Detection, Classification, and Segmentation of Intracranial Haemorrhages (January 2021)

Mohammad Al Fahim K (EE17B021)

***Abstract*— This work involves implementing a recently published research paper [1] on the detection and classification of Intracranial Haemorrhages (ICH) from CT scans. The deep learning model is a combination of CNN and LSTM. LSTMs are used to acquire information from neighbouring slices to improve the prediction of the model. We have also implemented Grad-CAM algorithm to produce heatmaps of the subclasses of ICH and to study the accuracy and reliability of the model published in the paper. To improve upon localizing and segmenting ICH regions, we trained a 2-D UNet++ model for semantic segmentation of ICH regions. To further improve segmentation DICE and Jaccard scores, we are currently training a 3-D UNet++ model to achieve better results by integrating inter-slice dependencies.**

***Index Terms*— Convolutional Neural Networks (CNN), Long-Short Term Memory Cells (LSTM), U-net, UNet++, 3-D UNet++, Intracranial Haemorrhage (ICH), Recurrent Neural Netwroks (RNN), Gradient based Class Activation Mapping (Grad-CAM), Fully Connected Network (FCN), Reciever Operating Characteristic curve (ROC), Area under ROC curve (AUC), Computerized Tomography (CT) scan, Jaccard index, DICE score.**

I. Introduction

Intracranial haemorrhages (ICH) account for roughly 10% of strokes in the United States. Although haemorrhagic strokes are less frequent than ischemic strokes (87%), the former ones present a higher mortality rate. Indeed, it seems that between 37% and 38% of haemorrhagic strokes result in death within 30 days. With approximately 795,000 strokes per year only in the United States, the number of yearly death cases caused by ICH is in the range of 30,000. Therefore, ICH is considered one of the most critical health conditions, demanding rapid intervention and intensive post-traumatic healthcare. Rapid intervention requires an urgent diagnosis for this life-threatening condition. Severe headache or loss of consciousness are neurological symptoms often associated with ICH. When a patient shows such symptoms, highly trained radiologists typically analyse CT scans of the patient's brain to find and determine the type of haemorrhage. However, the manual analysis performed by radiologists is complicated and usually time consuming, inherently and undesirably postponing the intervention. In these circumstances, fast and automatic detection, classification, and segmentation of ICHs is of utter importance. The 5 classes of ICH are Epidural Haemorrhage, Intraparenchymal Haemorrhage, Intraventricular Haemorrhage, Subarchanoid Haemorrhage, and Subdural Haemorrhage.

