the U-shape appearance [12]. It consists of down-sampling (contracting path) and up-sampling (expansive path) operations. The down-sampling path starts when an image is given as input and passes through two sequential 3x3 convolutions of 64 feature channels, followed by a ReLU (rectified linear unit) activation function. The process is then followed by a max pooling operation with a kernel size of 2x2 and stride 2, which will downsample (reduce) the image dimensions. The process then repeats itself doubling the number of feature channels in each layer until it reaches a feature space of 1024. Consequently, the up-sampling operations are carried up using 2x2 upconvolutions, followed by a concatenation of these resulting maps with their corresponding parallel's cropped output from the contracting path. The cropping for each concatenation is required because in every performed convolution, the border pixels are lost. After a concatenation is made, two 3x3 convolutional layers are applied (each including a ReLU activation). This up-sampling process is done repeatedly until the final layer, where a 1x1 kernel size convolutional operation is applied to segment into the different output classes.

C. Proposed method

We have extended the U-net baseline architecture to develop an improved model that can overcome the baseline model in terms of ICH segmentation. The main contributions are related to the inclusion of dropout layers, batch normalization layers, and the reduction of the number of feature channels at each step in the workflow.

D-Unet. This proposed architecture describes a U-net-like model in which the contracting path starts with the given image, passing through two sequential 3x3 convolutions of 32 feature channels, each followed by a batch normalization layer and a ReLU activation function. Then, a max pooling operation with kernel size 2x2 and stride 2 is applied for down-sampling the input image and reducing its dimensions. Moreover, there is a dropout layer with 0.2 probability to introduce random variation in the learning process and avoid overfitting [26]. This process is repeated, and the number of feature channels per layer is doubled until reaching a total of 512 feature channels in the final layer. Then, the expansive path begins with the application of 2x2 transposed convolutions operations, up-sampling the image, contrary to the contracting path, which uses max pooling operations. It then incorporates a concatenation of the resulting maps with their corresponding parallel's cropped output from the contracting path, followed by two 3x3 convolutions, each including a batch normalization layer and a ReLU activation function sequentially. The process is repeated until the final layer is reached, where a 1x1 kernel size convolution operation is applied to output the segmented lesion, as shown in Figure 2.

