6. Supplementary Material

6.1. CMBs per TBI case

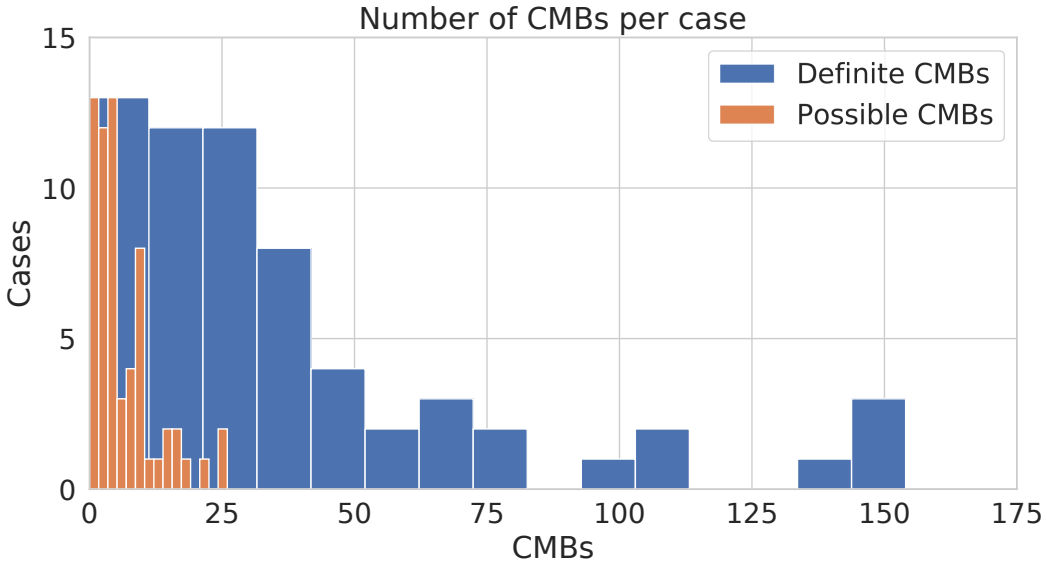

Fig. 7: Illustration of how many definite and possible CMBs were identified in cases of our dataset. A case is a single SWI of an individual.

6.2. *CMB Morphology*

6.2.1. *Metric Definitions*

- Diamater$_{max}$ is the length (in mm) of the longest line which can be drawn between any two voxels of the same CMB.
- Voxel size is nearly isotropic [0.98×0.98×1.00]. Therefore, Volume (in mm^3) is simply calculated as the number of voxels in any given CMB.
- We define the metric *Sphericalness* as a measure of how spherical an ellipsoid, in our case CMB, is. This is a simplified assumption of CMB shapes. It measures how much volume the CMB occupies in relation to a perfect sphere with Diamater$_{max}$. Linear CMBs tend to 0, while a perfectly spherical CMB has a Sphericalness of 1.0. Equation:

