PWML Detection in 3D Cranial Ultrasound Volumes using Over-segmentation and Multimodal Classification with Deep Learning

Flora Estermann [1], *Valérie Kaftandjian* [2], *Philippe Guy* [2], *Philippe Quetin* [3], *Philippe Delachartre* [1]

[1] Univ Lyon, INSA Lyon, Université Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Lyon, France

[2] Univ Lyon, INSA Lyon, Laboratoire Vibrations Acoustique (LVA), F-69621 Villeurbanne, France

[3] CH Avignon, France

ABSTRACT

Punctate white matter lesions (*PWML*) are the most common white matter injuries observed in preterm neonates. Automatic detection of these lesions could better assist doctors in diagnosis. Recent advances in deep learning have resulted in optimistic results on many MR biomedical image benchmark datasets, but few methods seem to tackle the detection of very small lesions in ultrasound images. In this paper, we propose a two-phase strategy. Firstly, we highlight the foreground information by aggregating the lesions in the ground truth along the coronal projection of the brain, then we train a segmentation network to detect PWML with the resulting over-segmented masks. Secondly, we introduce a novel deep architecture for multimodal classification, called 2.5D SC-Net, which is used to eliminate false alarms and improve specificity. Experimental results demonstrate the effectiveness of our method to detect PWML in ultrasound images, improving the recall by 15% compared to the best published models, while limiting the number of false alarms efficiently.

Index Terms— Deep Learning, Automatic Anomaly Detection, 3D Ultrasound Imaging, White Matter Injury, U-Net.

1. INTRODUCTION

Punctate white matter lesions (*PWML*) are responsible for neurodevelopmental sequelae in early childhood [1] [2]. Currently, MRI is the gold standard for lesion detection, but this procedure is expensive and not always accessible. As ultrasound is routinely performed on newborns, this modality could be of real interest for the detection of lesions and allow more children with PWML to benefit from medical care.

Many papers present unsupervised approaches using CNNs for breast tumors classification [3], and brain tumor segmentation [4], but few methods seem to tackle the detection of very small lesions. On the other hand, on the side of supervised methods, research on automatic detection of PWML in MR images was first tackled by Mukherjee et al. [5], who proposed a method that considers the correlation of pixels in 3D space. Besides, Liu et al. [6] proposed a spatiotemporal transformation structure in order to use the information between adjacent slices, which achieves higher performance than existing methods and consumes less computing resources than 3D CNNs. Despite the high contrast and low noise of MR images, the reported accuracy for the PWML detection task remains low with a Dice under 0.60 and a recall at 0.65. Finally, Erbacher et al. [7] started working on this task with cranial ultrasound (*cUS*) images and introduced a new model based on the 2D U-Net [8] and a soft attention model focusing on the PWML localization, called Priority U-Net, with the recall and the precision in the PWML detection task reaching 0.53 and 0.50 respectively.

As stated above, the detection and segmentation of PWML on cUS is extremely challenging. First, despite a higher resolution than MRI, US images are difficult to analyse because of their low contrast, the presence of speckle and the high variability related to the data acquisition process. Second, PWML are very tiny (in our dataset, the median volume of the lesions is less than $1mm^3$), resulting in the lesion area being much smaller than the whole brain area, so the data imbalance problem between positive and negative pixels makes it a huge challenge for deep learning models to learn. Finally, the presence of numerous artifacts in the brain, very similar to true lesions in terms of gray levels, sometimes location or even shape, makes the distinction between true lesions and false alarms even more complicated. Recently, Dakak et al. [9] developed an automatic approach to analyze discontinuities on industrial CT volumes by using a U-Net for segmentation followed by a classification network called CT-Casting Net. Although this paper does not discuss medical lesions, the common point is that discontinuities are very small and also similar to artefacts. Thus, it is expected that this 2-step approach can also be adequate in our present study.