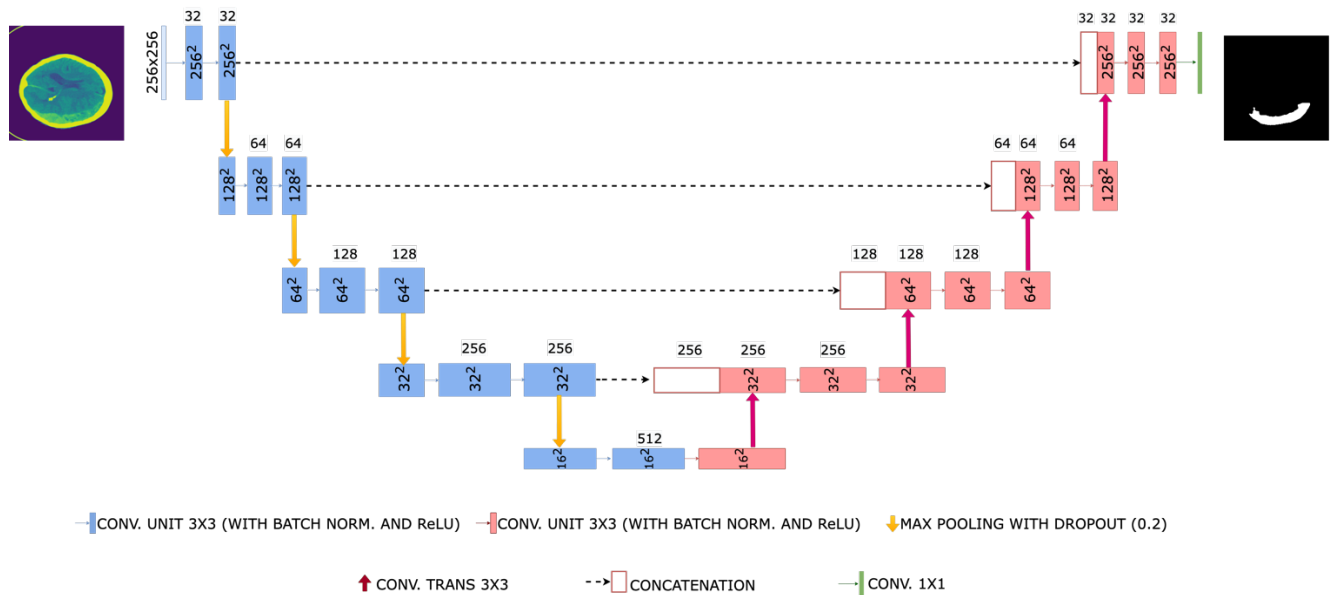


Figure 2. Workflow of the proposed D-Unet

D. Experimental setup

This section describes the experimental methodology employed to train, validate, and test the baseline U-net and the proposed D-Unet models.

1) Data processing and experimental data set creation. All images were transformed from NIFTI to PNG (Portable Network Graphic) file format and filtered to remove noise, such as bones, resulting in 2814 images. Since the employed database is class imbalanced, we created an experimental balanced data set, containing 318 images with and without ICH lesions. Also, each image was resized from 650x650 to 256x256 to decrease the processing time in the training task while conserving the lesions' predicting property. We also applied operations such as rotations, horizontal flips, width and height shifts, cuts and zoom to the images to augment the training set size on the fly. Since these operations were performed such that no more images were saved into the data set, our final created experimental data set consisted of a total of 636 CT images (318 with ICH and 318 with no ICH).

2) Training, validation, and test sets. We selected a random 10% of images of both classes (with and without ICH) from the experimental data set to form the test set. The remaining 90% was considered for feeding a stratified ten-fold cross validation [27] and, thus, to train and validate the proposed method. The use of the stratified cross-validation technique contributes to validating the models over several splits and avoiding overfitting in the training phase.

3) Model configuration. The hyperparameter configuration consisted of a static batch size of 32 and 1000 epochs for both architectures (normal U-net and D-Unet). Furthermore, we used a 10^{-4} learning rate and applied it to an Adam optimizer [28], since it has little memory requisites, is computationally efficient, and is optimal for data with large inputs, outputs, or parameters.

4) Assessment metrics. We computed the mean and standard deviation of Intersection Over Union (IoU) and Dice coefficient metrics over the 10 folds performed on each architecture to assess the proposed method. These metrics are largely used in image segmentation problems and vary from 0 to 1, with 0 being the lowest punctuation in terms of similarity between the real masks (ground truth) and the predicted masks. The IoU metric describes a ratio between the intersection of the predicted mask and the ground truth, and the union of these [29]. That means, it is obtained by dividing the intersection of the real and predicted masks by their union. On the other hand, the Dice coefficient is the result of two times the intersection of the predicted, and real masks divided by the total of pixels of the two masks [30]. Furthermore, we used the Dice loss as the loss function in the training step as in [31] to enhance our model's performance.

5) Selection criterion. We considered checkpoints every 100 epochs for a total of 10 inspected models per architecture (Unet and D-Unet) during the training stage (guarantees avoiding an exhaustive search in the whole learning space) to calculate the mean IoU and Dice coefficients over the 10 folds. The model with the maximum punctuation in both previously mentioned metrics is selected as the best model.

RESULTS AND DISCUSSION

This section presents and discusses the evaluation of two different methods on an experimental data set, containing 618 images with and without ICH lesions in conjunction with a stratified ten-fold cross-validation schema. The obtained results based on the mean of IoU, Dice, and Dice loss metrics can be read in Table I.

