CRoNE 2019
Conference Proceedings

Convolutional neural networks for detection intracranial hemorrhage in CT images

Juan Sebastian Castro[1], **Steren Chabert**[1,2], **Carolina Saavedra**[1], **Rodrigo Salas**[1,2]

[1]Universidad de Valparaíso; [2]Centro de Investigación y Desarrollo en Ingeniería en Salud

***For correspondence:**
juan.castro@postgrado.uv.cl (JSC);
rodrigo.salas@uv.cl (RS)

Present address: †Escuela de Ingeniería C. Biomédica, Universidad de Valparaíso, Chile; ‡Centro de Investigación y Desarrollo en Ingeniería en Salud, CINGS-UV, Universidad de Valparaíso, Chile.

Abstract Deep learning algorithms have recently been applied for image detection and classification, lately with good results in the medicine such as medical image analysis. This paper aims to support the detection of intracranial hemorrhage in computed tomography (CT) images using deep learning algorithms and convolutional neural networks (CNN). The motivation of this work is the difficulty of physicians when they face the task to identify intracranial hemorrhage, especially when they are in the primary stages of brain bleeding, making a misdiagnosis. A total of 491 CT studies were used to train and evaluate two convolutional neuronal networks in the task of classifying hemorrhage or non-hemorrhage. The proposed CNN networks reach 97% of recall, 98% accuracy and 98% of F1 measure.

Introduction

Intracranial hemorrhage (HIC) corresponds to bleeding inside the skull caused by a vascular rupture. Speed of diagnosis is crucial because the mortality reaches up to 60% after 30 days and 35% to 52% of patients die before a month after being diagnosed, and approximately half of these deaths occur within the first 24 hours (***Caceres and Goldstein, 2012***) (***Rodríguez-Yáñez et al., 2013***). This is a reason why HIC is considered a medical emergency and specialists must diagnose it properly and quickly. However, in general medicine settings and emergency rooms, up to 20% of patients with suspected HIC may be misdiagnosed, which is an indicator that bleeding cannot be reliably distinguished without the support of medical imaging techniques (***Gross et al., 2019***). Brain neuroimaging computed tomography (CT) for the diagnosis of intracranial hemorrhage, is the most reliable method during the first week after the onset of HIC. The visualization of intracranial hemorrhage in CT images depends on density, volume, location, relationship with the surrounding structures (***Cohen, 1992***), all previous properties make HIC diagnosis difficult. An automatic process for HIC detection in the triage workflow, would significantly decrease the time to diagnosis and expedite treatment.

Automatic or semi-automatic detection of intracerebral hemorrhages in CT images without contrast is a recent field of research that is follow by advances in artificial intelligence and image processing. Some of the models proposed for detection are based on K-means and Fuzzy K-means (***Bhadauria et al., 2013***) (***Zaki et al., 2011***) which in some cases are combined with the Otsu method for segmentation of regions of interest (***Loncaric et al., 1999***). Other authors propose models based on the intensity of pixel (***Liao et al., 2010***), level sets and weights of the histogram (***Shahangian and Pourghassem, 2016***), morphological operations (***Chan, 2007***). On the other hand, in recent years, the use of deep learning for image classification tasks has become popular, authors present models using convolutional neural networks (CNN) for the detection of intracranial hemorrhages

(***Chilamkurthy et al., 2018***) (***Helwan et al., 2018***), you can also find models that make use of deep learning for the segmentation of HIC or brain injuries in general (***Ito et al., 2019***) (***Kamnitsas et al., 2017***).

As described, the HIC is classified as a medical emergency in which survival is given by the speed and effectiveness of the diagnosis. So an algorithm that is used to support the diagnostic task must be precise and capable of generalizing, in these cases the best results have been obtained using techniques based on deep learning, that its speed after a previous training of the network. This work presents the use of convolutional neural networks for the task of classifying hemorrhage vs non-hemorrhage, 491 studies of computed tomography of the head without contrast were used with a total of 193,317 slices in which there are 4 types of intracranial hemorrhage and in addition to brains healthy.

This document is organized as follows. In Section II, the proposed method is presented. The results are presented in Section III. Discussion of results in Section IV. Finally, the main conclusions are presented in section V.