$$\text{Sphericalness} = \frac{Volume}{\frac{4\pi}{3} * (\frac{size}{2})^3}$$

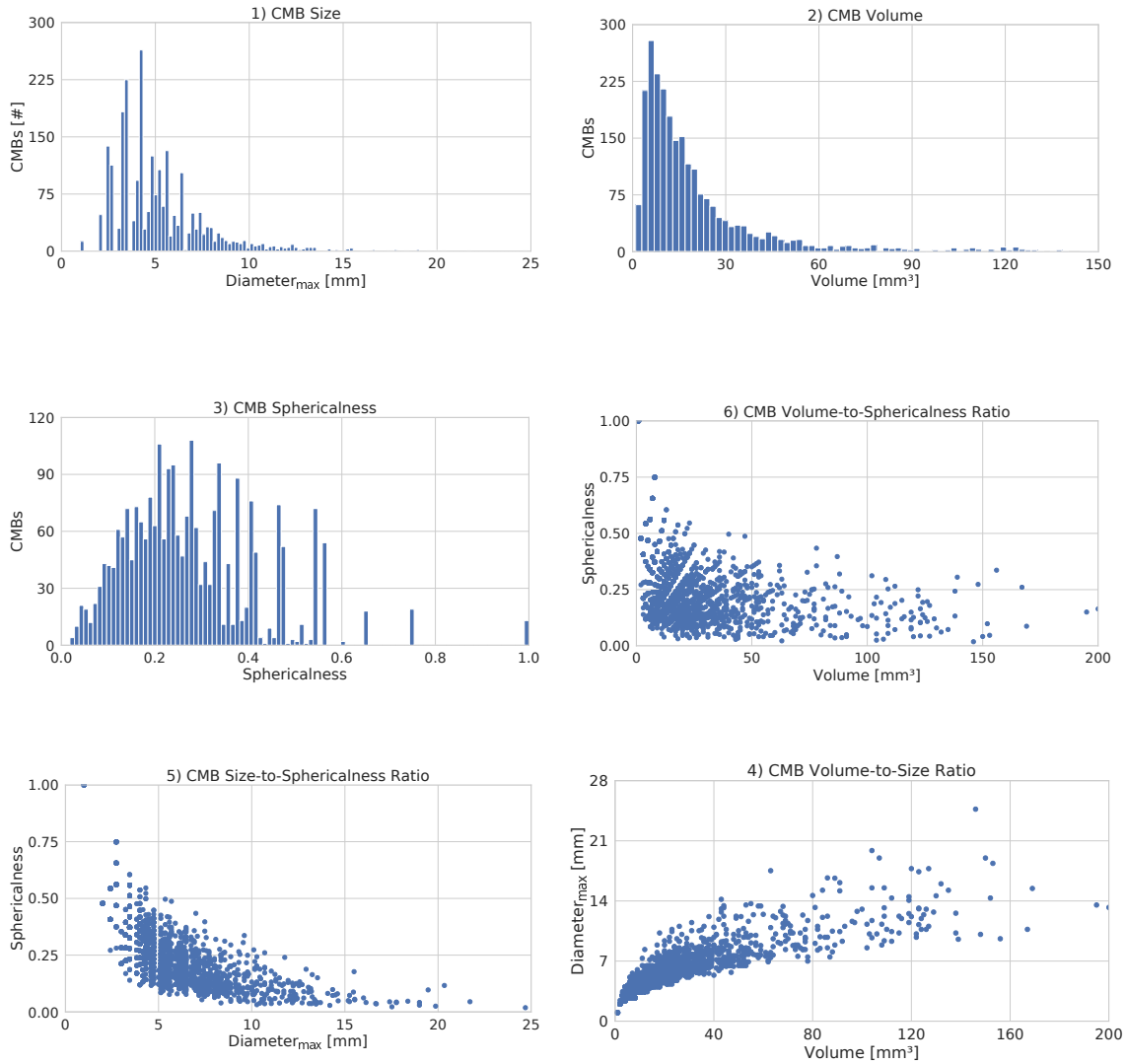

Fig. 8: 1.) Distribution of (definite) CMB sizes in the reference standard. 2.) Distribution of (definite) CMB volumes. 3.) Distribution of sphericalness of (definite) CMBs. 4.) Volume-to-size ratio of (definite) CMBs. 5.) Size-to-sphericalness ratio of (definite) CMBs. 6.) Volume-to-sphericalness ratio of (definite) CMBs.

6.3. Majority Vote

In Van den Heuvel et al. (2016), if $\geq$ 3 observers agreed on a lesion as *definite* it was considered a *definite* CMB. Alternatively, if $\geq$ 3 observers considered a focus *definite* or at least *possible*, it was considered a *possible* CMB in the majority vote. This thresholding was possible by automatically region-growing the annotated CMBs with their point annotations as seed.
For this study, we assigned each region-grown lesion a numerical value:

- *definite* = 1.0
- *possible* = 0.5.

Then the six individual maps were summed and thresholded. Since we used a majority vote of 6 for model evaluation, and a majority vote of 5 to rate the observers the thresholds were different. The lesion assessment was

- *definite*, if $\sum \geq 4.0$ and
- *possible*, if $\sum \geq 2.0$ in the case of the majority vote for model evaluation;

- *definite*, if $\sum \geq 3.0$ and
- *possible*, if $\sum \geq 1.5$ in the case of the majority vote for observer evaluation.

A problem with summing region-grown maps is that clusters of CMBs can "bleed" into each other, i.e., 2 or more nearby CMBs would appear as a single entity in the reference standard. Therefore we compared the clustered components with the summation map, and manually separated CMBs if the underlying summation map suggested multiple lesions.

