

Supplementary Material- MedLesSynth-LD: Lesion Synthesis using Physics-Based Noise Models for Robust Lesion Segmentation in Low-Data Medical Imaging Regimes

1 Noise distributions used in lesion simulations

Rician distribution probability density function (p.d.f.) can be expressed by

$$f(x|\nu, \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{(x^2 + v^2)}{2\sigma^2}\right) I_0\left(\frac{xy}{\sigma^2}\right) \quad (1)$$

where $\nu \geq 0$, the distance between the reference point and center of the bivariate distribution, $\sigma \geq 0$, scale and $I_0(z)$ is the modified Bessel function of first order.

Similarly **Gaussian distribution** p.d.f. can be written as

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (2)$$

where μ stands for mean and σ^2 for variance.

Speckle noise, the main source of noise in ultrasound, is predominantly multiplicative and occurs due to interference of scattered waves from tissue surfaces (bright and dark voxels due to constructive and destructive interference). In addition, the frequency of the sound waves also play an important role in determining the penetration depth. Lower frequencies have better penetration when compared to higher frequencies (though the latter provide better details in the scan) (Cardoso et al., 2012). We specifically chose **Perlin noise** to model this effect of frequency along with the speckle noise, since Perlin noise is formed by the composition of multiple scaled versions of the same noise functions with different amplitudes and frequency (called octaves) (Roos-Hoefgeest et al., 2023, Lawson et al., 2024). This is much like the interference of scattered waves since the composition of functions with varied amplitudes provides an effect similar to the constructive and destructive interference. At the same time, by choosing lower or higher octaves we can choose the frequencies that would best define the noise in various depths of scans. Hence, adding Perlin noise to data has been shown to approximate the speckle noise [2]. Here we make use of the multiple octaves present in Perlin to simulate the noise. Perlin noise construction has been described in [1,3]. Perlin noise is constructed by initially forming an n-dimensional grid, where each point on the grid is associated with a random vector r_g . For the grid vectors, we also determine the displacement vectors r_d based on the offset between grid corners and candidate points of interest. Later, dot products are performed between vectors r_g and the displacement vectors

r_d for candidate points. Finally, we interpolate the dot product values using a function which has its first derivative zero at the grid nodes.

We have shown sample noise profiles for the above noise distributions/models in Fig. S1.

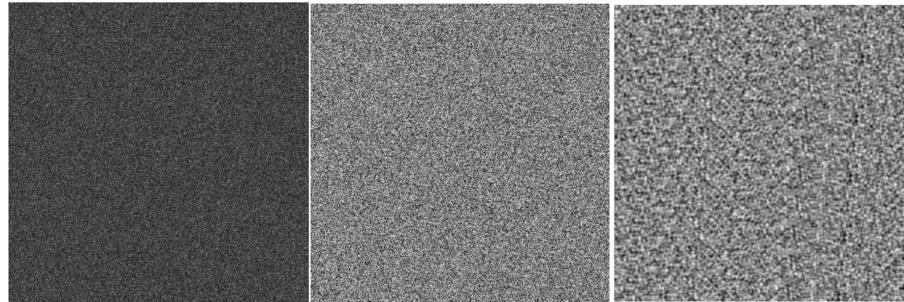


Fig. S1. Examples of the 3 noise patterns: (from left to right) Rician, Gaussian and Perlin.

2 Synthetic lesion simulation: diversity of shapes, spread and contrast by simulation components

By design, our method can simulate localised lesions with well-defined boundaries with perturbation in textures and intensities. For instance, the majority of tumour lesions, hemorrhagic and ischemic lesions fall under this category. Hence, we chose the datasets with the above category of lesions for evaluation. Both hyper and hypointense lesions can be generated based on the contrast parameter α (lower the α , lower the lesion intensity with respect to background and vice versa). Our method cannot simulate non-localised pathological signs (e.g., neurodegeneration signs such as increase in ventricle size and loss of sulci) or linear structures (e.g., blood vessels) that are tissue specific, which are extremely difficult to model as localised perturbations using combination of ellipsoids. Hence, we agree that generation of an infinite set of lesions is not possible, however we can generate lesions with diverse appearance with the use of anatomical priors such as intensity (hyper vs hypointense), texture (noisy vs smooth), shape (globular vs irregular) and spread (scattered vs dense/concentrated). The main aim of this work is to create such augmented data with approximate appearance of the lesions for the deep learning model for improved generalisation.

Fig. S2 illustrates the diversity in size and spread of lesions using the grid plot with varying scale and size for the same set of centroids, and a contrast chart showing lesions generated with varying intensity and contrast.

