UAI 2026 — CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE

Not Just How Much, But Where: Decomposing Epist ic Uncertainty into Per-Class ons

Mame Diarra Toure • David A.
Department of Mathematics and Stati: 602.00387

SELECTIVE RISK REDUCTION **34.7%** over MI baseline

OOD AUROC **0.815** best on MIMIC

1 **Novel Decomposition:** MI → per-class vector $C_k(x) = \sigma^2_k/(2\mu_k)$ with boundary correction

2 **Axiomatic Theory:** Boundedness, monotonicity, additivity, invariance — plus skewness diagnostic ρ_k

3 **Selective Prediction:** C_k reduces critical FNR by 34.7% over MI and 56.2% over variance on diabetic retinopathy

4 **OOD Detection:** $\sum_k C_k$ achi AUROC; per-class view re asymmetric shifts invisible

MOTIVATION

Safety-critical classifiers face **asymmetric failure costs**. A scalar mutual information (MI) of 0.3 nats carries entirely different implications depending on whether the model's confusion involves two benign classes or a benign and a safety-critical one.

Yet all standard Bayesian deep learning methods reduce epistemic uncertainty to a **single per-input scalar**:

$$MI(y; \omega \mid x) = H(\mu) - \mathbb{E}[H(p)]$$

This number tells us *how uncertain* the model is — not *which classes* drive that uncertainty. We fix this.

KEY GAP

A model unsure between Grade 0 (no disease) and Grade 3 (severe) should be treated very differently from one unsure between Grade 1 and Grade 2 — but MI cannot distinguish these.

DECOMPOSITION

Starting from a **2nd-order Taylor expansion** of the entropy, we derive:

$$C_k(x) = \sigma^2_k \,/\, (2\mu_k), \quad \mu_k = \mathbb{E}[p_k], \quad \sigma^2_k = Var[p_k]$$

The $1/\mu_k$ weight **corrects boundary suppression**, ensuring rare and common classes are comparable. By construction:

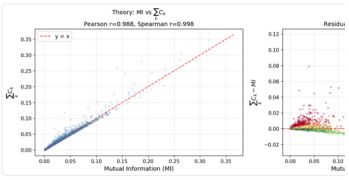
$$\sum_k C_k \approx MI \quad \text{CORE IDENTITY}$$

Skewness diagnostic: validity of the approximation is monitored via

$$\rho_k = |m_{3,k}| \,/\, (3\mu_k \cdot Var[p_k])$$

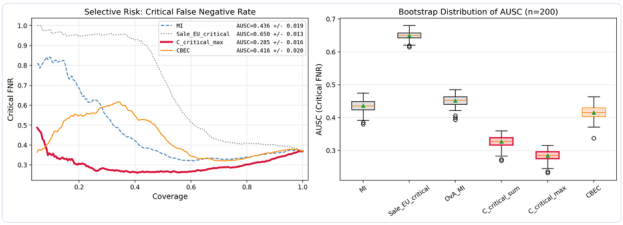
When $\rho_k < 0.3$, the Taylor approximation is reliable. Inputs violating this threshold are flagged for review.

THEORY VALIDATION



Pearson r = 0.988, Spearman r = 0.998 between MI (right) are driven by max ρ_k, confirming the

EXPERIMENT 1 — SELECTIVE PREDICTION (DIABETIC RETINOPATHY)



Left: Critical FNR vs. coverage. Right: Bootstrap AUSC distribution (n=200). C_k scores consistently dominate.

34.7% **Reduction in selective risk** $C_{critical,max}$ vs. MI (AUSC 0.285 vs. 0.436)

56.2% **Reduction vs. variance** $C_{critical,max}$ vs. Sale_EU_critical (AUSC 0.650)

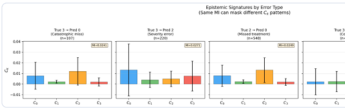
0.196 **CBEC score** on MC Dropout best overall AUSC for cross-boundary confusion

Setup: 4-class retinopathy grading (Grade 0–3). Abstain on least-certain inputs; measure critical false negative rate. Models: Low-rank BNN, MC Dropout.

KEY INSIGHT

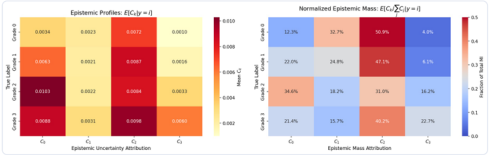
$C_{critical}$ ignores classes irrelevant to the risk definition — it only accumulates C_k for clinically critical grades. This targeted abstention is what MI cannot achieve.

ERROR SIGNATURES



Same MI can mask different C_k profiles: catastrophic m error (3→2) are distinguishable.

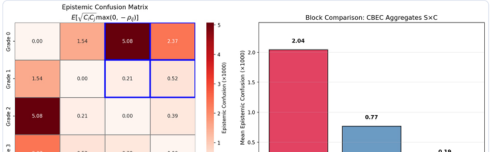
EXPERIMENT 2 — OOD DETECTION



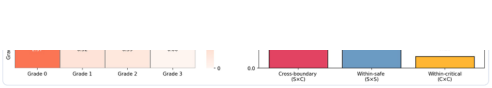
Epistemic profiles $E[C_k|y=i]$ (left) and normalised mass $E[C_k/\sum C_j|y=i]$ (right) on MIMIC-III. Grade 2 concentrates 34.6% on C_0 — a cross-boundary signal invisible to MI.

