Uncertainty-Aware Weakly Supervised Lesion Localization

Basma Nasir[1], Sadiya Imran[2], Uzma Saood[2], Zahra Khan[2], Muhammad Nawaz[3]

[1] Department of Computer Science, COMSATS University Islamabad, Pakistan
[2] Computer Science and Engineering, Indira Gandhi Delhi Technical University for Women (IGDTUW), India
[3] Faculty of Engineering and Information Technology, University of Technology Sydney, Australia



NeurIPS 2026 • MusIML Workshop

1 MOTIVATION

- **Pixel-level annotations** need expert delineation; image-level labels are cheap but carry no spatial information.
- **A single classifier's CAM** focuses on the most discriminative region and varies across stochastic passes.
- **Averaging** trusts every model equally everywhere; intersection discards valid evidence.

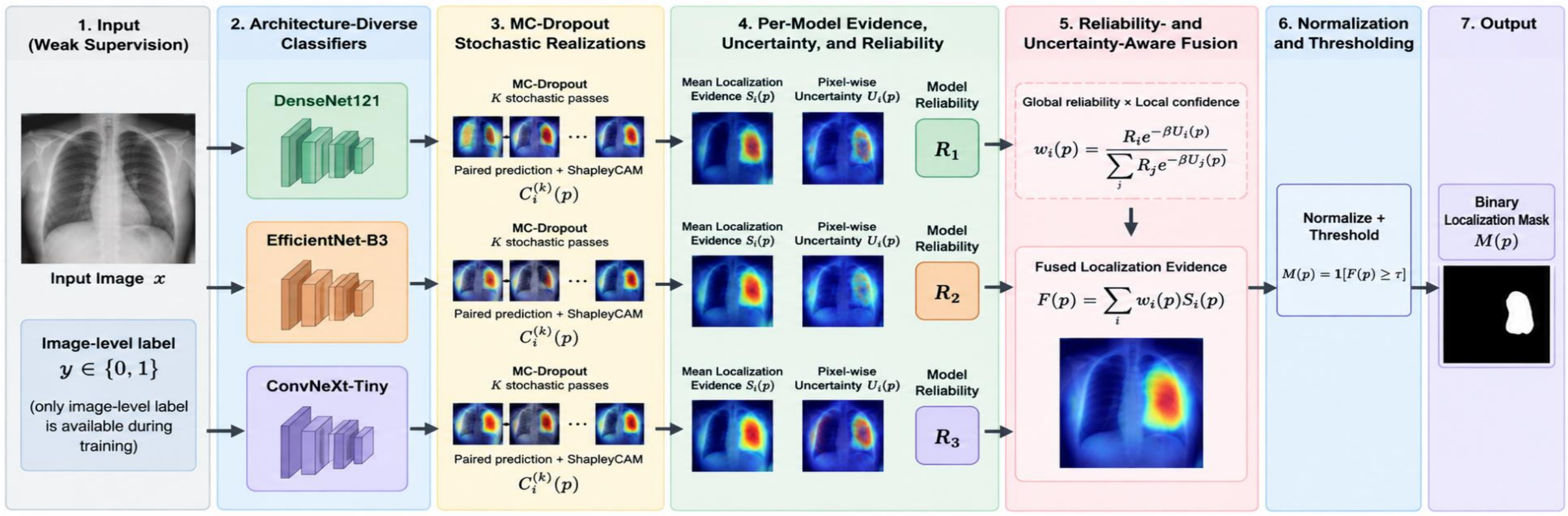
2 CONTRIBUTIONS

1. **Architecture-diverse readings**
   DenseNet121, EfficientNet-B3 and ConvNeXt-Tiny give complementary evidence.
2. **Dual reliability**
   Model-level calibration and pixel-level MC-Dropout uncertainty.
3. **Adaptive fusion**
   Both signals weight each pixel; masks come from image-level labels alone.

5 EXPERIMENTAL SETUP

Datasets	Fire/Non-Fire (true masks) and TBX11K chest X-rays (boxes only); spatial labels are withheld from training.
Training	DenseNet121, EfficientNet-B3, ConvNeXt-Tiny from ImageNet weights; patient-level splits; K = 50 MC-Dropout passes; 5 runs.
Comparisons	ResNet50 baseline, single-architecture ShapleyCAM, mean fusion, hard intersection.
Evaluation	Dice, IoU, precision, recall (TBX11K uses box-reference localization).

3 PROPOSED FRAMEWORK



An image with only an image-level label is passed through DenseNet121, EfficientNet-B3 and ConvNeXt-Tiny; MC-Dropout realizations yield paired predictions and stochastic ShapleyCAM maps, summarized per model as mean localization evidence $S_i(p)$ and pixel-wise uncertainty $U_i(p)$. Together with validation-derived model reliability R_i, they fuse into $F(p)$, then threshold to obtain the binary lesion mask $M(p)$.