II. Literary Review

Several papers have been published in the domain of detection, classification, and segmentation of ICHs. Yuh et al. [6] employed a threshold based method for detection and segmentation of ICHs and its regions. They reported 98% sensitivity and 59% specificity only in the detection of ICHs. Li et al. [7], [8] used support vector machines for detecting Subarchanoid Haemorrhages and used distance transform features and a Bayesian classifier for segmenting those regions. They reported 100% sensitvity, and 92% specificity only for the detection. Pervedello et al. [9] used CNNs for the detection of ICHs and reported 90% sensitivity, 85% specificity, and Area under ROC curve (AUC) of 0.91. Grewal et al. [10] used CNNs, specifically Densenet, and RNNs for employing inter-slice dependency and reported 88% sensitivity, 81% specificity, and 81% accuracy in the detection of ICHs. They also used auxillary tasks to Densenet for segmentation of ICH regions. Jnawali et al. [11] used an ensemble of CNNs for the detection of ICH and reported 77% sensitivity, 80% precision, and an AUC of 0.87. Chilamkurthy et al. [12] used CNN(ResNet18) and random forests for detection and classification, and reported 92% sensitivity, 70% specificity, and an average AUC of 0.93 over all classes. Arbashirani et al. [13] used 3-D CNNs for detection and reported an AUC of 0.846, 71.5% sensitivity, and 83.5% specificity. Ye et al. [14] used 3-D joint CNN-RNN and reported 98% sensitivity, 99% specificity, and an AUC of 1 in detection. In classifying, they reported 80% sensitivity, 93.2% specificity, and an AUC of 0.93 over all classes. They localized ICH regions by acquiring Attention maps of CNN using Grad-CAM method. Chan [15] trained knowledge based classifier for detection and segmentation and reported 100% sensitivity, and 84.1% specificity in detection. In segemntation, they reported 82.6% sensitivity. Shahangian et al. [16] trained support vector machines for detection and classification, and used distance regularized level set evolution for segmentation. They reported 92.46% accuracy in detection, 94.13% accuracy in classification, and DICE score of 0.58 in segmentation. Chang et al. [17] trained ROI-based CNNs for detection, classification, and segmentation of ICHs and reported 95% sensitivity, 97% specificity, and an AUC of 0.97(except IVH). In segmenting ICH regions, they reported an average DICE score of 0.85. Kuo et al. [18] trained FCNs for detection and segmentation and reported an average precision of 92.8% in detection, and 77.9% averge precision in segmentation. Cho et al. [19] trained a cascade of CNNs and dual FCNs(for segmentation) and reported 97.91% sensitvity, and 98.76% specificity in detection, and accuracies ranging from 70% to 90% in classification. In segmentation they achieved, 80.2% precision and 82.15% recall. Bhadauria et al. [20] used Fuzzy c-means clustering and region based active contour method for segmentation of ICHs and reported a Jaccard index of 0.78, DICE score of 0.87, 79.4% sensitivity, and 99.4% specificity. Nag et al. [21] used Autoencoder and active contour Chan-Vese model for segmentation and reported 71% sensitivity and a Jaccard index of 0.55. Muschelli et al. [22] used logistic regression, logistic regression with LASSO, generalized additive model, and random forest classifier for segmentation and reported a DICE score of 0.89. Kuang et al. [23] trained UNet and multi-region contour evolution for segmentation and reported a DICE score of 0.72. Gautam et al. [24] used fuzzy c-means clustering and wavelet based thresholding technique for segmentation, and reported a DICE score of 0.82. Prakash et al. [25] used distance regularized level set evolution for segmentation and reported an AUC of 0.88, 79.6% sensitivity, and 99.9% specificity

This work was done under the supervision of Prof. Subrahmanyam Gorthi (email: s.gorthi@iittp.ac.in).

Mohammad Al Fahim K is currently a final year electrical engineering undergraduate student at Indian Institute of Technology, Tirupati (email: ee17b021@iittp.ac.in).

III. Detection and Classification of ICH

A. Problem Statement

To find new ways to build upon current standards of detection and classification of ICH and its classes by implementing state-of-the-art method and improving upon it's shortcomings.

B. Methodology

1) Dataset: The dataset upon which the model was trained was published in the RSNA Intracranial Haemorrhage detection challenge in Kaggle. The dataset consists of 25,272 CT scans, 8,70,301 slices in dicom format. It was split into 7,28,513 slices for training, 24,290 slices for validation and 1,21,232 slices for testing. In the scan level, it was split in the ratio of 21,000:744:3528 CT scans. On an average, each CT scan has 35 slices approximately varying from 28 to 40 slices. The value of each pixel represents the density of the tissue measured in Hounsfield units (HU). The slices had a resolution of 512×512. Each slice was annotated with 6 labels, the first for 'any' representing the presence or absence of Intracranial Haemorrhage and the rest 5 labels for the Intracranial Haemorrhage subclasses: Epidural (EDH), Intraparenchymal (IPH), Intraventricular (IVH), Subdural (SDH) and Subarchanoid (SAH). Some slices were found to have more than one subclass.

2) Pre-Processing: Three HU windows were considered, each focusing on different type of tissues: Brain window (WC : 40, WW : 80), Subdural window (WC : 80, WW : 200) and Soft tissue window (WC : 40, WW : 380). The slices were reduced by linear interpolation to 256 X 256 and were inputted to the model by concatenating the results of the 3 windows for dimensions of $256 \times 256 \times 3$. They were normalized with mean of ([0.1738, 0.1433, 0.1970]) and standard deviation of ([0.3161, 0.2850, 0.3111]). To restrict the model from overfitting and improve its accuracy, the data was augmented by horizontal flipping, shifting, rotation, scaling, and brightness adjustment.