Table 1. Performance results of deep learning models

Architecture	Conv. layer (f)	Kernel size	Pool size per layer	Dropout factor	Batch size	Epochs (u)	Train IoU (u)	Train DICE (u)	Validation IoU (STD) (u)	Validation DICE (STD) (u)
U-net	(64, 128, 256, 512, 1024)	(3 × 3)	(2 × 2)	NA	32	100	0.61	0.75	0.45 (0.13)	0.61 (0.12)
						200	0.69	0.81	0.54 (0.15)	0.69 (0.13)
						300	0.73	0.84	0.61 (0.10)	0.75 (0.08)
						400	0.72	0.83	0.57 (0.13)	0.72 (0.11)
						500	0.74	0.85	0.60 (0.12)	0.74 (0.10)
						600	0.77	0.87	0.65 (0.04)	0.79 (0.03)
						700	0.80	0.88	0.63 (0.08)	0.77 (0.07)
						800	0.75	0.84	0.64 (0.08)	0.77 (0.06)
						900	0.80	0.88	0.60 (0.12)	0.74 (0.10)
						1000	0.81	0.89	0.64 (0.07)	0.77 (0.05)
D-UNet	(32, 64, 128, 256, 512)	(3 × 3)	(2 × 2)	0.2	32	100	0.31	0.46	0.30 (0.12)	0.44 (0.14)
						200	0.62	0.75	0.60 (0.11)	0.74 (0.09)
						300	0.68	0.80	0.59 (0.22)	0.71 (0.25)
						400	0.69	0.81	0.66 (0.06)	0.80 (0.04)
						500	0.71	0.82	0.65 (0.10)	0.78 (0.07)
						600	0.71	0.81	0.66 (0.07)	0.79 (0.05)
						700	0.72	0.82	0.71 (0.06)	0.83 (0.04)
						800	0.74	0.83	0.72 (0.05)	0.84 (0.03)
						900	0.75	0.84	0.70 (0.05)	0.82 (0.04)
						1000	0.74	0.83	0.66 (0.07)	0.79 (0.05)

Conv.- convolutional; f- number of filters per layer; u- units; Train IoU, Train DICE, Validation IoU, validation DICE - mean of training and validation metrics DICE and IoU over ten folds; STD - standard deviations the respective metric.

A. Performance evaluation in the training set

From Table 1, we can appreciate that both the Dice and IoU scores obtained by the proposed D-UNet were modest during the first epochs; nonetheless, we can observe these values start climbing and begin to reach certain stability between 500 and 600 epochs, which can be described as normal behavior since deep learning models require several iterations to appropriately adequate their internal parameters. Additionally, we can observe that even though the training values for both metrics are higher for the U-net model than the proposed D-UNet model, the validation scores do not present the same tendency, and even show better results for most

checkpoints for the proposed architecture. In this regard, the best performances in terms of high mean Dice and IoU validation scores and low standard deviations are shown in bold in Table I (one for each architecture). The mean scores show us the capacity of the models to properly segment ICH in comparison to the ground truth, whilst the standard deviations give us an idea of the closeness of all the collected values in each fold to the respective mean scores. The U-net architecture showed its best performance with 600 epochs, obtaining values of 0.65 for IoU and 0.79 for Dice, with 0.04 and 0.03 as their respective standard deviations. In the case of the DUnet architecture, the highest performance was obtained with 800 epochs, where the mean scores (0.72 and 0.84 for IoU and Dice respectively) were maximized with minimal standard deviations (0.05 for IoU and 0.03 for Dice).

Despite Table I shows that the scores of the original U-net grow faster than the proposed method, from the 200 epochs onwards, D-Unet starts showing more accurate predictions and its curve seems to exceed the one of U-net. This conveys that with more training iterations, D-Unet has the tendency to enhance ICH predictions and better the performance of the detector in comparison to its baseline model. In this context and comparing the best models of both architectures previously mentioned, we established that the best overall performance and the best-selected model was D-Unet with 800 epochs (underlined in Table I) since it outperforms the best model obtained from the normal U-net, with differences of 0.07 and 0.05 for IoU and Dice, respectively. This comparison was made based on the validation scores, being the most representative since the model needs to be tested on an unseen data set. Furthermore, these interesting scores, along with the low standard deviations of the best selected model, indicate us high generalization and prediction capacity.

