



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

When Views Disagree: Conflict-Aware Evidential Inference for Multi-View Mammography

Md Rayhan Ahmed* and Patricia Lasserre

The University of British Columbia, Irving K. Barber Faculty of Science
{md.ahmed, patricia.lasserre}@ubc.ca

Abstract. Multi-view mammography benefits from complementary craniocaudal and mediolateral oblique projections, yet differences in acquisition quality and anatomical coverage often lead to conflicting diagnostic evidence between views. Most existing multi-view learning approaches address this variability through deterministic feature fusion, implicitly assuming that disagreement can be resolved by stronger representations, an assumption that can yield brittle or overconfident predictions in safety-critical screening settings. We present Conflict-Evidential Inference Network (CEI-Net), an uncertainty-aware framework that reframes multi-view mammography classification as evidential reasoning under disagreement. Rather than enforcing consensus at the feature level, CEI-Net represents each view as a source of uncertain evidence and performs fusion by selectively moderating class hypotheses that are unsupported across views. This design allows disagreement to manifest as calibrated uncertainty when appropriate while still enabling confident decisions when reliable evidence is available from a single view. Empirical evaluation of the CBIS-DDSM and VinDr-Mammo benchmarks shows that this conflict-aware inference paradigm yields consistent improvements in discrimination and calibration over deterministic fusion methods, without adding auxiliary routing/quality networks. These findings suggest that explicitly reasoning about disagreement, rather than absorbing it, is critical for robust and interpretable multi-view decision-making in clinical screening. Our code is publicly available at <https://github.com/rayhan-ahmed91/CEI-Net>.

Keywords: Multi-view mammography · Evidential deep learning · Conflict-aware fusion · Uncertainty estimation · Medical image classification

1 Introduction

Multi-view learning is central to screening mammography, where craniocaudal (CC) and mediolateral oblique (MLO) views provide complementary anatomical and pathological evidence [1]. Early deep models exploited this complementarity through feature concatenation or averaging, improving over single-view baselines but implicitly assuming equal reliability across views [2, 3]. More structured fusion strategies have since been introduced, including anatomical correspondence

* Corresponding author: md.ahmed@ubc.ca

modeling in DIVF-Net [4] and attention-based cross-view interaction with Transformer architectures [3, 17]. Despite these advances, most methods merge views deterministically without modeling disagreement, risking overconfident predictions under degraded or conflicting views.

Recent work has further explored state-space models (SSMs), such as Mamba [5], for multi-view medical image classification due to their long-range modeling capacity and linear complexity. Methods such as XFMamba [6], Mammo-Mamba [8], and BI-Mamba [7] combine global SSM-based modeling with local attention or bidirectional view reasoning. In parallel, attention-based and context-aware mammography models improve region localization and robustness to acquisition variability [9, 10, 16, 18–20, 25]. However, these approaches primarily operate in deterministic feature space, absorbing disagreement into fused representations rather than separating reliable evidence from conflicting or uncertain evidence. Related multi-view frameworks such as MV-HFMD [21] similarly improve classification through hybrid fusion and distillation, but do not provide an explicit probabilistic mechanism for conflict-aware inference.

Evidential deep learning offers a principled way to represent predictive uncertainty through Dirichlet distributions [11, 12]. Existing multi-view evidential methods commonly aggregate view-level evidence using global Dempster–Shafer-style fusion rules [13]. While effective for uncertainty aggregation, such methods treat conflict coarsely and do not capture class-conditional disagreement, making it difficult to suppress evidence only for disputed diagnostic hypotheses. Orthogonal vision–language pretraining methods improve representation robustness through CLIP-style alignment [22–24], but they do not address uncertainty-aware evidential fusion during inference.

Motivated by these limitations, we propose the CEI-Net, a unified evidential framework for multi-view mammography that explicitly models epistemic uncertainty and inter-view disagreement. CEI-Net converts localized multi-view visual cues into evidential belief states, aggregates token-level evidence into view-specific Dirichlet distributions, and derives view reliability directly from predictive uncertainty. Fusion is then performed in evidence space through reliability-weighted, class-conditional conflict gating based on Jensen–Shannon divergence [15]. This allows CEI-Net to selectively attenuate disputed diagnostic hypotheses while preserving confident agreement when views are consistent. To the best of our knowledge, CEI-Net is the first multi-view mammography framework to perform class-conditional, uncertainty-aware evidential fusion without auxiliary routing or alignment networks, offering an interpretable alternative to existing CNN, Transformer, and SSM-based fusion models.