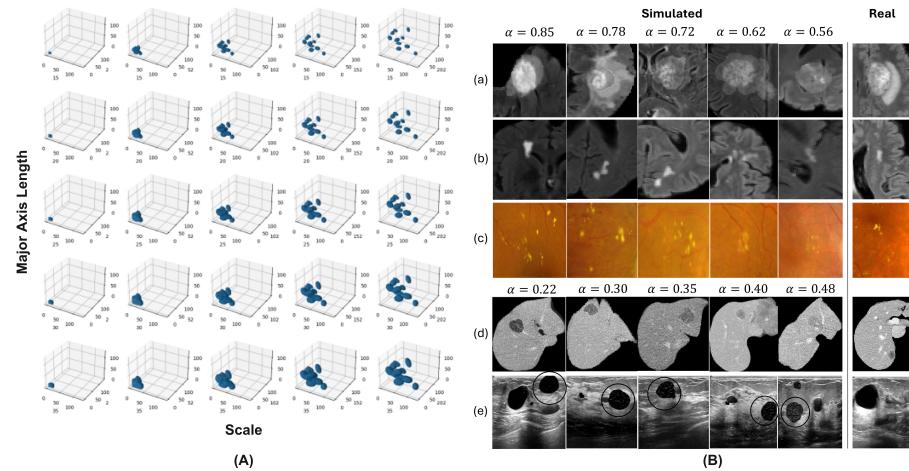


Fig. S2. Lesion simulation by changing scale, major axis length, intensity and contrast. (A) For a centroid cloud c_j^k , we increased the scale (left to right) and major axis length (top to bottom) illustrating the increase in spread and size (more blob-like) of lesions; (B) instances of simulated lesions with varying levels of contrast and intensities (shown for diverse sizes and scales). (a-c) hyperintense lesions where higher contrast was simulated with higher α^k values, (d-e) hypointense lesions where higher contrast was simulated with lower α^k values (simulated lesions shown in (e)).

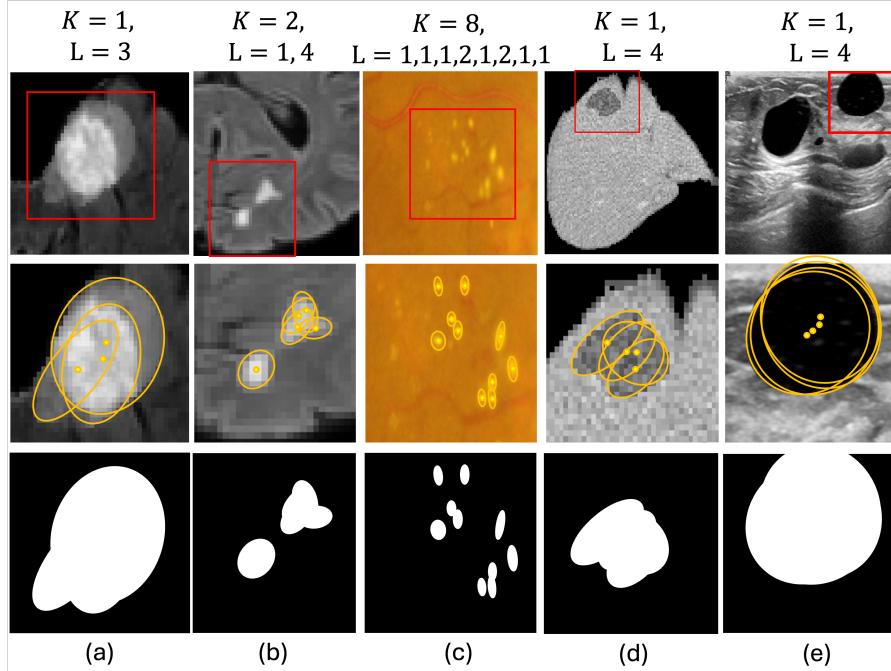


Fig. S3. Lesion shape generation in the simulation method using composite of ellipsoidal polygons. Top row: cropped regions showing simulated lesions; middle row: magnified regions, with K lesions, each formed by the union of L ellipsoidal polygons (k and j values indicated at the top of the figure); bottom row: binary mask $Mask^k$ formed by binarising the union of ellipsoidal polygons. The lesion texture (noise perturbed healthy texture) is placed within the $Mask^k$ area.

Fig. S3 shows the composite of ellipsoids that define a lesion shape in various modalities, where k is the number of lesions, and j is the number of ellipsoids formed from the centroid cloud c_j^k . For the figure, we had considered the subset of images from Fig. S2 have shown how irregular shaped lesions are formed from a combination of several ellipsoidal polygons. We can observe from the image that both globular (e.g., lesions on BraTS and BUSI in (a) and (e) respectively) and irregular shaped (e.g., lesions on WMH and LiTS in (b) and (d) respectively) can be obtained as long as the lesions are localised. When the centroid locations and ellipsoid size and orientation changes, different shapes of lesions are generated. For closer centroids and larger ellipsoids, we get larger, denser lesions (e.g., lesions on BraTS, LiTS and BUSI in (a), (d) and (e) respectively) and for farther centroids and smaller ellipsoids, we get small, diffuse and scattered lesions (e.g., lesions on lesions on WMH and IDRID in (b) and (c) respectively). Hence, for a range of centroids and sizes/orientation of ellipsoids, wider range of diverse lesions can be generated.

3 Comparison of semi-supervised lesion segmentation using synthetic lesions with baseline

Fig. S4 contains the Receiver Operating Characteristics (ROC) curve for all 3 settings (F_{sup} , DA and FT) which was further used to determine suitable threshold for inference stage. Table S1 reports the result of applying Wilcoxon signed ranked test on comparison of M_{seg} architectures (results of Table 2 from the main manuscript). Models are compared from left to right (arranged in the order of increasing DSC scores).