Methods and Materials

Data Base

The database chosen is known as **CQ500** (***Chilamkurthy et al., 2018***) and was provided by the Center for Advanced Research in Imaging, Neurosciences and Genomics (CARING) in New Delhi, India. This database is part of a set of head CT images taken by several radiologists in the center of New Dehli. The tomographs used in radiology centers to obtain the images vary between 16 to 128 cuts. The data was taken from the local PACS servers and anonymized according to the internal guidelines defined in HIPAA. Data were collected in two blocks (B1 & B2). Block B1 was collected by selecting all CT studies taken at the radiological center for 30 days beginning on November 20, 2017, Block B2 was selected from the remaining studies. Each of the selected studies was evaluated according to the following exclusion criteria:

1. Patients should not have any post-operative defects such as burr hole/shunt/clips.
2. They should have at least one CT study without axial cut contrast and a soft kernel reconstruction that contemplates the entire brain
3. Patients should not be less than 7 years old. If age information is not available, it will be estimated through bone degradation and cranial sutures.

The total of 491 studies were evaluated by three independent expert radiologists with 8, 12 and 20 years experience in the interpretation of cranial CT images. None of the 3 readers participated in the clinical care or diagnosis of the patients, nor did they have access to their medical history. Each of the radiologists independently assessed the studies in the CQ500 data set following the evaluation instructions, the order of presentation of the studies was randomized to minimize patient remember.

For each CT study the following information was recorded:

- The presence or absence of intracranial hemorrhage, and if its type (intracerebral, subarachnoid, epidural, subdural), state (chronic or non-chronic) and the affected hemisphere (right, left) is present
- The presence or absence of midline movement and mass effect
- The presence or absence of fractures. If present, if this is a cranial (partial) fracture

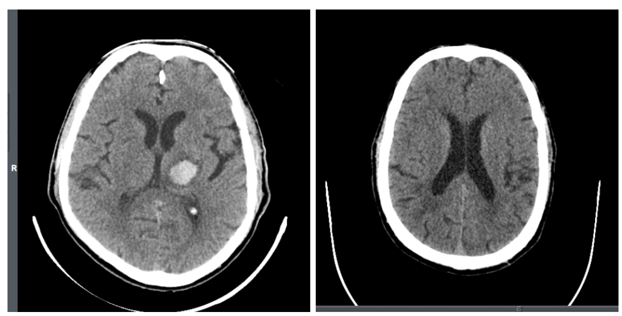
If the three evaluators did not reach a unanimous agreement for each of the studies and findings, the interpretation of the majority of the evaluators was used as the final diagnosis. The characteristics of the database used are found in the table 1.

Characteristic	CQ500 dataset
No. of scans	491 / 193.317 slices
Mean age	22.43
PREVALENCE	
No. of scans (percentage) with intracranial hemorrhage	205 (41.17%)
Intracerebral	162 (32.99%)
Subdural	53 (10.79%)
Extradural	13 (2.64%)
Subarachnoid	60 (12.21%)

Table 1. Dataset characteristics

Image Preprocessing

For the non-contrast CT series original dataset, first, we decided to remove the background image of all slices, since it does not provide any information to the classification algorithm. Next, instead of using the entire CT dynamic range, it was decided to windowed the densities of each slice using a window (level=50, width=80), to visualize only the brain parenchyma. An anisotropic filter was applied with values (kernel=0.02, time=10) also, the pixel values of each slice in the data set were normalized between 0 and 1. Finally, resizing to 256 x 256 pixels, all preprocessing was done before passing to the deep learning models. Figure 1 shows CT slices with the presence of hemorrhage and no presence after preprocessing stage.



(a) Intracranial Hemorrhage (b) Healthy

Figure 1. Two CT no contrast images after preprocesing

Convolutional Neural Network

Convolutional Neural Network (CNN) is a specialized network in processing grid topology data. The most common examples are 1-D grid data at regular time intervals, images and 2-D data with pixel grids. The name for this type of networks arises from the mathematical convolution operation that the network uses within its processing. In simple words, convolution is the operation between two functions with a real value argument which is typically denoted as:

$$S(t) = (x * w)(t)$$

In CNN terminology the first argument (x) is called the input and the second argument (w) is called the kernel. The output is usually denoted as a feature map (***Goodfellow et al., 2016***). The principal function of convolution is feature extraction from input images. Generally, a convolutional neural network is divided into three stages: first convolutional layers with an activation like ReLU (rectified linear unit), second a pooling layer for size reduction typically max-pooling. Finally, a flatten layer before to a fully connected layer to classification the features maps.