6.4. Normalization

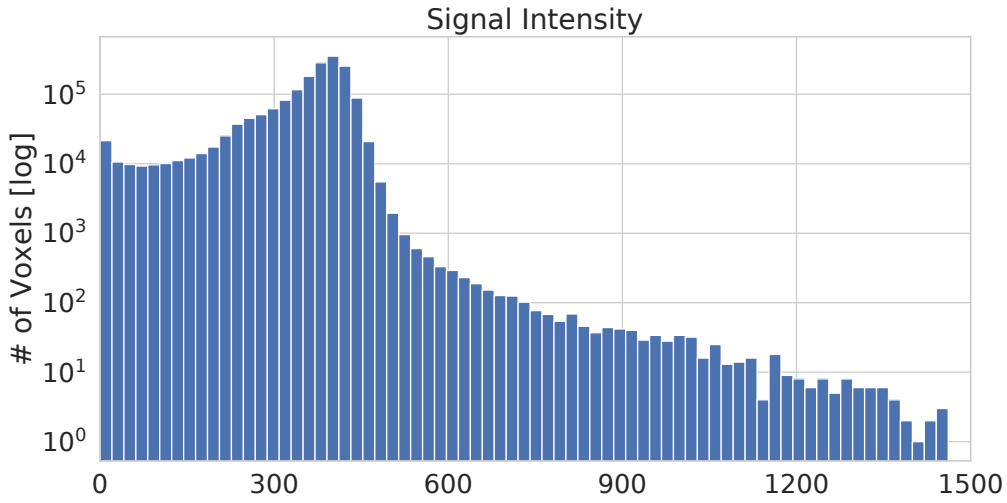

Fig. 9: Signal intensity distribution of an exemplary SWI scan, displayed on a log-scale. Only voxels within the brain mask were considered. The main body of the histogram corresponds to the brain parenchyma.

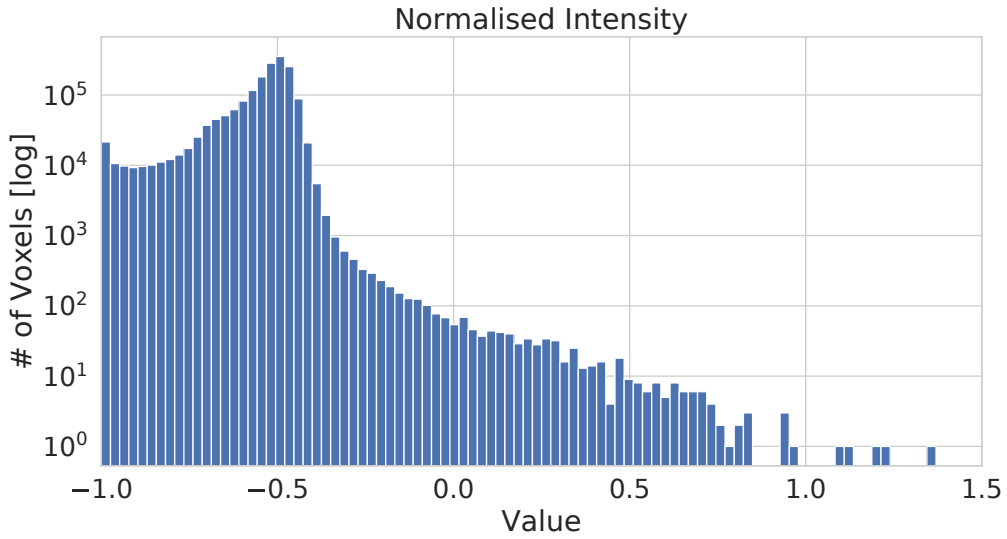

Fig. 10: Value distribution of the normalised SWI scan, displayed on a log-scale. Histogram peak is assigned the value −0.5, lower bound is −1.0, there is no upper bound.

6.5. Model Architectures

Table 4: Architecture of the Patch-CNN classification model - every CONV block consists of a 3D convolution layer, followed by 3D batch normalization layer.

Layer	Kernel size	Stride Stride	Output size	Feature volumes	Activation Function
INPUT	-	-	21^3	1	-
CONV-1	3^3	1	19^3	64	ReLU
CONV-2	3^3	1	17^3	64	ReLU
CONV-3	3^3	1	15^3	64	ReLU
CONV-4	3^3	1	13^3	128	ReLU
CONV-5	3^3	1	11^3	128	ReLU
CONV-6	3^3	1	9^3	128	ReLU
CONV-7	3^3	1	7^3	256	ReLU
CONV-8	3^3	1	5^3	256	ReLU
CONV-9	3^3	1	3^3	256	ReLU
DROPOUT	-	-	-	50%	-
CONV-10	3^3	1	1^3	2	Softmax

Table 5: Architecture of the Segmentation-CNN segmentation model - every CONV block consists of a 3D convolution layer, followed by 3D batch normalization layer.