In this work, we demonstrate that training the over-segmentation network with the expanded groundtruth helps to increase the number of detected lesions (higher recall). In addition, we introduce a novel architecture for multimodal classification, called 2.5D SC-Net, and show that performing a second stage of classification with more spatial context after segmentation improves the accuracy of the model by reducing the number of false alarms (higher precision).

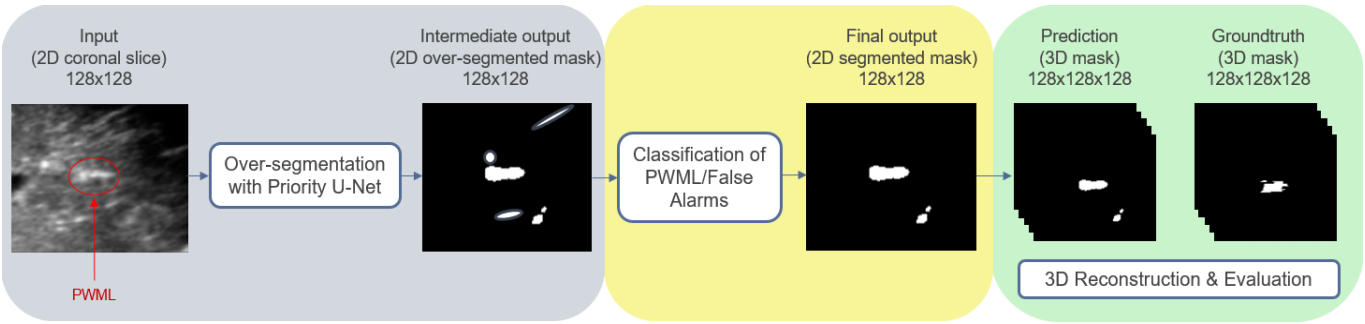

Fig. 1: Full pipeline : The over-segmentation of PWML with the Priority U-Net improves the recall but produces many false positives. A second stage of classification allows us to differentiate true lesions from false alarms and improves the precision of the intermediate output. 2D predictions are then concatenated and compared to the 3D groundtruth for final assessment.

2. METHODOLOGY

2.1. Over-segmentation with Priority U-Net

Before training the model, the lesions in the ground truth are expanded by aggregating the foreground pixels within a 5-slice sliding block along the coronal projection of the brain volume. It results in a label image with a higher percentage of foreground than the original label image, causing additional losses and helping to make training more effective.

Segmentation training is then performed on the modified ground truth with the Priority U-Net model (Fig. 1), which includes layers relying on 3D probabilistic maps derived from a spatial prior knowledge of PWMLs location and computed on the training dataset [7]. The output of the network usually includes many false positives, that will be removed during the next classification step.

2.2. Classification of PWML & False alarms

In parallel, we developed a novel multimodal classifier, called 2.5D SC-Net, trained on the joint fusion of the sagittal and coronal projections of the brain. Vignettes of size ($32\times32\times2$) centered on connected components (*CC*) from the 3D over-segmented mask are fed to the network as a 2-channel input (corresponding to each projection) to predict the class of the corresponding CC. The aim is to teach the network to differentiate true lesions from artifacts present in the brain, at a smaller scale but with more spatial context, while limiting the computational costs by using 2.5D instead of 3D volumes [6].

During training, features are extracted through convolutional blocks for each projection separatly, then joint fusion is performed by computing the weighted average on the flatten output of the feature extraction part of the network. Note that the weighting factors are fully learnt through the *Weighted Average Layer* (Fig 2).

During the testing phase, once the 3D intermediate mask is obtained after the first step of over-segmentation, 2D patches are extracted around the regions of interest (thumbnails from the image, centered around the connected components of the predicted mask) from the sagittal and coronal projections of the brain, concatenated and sent to the 2.5D SC-Net to identify the true PWML (Fig. 2). The intermediate mask is then corrected (*i.e.* the connected components predicted as false alarms are removed from the mask).