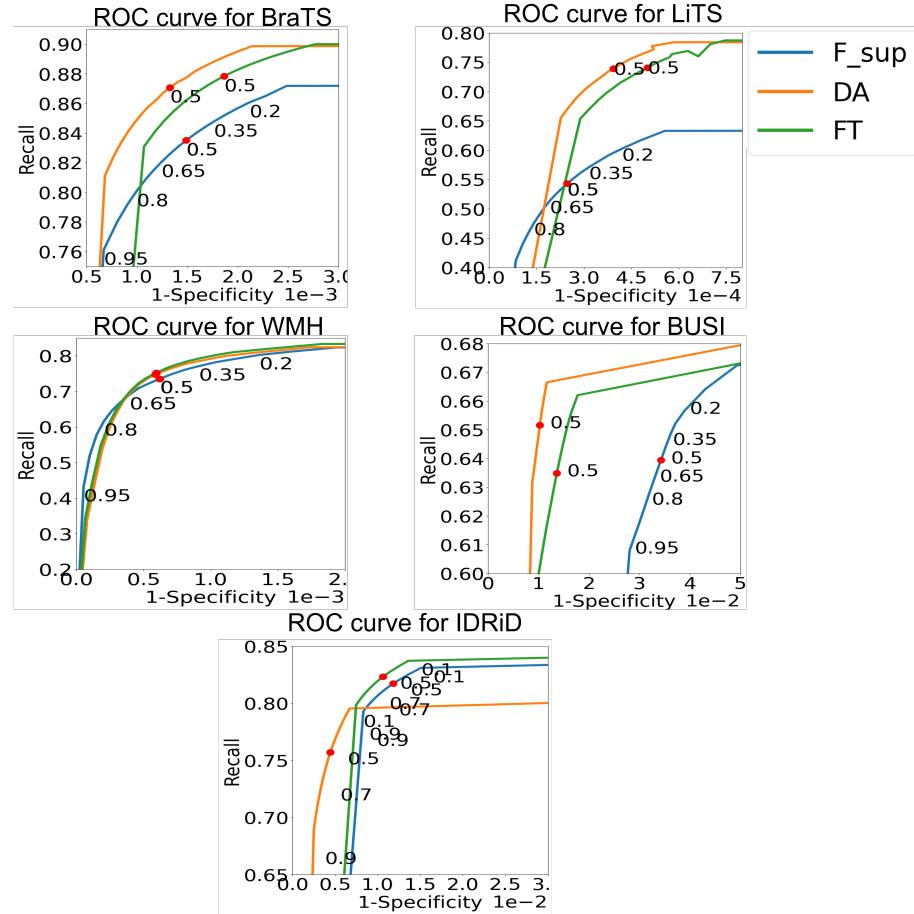


Fig. S4. ROC curve for F_{sup} (blue), DA (orange) and FT (green) performance for BraTS, WMH, LiTS, BUSI and IDRiD

4 Comparison of F_sup performances across various M_{seg} architectures : statistical evaluation

Table S1. Wilcoxon signed rank test values for comparison between the performance of candidate M_{seg} models.

Dataset	Slim UNetr	Half UNet	UNet	UNetr	UNet++
DSC					
BraTS	-	1.43x10-133	2.92x10-17	7.35x10-09	6.27x10-50
WMH	-	0.04x10-2	0.02	0.09	0.30
LiTS	-	2.28x10-7	3.27x10-7	0.05x10-2	0.01
BUSI	-	1.186x10-10	0.34	6.58x10-5	0.42
IDRID	-	0.001	0.30	6.10x10-5	6.10x10-05
Recall					
BraTS	-	1.74x10-153	5.34x10-84	3.42x10-117	5.20x10-30
WMH	-	0.79	0.14x10-02	0.73	0.57
LiTS	-	0.40	0.18x10-3	3.16x10-6	0.09x10-2
BUSI	-	2.32x10-36	5.14x10-24	8.18x10-37	3.09x10-15
IDRID	-	0.68	6.10x10-5	6.10x10-5	6.10x10-5
HD95 (norm)					
BraTS	-	5.32x10-80	0.84	0.13	7.64x10-25
WMH	-	0.47	0.76x10-2	0.68	0.26
LiTS	-	0.99	8.19x10-9	0.29x10-3	0.03
BUSI	-	0.35	0.03	3.2x10-4	0.53
IDRID	-	0.09	6.10x10-5	0.09x10-2	6.10x10-5
Specificity					
BraTS	-	1.97x10-176	3.34x10-7	2.09x10-30	1.29x10-14
WMH	-	0.05x10-2	0.57	0.23	0.42
LiTS	-	2.70x10-9	0.01	0.01x10-2	0.43x10-2
BUSI	-	4.96x10-10	8.27x10-18	9.19x10-51	1.51x10-24
IDRID	-	6.10x10-5	6.10x10-5	0.01	0.23

5 Effect of real-world and simulated data proportions on lesion segmentation: statistical evaluation

In this section, we report the Wilcoxon signed-rank test results for Fig.6 from the main manuscript depicting the significance between DA and F_sup (DA – F_sup) on different real data proportions in Table S2. Similar test results have been reported for the simulation factor in Table S3. UNet++ architecture was used for M_{seg} for all cases.