Layer	Kernel size	Stride Stride	Output size	Feature volumes	Activation Function
INPUT	-	-	21^3	1	-
CONV-1	3^3	1	19^3	64	ReLU
CONV-2	3^3	1	17^3	64	ReLU
CONV-3	3^3	1	15^3	64	ReLU
CONV-4	3^3	1	13^3	128	ReLU
CONV-5	3^3	1	11^3	128	ReLU
CONV-6	3^3	1	9^3	128	ReLU
CONV-7	3^3	1	7^3	256	ReLU
CONV-8	3^3	1	5^3	256	ReLU
CONV-9	3^3	1	3^3	256	ReLU
CONV-10	3^3	1	1^3	2	Softmax

Table 6: Architecture of the U-Net segmentation model - every CONV block consists of a 3D convolution layer, followed by 3D batch normalization layer.

Pathway	Layer	Input Input	Kernel size	Stride Stride	Input size ⊗ features	Output size	Feature volumes	Activation Function
	INPUT	-	-	-	$68^3 \otimes 1$	68^3	1	-
Encoding	CONV-1	INPUT	3^3	1	$68^3 \otimes 1$	66^3	32	ReLU
	CONV-2	CONV-1	3^3	1	$66^3 \otimes 32$	64^3	32	ReLU
	CONV-3	CONV-2	3^3	1	$64^3 \otimes 32$	62^3	64	ReLU
	CONV-4	CONV-3	3^3	1	$62^3 \otimes 64$	60^3	64	ReLU
	POOL-4	CONV-4	3^3	1	$60^3 \otimes 64$	30^3	64	-
	CONV-5	POOL-4	3^3	2	$30^3 \otimes 64$	28^3	128	ReLU
	CONV-6	CONV-5	3^3	1	$28^3 \otimes 128$	26^3	128	ReLU
	POOL-6	CONV-6	3^3	2	$26^3 \otimes 128$	13^3	128	-
	CONV-7	POOL-6	3^3	1	$13^3 \otimes 128$	11^3	128	ReLU
	CONV-8	CONV-7	3^3	1	$11^3 \otimes 128$	9^3	128	ReLU
Skip Connections	CONV-4-SKIP	CONV-4 (cropped)	1^3	1	$28^3 \otimes 64$	28^3	64	ReLU
	CONV-6-SKIP	CONV-6 (cropped)	1^3	1	$18^3 \otimes 128$	18^3	128	ReLU
Decoding	UPSAMPLE-1	CONV-8	2^3	-	$9^3 \otimes 1$	18^3	128	-
	CONV-9	UPSAMPLE-1 ⊗ CONV-6-SKIP	3^3	1	$18^3 \otimes 256$	16^3	128	ReLU
	CONV-10	CONV-9	3^3	1	$16^3 \otimes 128$	14^3	128	ReLU
	UPSAMPLE-2	CONV-10	2^3	-	$14^3 \otimes 128$	28^3	128	-
	CONV-11	UPSAMPLE-2 ⊗ CONV-4-SKIP	3^3	1	$28^3 \otimes 196$	26^3	128	ReLU
	CONV-12	CONV-9	3^3	1	$26^3 \otimes 128$	24^3	128	ReLU
	CONV-13	CONV-12	3^3	1	$24^3 \otimes 128$	22^3	128	ReLU
	CONV-14	CONV-13	3^3	1	$22^3 \otimes 128$	20^3	2	Softmax

6.6. Patient Demographics

Table 7: Patient demographics and counts of definite and possible CMBs in the Full Segmentation for the training and validation datasets.

