

Slice-level Detection of Intracranial Hemorrhage on CT Using Deep Descriptors of Adjacent Slices

Dat T. Ngo
Vingroup Big Data Institute
Hanoi, Vietnam

Thao T.B. Nguyen
Vingroup Big Data Institute
Hanoi, Vietnam

Hieu T. Nguyen
Vingroup Big Data Institute
Hanoi, Vietnam

Dung B. Nguyen
Vingroup Big Data Institute
Hanoi, Vietnam

Ha Q. Nguyen
Vingroup Big Data Institute
VinBigData JSC,
Hanoi, Vietnam

Hieu H. Pham
VinUni-Illinois Smart Health Center,
College of Engineering & Computer Science,
VinUniversity, Hanoi, Vietnam

Abstract—We propose for the first time a new strategy to train slice-level classifiers on CT scans based on the descriptors of the adjacent slices along the axis. In particular, each of which is extracted through a convolutional neural network (CNN). This method is applicable to CT datasets with per-slice labels such as the RSNA Intracranial Hemorrhage (ICH) dataset, which aims to predict the presence of ICH and classify it into 5 different sub-types. We obtain a single model in the top 4% best-performing solutions of the RSNA ICH challenge, where model ensembles are allowed. Experiments also show that the proposed method significantly outperforms the baseline model on CQ500. The proposed method is general and can be applied to other 3D medical diagnosis tasks such as MRI imaging. To encourage new advances in the field, we will make our codes and pre-trained model available upon acceptance of the paper.

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

Deep neural networks, in particular deep convolutional neural networks (CNNs) [1], [4], [24] became a key tool in medical imaging analysis [6], [20], [22], [26]. In particular, deep neural networks have showed excellent performance in 2D medical imaging data such as the interpretation of X-ray scans [12]–[14], [16]–[18], [21], [23]. However, training very deep networks on high-resolution 3D volumes of Computed Tomography (CT) scans requires a huge computing resource. For examples, each typical 3D CT scans study may contain from hundreds to thousands slices, which makes training deep neural networks on this modality challenging in terms of computing requirements (see Figure 1). In this study, we aim to develop an efficient 2D deep learning-based approach for 3D medical imaging analysis that is able to provide the same level of performance compared to 3D-based learning approaches. To tackle this challenge, we consider each 3D CT scan as a set of 2D images and propose a new strategy to train *slice-level* classifiers on CT based on the descriptors of the adjacent slices along the axis, each of which is extracted through a convolutional neural network (CNN). Extracted features are then concatenated to produce better representations for classification tasks. This method is applicable to CT datasets with

per-slice labels such as the RSNA Intracranial Hemorrhage (ICH) dataset [19], which aims to predict the presence of ICH and classify it into 5 different sub-types.

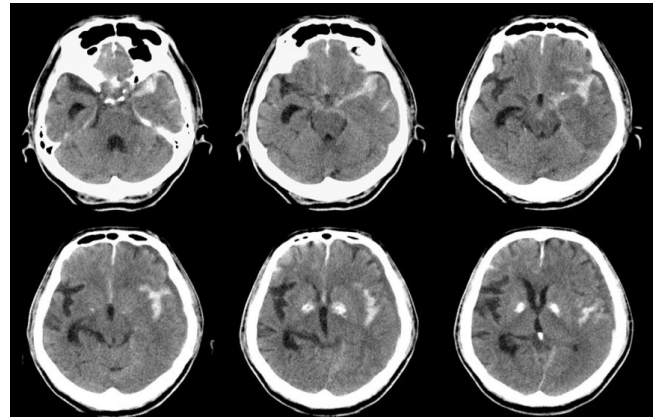


Fig. 1: Examples of head computed tomography scans. Each study may contain from hundreds to thousands slices, which makes training deep neural networks on this modality challenging in terms of computing requirements.