4 RELIABILITY- & UNCERTAINTY-AWARE FUSION

Which model to trust? → R_i | Where to trust it? → $U_i(p)$

01 Mean Evidence

$$S_i(p) = (1/K)\, \Sigma_k\, C_i^{(k)}(p)$$

Aggregates ShapleyCAM evidence across stochastic passes.

02 Pixel-Level Uncertainty

$$U_i(p) = Std_k\,[C_i^{(k)}(p)]$$

Measures spatial variability in localization evidence.

03 Reliability-Weighted Fusion

$$F(p) = \Sigma_i\, R_i \cdot S_i(p) \,/\, (U_i(p) + \varepsilon)$$

Combines globally reliable models with locally stable evidence.

04 Binary Mask Generation

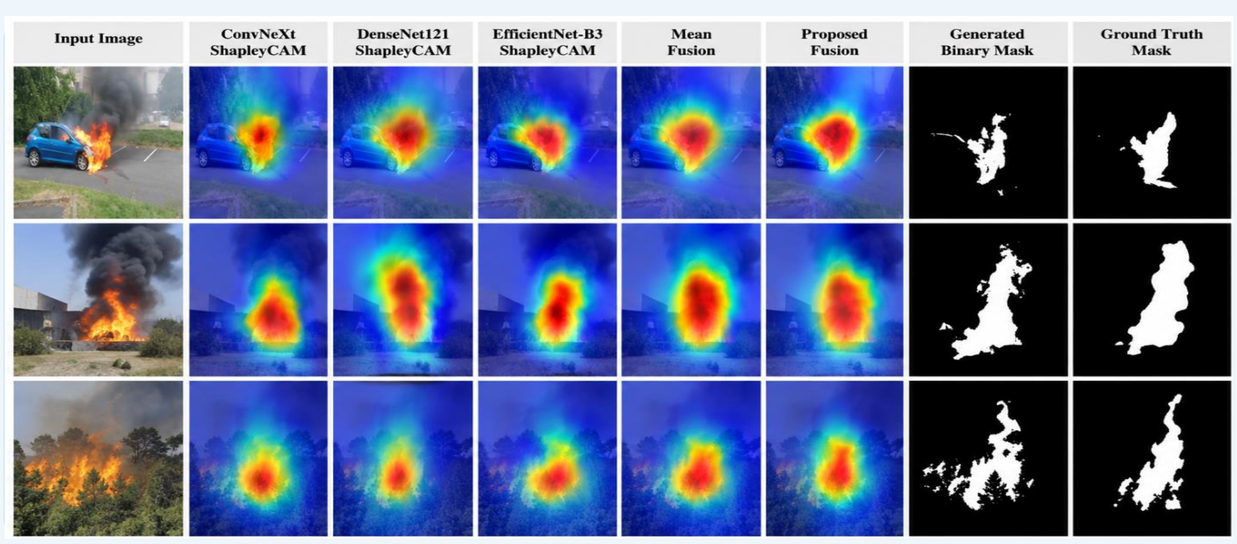
$$M(p) = 1[\tilde{F}(p) \geq \tau]$$

Thresholds normalized fused evidence into a candidate lesion mask.

$$R_i \propto 1 - ECE_i$$

Global model reliability from calibration, where ECE is the expected calibration error of model i [3].

6 QUALITATIVE RESULTS (FIRE)



Left to right: input, ConvNeXt-Tiny, DenseNet121, EfficientNet-B3, mean fusion, proposed fusion, generated mask, ground truth. Blue = low, red = high localization evidence.

7 RESULTS

Method	FIRE (TRUE MASKS)		TBX11K (BOX-REF.)	
	Dice ↑	IoU ↑	Dice ↑	IoU ↑
Single model†	0.65	0.48	0.53	0.37
Mean fusion	0.78	0.64	0.57	0.41
Hard intersection	0.76	0.61	0.60	0.44
Proposed	**0.87**	**0.76**	**0.65**	**0.50**

†ResNet50 on Fire, single-model ShapleyCAM on TBX11K. TBX11K uses rasterized boxes, so its Dice/IoU are box-reference; TBX11K values are averaged over 5 runs.

FIRE Dice	FIRE IoU	TBX11K Dice	TBX11K IoU
0.87	0.76	0.65	0.50

8 REFERENCES

1. Class re-activation maps for weakly-supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 969–978, 2022.
2. Huaiguang Cai. Shapley value-based class activation mapping for improved explainability in neural networks. The Visual Computer, 41(10), 2025. doi: 10.1007/s00371-025-03803-1.
3. Yun Liu, Yu-Huan Wu, Yunfeng Ban, Huifang Wang, and Ming-Ming Cheng. Rethinking computer-aided tuberculosis diagnosis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 2646–2655, 2020.