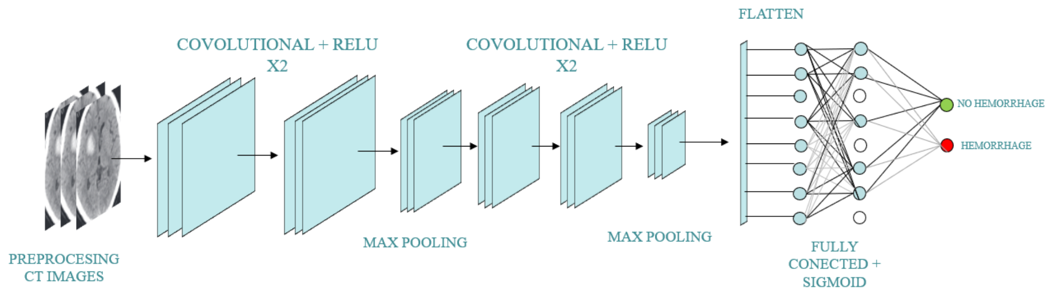


Figure 2. Proposed Convolutional Neural Network for hemorrhage detection

In this paper, two convolutional neural networks for hemorrhage detection were employed, first a simple custom CNN was develop maintaining parsimony. The network architecture of the first CNN is shown in figure 2, start with two convolutional layers with ReLu activation and kernel (3x3), next a max-pooling layer size reduction, then again two convolutional layers with the same characteristics than previous convolutional layers followed by a max-pooling layer with kernel (2x2), next a flatten layer to prepare the features maps to dense layers. Finally, two fully connected layers for classification was implemented to predict labels with sigmoid activation. We named our network as CNN4. Another CNN was employed for the hemorrhage detection task, we were decided to use a popular network VGG16 (***Zhang et al., 2015***) with a modification for binary classification (hemorrhage vs no-hemorrhage), VGG16 is one of the most used and reliable CNN tested in a variety of dataset like ImageNet or CMNIST (***Russakovsky et al., 2015***), that is compose of 5 blocks (convolutional + pooling) with 3 fully connected layers used for the classification task.

Training and evaluating models

In this study, the CNN models were trained using the preprocessing slices. A total of 193.317 slices of 491 CT scans were used to train and evaluate the CNN network. Then, two methods to train the models were proposed:

1. **Slices randomized:** All slices were randomized to train (0.85) and test (0.15) sets, regardless of independence between subjects. This means a part of the slices of a subject could be in the train set and another part of the slices could be in the test set.
2. **Subject randomized:** All slices were randomized to train (0.85) and test (0.15) sets, ensuring independence between subjects. This means all slices of one subject were sent to train or test.

Due to the need for a large amount of data from the deep learning networks, we decided to divide the dataset by 0.85 for training, of which 0.2 was used for validation during the training process. Each model was trained for 150 epochs with a batch size of 32, the best model was saved to be evaluated with the test set. Binary cross-entropy loss was used to assess performance over time.

Some metrics were obtained to evaluate the performance of CNN in the classification of hemorrhage vs non-hemorrhage. Receiver operating characteristic (ROC) curves were obtained for CNN4 and VGG16, with each of the proposed training methods. Accuracy, recall and F1 measure were also obtained for each of the algorithms.

Results

The ROC curves obtained from the performance evaluation of the VGG16 network (figure 4) show a much higher performance in the case of the randomized slices (A) method, reaching 0.989 of area under curve (AUC), as It also presents a recall (table 2) of 0.974 and an F1 measure of 0.971. These results contrast with those obtained for the method of subject randomized (Figure 4B) where the AUC is 0.783, with a recall that barely reaches 0.735 and an F1 measure of 0.758, concerning the

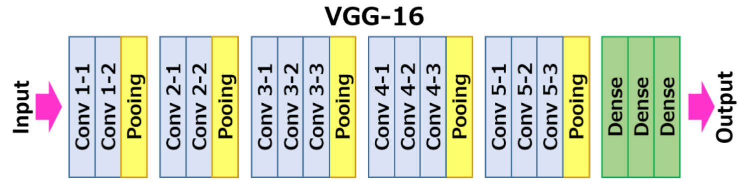
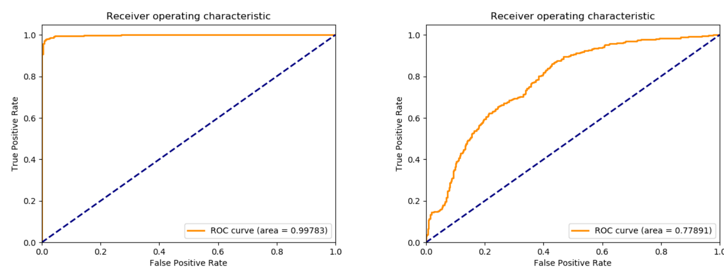


Figure 3. Architecture of the VGG16 Convolutional Neural Network

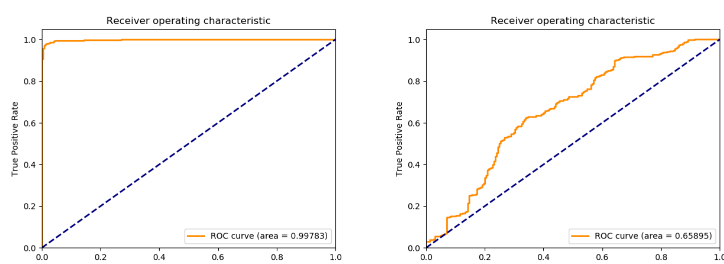
accuracy, a 0.707 for the classification of hemorrhage vs no hemorrhage in the test set.