The proposed method exploits a two-stage training scheme. In the first stage, we treat a CT scan simply as a set of 2D images and train a state-of-the-art CNN classifier [7] that was pretrained on ImageNet [10]. During the training process, each slice is sampled *together* with the 3 slices before and the 3 slices after it. By this way, each CNN model takes a batch size a multiple of 7 slices as input. In the second stage, the output *descriptors* of each block of 7 consecutive slices obtained from the first stage are stacked into an image and fed to another CNN model for final prediction of the middle slice. Then, we leverage the assemble learning to boost classification performance. To evaluate the effectiveness of the proposed approach, we train our deep learning framework on the RSNA dataset and additionally evaluated on the CQ500 dataset [2], which adopts the same a set of labels but

All correspondence to hieu.ph@vinuni.edu.vn (Hieu H. Pham)

only on study level. We obtain a single model¹ in the top 4% best-performing solutions of the RSNA ICH challenge, where model ensembles are allowed. Experiments also show that the proposed method significantly outperforms the baseline model on CQ500 [2]. To summarize, our key contribution is to introduce a simple 2D-based technique that allows to train a very deep neural network on 3D imaging data and providing better results compared to 3D-based approaches. The proposed method is simple, general can be applied for many other 3D medical imaging analysis tasks such as MRI imaging analysis. The rest of the paper is organized as follows. The proposed approach is presented in Section II. Section III describe the experiments and discusses the experimental results. Finally, section IV concludes the work as well as presents its perspectives.

II. PROPOSED APPROACH

Non-contrast head CT scans are the current standard imaging protocol for initial imaging of patients with stroke symptoms, including intracranial haemorrhage [3]. We aim in this study to develop and validate a 2D-based deep learning algorithms for automated detection of the key findings from head CT scan scans called intracranial haemorrhage. The proposed approach is validated on the RSNA Intracranial Hemorrhage (ICH) dataset. It is difficult to exploit high-performing deep neural networks for the ICH classification while keeping the full 3D resolution of the input CT. On the other hand, using slices as independent 2D images and ignoring the axial information of data causes detrimental effects to the algorithm's performance. Our method gets the best of both worlds: we apply transfer learning [15] on 2D images to perform classification per slice and then assemble the results of local slices to refine the prediction of the middle one. The proposed scheme, illustrated in Figure 2, consists of two stages: descriptor extraction and axial fusion. Below we describe these element in details.

A. Descriptor extraction

In this stage, we train a 2D CNN to classify individual slices of the CT scans, which are converted into 3-channel images using 3 different windows: brain ($l = 40, w = 80$), subdural ($l = 75, w = 215$), and bone ($l = 600, w = 2800$). The output of the network is a descriptor of size 6×1 that includes the probabilities of the 5 ICH sub-types and an additional class for any of them. During training, each slice is always sampled in a block of 7 that includes itself in the center and 6 neighboring slices. With this approach, we can take advantage of pre-trained models on ImageNet [5] to initialize the network. Specifically, a ResNet-50 [7] was used in our experiments. We followed the procedure in [8] and trained the network for 20 epochs with a batch size of 16×7 using Adam optimizer [9]. An initial learning rate of $5e - 4$ and the cosine annealing learning rate scheduler [11] were used. Several augmentation techniques such as cropping, resizing,

flips, rotations, distortions, gaussian noise, and CutMix [25] were applied to prevent the network from overfitting.

B. Axial fusion

This stage combines the descriptors of each 7 consecutive slices generated in stage 1 to exploit the axial information and to refine the prediction of the centered slice. In particular, we concatenate the 7 descriptors into a $7 \times 6 \times 1$ tensor and train a 3-layer CNN to output the final classification result for the representative slice in the middle of the block. This network contains only 2 convolution layers and 1 fully connected layer. The 2D convolution kernels help the model learn both the relationship between ICH predictions across local slices and the relationship between probabilities for the sub-types. The output of the fusion network can therefore be seen as a re-calibrated prediction of a single slice.

III. EXPERIMENTS AND RESULTS

A. Datasets and evaluation protocols

The RSNA and CQ500 datasets were used to verify the effectiveness of the proposed approach. Both of them contain non-contrast CT scans that are labeled with 5 sub-types of ICH: *intraparenchymal*, *intraventricular*, *subdural*, *extradural*, and *subarachnoid*. The only difference between the two datasets is that the labels of RSNA are per slice, while those of CQ500 are per CT scan. The whole RSNA dataset was split into 3 parts: a public training set (19,530 studies), a public testing set (2,214 studies), and a private testing set (3,518 studies). Each study is a CT scan of 20 to 60 slices of 512×512 pixels. The weighted log loss was used as the evaluation metric for this dataset, in which a weight of $2/7$ was used for the ICH label and a weight of $1/7$ was used for each of the 5 sub-types.