				Scan 1		Scan 2	
Patient	Sex	Age	Set	#def CMB_{FS}	#pos CMB_{FS}	#def CMB_{FS}	#pos CMB_{FS}
TBI#01	F	27	TRAINING	26	2	30	3
TBI#02	M	22	TRAINING	154	14	149	0
TBI#03	M	27	TRAINING	9	0	N.U.	N.U.
TBI#05	F	23	TRAINING	38	0	26	1
TBI#09	F	38	TRAINING	12	5	9	5
TBI#13	F	20	TRAINING	24	3	25	5
TBI#15	M	50	TRAINING	10	7	7	5
TBI#16	M	25	TRAINING	41	26	43	26
TBI#17	F	46	TRAINING	22	5	19	5
TBI#18	M	60	TRAINING	46	7	33	6
TBI#19	M	61	TRAINING	10	8	N.U.	N.U.
TBI#22	M	43	TRAINING	33	6	28	11
TBI#23	M	48	TRAINING	112	9	N.U.	N.U.
TBI#24	M	18	TRAINING	57	9	80	16
TBI#26	M	19	TRAINING	26	4	37	0
TBI#28	M	31	TRAINING	21	2	16	6
TBI#29	M	57	TRAINING	69	9	N.U.	N.U.
TBI#30	F	22	TRAINING	70	4	13	4
TBI#33	F	69	TRAINING	12	0	N.U.	N.U.
TBI#35	F	61	TRAINING	18	0	14	2
TBI#37	M	58	TRAINING	6	2	N.U.	N.U.
TBI#41	M	26	TRAINING	93	17	103	13
TBI#43	F	21	TRAINING	N.U.	N.U.	146	7
TBI#44	M	21	TRAINING	11	4	15	3
CONTROL#04	F	24	TRAINING	0	0	-	-
CONTROL#08	M	21	TRAINING	0	0	-	-
CONTROL#11	F	45	TRAINING	0	0	-	-
CONTROL#17	M	18	TRAINING	0	0	-	-
TBI#04	F	58	VALIDATION	7	0	N.U.	N.U.
TBI#07	M	19	VALIDATION	140	15	N.U.	N.U.
TBI#10	M	47	VALIDATION	8	2	N.U.	N.U.
TBI#11	M	27	VALIDATION	N.U.	N.U.	8	1
TBI#12	M	20	VALIDATION	43	10	N.U.	N.U.
TBI#14	M	30	VALIDATION	1	1	N.U.	N.U.
TBI#21	F	60	VALIDATION	9	2	N.U.	N.U.
TBI#27	F	21	VALIDATION	3	0	N.U.	N.U.
TBI#32	M	47	VALIDATION	N.U.	N.U.	44	9
TBI#36	M	50	VALIDATION	26	22	N.U.	N.U.
TBI#40	M	66	VALIDATION	N.U.	N.U.	16	4
TBI#40	M	26	VALIDATION	75	10	N.U.	N.U.
CONTROL#02	M	50	VALIDATION	0	0	-	-
CONTROL#05	M	31	VALIDATION	0	0	-	-
CONTROL#09	M	28	VALIDATION	0	0	-	-
CONTROL#15	M	62	VALIDATION	0	0	-	-

CMB: Cerebral Microbleed, #def: number of definite CMBs, #pos: number of possible CMBs, *FS*: Full Segmentation, N.U.: Not used in this study.

Table 8: Patient demographics and counts of definite and possible CMBs in the Full Segmentation and the Reference Standard for the test dataset.

Patient	Sex	Age	Set	#def CMB_{FS}	#pos CMB_{FS}	#def CMB_{RS}	#pos CMB_{RS}
TBI#06	M	54	TEST	17	2	2	10
TBI#08	M	48	TEST	28	2	19	25
TBI#20	F	40	TEST	64	18	30	38
TBI#25	M	23	TEST	27	4	5	15
TBI#31	M	22	TEST	62	10	19	37
TBI#35	M	21	TEST	33	9	14	18
TBI#38	M	38	TEST	24	1	7	12
TBI#39	F	19	TEST	36	2	15	22
TBI#42	M	20	TEST	17	4	5	6
TBI#45	F	25	TEST	36	5	22	23
CONTROL#01	F	22	TEST	0	0	0	0
CONTROL#03	M	30	TEST	0	0	0	0
CONTROL#06	M	49	TEST	0	0	0	0
CONTROL#07	F	22	TEST	0	0	0	0
CONTROL#10	F	26	TEST	0	0	0	0
CONTROL#12	M	62	TEST	0	0	0	0
CONTROL#13	F	53	TEST	0	0	0	0
CONTROL#14	M	34	TEST	0	0	0	0
CONTROL#15	M	62	TEST	0	0	0	0
CONTROL#18	M	21	TEST	0	0	0	0

CMB: Cerebral Microbleed, #def: number of definite CMBs, #pos: number of possible CMBs, *FS*: Full Segmentation, *RS*: Reference Standard.