0.815 **AUROC for $\sum_k C_k$** best OOD score on MIMIC (ICU vs. Newborn)

EPISTEMIC CONFUSION MATRIX (CBEC)

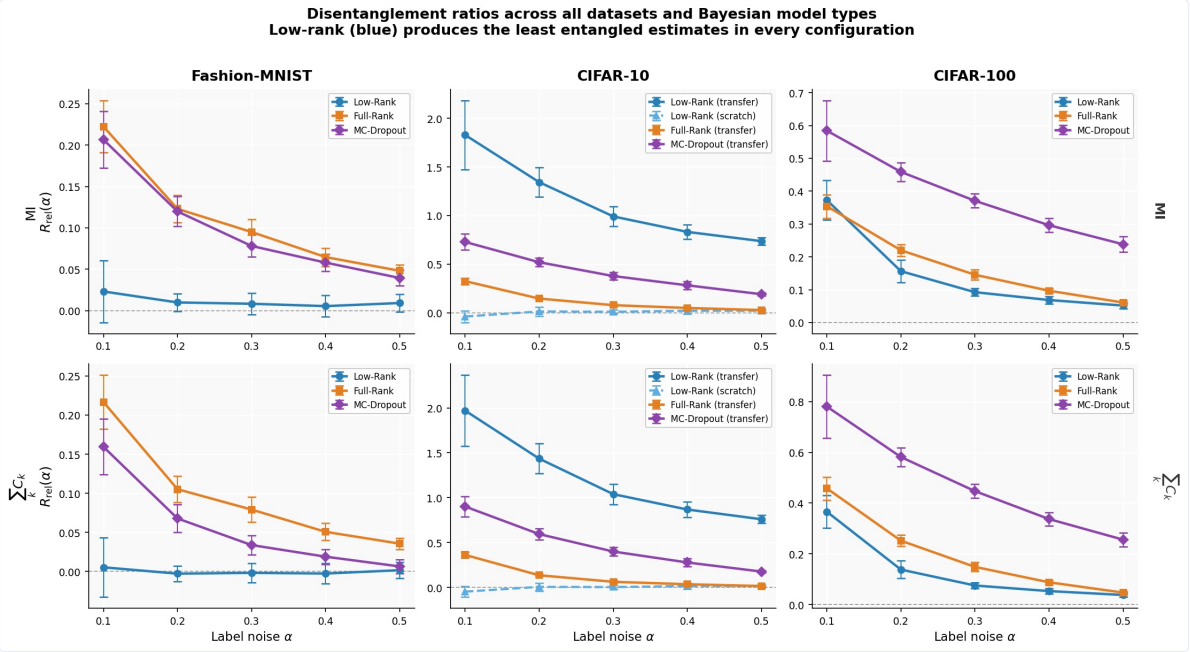


0.802 MI AUROC — $\sum_k C_k$ outperforms MI despite same scale

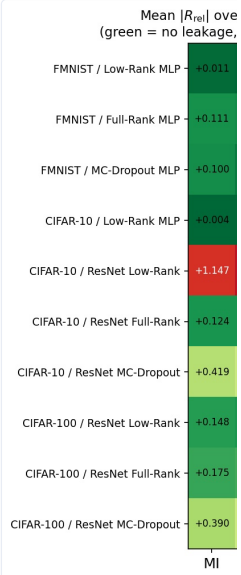


CBEC = $\mathbb{E}[N_i(C_iC_j) \cdot \max(0,-\rho_{ij})]$. Grade 0–Grade 2 pair shows highest cross-boundary epistemic confusion (5.08), while within-critical confusion is low (0.39).

EXPERIMENT 3 — DISENTANGLEMENT STUDY (LABEL NOISE)



Relative disentanglement ratio $R_{rel}(\alpha)$ across FMNIST, CIFAR-10, CIFAR-100 and three Bayesian model types. Low-rank BNN (blue) achieves near-zero R_{rel} under end-to-end training in every configuration.



Mean $|R_{rel}|$ over α∈[0.1, 0.5]. Green = end low-rank rows achieve the strong

<0.05 R_{rel} for Low-rank on FMNIST

1.97 R_{rel} for same mod learning — 40× w

KEY FINDING

How uncertainty is propagated network matters as much as ho
The posterior family dominates

KEY TAKEAWAYS

1. **C_k is a principled, closed-form extension of MI** that adds directionality: each component tells you how much epistemic mass the model places on class k.
2. **The skewness diagnostic ρ_k** is not just a correction — it is a first-order quality certificate for the posterior approximation, applicable to any Bayesian classifier.
3. **Posterior quality dominates metric choice**: end-to-end low-rank BNNs disentangl than the same architecture under transfer learning, regardless of whether MI or $\sum C$
4. **Practical recommendation:** report both $\sum_k C_k$ and the per-class profile. In high-carc ($K \geq 50$), use truncated summation or μ_k-weighted aggregation.

Mame Diarra Toure • mame.toure@mail.mcgill.ca • McGill University, Dept. Mathematics & Statistics
UAI 2026 • arXiv:2602.00387
Supervised by David A. Stephens • Sophie Germair