2.3. Class imbalance issue

As mentioned in the introduction, because the lesions are very small, the data imbalance problem between lesional pixels and background makes it a huge challenge for deep learning models to learn correctly. Therefore, the use of specific loss functions is considered to overcome this issue.

Self-Balancing Focal Loss: For a better training of the Priority U-Net, the Self-Balancing Focal Loss (SBFL) introduced by Liu et al. [6] is used in addition to the Dice Loss. The SBFL divides the whole loss into foreground loss and background loss part. It can automatically balance these losses and ultimately boosts the performance of the model.

$$SBFL_0 = -(1-y_{pred}) \times y_{pred} \times log(1-y_{pred}+\epsilon)$$

$$SBFL_1 = -y_{pred} \times (1-y_{pred}) \times log(y_{pred}+\epsilon)$$

$$\beta = \frac{0.4 \times sum(SBFL_0)}{sum(SBFL_0)+sum(SBFL_1)} + 0.5$$

$$SBFL = \beta \times SBFL_1 + (1-\beta) \times SBFL_0$$

Where y_{pred} is the intermediate output of the Priority U-Net and $SBFL_0$ and $SBFL_1$ are the focal loss of background and foreground pixels respectively. In order to segment PWML, we will focus on the segmentation of the lesion areas when balancing the loss of positive and negative samples. However, β should always be between 0.5 and 0.9 to ensure that the model does not only focus on the segmentation of positive areas. For that reason, we constrain β not to exceed 0.9 by applying a coefficient of 0.4 to the equation.

Weighted Cross-Entropy Loss: For the binary classification of the patches, we used the Weighted Cross-Entropy.