Table S2. Wilcoxon signed rank test values between F_sup and DA for different (20%, 40%, 60%, 75% and 100%) real data proportions.

Dataset	20%	40%	60%	75%	100%
DSC					
BraTS	9.64x10-48	0.01x10-2	0.01x10-2	3.29x10-7	1.41x10-53
WMH	0.79	0.72	0.05	0.09x10-2	0.04x10-2
LiTS	0.83	0.10	0.68	9.40x10-5	4.28x10-6
BUSI	0.04	0.41	4.15x10-3	4.62x10-11	6.78x10-25
IDRiD	0.067x10-1	0.03x10-1	0.12	0.026x10-1	6.10x10-5
Recall					
BraTS	3.17x10-34	7.08x10-62	2.01x10-65	8.02x10-73	4.71x10-26
WMH	0.68x10-2	0.23	0.92x10-2	0.85	0.05
LiTS	0.70x10-2	0.11	2.93x10-8	1.22x10-5	1.45x10-11
BUSI	2.54x10-21	2.52x10-29	1.92x10-8	1.24x10-5	0.09
IDRiD	0.01x10-2	0.01x10-2	6.10x10-5	6.10x10-5	0.02x10-1
HD95 (norm)					
BraTS	4.63x10-26	0.18	3.51x10-179	1.64x10-8	2.76x10-37
WMH	0.01	0.01	0.05	0.12	0.04x10-2
LiTS	0.37	0.90	0.06	0.01	3.56x10-6
BUSI	0.14	0.14	7.82x10-6	4.39x10-6	8.23x10-16
IDRiD	0.89	0.08	0.012	1	0.30
Specificity					
BraTS	5.82x10-130	1.78x10-79	1.46x10-75	7.106x10-42	0.04x10-2
WMH	0.03	0.79	0.34	0.38	0.34
LiTS	0.02x10-2	0.61	2.08x10-8	0.39	0.2x10-2
BUSI	0.25x10-2	6.57x10-10	4.40x10-12	1.25x10-25	1.56x10-42
IDRiD	6.10x10-5	6.10x10-5	6.10x10-5	6.10x10-5	6.10x10-5

Table S3. Wilcoxon signed rank test values between F_sup and DA for different factors of simulations (1x, 2x and 5x, where x is 100% real-world data).

Dataset	1x	2x	5x
DSC			
BraTS	4.37x10-5	0.22	1.41x10-53
WMH	0.03	0.03	0.04x10-2
LiTS	0.62x10-2	0.15x10-2	4.28x10-6
BUSI	1.25x10-5	1.11x10-9	6.78x10-25
IDRiD	0.01x10-2	0.03x10-2	6.10x10-5
Recall			
BraTS	1.26x10-99	2.76x10-84	4.71x10-26
WMH	0.09x10-2	0.03	0.05
LiTS	0.01x10-2	2.29x10-10	1.45x10-11
BUSI	3.25x10-4	3.76x10-4	0.09
IDRiD	0.03x10-1	0.23	0.02x10-1
HD95 (norm)			
BraTS	6.22x10-6	0.09	2.76x10-37
WMH	0.12	0.01	0.04x10-2
LiTS	0.24	0.34x10-2	3.56x10-13
BUSI	1.13x10-8	1.46x10-6	8.23x10-16
IDRiD	0.02x10-1	0.11	0.03
Specificity			
BraTS	4.25x10-97	2.76x10-58	0.04x10-2
WMH	0.24x10-2	0.97	0.34
LiTS	0.06	3.94x10-5	0.16x10-2
BUSI	2.38x10-13	1.24x10-22	1.56x10-42
IDRiD	0.03x10-1	0.01x10-2	6.10x10-5

Ablation study of the proposed lesion simulation: statistical evaluation

Table S4 contains the p-values from signed rank test between *create_pertGaussian* and *create_pert*, *create_pert* and *localise_pertShapeOnly*, *localise_pertShapeOnly* and *create_pert + localise_pert*, followed with comparison between *create_pert + localise_pert* and *create_pert + localise_pert + blend_intensity*. Actual comparison values are shown in the main manuscript Table 4.

Additionally, we also studied the effect of the loss function components (focal, Dice and focal + Dice). Table S5 and S6 contain loss ablation values and the corresponding p-values from the wilcoxon signed ranked test between Focal and Dice and Dice and Focal + Dice, respectively.

Table S4. Wilcoxon signed rank test values for the ablation study between different components of the proposed simulation method. $p < 0.05/p < 0.001$ shows significant difference between successive rows.