Meanwhile, CQ500 consists of 500 studies, from which 490 were selected for experiments while the rest 10 of them are noisy and were excluded from the dataset. We used CQ500 as another test set to validate the efficiency and robustness of the proposed algorithm, which was merely trained on the public training set of RSNA. On this test, the performance of our method is measured by area under the ROC curve (AUC). Note that the study-level probability of any class was taken as the maximum probability amongst all slices.

B. Experimental results

We report a weighted log loss of 0.05341 on the private testing set of RSNA, which ranks in top 4% over 1345 teams on the Kaggle leaderboard. Note that our result is provided by a single ResNet-50 model, while many other solutions in this competition exploit ensemble techniques. On CQ500, the proposed method achieves a mean AUC of 0.971. This is an improvement of around 2% compared to the baseline model [2]. Especially, our method provides better AUC scores over all disease labels as shown in Table I. These results strongly demonstrate the generalization capacity of our model, which was trained on a different dataset with a different labeling protocol.

¹Model weights and codes are available upon acceptance of the paper.

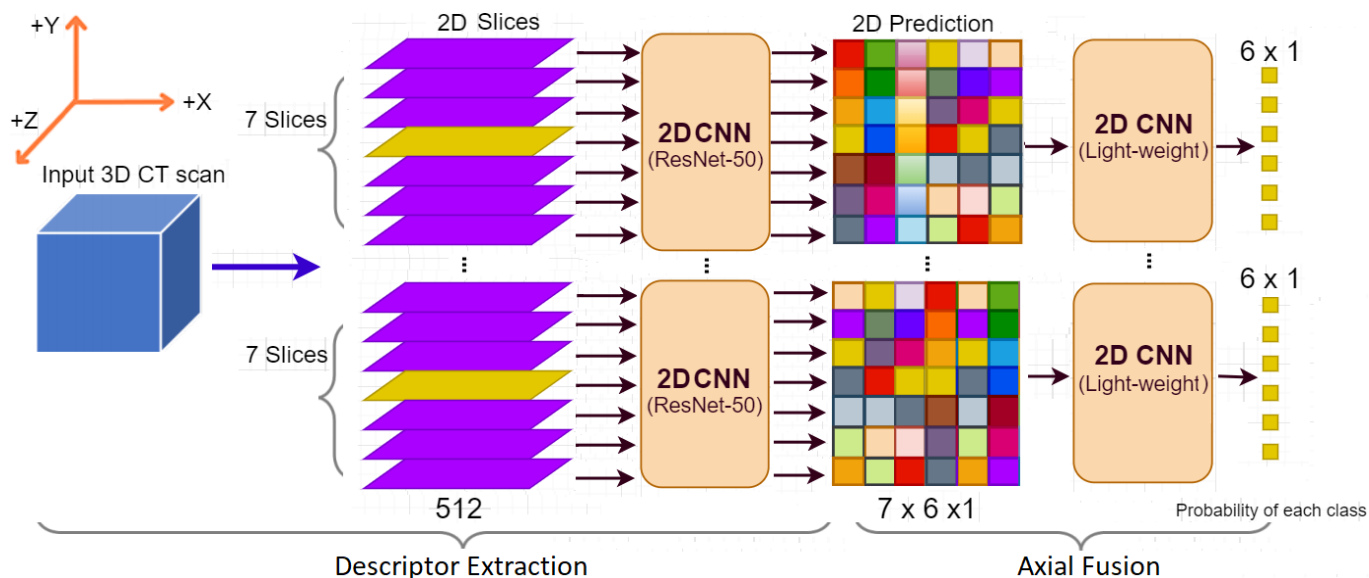


Fig. 2: Illustration of the proposed two-stage training procedure.

Findings	[2]	Ours
ICH (any subtypes)	0.9419	0.9612
Intraparenchymal	0.9544	0.9691
Intraventricular	0.9310	0.9832
Subarachnoid	0.9574	0.9596
Subdural	0.9521	0.9694
Extradural	0.9731	0.9814
Mean	0.9520	0.9710

TABLE I: Experimental results measured by AUC score on CQ500 dataset.

IV. CONCLUSION

Training very deep neural networks such as D-CNN model on 3D medical scans like CT or MRI is a challenging task. Three-dimensional computed tomography (3DCT) is a type of CT scanning which records multiple images over time, that requires huge computational requirements when D-CNNs are explored. We presented in this paper a novel two-stage training strategy for the task of slice-level classification on CT scans. The key idea is to sample each slice together with its neighbors and to refine the classification of it using the coarse descriptors of the whole group. This method was experimentally demonstrated to work well on ICH datasets like RSNA and CQ500. We believe that it can be easily extended to other 3D datasets of CT or MRI scans.