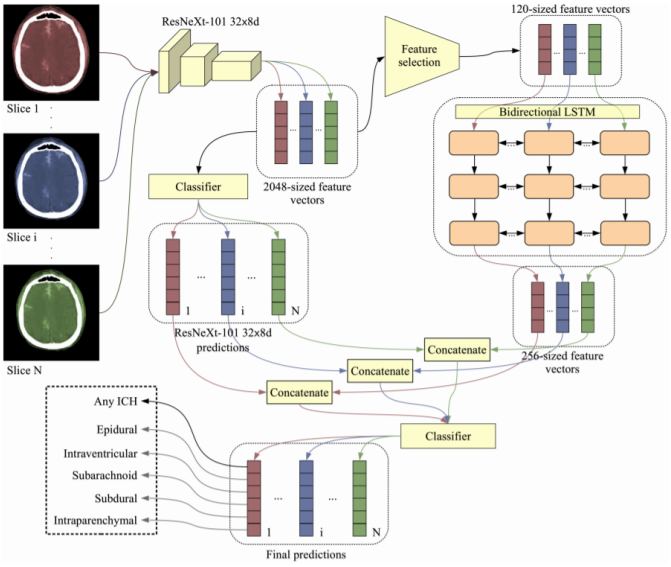

Fig. 1. From [1]. Model Architecture and Pipeline

3) Model Architecture: See figure 1. The CNN model was based on ResNext – 101 $32 \times 8d$ architecture which was pre-trained on Imagenet dataset. Each slice of a CT scan is inputted to the model individually and we retrieve the predicted class labels and a 2048×1 sized vector from the penultimate layer of the model. Using Principal Component Analysis algorithm, the 2048 sized vector is reduced to 120 sized vectors. Once we get the 120 sized vectors for all slices of a CT scan, they are all inputted simultaneously to the Bi-directional LSTM model to enhance the prediction. The advantage of Bi-directional LSTMs is that they can use information from the neighbouring vectors to improve the accuracy. The slices are inter-dependent because if an ICH is present in any one slice, it will also be present in its neighbouring slices. From the LSTMs, we get 256 sized vectors which are then concatenated to their corresponding slice-wise prediction labels. The concatenated vectors are then passed through a linear classifier to get the final slice-wise prediction labels.

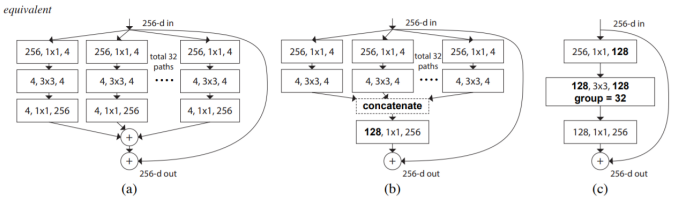

Fig. 2. From [3]. Equivalent building blocks of ResNeXt. (a): Aggregated residual transformations. (b): A block equivalent to (a), implemented as early concatenation. (c): A block equivalent to (a,b), implemented as grouped convolutions. A layer is denoted as (# input channels, filter size, # output channels).

In figure 2, the similarity is shown between (a) ResNext block, (b) inception-resnet block and (c) grouped convolution block. In resnxet – 101 $32 \times 8d$, 101 represents the number of ResNext blocks, 32 represents the cardinality or the total number of paths in a block and 8 represents the depth of the middle layer.

