

# GEM-FI

## Gated Evidential Mixtures with Fisher Modulation

*Marco Mustafa Mohammed · Fatemeh Daneshfar · Pietro Liò*

University of Kurdistan, Iran · University of Cambridge, UK

ICML 2026 · Seoul, South Korea

[github.com/Marcorazhan/GEM-FI](https://github.com/Marcorazhan/GEM-FI)

# Why Existing Methods Fall Short

## Evidential Deep Learning (EDL)

✓ **Single-pass inference (fast)**

- ✗ Overconfident on OOD inputs
- ✗ Poorly calibrated under shift
- ✗ Cannot capture multi-modal epistemic uncertainty

*Uncertainty concentrated  
in a narrow boundary band  
→ Overconfident far from support*

vs

## Density-Aware EDL (DAEDL)

✓ **Better distance awareness**

- ✗ Density cue is static & offline
- ✗ Not optimized end-to-end
- ✗ Still misses multi-modal epistemic structure

*Offline GDA density does not  
learn from data jointly  
→ Brittle under representation shift*

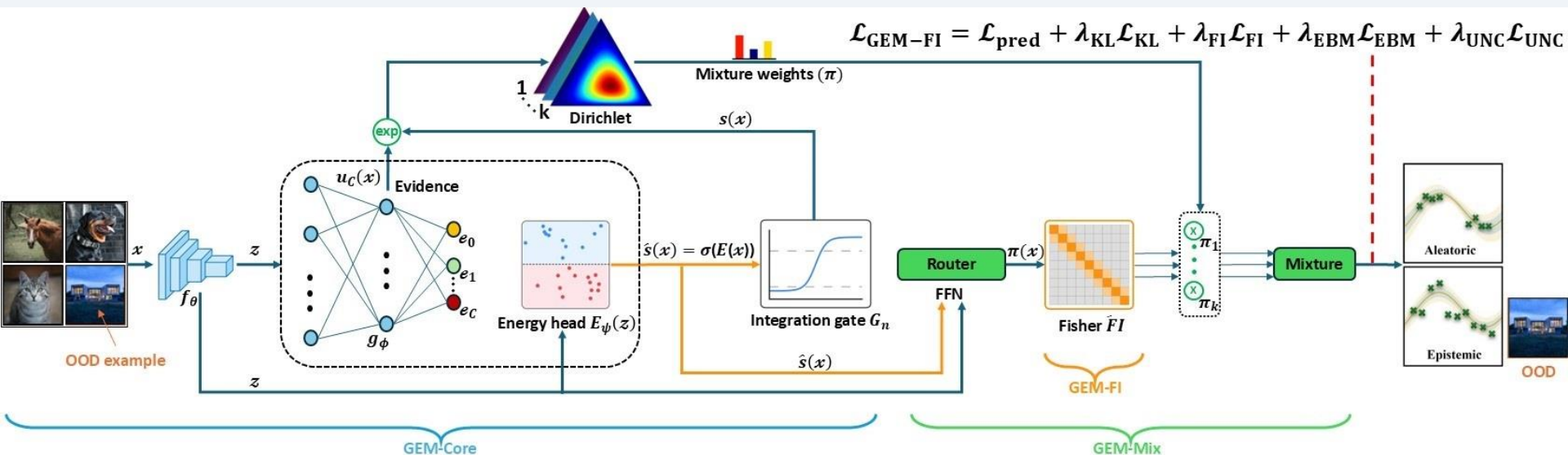
→ We need an in-model, end-to-end learnable support signal that gates evidence and captures multi-modal epistemic uncertainty.

# GEM-FI Architecture

Energy gating + mixture routing + Fisher stabilization in a single pass

## Key idea

- The backbone extracts features  $z$  and outputs evidential logits.
- GEM-CORE learns an energy  $E(x)$  and maps it to a gate  $s(x)$  that suppresses evidence off support.
- GEM-MIX routes inputs across  $K$  evidential heads; GEM-FI uses Fisher regularization to prevent head collapse.