6.7. Post-processing

Figure 11 illustrates that our segmentation methods (Segmentation-CNN, U-Net) predict border gradients which allows for separation of close CMBs, instead polarized/binary predictions. This is due to the inclusion of boundary loss in their training.

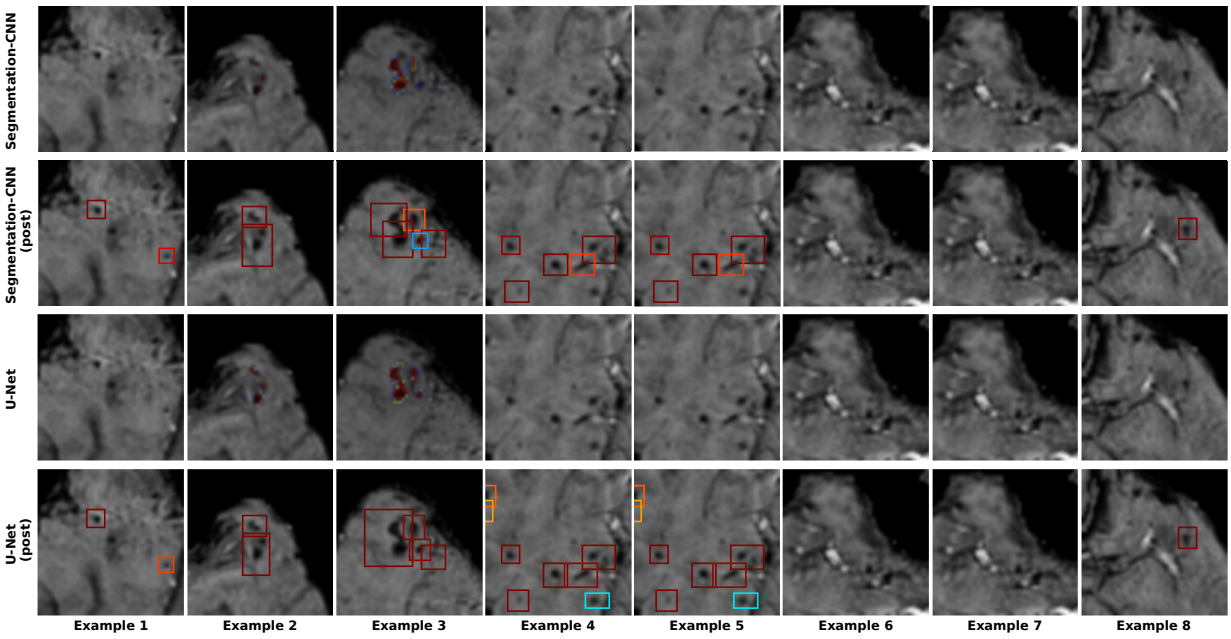

Fig. 11: Illustration of how post-processing separates large predictions with uneven probability distributions into individual lesions.

Figure 13 show the benefit of the proposed post-processing procedure on the performance of the Segmentation-CNN and U-Net. The separation of predictions into individual lesions improves the sensitivity.

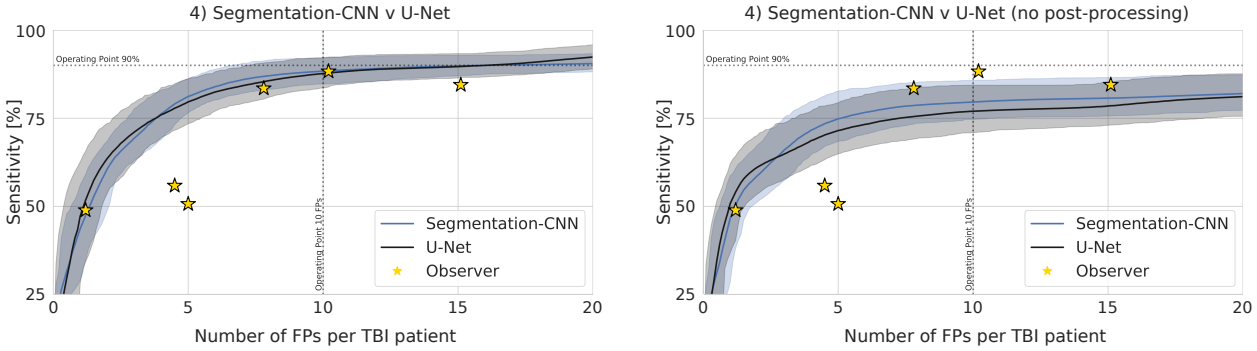

Fig. 12: Performance (bootstrapped at 1000 random samples of the available test set) of individual models in comparison with observers using FROC curve of Sensitivity over average FPs per TBI case.