TABLE I
Results

Class	AUC(scan)	AUC(slice)	Accuracy(scan)	Accuracy(slice)
any	0.9822	0.9825	0.9516	0.9566
EDH	0.9146	0.9767	0.9879	0.9967
IPH	0.9825	0.9910	0.9368	0.9804
IVH	0.9882	0.9965	0.9651	0.9901
SAH	0.9633	0.9780	0.9153	0.9741
SDH	0.9493	0.9678	0.9220	0.9623

4) Results: See table I. These results were obtained from the CNN's 3rd epoch and LSTM's 5th epoch of training, and achieved a validation loss of 0.0647. In addition to this, we have also drawn heatmaps for each of the ICH subclasses using Grad-CAM [4]. Grad-CAM stands for Gradient Class Activation Method.

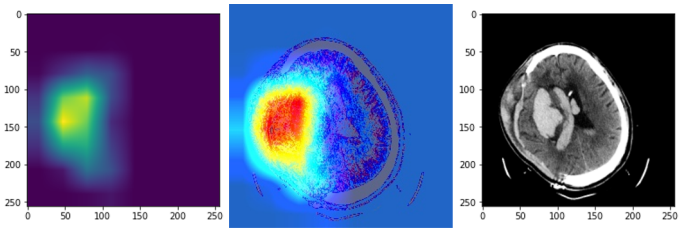

Fig. 3. Epidural Haemorrhage. (*from left*) Heatmap, Heatmap imposed over brain CT sample, Brain CT sample

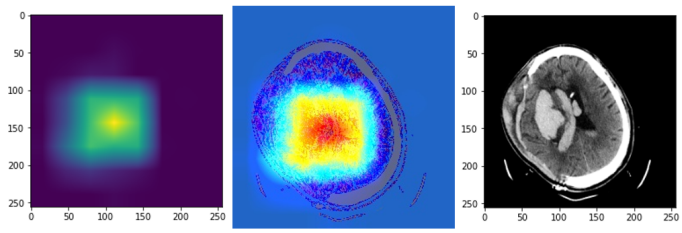

Fig. 4. Intraparenchymal Haemorrhage. (*from left*) Heatmap, Heatmap imposed over brain CT sample, Brain CT sample

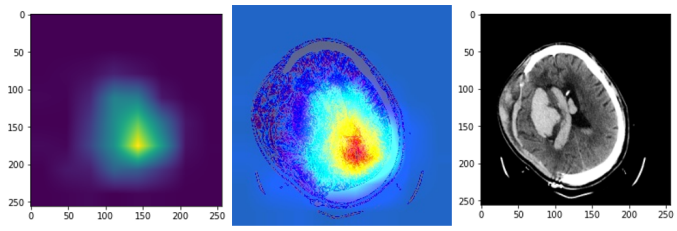

Fig. 5. Intraventricular Haemorrhage. (*from left*) Heatmap, Heatmap imposed over brain CT sample, Brain CT sample

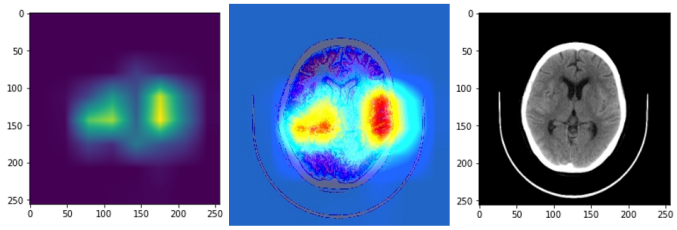

Fig. 6. Subarchanoid Haemorrhage. (*from left*) Heatmap, Heatmap imposed over brain CT sample, Brain CT sample

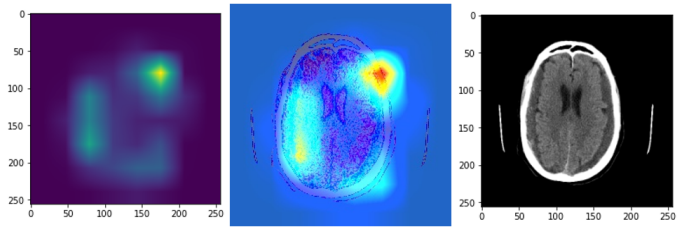

Fig. 7. Subdural Haemorrhage. (*from left*) Heatmap, Heatmap imposed over brain CT sample, Brain CT sample

IV. 2-D Segmentation of ICH

A. Problem Statement

Since the above discussed model was not particularly trained for localizing ICH regions, we trained a UNet++ model for semantic segmentation of ICH regions.