# GEM Family: Three Progressions

## ① GEM-CORE

### Energy-to-Gate Learning

Learns feature-level energy  $E(x)$  and maps it to a bounded per-class gate  $s(x)$  that directly modulates Dirichlet evidence — end-to-end.

*Off-support → smaller gate → conservative prediction*

## ② GEM-MIX

### Mixture of Beliefs

Adds  $K$  lightweight evidential heads with learned routing weights  $\pi(x)$ . Captures multi-modal epistemic structure without ensembling.

*Single-pass · No extra forward passes*

## ③ GEM-FI

### Fisher Modulation

Fisher-informed regularizer  $\mathcal{L}_{\text{FI}}$  penalizes high-sensitivity head allocations and stabilizes mixture routing to prevent head collapse.

*Smoother boundaries · Stronger OOD separation*

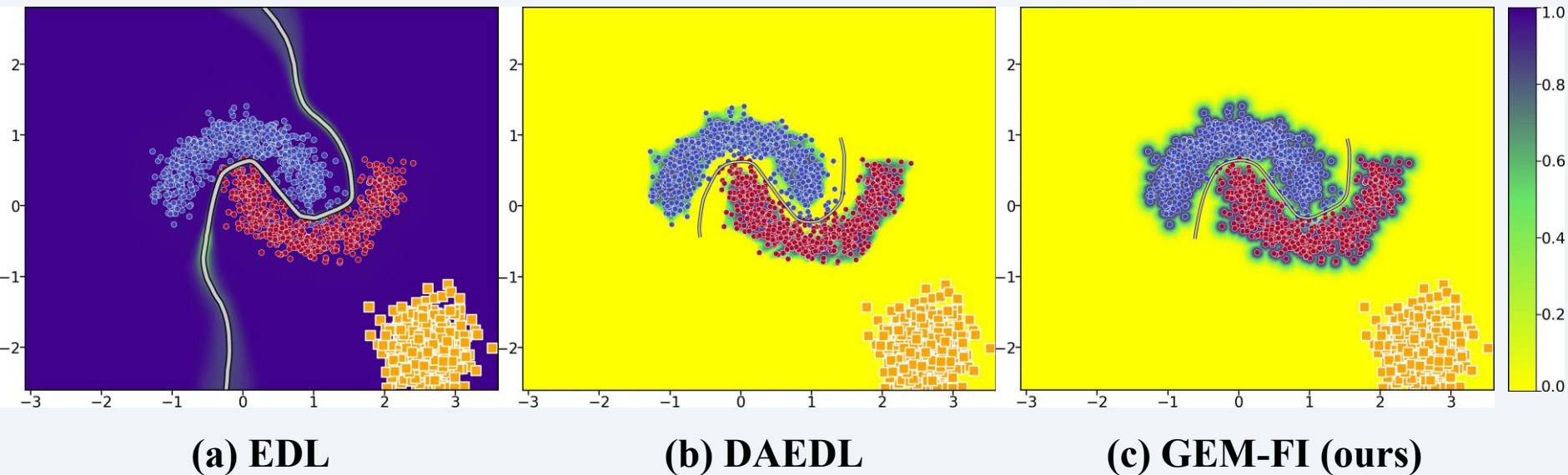
$$\text{Total loss: } \mathcal{L}_{\text{GEM-FI}} = \mathcal{L}_{\text{pred}} + \lambda_{\text{KL}} \mathcal{L}_{\text{KL}} + \lambda_{\text{FI}} \mathcal{L}_{\text{FI}} + \lambda_{\text{EBM}} \mathcal{L}_{\text{EBM}} + \lambda_{\text{UNC}} \mathcal{L}_{\text{UNC}}$$