6.8. Possible CMBs

Figure 3 illustrates how true positives (TPs) and FPs were counted for the FROC given the presence of possible CMBs in the reference standard, and in the case of multiple model predictions within a single reference lesion mask.
If we change how possible CMBs are treated with regard to evaluation, performance is changed.

Option 1:. All possible CMBs are removed from evaluation. Any prediction that would fall in the space of a possible CMB is counted as a FP. This also means that designated possible CMBs by the observers were not counted to their metrics.

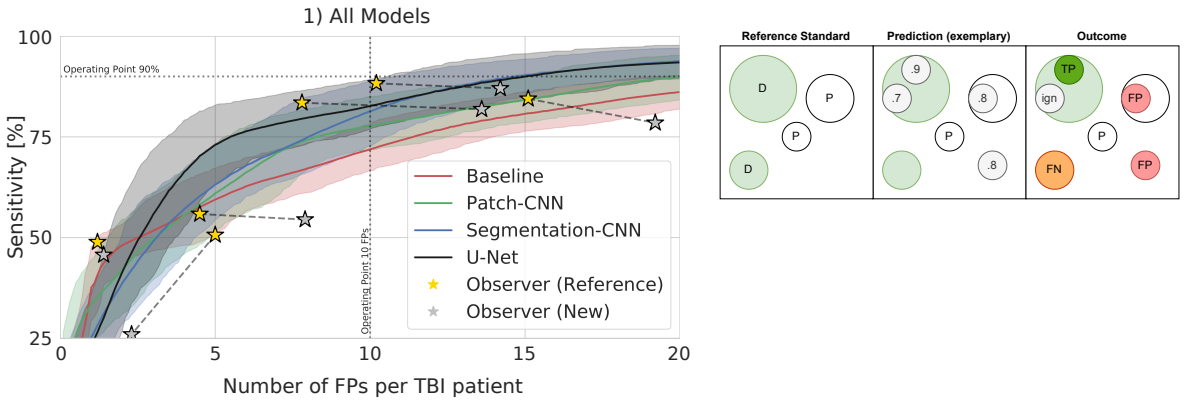

Fig. 13: Performance (bootstrapped at 1000 random samples of the available test set) of individual models in comparison with observers using FROC curve of Sensitivity over average FPs per TBI case.

Given option 1, sensitivity of all models decreases by 5-8% at the operating point of 10 FPs due to the increase in FPs re-aligning the prediction confidence threshold. The **U-Net** is less prone to identify possible CMBs (with high confidence). Of note, there is no measurable difference for the segmentation models at 15 FPs between the regular evaluation method and option 1. Metrics for the observers are changed as well: loss of sensitivity ranges from 1-6% with an increase of 3-6 FPs (except for one outlier).

Option 2:. All possible CMBs are considered definite CMBs. Any not-predicted CMB would count as a false negative. This also means that designated possible CMBs were counted as definite CMBs.

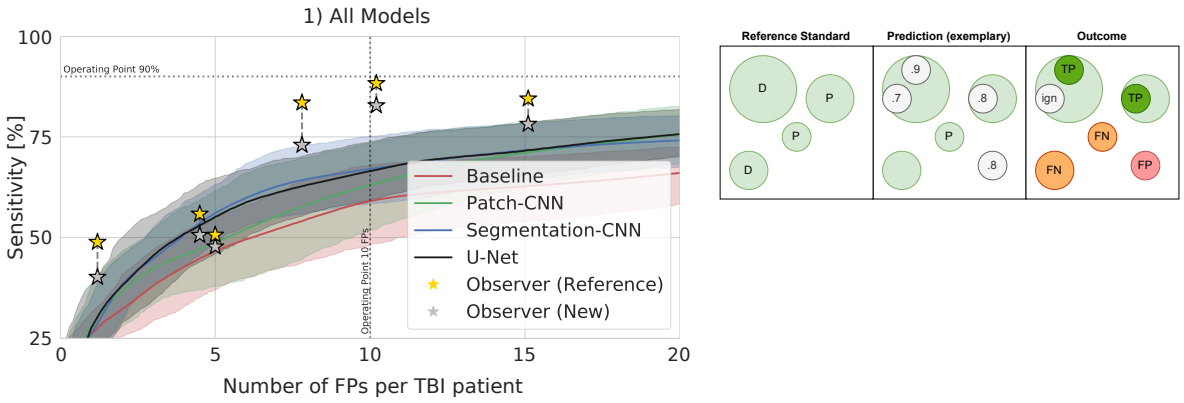

Fig. 14: Performance (bootstrapped at 1000 random samples of the available test set) of individual models in comparison with observers using FROC curve of Sensitivity over average FPs per TBI case.