B. Methodology

1) Dataset: M. D. Hassayeni et al. published a dataset containing 82 CT scans, including 36 CT scans with ICH, in [2]. The average age of the subjects from which the CT scans were collected was 28 years. The number of slices per CT scan is not constant and it varies from 33 to 60 slices. Among the 82 subjects, 46 were males and the remaining were females. The dataset consisted of the CT scans, their corresponding masks of ICH regions, and demographic information of the subjects, which also included the classes of the ICH, and, the presence of skull fractures.

2) Pre-Processing: The CT scans and their masks were in the *NIfTI* format and their resolutions were 512×512. The CT scans were first processed using a brain window [level= 40, width= 120]. Each slice of a CT scan was then cropped into 128×128 crops by striding over 64 pixels for each crop, resulting in 49 crops per CT slice. This was done because of the small size of the ICH regions in the masks. The input dimensions for the model were $128 \times 128 \times 3$. Since the UNet model used in [2] requires only $128 \times 128 \times 1$ and the UNet++ model on which we trained requires $128 \times 128 \times 3$, in our 1st apporach, the brain window processed crop was copied over to the remaining 2 layers. In our 2nd approach, instead of copying the brain window processed crop over to the remaining layers, we processed the CT scan on Subdural window [level= 80, width= 200], and Soft tissue window [level= 40, width= 380], and concatenated the results of the 3 windows along the depth to result with the model's input dimensions.

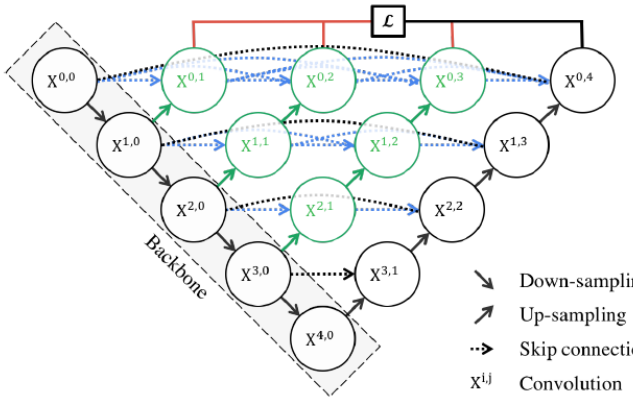

Fig. 8. From [5]. UNet++ model architecture

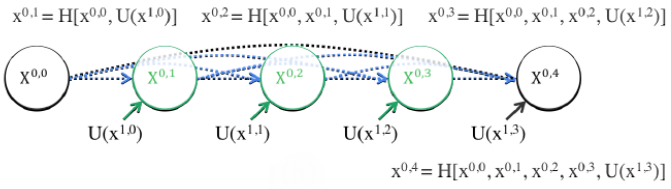

Fig. 9. From [5]. Single UNet++ layer

3) Model Architecture: The UNet++ is much different than the UNet by employing skip connections between nodes along the same horizontal layer. Because of the nested skip connection design, the UNet++ can achieve higher results than the UNet in [2]. Each node consists of convolution with a 3×3 kernel followed by batch normalization and passing through a ReLu activation layer. In our work, we employed many different backbones like Resnet50, VGG16, Densenet121, Densenet169, and InceptionResnetV2 for comparison of their results. The models were trained by both randomly intialized weights and pre-trained weights: ImageNet. The 2 loss functions used were Binary Cross-Entropy and Jaccard Distance. All the models were trained for 100 epochs in Keras and Tensorflow backend.