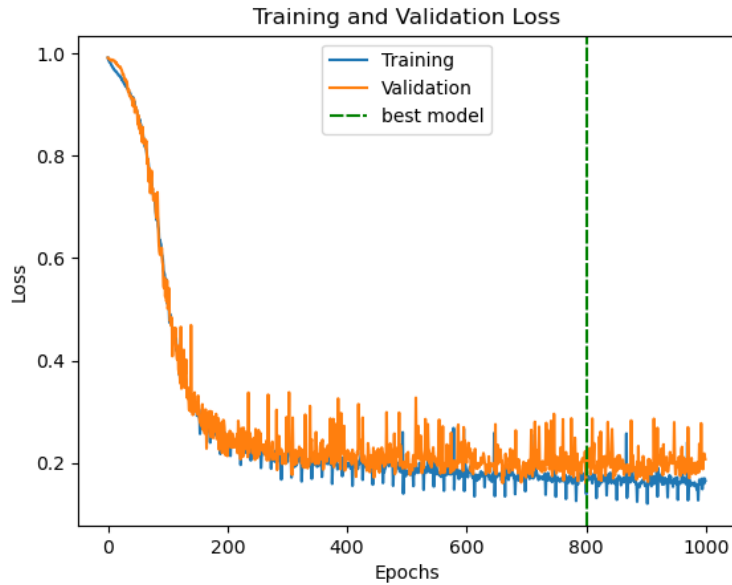


Figure 3. Performance of the best-chosen model regarding our Dice loss function

In Figure 3, we can see the behavior of the best selected model through a training and validation loss graph. This graph indicates us that the training and validation loss values decreased in a similar manner and reached a value of less than 0.2. This similarity of the curves shows us that overfitting was avoided successfully, meaning that the usage of dropout layers and the minimization of the model’s complexity helped in this training process.

B. Performance evaluation in the test set

We used the D-Unet trained with 800 epochs, which was the best model selected in the training step, to test its generalization power on an external test set, containing XXX images with and without ICH. The obtained results of 0.86 and 0.89 for the IoU and Dice coefficient highlighted the promising performance of the proposed method in segmenting ICH lesions on unseen CT scans. An example of successful ICH segmentation cases of the test set is shown in Fig. 4.

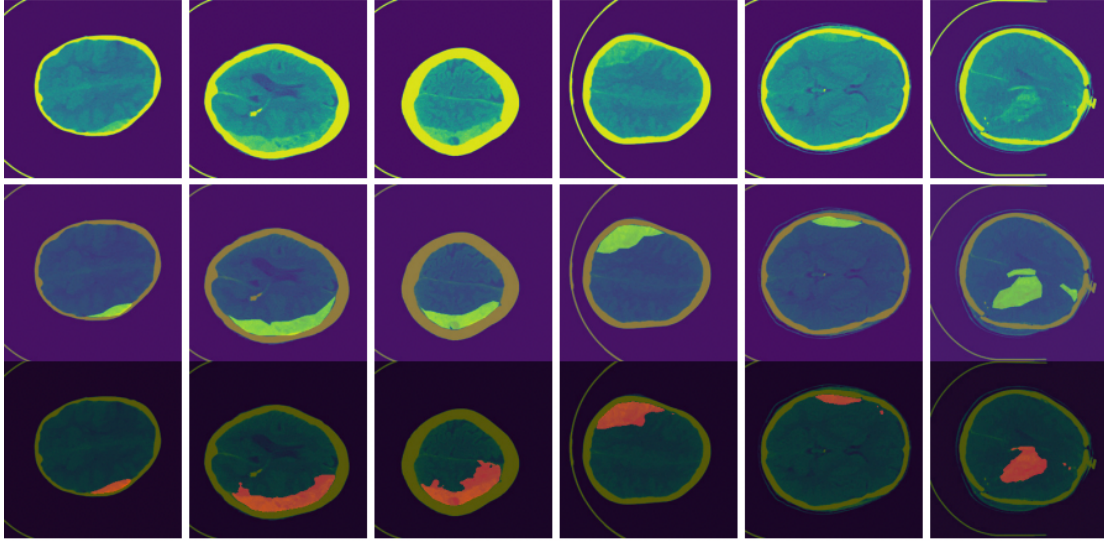


Figure 4. Images from our test set. The first row contains the original CT images, the second row contains the same CT images with their respective ground truth, and the third row contains the images with our predicted masks.