# Boundary and OOD Intuition

Two-moons example: smoother uncertainty and safer OOD behavior

## Takeaway

- EDL keeps uncertainty in a narrow band and remains overconfident far from support.
- DAEDL improves distance awareness, but its density cue is static and can still miss OOD regions.
- GEM-FI produces smoother uncertainty near curved boundaries and lower confidence on off-support inputs

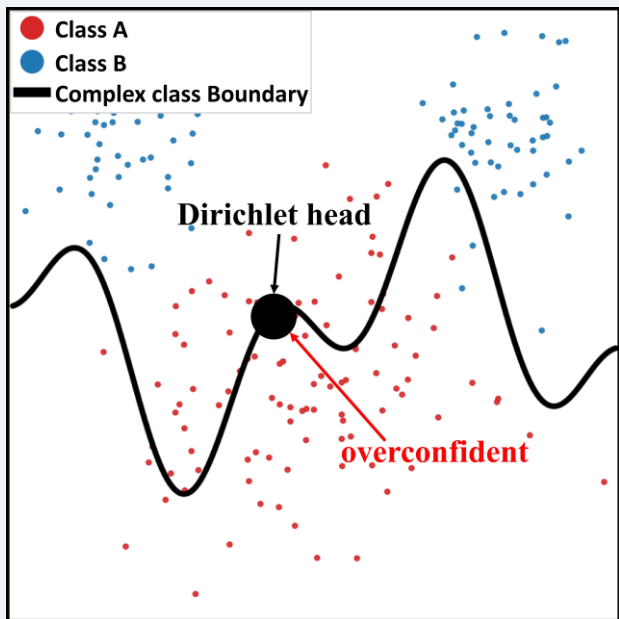


# Why Multi-Head + Fisher?

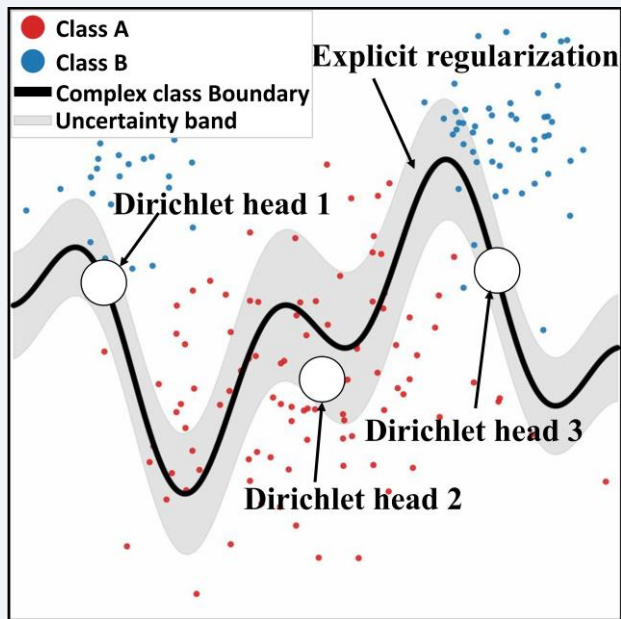
Retaining epistemic uncertainty on complex non-convex boundaries

## Interpretation

- A single Dirichlet head can collapse to an overconfident allocation on a complex boundary.
- Multiple routed heads specialize on different regions of the boundary.
- Fisher modulation stabilizes routing and preserves a wider uncertainty band.



(a) DAEDL



(b) GEM-FI (ours)

