# DEEP LEARNING-BASED ASSESSMENT OF CEREBRAL MICROBLEEDS IN COVID-19

*Neus Rodeja Ferrer* ⋆  *Malini Vendela Sagar* †  *Kiril Vadimovic Klein* ⋆
*Christina Kruuse* †  *Mads Nielsen* ⋆  *Mostafa Mehdipour Ghazi* ⋆

⋆ Department of Computer Science, University of Copenhagen
† Department of Neurology, Copenhagen University Hospital

## ABSTRACT

Cerebral Microbleeds (CMBs), typically captured as hypointensities from susceptibility-weighted imaging (SWI), are particularly important for the study of dementia, cerebrovascular disease, and normal aging. Recent studies on COVID-19 have shown an increase in CMBs of coronavirus cases. Automatic detection of CMBs is challenging due to the small size and amount of CMBs making the classes highly imbalanced, lack of publicly available annotated data, and similarity with CMB mimics such as calcifications, irons, and veins. Hence, the existing deep learning methods are mostly trained on very limited research data and fail to generalize to unseen data with high variability and cannot be used in clinical setups. To this end, we propose an efficient 3D deep learning framework that is actively trained on multi-domain data. Two public datasets assigned for normal aging, stroke, and Alzheimer's disease analysis as well as an in-house dataset for COVID-19 assessment are used to train and evaluate the models. The obtained results show that the proposed method is robust to low-resolution images and achieves 78% recall and 80% precision on the entire test set with an average false positive of 1.6 per scan.

***Index Terms***— Deep learning, cerebral microbleeds, COVID-19, susceptibility-weighted imaging, precision-recall

## 1. INTRODUCTION

Cerebral Microbleeds (CMBs) are small foci of chronic blood products in normal brain tissue leaked from cerebral small vessels [1]. They are recognized as neurological findings in patients with cerebrovascular disease [2], dementia [3], normal aging [4], and more recently, COVID-19 [5, 6]. CMB size can range from 2 mm to 5-10 mm in diameter [1]. These haemosiderin deposits are superparamagnetic, i.e., they have high magnetic susceptibility. Therefore, they produce inhomogeneity in the magnetic field of the surrounding area, making the decay of the local magnetic resonance imaging (MRI) signal faster. Hence, CMBs are typically seen as hypointense foci in MRI sequences sensitive to susceptibility effects such as gradient-recalled echo (GRE) T2*-weighted imaging (T2*WI) or susceptibility-weighted imaging (SWI).

Compared to GRE T2*WI, SWI increases the CMB detection rate from 20% to 40% [7].

Deep learning methods applied to CMB segmentation [8, 9, 10, 11] mostly consist of two-stage pipelines to search for potential positive objects and to differentiate true positives (TP) from false positives (FP). The state-of-the-art methods in [8] uses a two-stage approach based on 3D-FRST [12] and 3D-ResNet [13] trained on high-resolution images and phases to achieve 95% sensitivity, 71% precision, and an average FP of 1.6 per scan. However, the accuracies drop to 73% sensitivity, 67% precision, and an average FP of 1.9 per scan, when applied to low-resolution scans [11]. Due to learning from only 1-2 different datasets, these methods usually fail to generalize to data with other characteristics obtained in uncontrolled clinical settings.

Data augmentation is a common technique for addressing some of the domain shift issues in medical imaging [14], but it is often not enough to cope with all differences in demographics, comorbidities, and acquisition parameters to achieve high generalization accuracy. Therefore, sometimes there is a need for including data with high variability in the training set, especially when the training data is small. However, this requires the acquisition and annotation of some MRI scans from different sites and parameters, which is a time-consuming and labor-intensive task and can cause data privacy challenges.

In this work, a robust deep-learning model is trained and applied to assess COVID-19 CMBs in a retrospective cohort study data, obtained from the capital region of Denmark [15] and weakly labeled as case-control through matching. MultiResUNet architecture [16] is trained on different heterogeneous datasets using the three orthogonal planes of each scan for CMB segmentation and detection. Two public datasets assigned for normal aging, stroke, and Alzheimer's disease analysis [17, 10] as well as an in-house dataset for COVID-19 assessment are used to train and evaluate the models. The obtained results show that the proposed method is robust to low-resolution images and achieves 78% sensitivity and 80% precision on the entire test set with an average FP of 1.6 per scan. To show the stability of the method on heterogeneous data, the pipeline is successfully applied to distinguish between the two weakly labeled groups.