2 Method

Problem Formulation and methodology overview. Screening mammography relies on complementary CC and MLO views that may differ in diagnostic reliability due to artifacts, incomplete coverage, and positioning variability, often yielding asymmetric or conflicting evidence across views. Given paired inputs

$(x^{(1)}, x^{(2)})$ and a binary label $y \in \{0, 1\}$, our objective is to infer a calibrated belief state that represents both class probabilities and epistemic uncertainty under inter-view disagreement. We adopt a Dirichlet evidential formulation in which each view produces class evidence rather than a single deterministic probability vector. Our model encodes each view with a shared ConvNeXt backbone [14], refines the last-stage feature map with a lightweight residual refinement module, extracts a compact set of salient regional tokens using attention-guided and diversity-aware Top- M selection with foreground-constrained masking, and aggregates token evidence into a per-view Dirichlet distribution. We further introduce a token-feedback refinement mechanism that scatters selected token information back into the feature map to improve subsequent tokenization. Finally, the two view-level Dirichlets are fused by Conflict-Aware Evidential Opinion Pooling, which combines uncertainty-derived reliability weighting with class-conditional conflict gating based on a Jensen-Shannon (JS) divergence decomposition. During training, we optionally regularize the learned conflict gate with a lightweight calibration term, without introducing any auxiliary routing or quality-prediction network.

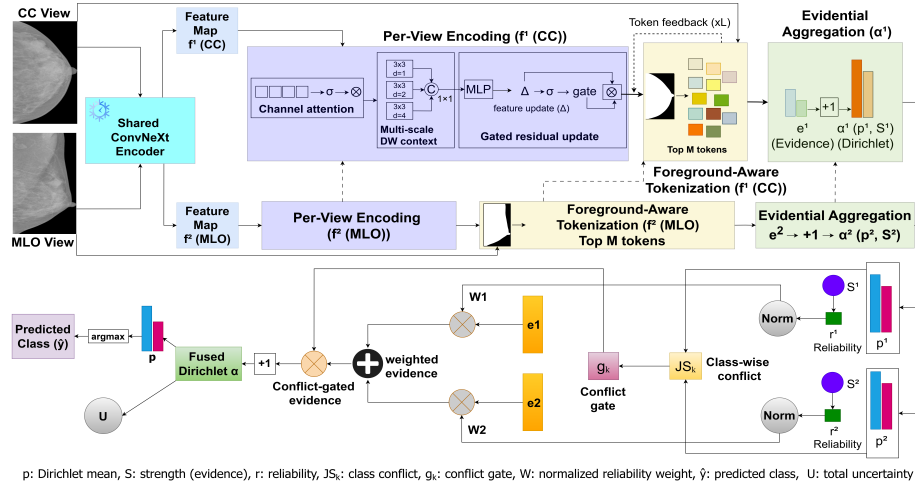


Fig. 1. Architecture overview of the proposed CEI-Net framework. CEI-Net performs foreground-aware tokenization per view, converts tokens to Dirichlet evidence, and fuses views via reliability-weighted, class-conditional conflict-gated opinion pooling to produce calibrated predictions with uncertainty.

Per-View Encoding, Foreground-Aware Tokenization, and Evidential Aggregation. Each view $x^{(v)}$ is encoded by a shared ConvNeXt backbone $\mathcal{E}$, producing $f^{(v)} \in \mathbb{R}^{C \times H' \times W'}$. A lightweight residual refinement module $\mathcal{R}$ (channel recalibration + multi-scale depthwise context) yields $\hat{f}^{(v)} = \mathcal{R}(f^{(v)})$. To suppress background-dominated attention, we derive a foreground prior by thresholding the de-normalized grayscale image to a binary mask $m^{(v)}$, resized to (H', W') ; masking is applied only when foreground coverage exceeds ρ , with logits filled by a large negative constant for stability.

Tokenization uses attention-guided, diversity-aware Top- M selection. A 1×1 scorer ϕ produces logits $a^{(v)}$ and attention $\pi^{(v)}$. We form a candidate set from the top-min($H'W'$, κM) locations and enforce spatial diversity via farthest-point sampling on grid coordinates, followed by foreground-constrained refill. Tokens are $T^{(v)} = \{t_m^{(v)}\}_{m=1}^M, t_m^{(v)} \in \mathbb{R}^C$.