Methods	BraTS	WMH	LiTS	BUSI	IDRID
DSC (\uparrow)					
<i>create_pertGaussian</i>	-	-	-	-	-
<i>create_pert</i>	0.02x10-2	0.02	1.17x10-6	0.02	0.03
<i>localise_pertShapeOnly</i>	3.77x10-13	0.20	4.15x10-6	0.92	0.52
<i>localise_pert</i>	0.01	0.90	0.52	5.47x10-7	0.25
<i>blend_intensity</i>	4.58x10-5	0.38	0.51	0.10	0.85
Recall (\uparrow)					
<i>create_pertGaussian</i>	-	-	-	-	-
<i>create_pert</i>	5.72x10-61	0.52	1.89x10-7	5.50x10-17	0.89
<i>localise_pertShapeOnly</i>	2.05x10-37	0.34	5.52x10-8	8.01x10-14	0.93
<i>localise_pert</i>	0.44	0.42	0.32	7.48x10-8	0.36
<i>blend_intensity</i>	1.04x10-11	0.09	1.86x10-9	0.05x10-1	0.80
HD95 (norm) (\downarrow)					
<i>create_pertGaussian</i>	-	-	-	-	-
<i>create_pert</i>	2.63x10-198	0.95	3.56x10-13	8.30x10-5	0.45
<i>localise_pertShapeOnly</i>	1.02x10-11	0.65	8.96x10-8	0.05x10-1	0.14
<i>localise_pert</i>	2.63x10-198	0.04x10-2	3.56x10-13	1.71x10-65	0.01
<i>blend_intensity</i>	0.02x10-2	0.09	3.56x10-13	0.03	0.85
Specificity (\uparrow)					
<i>create_pertGaussian</i>	-	-	-	-	-
<i>create_pert</i>	1.87x10-46	0.42	0.07	9.17x10-24	0.04
<i>localise_pertShapeOnly</i>	9.91x10-72	0.85	0.04x10-1	2.64x10-30	0.68
<i>localise_pert</i>	0.53	0.33	0.19	0.32	0.08
<i>blend_intensity</i>	2.86x10-6	0.34	7.06x10-7	0.01x10-1	6.1x10-5

Table S5. Ablation study on the loss function components on BraTS, WMH, LiTS, BUSI and IDRID. Mean (standard deviation) are reported, with the best scores indicated in bold, (\uparrow)/(\downarrow) higher/lower or better. */** indicates $p<0.05/p<0.001$ showing significant difference between successive rows. Actual p-values are reported in Table S4 in supplementary material).

Loss	BraTS	WMH	LiTS	BUSI	IDRID
DSC (\uparrow)					
Focal	0.74(0.19)	0.44(0.26)	0.36(0.33)	0.38(0.33)	0.65(0.08)
Dice	0.80(0.17)**	0.66(0.13)**	0.44(0.30)**	0.43(0.34)	0.67(0.07)
Focal + Dice	0.81(0.17)**	0.67(0.13)	0.45(0.31)	0.49(0.31)**	0.68(0.06)
Recall (\uparrow)					
Focal	0.82(0.22)	0.37(0.24)	0.36(0.35)	0.44(0.37)	0.67(0.18)
Dice	0.82(0.20)**	0.75(0.16)**	0.42(0.32)*	0.55(0.42)**	0.77(0.18)**
Focal + Dice	0.83(0.21)**	0.75(0.15)	0.54(0.34)**	0.64(0.37)**	0.82(0.15)**
HD95 (\uparrow)					
Focal	0.05(0.09)	0.22(0.31)	0.36(0.44)	0.26(0.28)	0.01(0.01)
Dice	0.05(0.09)**	0.02(0.01)**	0.22(0.39)**	0.26(0.29)**	0.01(0.01)
Focal + Dice	0.04(0.08)**	0.02(0.01)*	0.18(0.37)**	0.15(0.21)**	0.01(0.005)*
Specificity (\uparrow)					
Focal	0.99636	0.99905	0.99964	0.96216	0.99328
Dice	0.99785**	0.99907**	0.99985	0.96564*	0.99022**
Focal + Dice	0.99849**	0.99924	0.99924**	0.97593**	0.98811**

Table S6. Wilcoxon signed rank test values for the ablation study between different components of the loss function for training the segmentation model. $p<0.05/p<0.001$ shows significant difference between successive rows.

Loss	BraTS	WMH	LiTS	BUSI	IDRID
DSC (\uparrow)					
Focal	-	-	-	-	-
Dice	4.27x10-107	0.049x10-2	0.07x10-2	2.25	0.23
Focal + Dice	6.65x10-7	0.18	3.75	0.01x10-1	0.89
Recall (\uparrow)					
Focal	-	-	-	-	-
Dice	1.60x10-23	0.05x10-2	0.01	2.36x10-28	0.06x10-1
Focal + Dice	1.40x10-13	0.91	2.55x10-5	1.57x10-8	6.10x10-5
HD95 (\uparrow)					
Focal	-	-	-	-	-
Dice	2.60x10-45	0.05x10-2	3.74x10-8	0.02x10-2	0.56
Focal + Dice	2.63x10-198	0.04	3.56x10-13	1.71x10-65	0.02
Specificity (\uparrow)					
Focal	-	-	-	-	-
Dice	3.84x10-84	0.05x10-1	0.09	0.02	0.04x10-2
Focal + Dice	1.41x10-23	0.11	1.22x10-6	1.05x10-5	0.05x10-1