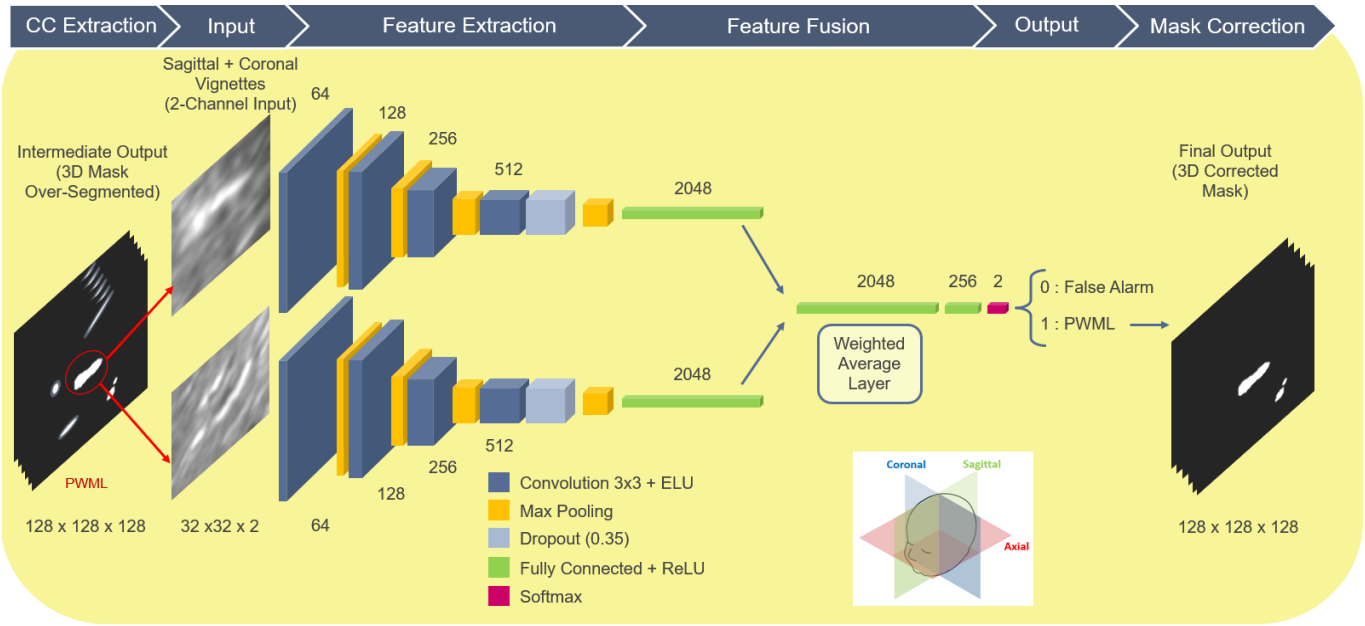

Fig. 2: Classification of PWML/False Alarms with 2.5D SC-Net. The classification helps to differentiate between actual lesions and other brain artifacts at a smaller scale. 2 projections of the brain (sagittal and coronal) are given as a 2-channel input to the network. The previous mask obtained after the over-segmentation is corrected with the output of the classification.

This loss function applies a scaling parameter α to Binary Cross Entropy, allowing us to penalize false positives or false negatives more harshly. When α is greater than 1, the model penalizes more on false negatives, hence helping increase the recall. On the other hand, when α is less than 1, the model penalizes more on false positives, hence increasing precision.

$$L_{BCE} = -y_{true} \times log(y_{pred}) - (1 - y_{true}) \times log(1 - y_{pred})$$

$$L_{WBCE} = -\alpha y_{true} \times log(y_{pred}) - (1 - y_{true}) \times log(1 - y_{pred})$$

Because our training dataset contained more negative than positive examples, but as we wanted to limit the trade-off between precision and recall, we performed a grid-search over 10 runs of 10-fold cross-validation and finally set α to 1.5.

3. EXPERIMENTS & RESULTS

3.1. Dataset

The 2D images are extracted from 54 reconstructed US brain volumes (including 29 with PWML) from 45 preterm babies whose mean age at birth was 31.6 ± 2.5 gestational weeks. In total, the dataset without preprocessing contains 414 lesions. The smallest lesion barely reaches $0.02mm^3$, while the largest is more than $41mm^3$. The median lesion size is $0.72mm^3$, which is extremely tiny. Besides, PWML have quite varied contrasts and do not really have specific shapes (punctate, ovoid or sometimes linear) 3. They are usually located in the center of hemispheres, near the lateral ventricles.

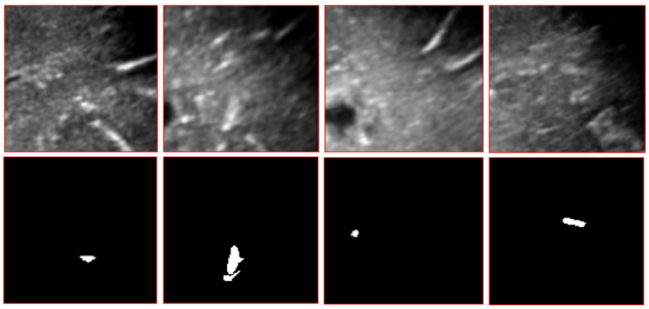

Fig. 3: PWML examples from the Brain US Dataset (Top row : US images. Bottom row : groundtruth masks). PWML have varied contrasts and shapes, and are often difficult to distinguish from peripheral vessels or arteries in cross-section.

3.2. Data Preprocessing

As the acquisition process of the ultrasound images is performed manually by the pediatrician along the anterior/posterior axis of the brain, the brain scan does not always results in the same number of dynamic sequences (DICOM). In order to recover a complete volume, we completed this process by a reconstruction algorithm [10]. After that, a first preprocessing phase consists of extracting a sub-volume of size $128 \times 128 \times 128$ in the top-right hemisphere, periventricular region of the brain for each patient.