## 2. METHOD

Our study design comprises strategic data selection and pre-processing, data augmentation and training deep learning models, postprocessing for classifier fusion, and standard testing for the detection and group differentiation capability. The utilized deep learning framework includes MultiResUNet models [16] trained individually on axial, sagittal, and coronal views, which are later fused for the 3-views agreement.

### 2.1. Data processing

Two public datasets [17, 10] and in-house data with SWI images are used in this study; the first data (DS1) is obtained from [17] and contains 57 scans from 30 patients with real CMBs (DS1r) and 3,700 scans from 118 patients with synthetic CMBs (DS1s); the second data (DS2) is obtained from [10] and contains 20 scans from 20 subjects with real CMBs; the third dataset (DS3) is obtained from different hospitals of the capital region of Denmark and contains 40 scans from 40 patients with CMBs and 22 scans from 22 patients with no CMBs (DS3n). Moreover, 40 scans with no CMBs, matched with those 40 scans with CMBs, are used for testing the group distinguishability of the model.

The DS1 scans were acquired in a spatial resolution of $0.93 \times 0.93 \times 1.75$ mm$^3$ using a 3T scanner with 20 ms echo time. The DS2 scans were obtained in a resolution of $0.45 \times 0.45 \times 1$ mm$^3$ using a 3T scanner with 24 ms echo time. Finally, the DS3 scans were captured from different sites and scanners (1.5T and 3T) in various resolutions (from $0.2 \times 0.2 \times 1$ mm$^3$ to $1 \times 1 \times 6$ mm$^3$), to increase our model's robustness and generalization accuracy.

Since the DS1 and DS2 only provide the CMB center positions on the SWI images, to train a multiview deep network, we need to create volumetric masks for the CMBs. We select patches with an isotropic voxel size of 10 mm$^3$ as the maximum CMB size according to [1] around the CMB centers and estimate the fraction of microbleeds for each pixel using

$$\alpha_{\text{CMB}} = \frac{I_{pixel} - I_{mean}}{I_{center} - I_{mean}},$$

where $I_{pixel}$ is the current pixel's intensity, $I_{center}$ is the center pixel's intensity, and $I_{mean}$ is the mean intensity of the surrounding pixels in a 2 mm neighborhood. Pixels with $\alpha > 0.65$ and $\alpha > 0.52$ were labeled as CMBs in DS1 and DS2, respectively. The different $\alpha$ values were obtained due to the difference in echo time or tissue contrast.

The DS3 scans were annotated by clinical experts and weakly labeled by probabilities estimated using the electronic healthcare records [15] based on logistic regression and odds ratio according to the UK Biobank study [18]. For testing group differentiation capability, we selected 40 scans with $p_{\text{CMB}} > 0.3$, annotated them, and matched them on 40 unannotated scans with $p_{\text{CMB}} < 0.01$ based on the magnetic field strength, echo time, slice thickness, and scanner model.

### 2.2. CMB segmentation and detection

A 3D framework is built based on 2D convolutional neural networks (CNNs) to segment and count the number of CMBs. The pipeline consists of MultiResUNet models [16], trained individually on three orthogonal views of the MRI scans and fused later for the mask volume reconstruction. The volumes are first reshaped to $256 \times 256 \times 256$ with an isotropic voxel size of 1 mm$^3$. To benefit from the 3D information, thickened 2D slices are then formed by three consecutive images of each MRI scan as inputs to the 3-channel CNNs.

Several MRI-specific transforms are randomly applied to the training images for data augmentation [14], including elastic deformation, bias field, rotation, flipping, blurring, motion ghosting, Gibbs ringing, and additive-multiplicative noise. Moreover, intensity normalization and contrast adjustment are applied to the network inputs. Finally, the output probability maps of the central slices from different views are fused by multiplying the three maps to obtain the label masks based on a full agreement between the three predictions. Any quantitative measures like the counts, position, and size of the detected CMBs can then be computed using the output map.