Additional visual results comparing F_sup, DA and FT

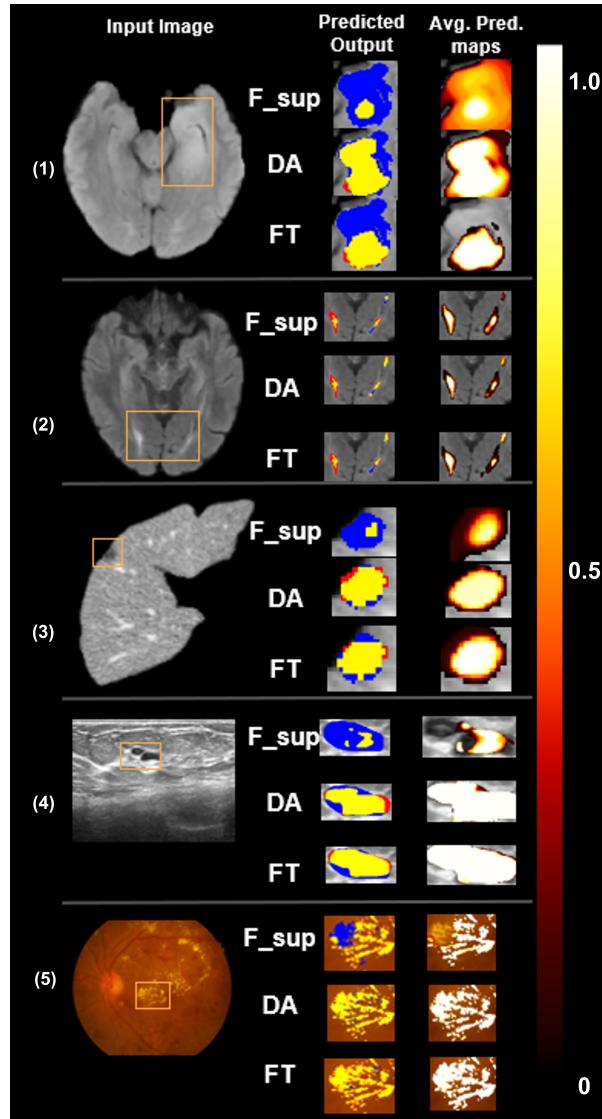


Fig. S5. Comparison of F_sup, DA and FT lesion segmentation with UNet++ architecture. (a) Input image is shown with (b) the segmentation results and (c) average prediction maps ($Pred_{avg}$) showing the confidence of prediction (Sec. 2.5) for BraTS, WMH, LiTS, BUSI and IDRiD (1-5). For each modality, F_sup, DA and FT results are shown, where in column (b), yellow, blue and red show true positive, false negative and false positive voxels respectively on the predicted maps.

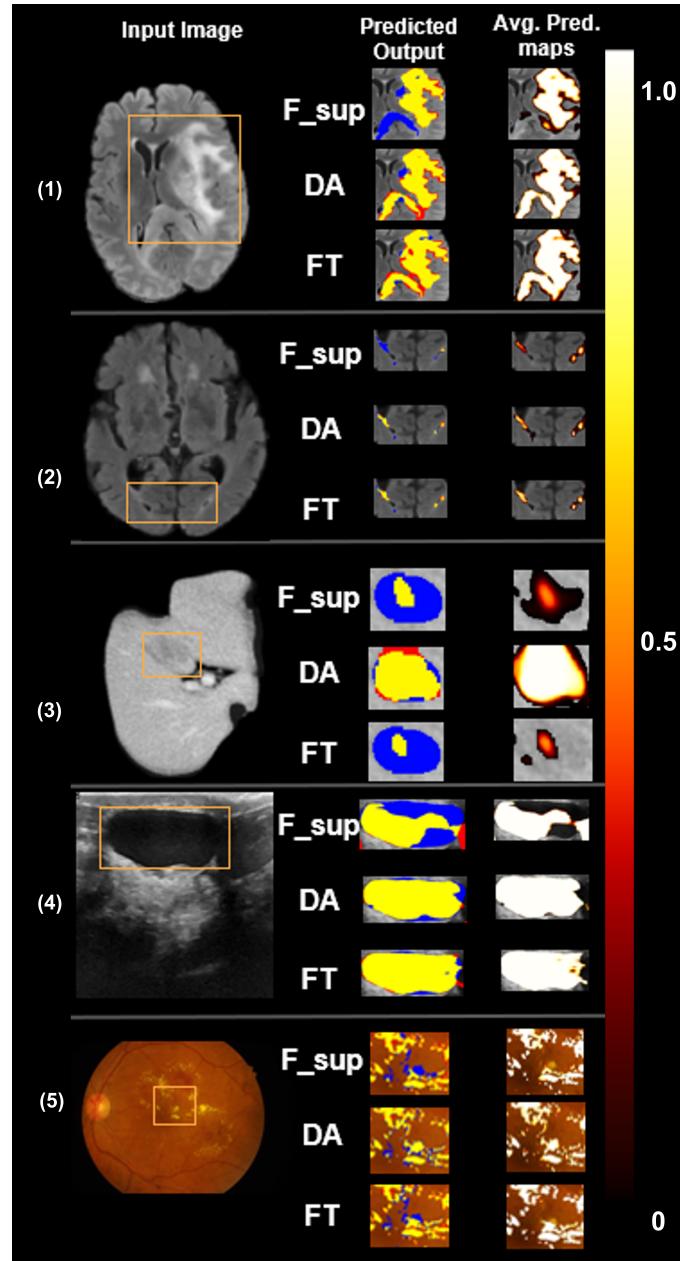


Fig. S6. Comparison of **F_{sup}**, **DA** and **FT** lesion segmentation with UNet++ architecture. (a) Input image is shown with (b) the segmentation results and (c) average prediction maps (Pred_{avg}) showing the confidence of prediction (Sec. 2.5) for BraTS, WMH, LiTS, BUSI and IDRiD (1-5). For each modality, **F_{sup}**, **DA** and **FT** results are shown, where in column (b), yellow, blue and red show true positive, false negative and false positive voxels respectively on the predicted maps.