To stabilize extraction, token-feedback $\mathcal{F}$ scatters projected tokens back to their indices, smooths the sparse map, and injects it into $\hat{f}^{(v)}$ via a gated residual update for a small fixed number of iterations ($L=2$). This parameter-light reweighting reuses the same tensor and indices, introducing no new heads or stages and preserving resolution and receptive field. Tokens are mapped to non-negative evidence via a lightweight MLP with Softplus($\cdot$) and floor δ , aggregated as $e^{(v)} = \sum_m e_m^{(v)}$, yielding $\alpha^{(v)} = e^{(v)} + 1$. The Dirichlet mean $p^{(v)}$ and strength $S^{(v)}$ are used for disagreement and reliability estimation.

Class-Conditional Cross-View Disagreement. To explicitly model conflicting evidence between mammographic views prior to fusion, we quantify disagreement at the level of individual class hypotheses. Let $p^{(1)}$ and $p^{(2)}$ denote the Dirichlet mean class probability vectors inferred from the two views. The class-conditional JS disagreement for class k is defined as

$$\text{JS}_k = \frac{1}{2} \left[p_k^{(1)} \log \frac{p_k^{(1)}}{m_k} + p_k^{(2)} \log \frac{p_k^{(2)}}{m_k} \right], \quad m = \frac{1}{2} (p^{(1)} + p^{(2)}). \quad (1)$$

The vector $\text{JS} \in \mathbb{R}^K$ captures hypothesis-level conflict between views, and $\sum_k \text{JS}_k$ recovers the scalar JS divergence between $p^{(1)}$ and $p^{(2)}$. When only a single view is available, we set $\text{JS}_k = 0$, so the available view receives unit normalized weight and its evidential prediction is passed to inference without conflict-induced attenuation.

Uncertainty-Derived Reliability Weighting. Because token-level aggregation yields view-specific evidence of varying strength and coverage, we weight each view's contribution according to its evidential reliability. Let $S^{(v)} = \sum_j \alpha_j^{(v)}$ denote the Dirichlet strength of view v , from which epistemic uncertainty is given by $u^{(v)} = K/S^{(v)}$. We define a bounded reliability score as

$$r^{(v)} = \frac{1}{1 + u^{(v)}} = \frac{S^{(v)}}{S^{(v)} + K}, \quad (2)$$

which increases monotonically with committed evidence and lies in $(0, 1)$. When a view-presence mask is available, we apply it directly at the reliability level by setting $r^{(v)} \leftarrow r^{(v)} \mathbb{I}[\text{view } v \text{ present}]$, ensuring that missing views contribute zero weight. The normalized fusion weights are

$$w^{(v)} = \frac{r^{(v)}}{r^{(1)} + r^{(2)}}. \quad (3)$$

Consequently, under asymmetric visibility, a high-strength view can dominate the pooled evidence when its counterpart is weak or uncertain; the subsequent

class-wise JS gate modulates only disputed pooled hypotheses rather than requiring CC/MLO consensus. With only a single available view, fusion degenerates to its Dirichlet estimate.

Conflict-Aware Evidential Opinion Pooling. We fuse directly in evidence space using view reliability and class-conditional disagreement. Let $e^{(v)} = \alpha^{(v)} - \mathbf{1}$ denote view-level evidence. Reliability-weighted pooling gives $\tilde{e} = w^{(1)}e^{(1)} + w^{(2)}e^{(2)}$, followed by a class-wise conflict gate $g_k = \exp(-\gamma \text{JS}_k)$ with learnable scalar $\gamma > 0$. The fused evidence is $e_k = g_k \tilde{e}_k$ and $\alpha_k = e_k + \mathbf{1}$. Conflict is applied only when both views are present; otherwise $\text{JS} = 0$ and fusion reduces to single-view reliability weighting (or $\alpha = \mathbf{1}$ if both are absent). Class-wise gating suppresses only disputed hypotheses, reducing total evidence, and increasing epistemic uncertainty under strong disagreement. The gate is a scalar applied to closed-form JS divergence and introduces no view-quality subnetworks, routing modules, or additional encoders. Fusion remains linear in evidence space with element-wise gating, preserving parameter count and asymptotic complexity relative to standard evidential pooling while improving calibration sensitivity to inter-view conflict. Importantly, the conflict gate attenuates evidence strength rather than directly forcing a benign prediction. When both views are reliable but support opposing class hypotheses, the class-wise gate reduces only the disputed pooled evidence and thereby decreases Dirichlet strength rather than biasing the prediction toward benign. Therefore, severe CC-MLO disagreement produces lower Dirichlet strength and higher epistemic uncertainty, allowing such cases to be treated as uncertain/high-risk rather than confidently negative.