From Figure 4, it is possible to observe that the proposed method approximates the lesion's area close to the ground truth on each sample. Furthermore, it identified and segmented different types of ICHs, independently of their shape, size, and location, such as Epidural Hemorrhages (see Figure 4, columns 1, 4, and 5), Subdural Hemorrhages (Figure 4, columns 2 and 3), and Subarachnoid, Intraparenchymal and Intraventricular hemorrhages (Figure 4, column 6). However, small, and dispersed ICHs are still challenging. That could be explained by the initial resize of the images in the experimental data set. The loss of image information during this process may affect the learning process of the proposed method.

Overall, the obtained results of the proposed method led us to believe that it is competitive in comparison to its baseline model. Despite the limited data set size, the proposed architecture was trained with, the yielded results on the test set suggest proper ICH segmentation and strong generalization capacity.

CONCLUSIONS AND FUTURE WORK

We proposed a new deep-learning method called D-Unet based on the standard U-net architecture to successfully detect and segment ICH lesions on a data set of CT scan images belonging to 82 patients. Both the D-Unet and U-net were trained under the same experimental conditions using a stratified ten-fold cross-validation schema, and the obtained mean and standard deviation of IoU and DICE coefficient scores of 0.72 (0.05) and 0.84 (0.03) for the D-Unet, and 0.65 (0.04) and 0.79 (0.03) for the U-net, demonstrated that D-Unet tends to perform better than its baseline method. Also, the best-selected D-Unet model with 800 epochs in the training stage was validated in an external test set, reaching scores of 0.86 for IoU and 0.89 for DICE. This performance evaluation on the test data set confirmed the model's quality and generalization capacity, making it successful in detecting and segmenting ICH of different types, shapes, sizes, and locations.

In the future, we aim to implement other deep-learning variations of the proposed method, improving the architecture and parameter configurations. Additionally, we will explore alternatives in which the combination of other deep-learning models, such as recurrent-residual convolutional neural networks with U-net could enhance the performance of the proposed method. Finally, we want to search for different and larger databases, including not only CT scan images but also magnetic resonance images, to repeat the entire process in search of more generalizable models.

ACKNOWLEDGMENT

Authors thank to the Applied Signal Processing and Machine Learning Research Group of USFQ for providing the computing infrastructure (NVidia DGX workstation) to implement and execute the developed source code. Publication of this article was funded by the Academic Articles Publication Fund of Universidad San Francisco de Quito USFQ.

REFERENCES

- [1] C. Clinic, “Hemorragia intracraneal, hemorragia cerebral, y apoplej’ia hemorragica,” 2022 [Online]. [Online]. Available: <http://www.clevelandclinic.org/health/shic/html/s14480.asp>
- [2] H. Brody, “Medical imaging,” *Nature*, vol. 502, no. 7473, pp. S81–S81, 2013.
- [3] D. Harpaz, E. Eltzov, R. C. Seet, R. S. Marks, and A. I. Tok, “Point-of-care-testing in acute stroke management: an unmet need ripe for technological harvest,” *Biosensors*, vol. 7, no. 3, p. 30, 2017.
- [4] G. Liu, H. Bao, and B. Han, “A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis,” *Mathematical Problems in Engineering*, vol. 2018, 2018.
- [5] G. E. Hinton, “Deep belief networks,” *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.
- [6] R. Salakhutdinov and H. Larochelle, “Efficient learning of deep boltzmann machines,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings*, 2010, pp. 693–700.
- [7] L. Mou, P. Ghamisi, and X. X. Zhu, “Deep recurrent neural networks for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, 2017.
- [8] Q. D. Buchlak, M. R. Milne, J. Seah, A. Johnson, G. Samarasinghe, B. Hachey, N. Esmaili, A. Tran, J.-C. Leveque, F. Farrokhi et al., “Charting the potential of brain computed tomography deep learning systems,” *Journal of Clinical Neuroscience*, vol. 99, pp. 217–223, 2022.
- [9] X. W. Gao, R. Hui, and Z. Tian, “Classification of ct brain images based on deep learning networks,” *Computer methods and programs in biomedicine*, vol. 138, pp. 49–56, 2017.