Acknowledgements. This study was supported by VinBigData JSC.

REFERENCES

- [1] W. Alakwaa, M. Nassef, and A. Badr. Lung cancer detection and classification with 3d convolutional neural network (3d-cnn). *International Journal of Advanced Computer Science and Applications*, 8(8), 2017.
- [2] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. Campeau, V. Venugopal, V. Mahajan, P. Rao, and P. Warier. Development and validation of deep learning algorithms for detection of critical findings in head CT scans. *Lancet*, 392:2388–2396, 2018.
- [3] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier. Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study. *The Lancet*, 392(10162):2388–2396, 2018.
- [4] B. T. Dao, T. V. Nguyen, H. H. Pham, and H. Q. Nguyen. Phase recognition in contrast-enhanced ct scans based on deep learning and random sampling. *Medical Physics*, 2022.
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proc. CVPR*, pages 248–255, 2009.
- [6] R. Gruetzemacher, A. Gupta, and D. Paradise. 3d deep learning for detecting pulmonary nodules in ct scans. *Journal of the American Medical Association*, 25(10):1301–1310, 2018.
- [7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. CVPR*, pages 770–778, 2016.
- [8] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li. Bag of tricks for image classification with convolutional neural networks. In *Proc. CVPR*, pages 558–567, 2019.
- [9] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, pages 1–15, 2015.
- [10] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. NeurIPS*, pages 1097–1105, 2012.
- [11] I. Loshchilov and F. Hutter. SGDR: Stochastic gradient descent with warm restarts. In *Proc. ICLR*, pages 1–16, 2017.
- [12] H. T. Nguyen, H. H. Pham, N. T. Nguyen, H. Q. Nguyen, T. Q. Huynh, M. Dao, and V. Vu. Vindr-spinexr: A deep learning framework for spinal lesions detection and classification from radiographs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 291–301. Springer, 2021.
- [13] H. T. Nguyen, S. B. Tran, D. B. Nguyen, H. H. Pham, and H. Q. Nguyen. A novel multi-view deep learning approach for bi-rads and density assessment of mammograms. *arXiv preprint arXiv:2112.04490*, 2021.
- [14] N. H. Nguyen, H. Q. Nguyen, N. T. Nguyen, T. V. Nguyen, H. H. Pham, and T. N.-M. Nguyen. A clinical validation of vindr-cxr, an ai system for detecting abnormal chest radiographs. *arXiv preprint arXiv:2104.02256*, 2021.
- [15] S. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2009.
- [16] H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, and H. Q. Nguyen. Interpreting chest x-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing*, 437:186–194, 2021.

- [17] H. H. Pham, H. Q. Nguyen, K. Lam, L. T. Le, D. B. Nguyen, H. T. Nguyen, T. T. Le, T. V. Nguyen, M. Dao, and V. Vu. An accurate and explainable deep learning system improves interobserver agreement in the interpretation of chest radiograph. *medRxiv*, 2021.
- [18] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.
- [19] RSNA. RSNA Intracranial Hemorrhage Detection. <https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/overview>, 2019. [Online].
- [20] S. Serte and H. Demirel. Deep learning for diagnosis of covid-19 using 3d ct scans. *Computers in biology and medicine*, 132:104306, 2021.
- [21] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh. Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, 9(1):1–12, 2019.
- [22] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás. 3d deep learning on medical images: a review. *Sensors*, 20(18):5097, 2020.
- [23] T. T. Tran, H. H. Pham, T. V. Nguyen, T. T. Le, H. T. Nguyen, and H. Q. Nguyen. Learning to automatically diagnose multiple diseases in pediatric chest radiographs using deep convolutional neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3314–3323, 2021.
- [24] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018.
- [25] S. Yun, D. Han, S. Oh, S. Chun, J. Choe, and Y. Yoo. CutMix: Regularization strategy to train strong classifiers with localizable features. In *Proc. ICCV*, pages 6023–6032, 2019.
- [26] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang. Deep learning-based detection for covid-19 from chest ct using weak label. *MedRxiv*, 2020.