Datasets and Evaluation Metrics. We evaluated CEI on two public screening mammography benchmarks: CBIS-DDSM [26] and VinDr-Mammo [27]. For both datasets, paired CC and MLO views are constructed, discarding incomplete pairs. For CBIS-DDSM, following standard practice, we restrict evaluation to mass cases and form CC/MLO pairs at the abnormality level by grouping images by patient, laterality, and abnormality identifier. The task is binary classification of benign versus malignant masses. VinDr-Mammo is a large-scale full-field digital mammography dataset; CC/MLO pairs are formed at the breast level with binary labels derived using a BI-RADS threshold of ≥ 4 . Duplicate images within the same view are resolved by selecting the instance with the largest foreground breast area after cropping, while preserving both CC and MLO for fusion. We split each dataset into disjoint train and test partitions at the patient or study level (80:20) to prevent identity leakage, and additionally carve out a held-out validation subset from the training split (15%) for model selection and calibration only. All images are preprocessed using dataset-specific breast cropping and resized to 224×224 . Performance is evaluated using Area Under the Receiver Operating Characteristic curve (AUROC), while calibration is assessed via Expected Calibration Error (ECE) and Negative Log-Likelihood (NLL). We isolate uncertainty-aware modeling from post-hoc temperature scaling, which is fitted on validation data and applied only at evaluation. Temperature scaling is fitted only to validation predictions after training and is used solely for

evaluation calibration, whereas CEI-Net uncertainty is produced intrinsically by Dirichlet evidence strength and JS-modulated conflict gating.

Experimental Details. Experiments are implemented in PyTorch and conducted on an NVIDIA RTX A6000 GPU with deterministic seeding. Views are processed by a shared backbone, and top- M spatial tokens ($M=8$) are selected per view. Token-level Dirichlet evidence is accumulated into a view-level Dirichlet, and the resulting view distributions are fused using CEI. Training minimizes an evidential cross-entropy loss with KL regularization to a uniform Dirichlet,

$$\mathcal{L}_{\text{cls}} = \sum_k y_k (\psi(S) - \psi(\alpha_k)) + \lambda_{\text{KL}} \eta(e) \text{KL}(\text{Dir}(\alpha) \parallel \text{Dir}(\mathbf{1})), \quad (4)$$

where $\alpha \in \mathbb{R}_+^K$ denotes the fused Dirichlet concentration parameters, $y \in \{0, 1\}^K$ is the one-hot label, $S = \sum_k \alpha_k$, $\psi(\cdot)$ is the digamma function, and $\eta(e) = \min(1, e/T_{\text{anneal}})$ anneals the KL term over the first $T_{\text{anneal}}=10$ epochs. Class imbalance is handled via positive-class reweighting. Auxiliary evidential losses on the per-view Dirichlets are included with weight $\lambda_{\text{aux}}=0.25$ to stabilize view-level evidence learning. When enabled, we regularize the learned conflict gate using a small calibration loss that enforces monotonic consistency with normalized JS disagreement, applied only when both views are present. Optimization uses Adam with cosine learning-rate scheduling and early stopping on validation AUROC; models are trained for 100 epochs with batch size 8, and the best checkpoint is used for evaluation.