- [10] M. F. Sharrock, W. A. Mould, H. Ali, M. Hildreth, I. A. Awad, D. F. Hanley, and J. Muschelli, "3d deep neural network segmentation of intracerebral hemorrhage: development and validation for clinical trials," *Neuroinformatics*, vol. 19, no. 3, pp. 403–415, 2021.
- [11] A. Gautam and B. Raman, "Towards effective classification of brain hemorrhagic and ischemic stroke using cnn," *Biomedical Signal Processing and Control*, vol. 63, p. 102178, 2021.
- [12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [13] S. Jiang, L. Guo, G. Cheng, X. Chen, C. Zhang, and Z. Chen, "Brain extraction from brain mri images based on wasserstein gan and o-net," *IEEE Access*, vol. 9, pp. 136 762–136 774, 2021.
- [14] Z. Kuang, X. Deng, L. Yu, H. Wang, T. Li, and S. Wang, " ψ -net: Focusing on the border areas of intracerebral hemorrhage on ct images," *Computer Methods and Programs in Biomedicine*, vol. 194, p. 105546, 2020.
- [15] A. Majumdar, L. Brattain, B. Telfer, C. Farris, and J. Scalera, "Detecting intracranial hemorrhage with deep learning," in *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2018, pp. 583–587.
- [16] ISO, "Iso/iec tr 24372:2021(en) - international organization for standardization," 2021. [Online]. Available: <https://www.iso.org/obp/ui/#!iso:std:78508:en>
- [17] Y. Haik, S. Sivaloganathan, and T. M. Shahin, *Engineering design process*. Cengage Learning, 2015.

- [18] M. Hssayeni, “Computed tomography images for intracranial hemorrhage detection and segmentation,” 2022 [Online]. [Online]. Available: <https://physionet.org/content/ct-ich/1.3.1/>
- [19] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., “Imagenet large scale visual recognition challenge,” *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [21] K. Noda, Y. Yamaguchi, K. Nakadai, H. G. Okuno, and T. Ogata, “Audio-visual speech recognition using deep learning,” *Applied Intelligence*, vol. 42, no. 4, pp. 722–737, 2015.
- [22] R. Refianti, A. B. Mutiara, and R. P. Priyandini, “Classification of melanoma skin cancer using convolutional neural network,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, 2019.
- [23] W. Kopp, R. Monti, A. Tamburrini, U. Ohler, and A. Akalin, “Deep learning for genomics using janggu,” *Nature communications*, vol. 11, no. 1, pp. 1–7, 2020.
- [24] I. Koprinska, D. Wu, and Z. Wang, “Convolutional neural networks for energy time series forecasting,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.
- [25] M. Saha, R. Mukherjee, and C. Chakraborty, “Computer-aided diagnosis of breast cancer using cytological images: A systematic review,” *Tissue and Cell*, vol. 48, no. 5, pp. 461–474, 2016.

- [26] H. H. Tan and K. H. Lim, "Vanishing gradient mitigation with deep learning neural network optimization," in 2019 7th international conference on smart computing & communications (ICSCC). IEEE, 2019, pp. 1–4.
- [27] M. Stone, "Cross-validators: choice and assessment of statistical predictions," *Journal of the royal statistical society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.
- [28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.
- [29] Hasty, "Intersection over union (iou)," 2022 [Online]. [Online]. Available: <https://hasty.ai/docs/mp-wiki/metrics/iou-intersection-over-union#:~:text=To%20define%20the%20term%2C%20in,matches%20the%20ground%20truth%20data>.
- [30] Y. Varun, "Understanding dice coefficient," 2020 [Online]. [Online]. Available: <https://www.kaggle.com/code/yerramvarun/understanding-dice-coefficient>
- [31] L. A. Erazo, N. Perez, D. Benítez, D. Riofrío, R. F. Moyano, and M. Baldeon-Calisto, "U-net variations for spontaneous intracranial hemorrhages detection on ct images," in 2022 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI), 2022, pp. 1–6.