## 3. EXPERIMENTS AND RESULTS

### 3.1. Experimental setup

Scans from different subjects are randomly partitioned into 70%, 10%, and 20% for training, validation, and testing. To balance different datasets for training, only 10% of DS3n slices and 5% of DS1s slices are used, optimized using a grid search method on the validation set.

The networks, pretrained on medical data for skin lesion segmentation [16], are finetuned for CMB segmentation using the Adam Optimizer with cross-entropy loss function and a weight decay of $10^{-6}$. The base learning rate is set to $5 \times 10^{-4}$ and decreased by 10% every 10 epochs. The hyperparameter values are obtained using a grid search algorithm on the validation set.

### 3.2. Results and discussion

The test prediction results of our model on different datasets are summarized in Table 1. As can be seen, our framework achieves a Dice similarity score (DSC) of 0.79, sensitivity of 0.78, precision of 0.80, and an average FP of 1.64 per scan. Note that CMB segmentation and detection in the wild (DS3) is very challenging, as the data is acquired from uncontrolled clinical settings with different imaging protocols. The high variations in the image parameters, anatomies, and patient conditions can affect the model's accuracy and robustness. The proposed method, trained on a few annotated DS3 scans, has achieved comparable results in the DS3 test set.

The test results of the DS3 show lower sensitivity and higher precision. The higher FN rate can be due to the dif-

**Table 1**. CMB segmentation and detection results on different test sets using the proposed framework.

| Dataset | TP/scan | FP/scan | FN/scan | DSC | Sensitivity | Precision |
|---------|---------|---------|---------|-----|-------------|-----------|
| DS1r | 3.58 | 2.00 | 1.08 | 0.70 | 0.77 | 0.64 |
| DS1s | 8.11 | 1.75 | 1.72 | 0.82 | 0.83 | 0.82 |
| DS2 | 1.00 | 0.00 | 1.00 | 0.67 | 0.50 | 1.00 |
| DS3 | 8.57 | 1.43 | 5.00 | 0.73 | 0.63 | 0.86 |
| DS3n | 0.00 | 0.00 | 0.00 | 1.00 | NA | NA |
| All | 6.75 | 1.64 | 1.92 | 0.79 | 0.78 | 0.80 |

**Table 2**. Test CMB segmentation and detection results of different state-of-the-art models on various scan resolutions.

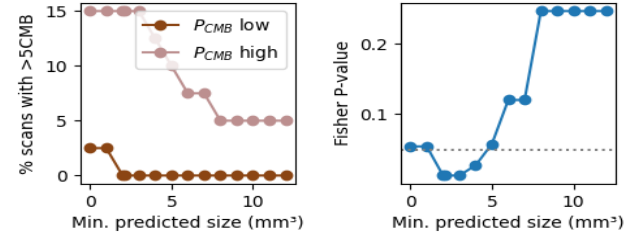| Model | Type | Phase | FP/scan | Sensitivity | Precision |
|-------|------|-------|---------|-------------|-----------|
| YOLO + 3D-CNN [11] | LR | Yes | 1.89 | 0.78 | 0.67 |
| YOLO + 3D-CNN [11] | HR | Yes | 1.42 | 0.88 | 0.62 |
| 3D-FRST + 3D-ResNet [8] | HR | Yes | 1.6 | 0.95 | 0.71 |
| 3D-FCN + 3D-CNN [10] | HR | No | 2.74 | 0.91 | 0.44 |
| The proposed | LR | No | 1.92 | 0.80 | 0.80 |
| The proposed | HR | No | 0.30 | 0.49 | 0.87 |

ference in CMB distribution per scan compared to the other datasets, where many scans could have a single CMB while many others have a large number of CMBs. The higher precision can also be due to the higher resolution of DS3 scans compared to the alternatives. In addition, as expected, no CMBs were detected on DS3n, indicating that our method can be applied to distinguish between healthy and diseased scans. Finally, the lower accuracy obtained on DS2 can be because it has the smallest dataset size in terms of the number of scans and CMBs per scan.