TABLE II
RESULTS

Backbone, weight, and loss function	Jaccard(Avg.)	DICE(Avg.)
Resnet50, imagenet, BCE	0.312	0.403
Resnet50, RI, BCE	0.299	0.383
VGG16, RI, BCE	0.284	0.369
Densenet121, imagenet, BCE	0.321	0.415
Densenet169, imagenet, BCE	0.323	0.413
InceptionResnetV2, imagenet, BCE	0.203	0.279
Densenet169, imagenet, JD	**0.338**	**0.433**

4) Results: In table II, BCE stands for Binary Cross-Entropy loss and JD stands for Jaccard Distance loss function. RI stands for random initialization. In table III, a comparison of results between UNet++ and UNet of [2] is shown. The 1st approach provides better results than the UNet of [2] and the highest Jaccard and Dice scores were observed by training the model with a Densenet169 backbone, pre-trained imagenet weights and Jaccard Distance loss function. The 2nd approach did not segment any ICH region and predicted only black pixels. This can be attributed to the differences in the layers of the input tensors caused by using different windows for the different layers.

TABLE III
RESULTS

Model	Jaccard(Avg.)	DICE(Avg.)	Sens.(Avg.)	Spec.(Avg.)
UNet++	**0.338**	**0.433**	0.673	**0.727**
UNet [2]	0.218	0.315	**0.972**	0.504

V. CURRENT AND FUTURE WORK

For improving upon the Jaccard and DICE scores of the 2D-UNet++, we intend to integrate interslice dependency of ICH regions by training a **3D UNet++** over the dataset published in [2]. The model has been prepared in PyTorch, but because of the model's complexity and computation of huge 3-D tensors ($512 \times 512 \times \#slices$), training the model has always faced GPU memory allocation issues. We intend to bypass this by training it on the CPU or by adding up the tensors in the skip connections instead of concatenating them.

REFERENCES

[1] Mihail Burduja, Radu Tudor Ionescu, and Nicolae Verga, "Accurate and Efficient Intracranial Hemorrhage Detection and Subtype Classification in 3D CT Scans with Convolutional and Long Short-Term Memory Neural Networks" *MDPI*, August. 2020.

[2] M. D. Hssayeni, M. S. Croock, A. D. Salman, Al-khafaji Hassan Falah, Z. A. Yahya, and B. Ghoraani, "Intracranial Hemorrhage Segmentation Using a Deep Convolutional Model," Data, vol. 5, no. 1, p. 14, Feb. 2020.

[3] Saining Xie, Ross Girshick, Piotr Dollar, Zhuowen Tu, Kaiming He;"Aggregated Residual Transformations for Deep Neural Networks", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1492-1500

[4] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra;"Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization", Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618-626

[5] Zhou Z., Rahman Siddiquee M.M., Tajbakhsh N., Liang J. (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov D. et al. (eds) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science, vol 11045. Springer, Cham. https://doi.org/10.1007/978-3-030-00889-5_1

[6] Yuh, E.L.; Gean, A.D.; Manley, G.T.; Callen, A.L.; Wintermark, M. Computer-aided assessment of head computed tomography (CT) studies in patients with suspected traumatic brain injury. Journal of neurotrauma 2008, 25, 1163–1172

[7] Li, Y.; Wu, J.; Li, H.; Li, D.; Du, X.; Chen, Z.; Jia, F.; Hu, Q. Automatic detection of the existence of subarachnoid hemorrhage from clinical CT images. Journal of medical systems 2012, 36, 1259–1270.

[8] Li, Y.H.; Zhang, L.; Hu, Q.M.; Li, H.W.; Jia, F.C.; Wu, J.H. Automatic subarachnoid space segmentation and hemorrhage detection in clinical head CT scans. International journal of computer assisted radiology and surgery 2012, 7, 507–516.

[9] Prevedello, L.M.; Erdal, B.S.; Ryu, J.L.; Little, K.J.; Demirer, M.; Qian, S.; White, R.D. Automated critical test findings identification and online notification system using artificial intelligence in imaging. Radiology 2017, 285, 923–931.

[10] Grewal, M.; Srivastava, M.M.; Kumar, P.; Varadarajan, S. RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans. Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. IEEE, 2018, pp. 281–284.