Assuming option 2, all models score 21-23% lower sensitivity. This shows that all models tend to predict possible CMBs with lower confidence. FP counts of observers are unaffected in this case, but their sensitivity is 5-11% lower compared to our regular method (with the previous outlier only losing 2.9%).

6.9. nnUNet

We performed preliminary experiments to investigate the potential of the nnUnet Isensee et al. (2018, 2019a) in the detection of CMBs in moderate-to-severe TBI. The nnUnet has garnered a lot of attention in recent years as an out-of-the-box solution to many image analysis problems in both domains: natural and medical images. During training, the nnUnet optimises its own structure and loss function, tasks which usually fall on the researcher.

Methods. The nnUnet follows a different training approach to our models. For our models, training and validation data was split (41+4 TBI and healthy SWI scans for training, 12+4 SWI scans for validation), and a single model was trained.
On the other hand, for the nnUnet several folds were trained, in our case 4 folds, on a combination of our training and validation data. Each fold has a validation set of 2 healthy and $13-14$ TBI scans. As a result, every scan is used both as training and validation for the eventual nnUnet-ensemble.
For the predictions of the nnUnet on the test data, an ensemble (averaged prediction) was calculated of the 4 folds.

Results. The nnUnet-ensemble achieved similar sensitivity to our proposed models in the target range of 10FPs (see Figure 15. Compared to our models, it maintains most of its sensitivity at lower FPs counts.

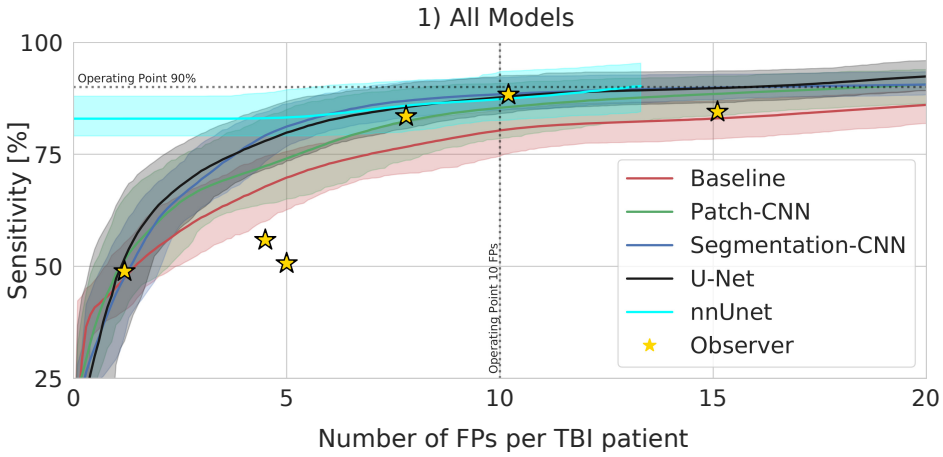

Fig. 15: Performance (bootstrapped at 1000 random samples of the available test set) of the nnUnet ensemble in comparison with the other models and observers using FROC curve of Sensitivity over average FPs per TBI case.

Discussion. The nnUnet shows a lot of promise for the CMB detection in moderate-to-severe TBI patients. It achieves similar sensitivity and better precision than our proposed models. However, it has to be acknowledged that the nnUnet ensemble has a significantly higher computational complexity and therefore capability than our models. Each nnUnet (4 folds) in the ensemble has 10 times the parameter count compared to our models. An important aspect of our analysis was to maintain similar computational power between the models. Under this constraint, the comparison off the nnUnet with any of our models would not be fair.
On the other hand, we proposed certain methods, e.g. boundary loss, that were not available in our nnUnet implementation. Boundary loss enabled more gradual predictions, while the individual nnUnet folds predicted binary outcomes (over-fitting). Adding these to the nnUnet would present an interesting avenue of further research.