# Results: CIFAR-10 Benchmarks

**93.75%**

Test Accuracy

*DAEDL: 91.11% (+2.64 pp)*

**6.81**

Brier  $\times$  100

*DAEDL: 14.27 (-7.46)*

**99.94%**

Misclass. AUPR

*DAEDL: 99.08% (+0.86)*

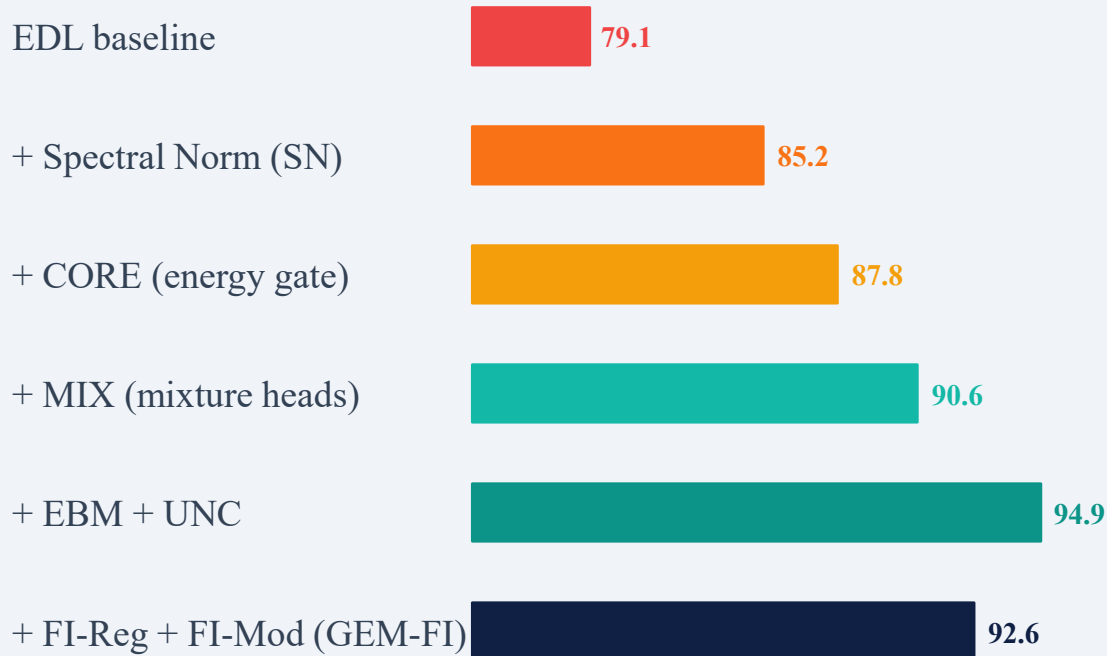
OOD Detection — AUPR (Epistemic)  $\uparrow$  higher is better

Method	MNIST $\rightarrow$ KMNIST	MNIST $\rightarrow$ FMNIST	CIFAR-10 $\rightarrow$ SVHN	CIFAR-10 $\rightarrow$ CIFAR-100
EDL	96.31	97.84	79.32	84.80
DAEDL	99.92	99.87	85.54	88.19
Re-EDL	99.03	99.65	89.89	88.30
<b>GEM-FI (ours)</b>	<b>99.96</b>	<b>99.99</b>	<b>92.59</b>	<b>90.20</b>

*All results averaged over 5 random seeds.  $\uparrow$  higher is better.*

# Ablation: Every Component Matters

CIFAR-10 → SVHN (Epistemic AUPR ↑)



## Key Findings

- ① SN alone already stabilizes training and improves OOD (85.2 vs 79.1)
- ② CORE adds end-to-end gating on top of SN for further gain
- ③ MIX captures multi-modal epistemic structure without ensemble cost
- ④ FI components need EBM+UNC to work best — FI refines, not bootstraps
- ⑤ Removing SN from full model degrades both accuracy and uncertainty

# GEM-FI: Takeaways

## Single-Pass Inference

All uncertainty estimation in one forward pass — no ensembling needed.

## Multi-Modal Epistemic Uncertainty

Mixture of evidential heads captures complex boundary structure with Fisher stabilization.

## End-to-End Density Gating

Learned energy head gates Dirichlet evidence based on feature-space support — jointly optimized.

## State-of-the-Art Results

93.75% accuracy · Brier 6.81 · OOD AUPR 92.59 on CIFAR-10→SVHN