[11] Jnawali, K.; Arbabshirani, M.R.; Rao, N.; Patel, A.A. Deep 3D convolution neural network for CT brain hemorrhage classification. Medical Imaging 2018: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2018, Vol. 10575, p. 105751C.

[12] Chilamkurthy, S.; Ghosh, R.; Tanamala, S.; Biviji, M.; Campeau, N.G.; Venugopal, V.K.; Mahajan, V.; Rao, P.; Warier, P. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. The Lancet 2018.

[13] Arbabshirani, M.R.; Fornwalt, B.K.; Mongelluzzo, G.J.; Suever, J.D.; Geise, B.D.; Patel, A.A.; Moore, G.J. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. npj Digital Medicine 2018, 1, 9.

[14] Ye, H.; Gao, F.; Yin, Y.; Guo, D.; Zhao, P.; Lu, Y.;Wang, X.; Bai, J.; Cao, K.; Song, Q.; others. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. European radiology 2019, pp. 1–11.

[15] Chan, T. Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain. Computerized Medical Imaging and Graphics 2007, 31, 285–298.

[16] Shahangian, B.; Pourghassem, H. Automatic brain hemorrhage segmentation and classification algorithm based on weighted grayscale histogram feature in a hierarchical classification structure. Biocybernetics and Biomedical Engineering 2016, 36, 217–232.

[17] Chang, P.; Kuoy, E.; Grinband, J.; Weinberg, B.; Thompson, M.; Homo, R.; Chen, J.; Abcede, H.; Shafie, M.; Sugrue, L.; others. Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. American Journal of Neuroradiology 2018, 39, 1609–1616.

[18] Kuo, W.; Häne, C.; Yuh, E.; Mukherjee, P.; Malik, J. Cost-Sensitive Active Learning for Intracranial Hemorrhage Detection. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018; Frangi, A.F.; Schnabel, J.A.; Davatzikos, C.; Alberola-López, C.; Fichtinger, G., Eds.; Springer International Publishing: Cham, 2018; pp. 715–723.

[19] Cho, J.; Park, K.S.; Karki, M.; Lee, E.; Ko, S.; Kim, J.K.; Lee, D.; Choe, J.; Son, J.; Kim, M.; others. Improving Sensitivity on Identification and Delineation of Intracranial Hemorrhage Lesion Using Cascaded Deep Learning Models. Journal of digital imaging 2019, 32, 450–461.

[20] Bhadauria, H.; Dewal, M. Intracranial hemorrhage detection using spatial fuzzy c-mean and region-based active contour on brain CT imaging. Signal, Image and Video Processing 2014, 8, 357–364.

[21] Nag, M.K.; Chatterjee, S.; Sadhu, A.K.; Chatterjee, J.; Ghosh, N. Computer-assisted delineation of hematoma from CT volume using autoencoder and Chan Vese model. International journal of computer assisted radiology and surgery 2018, pp. 1–11.

[22] Muschelli, J.; Sweeney, E.M.; Ullman, N.L.; Vespa, P.; Hanley, D.F.; Crainiceanu, C.M. PItcHPERFeCT: Primary intracranial hemorrhage probability estimation using random forests on CT. NeuroImage: Clinical 2017,14, 379–390.

[23] Kuang, H.; Menon, B.K.; Qiu, W. Segmenting Hemorrhagic and Ischemic Infarct Simultaneously From Follow-Up Non-Contrast CT Images in Patients With Acute Ischemic Stroke. IEEE Access 2019, 7, 39842–39851.

[24] Gautam, A.; Raman, B. Automatic Segmentation of Intracerebral Hemorrhage from Brain CT Images. In Machine Intelligence and Signal Analysis; Springer, 2019; pp. 753–764.

[25] Prakash, K.B.; Zhou, S.; Morgan, T.C.; Hanley, D.F.; Nowinski, W.L. Segmentation and quantification of intra-ventricular/cerebral hemorrhage in CT scans by modified distance regularized level set evolution technique. International journal of computer assisted radiology and surgery 2012, 7, 785–798.