3 Result Analysis

Table 1 shows that we observed consistent improvements in both discrimination and calibration with the proposed CEI models. On CBIS-DDSM, we found that CEI-S increases AUROC from the strongest prior baseline (Mammo-Clustering, 0.803) to 0.855, corresponding to a relative gain of approximately 6.5%. At the same time, ECE is reduced from 0.304 to 0.109, which represents a 64% decrease in calibration error. On VinDr-Mammo, CEI-S yields a relative AUROC improvement of approximately 2.9% along with a 69% reduction in ECE. We also found that these gains are achieved without increasing model capacity. CEI-Net-S (57.69M parameters) outperforms the substantially larger XFMamba (90.5M) in both AUROC and ECE, and the lightweight CEI-T already surpasses several CNN and Transformer baselines. Improvements of CEI-Net-S were found to be statistically significant under two-sided paired t-tests across identical random seeds ($p < 0.05$) for both AUROC and ECE on CBIS-DDSM and for ECE on VinDr-Mammo. These results confirm that reliability and conflict-aware evidential fusion improves both accuracy and calibration in multi-view prediction

Fig. 2 provides a spatial interpretation of CEI-Net under inter-view disagreement. In the exemplar panels, the selected Top- M tokens are overlaid on the mammograms and distinguished as malignant-supporting, benign-supporting, and most-discrepant tokens, linking localized image evidence to the per-view beliefs, class-wise JS disagreement, and conflict gates. The JS histograms in

Table 1. AUROC, ECE, and model size on CBIS-DDSM and VinDr-Mammo (mean $\pm$ std over 4 runs). $\dagger$ indicates statistically significant improvement over strongest prior baseline (Mammo-Clustering) ($p < 0.05$, paired t-test).

Method	Year	Backbone	Params (M)	CBIS-DDSM		VinDr-Mammo	
				AUROC	ECE	AUROC	ECE
DIVE-Net [4]	2023	ResNet-34	43.25	0.718 ± 0.005	0.264 ± 0.008	0.744 ± 0.002	0.340 ± 0.004
MV-Swin-T [3]	2024	Swin-T	58.32	0.719 ± 0.003	0.288 ± 0.012	0.722 ± 0.008	0.294 ± 0.007
MV-HFMD [21]	2024	ViT	54.07	0.683 ± 0.002	0.358 ± 0.005	0.703 ± 0.003	0.311 ± 0.002
BI-Mamba [7]	2024	Vision-Mamba	37.39	0.657 ± 0.002	0.320 ± 0.003	0.682 ± 0.007	0.426 ± 0.002
MV-TRAMS [25]	2025	ResNet-34	22.62	0.739 ± 0.004	0.282 ± 0.005	0.766 ± 0.002	0.303 ± 0.006
XFMamba [6]	2025	Vision-Mamba	90.54	0.783 ± 0.002	0.229 ± 0.006	0.775 ± 0.004	0.288 ± 0.013
Mammo-Clustering [9]	2025	Vision-Mamba	9.80	0.803 ± 0.002	0.304 ± 0.004	0.806 ± 0.003	0.234 ± 0.018
CEI-Net-T(iny)	2026	ConvNeXt-T	36.06	0.822 ± 0.003	0.118 ± 0.003	0.813 ± 0.004	0.079 ± 0.003
CEI-Net-S(mall)	2026	ConvNeXt-S	57.69	$0.855 \pm 0.005^\dagger$	$0.109 \pm 0.002^\dagger$	0.829 ± 0.005	$0.073 \pm 0.004^\dagger$

Fig. 2(a–b) are strongly concentrated near zero, indicating that CC and MLO predictions are typically consistent, with VinDr-Mammo showing a heavier tail consistent with greater acquisition variability. As JS increases, both error rate and NLL rise in CBIS-DDSM (Fig. 2(c)), demonstrating that stronger disagreement is associated with reduced predictive reliability. VinDr-Mammo exhibits a more gradual trend (Fig. 2(d)), consistent with reliability weighting reducing the influence of low-confidence views before conflict propagates to the fused belief. Fig. 2(e) shows a disagreement case in which the CC view provides stronger malignant evidence than the MLO view, producing a moderated fused prediction with elevated epistemic uncertainty. The nonzero class-wise JS disagreement reduces the corresponding conflict gate, indicating selective attenuation of disputed evidence rather than suppression of the malignant hypothesis. In contrast, Fig. 2(f) shows an agreement case with consistent CC/MLO predictions, near-zero JS disagreement, and fully open gates that preserve shared evidence. These examples demonstrate that CEI-Net links conflict-aware fusion to localized image evidence while moderating confidence under disagreement and preserving confidence when views agree.