In order to compare with MRI, a first filtering is performed on the size of the lesions for each volume to limit the number of lesions that are too small and to make the problem

less complex. As a result, only 90% of the lesional volume is kept for each patient, which allows us to get rid of the tiniest lesions, that are usually not even visible in the MRI.

For the first over-segmentation step, lesions in the ground truth are expanded as described in the section 2.1, and 2D images of size 128×128 containing PWML are extracted along the coronal projection of the brain to train the Priority U-Net. Horizontal flipping is randomly applied with a probability of 0.5 for data augmentation.

For the second classification step, sagittal and coronal patches of size 32×32 are extracted from the original volumes and used to train the 2.5D SC-Net (Fig. 2), including various examples with real lesions, artifacts/false alarms and normal examples. Random affine transforms (rotation, shearing, scaling, and translation) and flipping were employed for data augmentation.

3.3. Experimental Setup

The proposed pipeline was implemented in Python 3.9 with Keras and TensorFlow backend. All the models were trained and tested with GPU. For each model, we performed a 10-fold cross validation with 3 patients in the validation set and 22 patients in the training set.

The Priority U-Net was trained for 25 epochs with the Self-balancing Focal Loss and the Dice Loss, whereas the 2.5D SC-Net was trained for approximately 20 epochs using the Weighted-binary Cross-entropy Loss. The batch size is 4 for the segmentation network and 32 for the classification network. The initial learning rate was fixed at 10e-1 with the Adam optimizer and automatically decreased by a factor 0.1 when validation loss did not improved after 10 epochs.

3.4. Results

Table 1: Final results of the proposed approach (Priority U-Net with Highlighted Foreground + 2.5D SC-Net) compared to state-of-the-art algorithms. All these results are the medians of 10-folds cross-validation (± std).

Model	Recall	Precision	Dice *(TP)*
U-Net [8]	53.41 (± 12)	54.12 (± 13)	49.31 (± 15)
Priority U-Net [7]	58.22 (± 25)	47.76 (± 13)	52.76 (± 16)
Priority U-Net (HF)	**73.32** (± 12)	47.72 (± 9)	58.79 (± 13)
Priority U-Net (HF) + CT-Casting Net [9]	**68.91** (± 14)	**56.00** (± 7)	56.27 (± 14)
Priority U-Net (HF) + 2.5D SC-Net	**72.50** (± 16)	**56.30** (± 9)	**58.65** (± 7)

To quantitatively assess the quality of the PWML detection produced by the target pipeline, we employed 3 criteria to evaluate each model : the Recall and the Precision for the detection task but also the Dice on true positives *(TP)* to get an overview of the segmentation ability of the model. For each of these metrics, the value closer to 1 the better. The quantitative results are shown in Table 1. Note that the high variability in results may be explained by the limited number of patients available for each validation fold.

The evolution of the results after training the Priority U-Net with the expanded labels (*HF* for highlighted foreground) seems to support the fact that expanding the size of the lesion groundtruth during training actually helps the segmentation network to detect more PWML and smaller objects in general, significantly increasing the recall from 58 to 73%. After performing classification with the proposed 2.5D SC-Net, the precision improves from 48 to 56%, which compensates for the degradation observed with the over-segmentation. Besides, the Dice also increases from 53 to 59%. Please note that the network optimization was more focused on the detection task, and improvements are currently under study for the segmentation task.

In addition, we also show that giving more spatial context as a 2-channel input and implementing joint fusion in the multimodal classifier architecture also helps to preserve the recall because when classification was performed with the CT-Casting Net on a single projection, the recall dropped from 73 to 69% after classification, whereas it decreases by less than 1% with the 2.5D SC-Net.

As a result, our model achieves better performances for PWML detection in US images compared to other methods, with a higher sensitivity to very small lesions.