In the case of the CNN4 algorithm, the ROC curves obtained (figure 5) also show a good network performance for the randomized slices method (figure 5 A), with an area under the curve of 0.982, as well as a recall of 0.972, F1 measure of 0.972 and an accuracy of 0.981, which are very similar to those obtained with VGG16. On the other hand, the performance obtained with the subject randomized method in the CCN4 network were: AUC 0.658, recall 0.721, F1 measure 0.687 and accuracy of 0.598, as well as their respective ROC curve (figure 5 B). All the performance metrics obtained for the test set can be found in table 2.



(a) VGG16 trained with slices randomized (b) VGG16 trained with subject randomized

Figure 4. ROC curves for VGG16 network trained with the two proposal methods for detection intracranial hemorrhage. Area Under Curve (AUC) is also presented



(a) CNN4 trained with slices randomized (b) CNN4 trained with subject randomized

Figure 5. ROC curves for CNN4 network trained with the two proposal methods for detection intracranial hemorrhage. Area Under Curve (AUC) is also presented

Discussion

The ROC curves obtained in figure 4 and figure 5 for both the VGG16 network and the proposed CNN4 network, present excellent results for the classification of hemorrhage vs. non-hemorrhage using the training method of slices randomized, reaching 0.98 of AUC in both cases. This represents

Mode	Accuracy	Recall	F1_measure	Area Under Curve (ROC)
VGG-16				
Slices Randomized	0.968	0.974	0.971	0.989
Subject Randomized	0.707	0.735	0.758	0.783
CNN - 4				
Slices Randomized	0.981	0.972	0.982	0.982
Subject Randomized	0.598	0.721	0.687	0.658

Table 2. Performance of algorithms for CQ500 dataset

a high recall and specificity of both algorithms, this is confirmed with the f1 measure that in both cases exceeds 0.97, this metric is the compromise between recall and accuracy. This performance is explained by the nature of the images used for the classification of hemorrhage vs non-hemorrhage. A CT scan is composed of several slices that represent the skull in this case in different positions in a cut (axial for this study), as all slices belong to the same patient contain similarities. On the other hand, there is no difference between the performance of the two classification algorithms proposed with the randomized slices method, the results show that the CNN4 network despite being simpler than the VGG16 can have a similar performance.

The results presented for the subject randomized method in figure 4 and figure 5 for both VGG16 and CNN4 networks do not present the expected performance. Table 2 shows the metrics obtained, which highlights the recall that exceeded 0.72 for both networks evaluated, the performance obtained by the VGG16 being outstanding, which in this case exceeds more than 10 percentage points in the metrics of AUC and F1 measure to the CNN4 network. So for this subject randomized training method, the VGG16 network has better results than CNN4. On the other hand, although the performance of both VGG16 and CNN4 networks was not as expected, the subject randomized method has the advantage of preserving the independence of the data and therefore having a better capacity to generalize.

Regarding the performance of the networks compared to the state of the art, it can be determined that the randomized slices method with a recall of 0.974 for the VGG16 and 0.972 for CNN4, is very similar to the performance obtained by other studies (***Chilamkurthy et al., 2018***) (***Helwan et al., 2018***). And although the performance obtained with the subject randomized method does not reach these recall levels, it still exhibits outstanding performance over traditional medical image processing techniques (***Bhadauria et al., 2013***) (***Zaki et al., 2011***).

Conclusion

In this paper, two convolutional neural networks were proposed for the task of classification of intracranial hemorrhage vs. non-hemorrhage, a popular VGG16 network, and a CNN4 own network. Two different training methods were also proposed (slices randomized and subject randomized) where the second ensures the independence of the data. The results show an outstanding performance for the first training method in the classification task, on the other hand, the second training method proposed is at the level of the classic medical image processing techniques. With this, it can be concluded that convolutional neural networks are a useful tool for the identification of intracranial hemorrhages in computed tomography images and can be used as a support in the diagnosis of this type of pathologies. Additionally, it was found that the method for choosing the train set and test set is influential for the performance of deep learning algorithms. Therefore, a greater study of the independence of the data in the use of computed tomography images is required for classification through convolutional neural networks.