Ablation Study. Table 2 shows a clear progression from deterministic fusion to the full CEI-Net formulation. Deterministic probability averaging performs worst, while equal-weight evidential fusion improves both AUROC and calibration. Conflict gating provides additional gains, and reliability weighting yields the largest single-module reduction in ECE (to 0.173 on CBIS-DDSM and 0.340 on VinDr-Mammo). Removing tokenized evidence degrades both AUROC and ECE, confirming the importance of localized evidence aggregation. Under missing-view conditions, performance drops gracefully. CC-only inference performs better than MLO-only on both datasets, suggesting that CC carries stronger standalone signal in these splits; the similarity between deterministic averaging and CC-only performance reflects the naive equal-weight baseline rather than CEI-Net collapsing to CC. Full CEI-Net-S consistently outperforms both single-view settings, confirming that the two views provide complementary evidence rather than redundant information. The ablation results show that the gains are not attributable to a single component: conflict gating yields larger AUROC improvements in high- JS settings, whereas reliability weighting produces

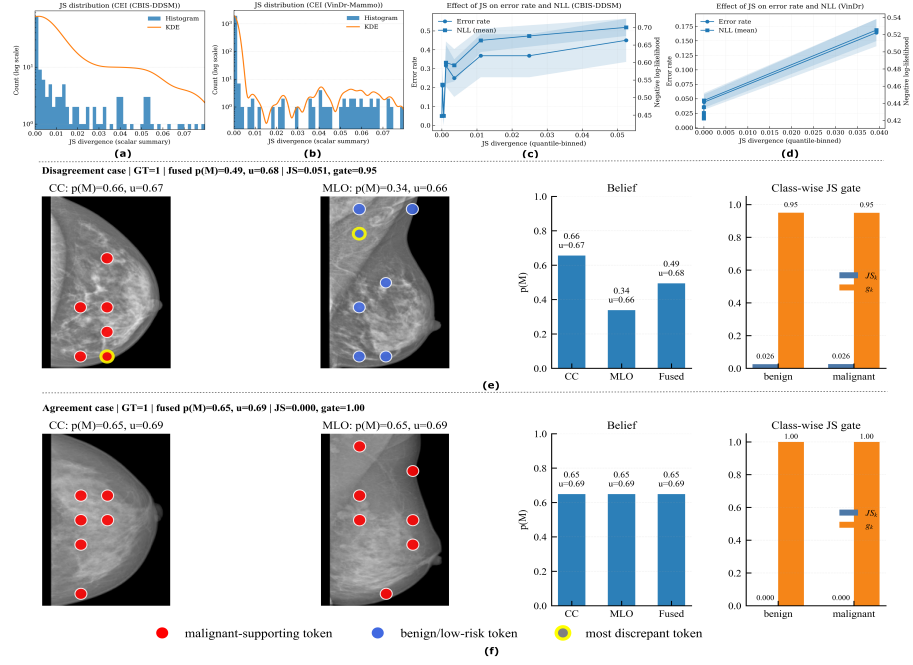


Fig. 2. Inter-view disagreement and its impact on predictive reliability.

Table 2. Ablation, missing-view robustness, and cross-dataset generalization.

Method	CBIS-DDSM			VinDr-Mammo		
	AUROC $\uparrow$	ECE $\downarrow$	NLL $\downarrow$	AUROC $\uparrow$	ECE $\downarrow$	NLL $\downarrow$
<i>In-dataset evaluation</i>						
Deterministic Prob-Avg	0.797	0.245	0.741	0.790	0.462	0.576
Evidential (Equal Weights)	0.805	0.195	0.670	0.795	0.368	0.511
JS Gating Only	0.812	0.182	0.645	0.802	0.344	0.502
Reliability Only	0.819	0.173	0.632	0.816	0.340	0.499
No Tokenization	0.812	0.178	0.638	0.808	0.359	0.513
<i>Missing-view at test time (no retraining)</i>						
CEI-Net-S (CC only)	0.796	0.135	0.581	0.804	0.300	0.212
CEI-Net-S (MLO only)	0.770	0.154	0.625	0.775	0.326	0.248
CEI-Net-S (Full)	0.855	0.109	0.498	0.829	0.073	0.145
<i>Cross-dataset generalization (thr=0.5, no fine-tuning)</i>						
Method	CBIS-DDSM $\rightarrow$ VinDr-Mammo			VinDr-Mammo $\rightarrow$ CBIS-DDSM		
	AUROC $\uparrow$	ECE $\downarrow$	NLL $\downarrow$	AUROC $\uparrow$	ECE $\downarrow$	NLL $\downarrow$
Deterministic Prob-Avg	0.503	0.612	0.941	0.680	0.154	0.701
Evidential (Equal Weights)	0.514	0.558	0.866	0.689	0.121	0.648
CEI-Net-S (Full)	0.522	0.126	0.373	0.698	0.082	0.605

the largest reduction in ECE, indicating complementary roles in evidential fusion (Table 2). Only their joint use in CEI-Net-S improves AUROC, ECE, and NLL simultaneously.