Due to the different nature of the used datasets in the literature, we classify them as low-resolution (LR) and high-resolution (HR) SWI-based studies for a fair comparison. The HR studies on CMB segmentation involve [11, 8, 10] with scans of spacings $0.5 \times 0.5 \times 2$ mm$^3$, $0.45 \times 0.5 \times 1.2$ mm$^3$ to $0.54 \times 1.07 \times 2.65$ mm$^3$, and $0.45 \times 0.45 \times 2$ mm$^3$, respectively, while the LR one is obtained from [11] with scans of spacing $0.8 \times 0.8 \times 2$ mm$^3$. In contrast, our LR test set contains scans with spacings of $0.89 \times 0.89 \times 2$ mm$^3$ (DS3) and $0.93 \times 0.93 \times 1.75$ mm$^3$ (DS1), while the HR test set includes scans with spacings of $0.3 \times 0.3 \times 2$ mm$^3$ to $0.34 \times 0.34 \times 4$ mm$^3$ (DS3) and $0.45 \times 0.45 \times 1$ mm$^3$ (DS2). It should also be noted that [11, 8] benefit from the phase images for segmentation. Table 1 shows the prediction results of different models on various scan resolutions of the test sets. As seen, our model achieves better results on the LR subset while obtaining poor sensitivity on the HR test set. The reason for such behavior compared to the state-of-the-art could be because the scan resolutions in different studies do not match exactly, our model is trained on much more LR scans (83% LR scans vs. 17% HR scans), and we do not use phase images for CMB segmentation and detection. Moreover, the HR dataset of DS2 used in this study only includes a small publicly available subset of the one used in [10].

Finally, to see how the weakly labeled groups of our COVID-19 study are statistically significantly different in terms of the number of detected CMBs, we use two unannotated scan subsets labeled as non-CMB ($p_{CMB} < 0.01$) and CMB ($p_{CMB} > 0.3$). On average per scan, we found 0.62 CMB in the non-CMB scans and 2.60 CMB in the CMB scans, which were statistically significantly different with a $p = 0.001$ using the Wilcoxon signed-rank test. Next, we set clinical criteria for illness based on 5 CMBs per scan and



**Fig. 1**. The CMB frequency of the two groups and their statistical significance for the filtered CMB size. As seen, the two groups with more than 5 CMB are significantly different for the CMB size smaller than 5 mm$^3$.

attempt to distinguish between the groups with more or less than 5 CMBs. To minimize the FP effects, we filter the predicted CMBs by the minimum clinical size of 4.2 mm$^3$ with a 2 mm diameter. Figure 1 shows the difference between the two groups for the filtered CMB size, where the CMB frequency of the two filtered groups is statistically significantly different with a $p = 0.013$ using Fisher's exact test.

## 4. CONCLUSION

In this work, we developed a deep learning-based framework based on MultiResUNet architecture using different heterogeneous datasets and MRI-specific data augmentation for CMB segmentation and detection and assessed COVID-19 CMBs in a retrospective cohort study. The proposed pipeline was tested on uncontrolled clinical data and outperformed the state-of-the-art methods, especially when applied to low-resolution images. Moreover, the proposed pipeline was able to distinguish between two weakly labeled groups of the heterogeneous COVID-19 data, and the results were statistically significantly different in terms of the number of detected CMBs.

## 5. ACKNOWLEDGMENTS

# 6. REFERENCES

[1] S. M. Greenberg, M. W. Vernooij, C. Cordonnier, A. Viswanathan, R. Al-Shahi Salman, S. Warach, L. J. Launer, M. A. Van Buchem, and M. M. Breteler, "Cerebral microbleeds: A guide to detection and interpretation," *The Lancet Neurology*, vol. 8, no. 2, pp. 165–174, 2009.

[2] A. Charidimou and D. J. Werring, "Cerebral microbleeds and cognition in cerebrovascular disease: An update," *Journal of the Neurological Sciences*, vol. 322, no. 1-2, pp. 50–55, 2012.