References

Bhadauria HS, Singh A, Dewal ML. An integrated method for hemorrhage segmentation from brain CT Imaging. Computers & Electrical Engineering. 2013 Jul; 39(5):1527–1536. https://linkinghub.elsevier.com/retrieve/pii/S0045790613000955, doi: 10.1016/j.compeleceng.2013.04.010.

Caceres JA, Goldstein JN. Intracranial Hemorrhage. Emergency Medicine Clinics of North America. 2012 Aug; 30(3):771–794. https://linkinghub.elsevier.com/retrieve/pii/S0733862712000272, doi: 10.1016/j.emc.2012.06.003.

Chan T. Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain. Computerized Medical Imaging and Graphics. 2007 Jun; 31(4-5):285–298. https://linkinghub.elsevier.com/retrieve/pii/S0895611107000183, doi: 10.1016/j.compmedimag.2007.02.010.

Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, Mahajan V, Rao P, Warier P. Development and Validation of Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans. arXiv:180305854 [cs]. 2018 Mar; http://arxiv.org/abs/1803.05854, arXiv: 1803.05854.

Cohen W. Computed tomography of intracranial hemorrhage. Radiologic Clin North Amer. 1992; 2:75–87.

Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016. http://www.deeplearningbook.org.

Gross BA, Jankowitz BT, Friedlander RM. Cerebral Intraparenchymal Hemorrhage: A Review. JAMA. 2019 Apr; 321(13):1295. http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2019.2413, doi: 10.1001/jama.2019.2413.

Helwan A, El-Fakhri G, Sasani H, Uzun Ozsahin D. Deep networks in identifying CT brain hemorrhage. Journal of Intelligent & Fuzzy Systems. 2018 Aug; 35(2):2215–2228. http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JIFS-172261, doi: 10.3233/JIFS-172261.

Ito R, Nakae K, Hata J, Okano H, Ishii S. Semi-supervised deep learning of brain tissue segmentation. Neural Networks. 2019 Aug; 116:25–34. https://linkinghub.elsevier.com/retrieve/pii/S0893608019300954, doi: 10.1016/j.neunet.2019.03.014.

Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis. 2017 Feb; 36:61–78. https://linkinghub.elsevier.com/retrieve/pii/S1361841516301839, doi: 10.1016/j.media.2016.10.004.

Liao CC, Xiao F, Wong JM, Chiang IJ. Computer-aided diagnosis of intracranial hematoma with brain deformation on computed tomography. Computerized Medical Imaging and Graphics. 2010 Oct; 34(7):563–571. https://linkinghub.elsevier.com/retrieve/pii/S0895611110000388, doi: 10.1016/j.compmedimag.2010.03.003.

Loncaric S, Dhawan AP, Cosic D, Kovacevic D, Broderick J, Brott T. Quantitative intracerebral brain hemorrhage analysis. In: Hanson KM, editor. San Diego, CA; 1999. p. 886–894. http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=980754, doi: 10.1117/12.348648.

Rodríguez-Yáñez M, Castellanos M, Freijo MM, López Fernández JC, Martí-Fàbregas J, Nombela F, Simal P, Castillo J, Díez-Tejedor E, Fuentes B, Alonso de Leciñana M, Álvarez Sabin J, Arenillas J, Calleja S, Casado I, Dávalos A, Díaz-Otero F, Egido JA, Gállego J, García Pastor A, et al. Guías de actuación clínica en la hemorragia intracerebral. Neurología. 2013 May; 28(4):236–249. https://linkinghub.elsevier.com/retrieve/pii/S0213485311001447, doi: 10.1016/j.nrl.2011.03.010.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al. Imagenet large scale visual recognition challenge. International journal of computer vision. 2015; 115(3):211–252.

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(1):217–232. https://linkinghub.elsevier.com/retrieve/pii/S0208521615300036, doi: 10.1016/j.bbe.2015.12.001.

Zaki WMDW, Fauzi MFA, Besar R, Ahmad WSHMW. Abnormalities detection in serial computed tomography brain images using multi-level segmentation approach. Multimedia Tools and Applications. 2011 Aug; 54(2):321–340. http://link.springer.com/10.1007/s11042-010-0524-0, doi: 10.1007/s11042-010-0524-0.

Zhang X, Zou J, He K, Sun J. Accelerating very deep convolutional networks for classification and detection. IEEE transactions on pattern analysis and machine intelligence. 2015; 38(10):1943–1955.