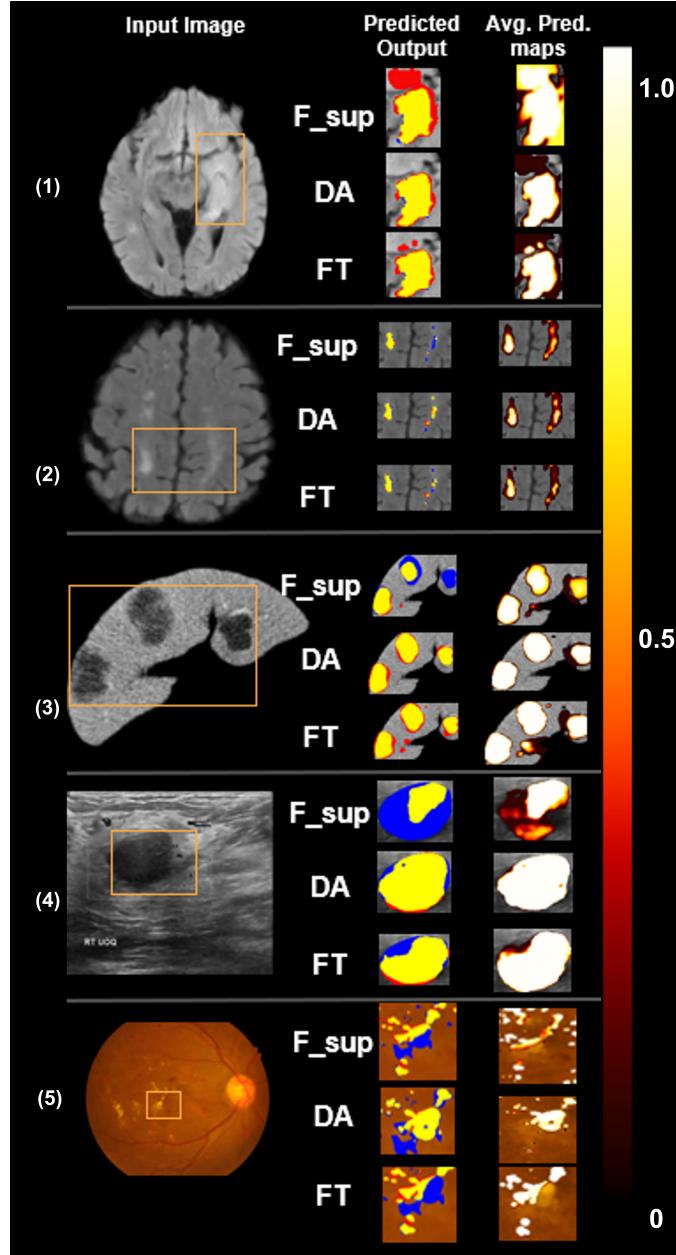


Fig. S7. Comparison of F_{sup}, DA and FT lesion segmentation with UNet++ architecture. (a) Input image is shown with (b) the segmentation results and (c) average prediction maps ($Pred_{avg}$) showing the confidence of prediction (Sec. 2.5) for BraTS, WMH, LiTS, BUSI and IDRiD (1-5). For each modality, F_{sup}, DA and FT results are shown, where in column (b), yellow, blue and red show true positive, false negative and false positive voxels respectively on the predicted maps.