[3] C. Qiu, M. F. Cotch, S. Sigurdsson, P. V. Jonsson, M. K. Jonsdottir, S. Sveinbjörnsdottir, G. Eiriksdottir, R. Klein, T. B. Harris, M. A. Van Buchem, V. Gudnason, and L. J. Launer, "Cerebral microbleeds, retinopathy, and dementia: The AGES-Reykjavik study," *Neurology*, vol. 75, no. 24, pp. 2221–2228, 2010.

[4] M. H. Zwartbol, R. Ghaznawi, K. Blom, H. J. Kuijf, T. Witkamp, J. Hendrikse, and M. I. Geerlings, "Cerebral microbleeds on 7 tesla MRI in preclinical Alzheimer's disease: The Medea-7T study," *Alzheimer's & Dementia*, vol. 16, no. S4, pp. e044763, 2020.

[5] F. Lersy, T. Willaume, J. C. Brisset, O. Collange, J. Helms, F. Schneider, A. Chammas, A. Willaume, N. Meyer, M. Anheim, F. Cotton, and S. Kremer, "Critical illness-associated cerebral microbleeds for patients with severe COVID-19: etiologic hypotheses," *Journal of Neurology*, vol. 268, no. 8, pp. 2676–2684, 2021.

[6] E. Lin, J. E. Lantos, S. B. Strauss, C. D. Phillips, T. R. Campion, B. B. Navi, N. S. Parikh, A. E. Merkler, S. Mir, C. Zhang, H. Kamel, M. Cusick, P. Goyal, and A. Gupta, "Brain imaging of patients with COVID-19: Findings at an academic institution during the height of the Outbreak in New York City," *American Journal of Neuroradiology*, vol. 41, no. 11, pp. 2001–2008, 2020.

[7] S. Haller, M. W. Vernooij, J. P. Kuijer, E. M. Larsson, H. R. Jäger, and F. Barkhof, "Cerebral microbleeds: Imaging and clinical significance," *Radiology*, vol. 287, no. 1, pp. 11–28, 2018.

[8] S. Liu, D. Utriainen, C. Chai, Y. Chen, L. Wang, S. K. Sethi, S. Xia, and E. M. Haacke, "Cerebral microbleed detection using susceptibility weighted imaging and deep learning," *NeuroImage*, vol. 198, pp. 271–282, 2019.

[9] Y. Chen, J. E. Villanueva-Meyer, M. A. Morrison, and J. M. Lupo, "Toward automatic detection of radiation-induced cerebral microbleeds using a 3D deep residual network," *Journal of Digital Imaging*, vol. 32, no. 5, pp. 766–772, 2019.

[10] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P. A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.

[11] M. A. Al-masni, W. R. Kim, E. Y. Kim, Y. Noh, and D. H. Kim, "Automated detection of cerebral microbleeds in MR images: A two-stage deep learning approach," *NeuroImage: Clinical*, vol. 28, 2020.

[12] G. Loy and A. Zelinsky, "Fast radial symmetry transform for detecting points of interest," in *Computer Vision – ECCV 2002*, Berlin, Heidelberg, 2002, pp. 358–368, Springer Berlin Heidelberg.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on CVPR*, 2016.

[14] M. Mehdipour Ghazi and M. Nielsen, "FAST-AID Brain: Fast and accurate segmentation tool using artificial intelligence developed for brain," *arXiv preprint arXiv:2208.14360*, 2022.

[15] E. Jimenez-Solem, T. S. Petersen, C. Hansen, C. Hansen, C. Lioma, C. Igel, et al., "Developing and validating COVID-19 adverse outcome risk prediction models from a bi-national European cohort of 5594 patients," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.

[16] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.

[17] S. Momeni, A. Fazlollahi, P. Yates, C. Rowe, Y. Gao, A. W. C. Liew, and O. Salvado, "Synthetic microbleeds generation for classifier training without ground truth," *Computer Methods and Programs in Biomedicine*, vol. 207, pp. 106127, 2021.

[18] D. Lu, J. Liu, A. MacKinnon, D. Tozer, and H. Markus, "Prevalence and risk factors of cerebral microbleeds: Analysis from the UK Biobank," *Neurology*, vol. 97, no. 15, pp. e1493–e1502, 2021.