Cross-dataset generalization. Under domain shift, CEI-Net-S consistently provides the best calibration and NLL, while also improving AUROC. During

training on CBIS and VinDr testing, AUROC increases from 0.503 to 0.522 and ECE decreases from 0.612 to 0.126. For VinDr→CBIS, CEI-Net-S achieves the highest AUROC (0.698) and lowest ECE (0.082). These results indicate improved robustness primarily through better uncertainty estimation under dataset shift.

4 Conclusion

We presented CEI-Net, a multi-view mammography framework that performs Dirichlet-space evidential fusion using reliability from uncertainty and class-conditional JS conflict gating. By aggregating evidence over salient regional tokens, CEI-Net interprets inter-view disagreement as insufficient reliable evidence rather than noise, improving calibration, robustness, and fusion interpretability. Experiments on CBIS-DDSM and VinDr-Mammo show consistent gains over CNN, Transformer, and SSM baselines with graceful degradation under missing views and stable performance across random initializations. Although missing-view robustness is evaluated, future work will test view mispairing, acquisition corruption, and imperfect breast cropping in prospective screening workflows. Future work will extend CEI-Net to multi-class assessment, high-resolution patch, integrate vision-language priors, and develop uncertainty-aware clinical decision support for large-scale screening.

Acknowledgement

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number ALLRP 561264 - 21].

Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), [numéro de référence ALLRP 561264 - 21].

Disclosure of Interest

The authors declare that they have no competing interests.

References

1. Thrun, S., Hansen, S., Sun, Z., Blum, N., Salahuddin, S. A., Wickstrøm, K., Wetzer, E., Jenssen, R., Stille, M., Kampffmeyer, M.: Reconsidering explicit longitudinal mammography alignment for enhanced breast cancer risk prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 495–505. Springer (2025).
2. Truong, T.T.N., Nguyen, H.T., Lam, T.B., Nguyen, D.V.M., Nguyen, P.H.: Delving into ipsilateral mammogram assessment under multi-view network. In: International Workshop on Machine Learning in Medical Imaging (MLMI 2023), LNCS, pp. 367–376. Springer, Cham (2023).