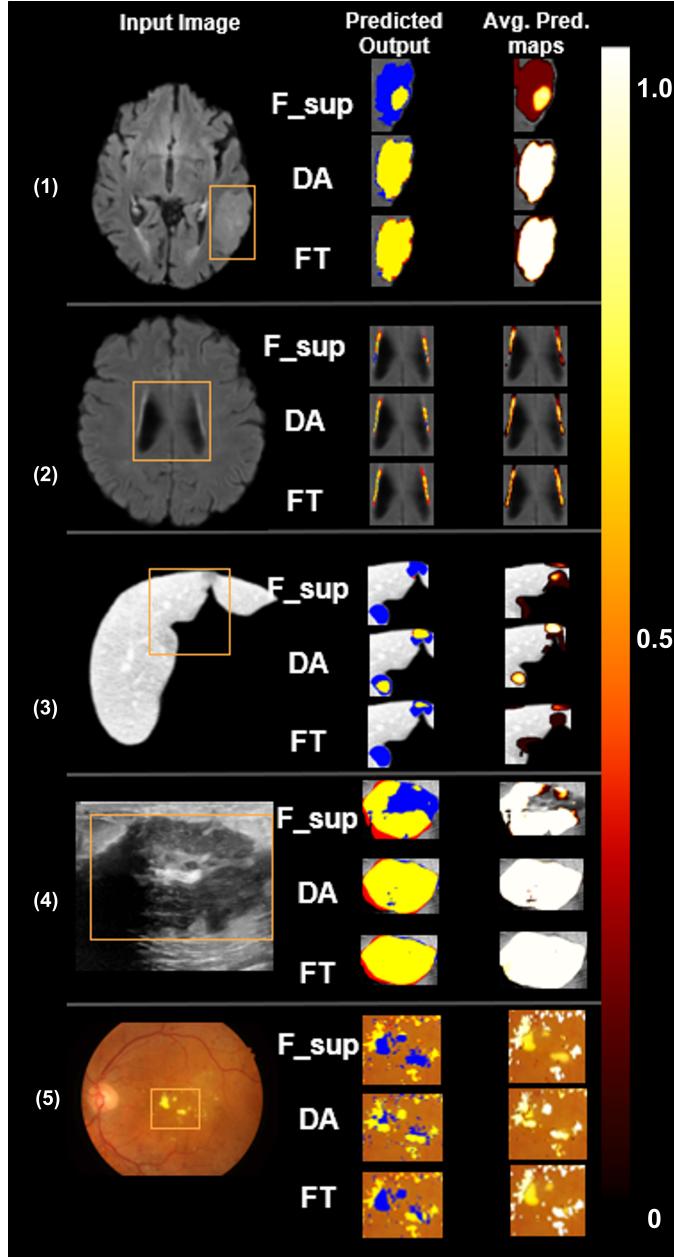


Fig. S8. Comparison of **F_{sup}**, **DA** and **FT** lesion segmentation with UNet++ architecture. (a) Input image is shown with (b) the segmentation results and (c) average prediction maps ($Pred_{avg}$) showing the confidence of prediction (Sec. 2.5) for BraTS, WMH, LiTS, BUSI and IDRiD (1-5). For each modality, **F_{sup}**, **DA** and **FT** results are shown, where in column (b), yellow, blue and red show true positive, false negative and false positive voxels respectively on the predicted maps.

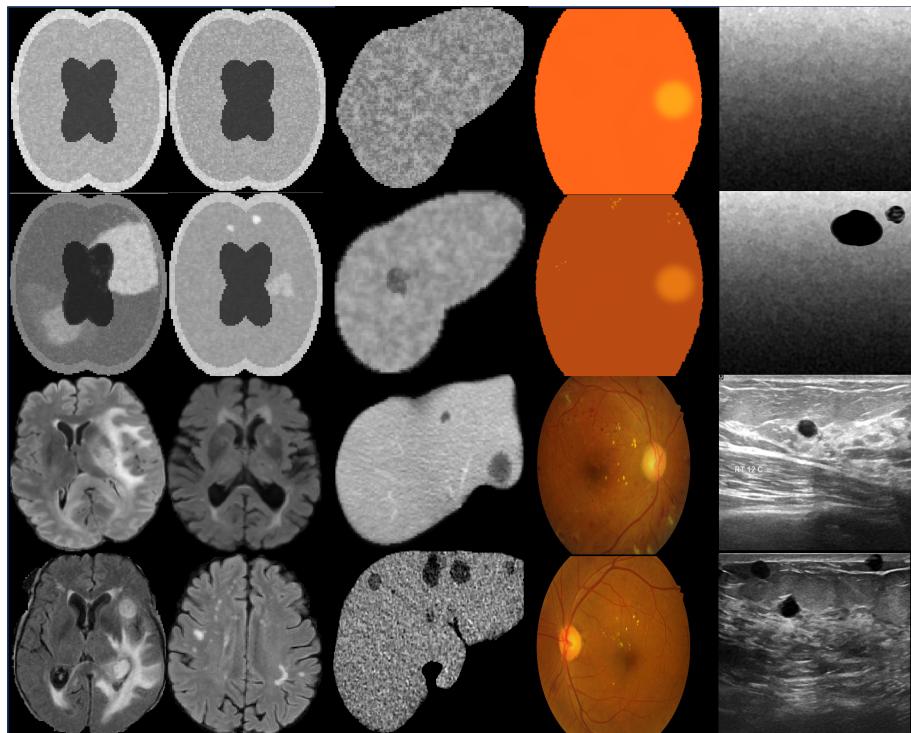


Fig. S9. Examples of real and simulated lesions on brain MRI, liver CT, retinal fundus imaging and breast ultrasound (left-to-right). Top 2 rows: simulated lesions on the generic simulation map GS_{map} , used for pretraining (explained in Sec. 2.2 in the main manuscript). Third row: Real-world lesions; Fourth row: simulated lesions generated using the proposed method; Lesions were generated using underlying noise models (Rician, Gaussian, and Perlin, from left to right) of the individual modalities.

References

1. Perlin, K.: Improving noise. In: Proc. of the 29th annual SIGGRAPH. pp. 681–682 (2002)
2. Roos-Hoefgeest, S., Roos-Hoefgeest, M., Álvarez, I., González, R.C.: Simulation of laser profilometer measurements in the presence of speckle using perlin noise. Sensors **23**(17), 7624 (2023)
3. Wikipedia contributors: Perlin noise — Wikipedia, the free encyclopedia (2024), https://en.wikipedia.org/w/index.php?title=Perlin_noise&oldid=1230993513, [Online; accessed 22-July-2024]