4. DISCUSSION AND CONCLUSION

In this paper, we first preprocessed the label images to get a higher percentage of foreground, before training a network to perform an over-segmentation of the original lesions. This first step was followed by a second step of multimodal classification to limit the number of missclassified artifacts in the brain. Our work shows that expanding the groundtruth allows a better detection of very small lesions such as PWML in US images. We also demonstrate that focusing on the lesions at a smaller scale and with more spatial context helps to improve the specificity of the model as well. At the end of the proposed pipeline, we reach a higher recall and precision (72% and 56% respectively) than those obtained with other state-of-the-art techniques.

Many previous studies have focused on MRI-based lesion detection, and very few have addressed problems on a scale as small as PWML. While relatively few have conducted this task on US images, this work highlights once again the possibility of detecting brain lesions through ultrasound imaging.

Nevertheless, we anticipate that results might be improved by enriching the database with more patients, or exploiting even more spatial context using 3D methods or attention modules. Improvements on the segmentation results are currently under study and will be the object of future works as well.

5. ACKNOWLEDGMENTS

This work was supported by the LABEX CELYA (ANR-10-LABX-0060) operated by the French National Research Agency (ANR) and the University of Lyon, within INSA Lyon.

6. COMPLIANCE WITH ETHICAL STANDARDS

The data from human subjects used in this work were obtained and treated in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committees of the institutions involved in creating the PWML database, from which these data were accessed. The authors have no relevant financial or non-financial interests to disclose.

7. REFERENCES

[1] Nguyen A.L.A., Ding Y., Suffren S., Londono I., Luck D., and Lodygensky G.A., "The brain's kryptonite: Overview of punctate white matter lesions in neonates," *International Journal of Developmental Neuroscience*, vol. 77, no. 1, pp. 77–88, October 2019.

[2] De Bruijn C.A.M., Di Michele S., Tataranno M.L., Ramenghi L.A., Rossi A., Malova M., Benders M., Van Den Hoogen A., and Dudink J., "Neurodevelopmental consequences of preterm punctate white matter lesions: a systematic review," *Pediatric Research*, September 2022.

[3] Jiao Z., Gao X., Wang Y., and Li J., "A parasitic metric learning net for breast mass classification based on mammography," *Pattern Recognition*, vol. 75, pp. 292–301, 2018.

[4] Liu Y., Li J., Wang Y., Wang M., Li X., Jiao Z.and Yang J., and Gao X., "Refined-segmentation r-cnn: A two-stage convolutional neural network for punctate white matter lesion segmentation in preterm infants," *arXiv e-prints*, vol. abs/1906.09684, pp. 1265–1274, June 2019.

[5] Mukherjee S., Cheng I., Miller S., Guo T., Chau V., and Basu A., "A fast segmentation-free fully automated approach to white matter injury detection in preterm infants," *Medical and Biological Engineering and Computing*, vol. 57, no. 1, pp. 71–87, January 2019.

[6] Liu Y., Li J., Wang M., Jiao Z., Yang J., and Li X., "Trident segmentation cnn: A spatiotemporal transformation cnn for punctuate white matter lesions segmentation in preterm neonates," *Med Biol Eng Comput*, 2019.

[7] Erbacher P., Lartizien C., Martin M., Foletto Pimenta P., Quetin P., and Delachartre P., "Priority u-net: Detection of punctuate white matter lesions in preterm neonate in 3d cranial ultrasonography," *Proceedings of the Third Conference on Medical Imaging with Deep Learning (MIDL)*, 2020.

[8] Ronneberger O., Fischer P., and Brox T., "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, pp. 234–241, 2015.

[9] Dakak A.R., Kaftandjian V., and Duvauchelle P.and Bouvet P., "Deep learning-based defect detection in industrial ct volumes of castings," *Insight Non-Destructive Testing and Condition Monitoring*, vol. 64, no. 11, November 2022.

[10] Martin M., Sciolla B., Sdika M., Wang X., Quetin P., and Delachartre P., "Automatic segmentation of the cerebral ventricle in neonates using deep learning with 3d reconstructed freehand ultrasound imaging," *IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4, 2018.