3. Sarker, S., Sarker, P., Bebis, G., Tavakkoli, A.: MV-Swin-T: Mammogram classification with multi-view Swin Transformer. In: IEEE International Symposium on Biomedical Imaging (ISBI 2024), pp. 1–5. IEEE, (2024).
4. Nguyen, T.-H., Kha, Q.H., Truong, T.N.T., Lam, B.T., Ngo, B.H., Dinh, Q.V., Le, N.Q.K.: Towards robust natural-looking mammography lesion synthesis on ipsilateral dual-views breast cancer analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2023), pp. 2564–2573. IEEE (2023).
5. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: First Conference on Language Modeling (2024).
6. Zheng, X., Chen, X., Gong, S., Griffin, X., Slabaugh, G.: XFMamba: Cross-fusion Mamba for multi-view medical image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2025), LNCS, pp. 672–682. Springer, Cham (2025).
7. Yang, Z., Zhang, J., Wang, G., Kalra, M.K., Yan, P.: Cardiovascular disease detection from multi-view chest X-rays with Bi-Mamba. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2024), LNCS, pp. 134–144. Springer, Cham (2024).
8. Bayatmakou, F., Taleei, R., Simone, N., Mohammadi, A.: Mammo-Mamba: A hybrid state-space and transformer architecture with sequential mixture of experts for multi-view mammography. arXiv preprint arXiv:2507.17662 (2025).
9. Yang, S., Zhang, C., Liang, X., Zang, Q., Yu, J., Zeng, L., Luo, X., Xing, Y., Pan, X., Li, Q., et al.: Mammo-Clustering: Context clustering based multi-view tri-level information fusion for lesion location and classification in mammography. arXiv preprint arXiv:2507.18642 (2025).
10. Ibragimov, A., Senotrusova, S., Litvinov, A., Ushakov, E., Karpulevich, E., Markin, Y.: MamT 4: Multi-view Attention Networks for Mammography Cancer Classification. In: IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC 2024), pp. 1965–1970. IEEE (2024).
11. Sensoy, M., Kaplan, L., Kandemir, M.: Evidential deep learning to quantify classification uncertainty. In: Advances in Neural Information Processing Systems (NeurIPS 2018), vol. 31. Curran Associates, Inc. (2018).
12. Mahmud, K.M.F., Qasem, A., Staley, J.M., Yoder, R., Aripoli, A., Stecklein, S.R., Sharma, P., Zhou, Z.: ER2Net: An evidential reasoning rule-enabled neural network for reliable triple-negative breast cancer tumor segmentation in magnetic resonance imaging. *Journal of Medical Imaging* **13**(1), 014005 (2026).
13. Gudhe, N.R., Sudah, M., Mannermaa, A., Kosma, V.-M., Behravan, H.: Multi-view deep evidential fusion neural network for assessment of screening mammograms. (2024).
14. Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S.: A ConvNet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11976–11986. IEEE (2022).
15. Menéndez, M. L., Pardo, J. A., Pardo, L., Pardo, M. del C.: The Jensen–Shannon divergence. *Journal of the Franklin Institute* **334**(2), 307–318 (1997).
16. Yang, S., Zhang, C., Zang, Q., Yu, J., Zeng, L., Luo, X., Xing, Y., Pan, X., Li, Q., Liang, X., et al.: Mammo-Clustering: A multi-views tri-level information fusion context clustering framework for localization and classification in mammography. arXiv preprint arXiv:2409.14876 (2024).
17. Van Tulder, G., Tong, Y., Marchiori, E.: Multi-view analysis of unregistered medical images using cross-view transformers. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2021), LNCS, pp. 104–113. Springer, Cham (2021).

18. Ahmed, M. R., Lasserre, P.: FusionSegNet: A hierarchical multi-axis attention and gated feature fusion network for breast lesion segmentation with uncertainty modeling in ultrasound imaging. *Information Fusion*, 103399 (2025).
19. Lamprou, C., Katsikari, K., Rahmani, N., Hadjileontiadis, L.J., Seghier, M., Alshehhi, A.: StethoNet: Robust breast cancer mammography classification framework. *IEEE Access* (2024). IEEE.
20. Ahmed, M. R., Shehata, M. S., Lasserre, P.: Integrating lightweight convolutional neural network with entropy-informed channel attention and adaptive spatial attention for OCT-based retinal disease classification. *Computers in Biology and Medicine* **190**, 110071 (2025).
21. Black, S., Souvenir, R.: Multi-view classification using hybrid fusion and mutual distillation. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2024)*, pp. 270–280. IEEE (2024).
22. Wei, H., Fan, H.: MM-CLIP: Multi-granularity Mammography CLIP for Breast Cancer Diagnosis. In: *Proceedings of the 9th International Conference on Digital Signal Processing (DSP 2025)*, pp. 92–99 (2025).
23. Yan, Z., Zhang, W., Yang, Y., Saha, A.: Multi-Distribution Mammogram Classification: Leveraging CLIP for enhanced generalization across diverse datasets. In: *IEEE 22nd International Symposium on Biomedical Imaging (ISBI 2025)*, pp. 1–5. IEEE (2025).
24. Zheng, S.-F., Lee, H., Kooi, T., Diba, A.: MV-MLM: Bridging multi-view mammography and language for breast cancer diagnosis and risk prediction. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2025)*, pp. 1213–1222. IEEE (2025).
25. Nguyen, H.T., Lam, T.B., Truong, T.T.N., Duong, T.D., Dinh, V.Q.: MV-Trams: An Efficient Tumor Region-Adapted Mammography Synthesis Under Multi-View Diagnosis. In: *SSRN Preprint*, (2025).
26. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data* 4(1), 1–9 (2017). Springer, Nature Publishing Group.
27. Pham, H.H., Nguyen, H.T., Nguyen, H.Q.: VinDr-Mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography. *Sci Data*, (2022). Springer, Nature Publishing Group.