Density-Informed Pseudo-Counts for Calibrated Evidential Deep Learning

Pietro Carlotti[1,*], Nevena Gligić[1,*], Arya Farahi[1]
[1]*Department of Statistics and Data Sciences, The University of Texas at Austin*
**Equal contribution.*

The University of Texas at Austin
Department of Statistics and Data Sciences
College of Natural Sciences

CosmicAI
Funded by the U.S. National Science Foundation and the Simons Foundation

Problem

Suppose data is generated according to:

$$(Y_i, X_i) \overset{iid}{\sim} P^*_{Y,X}, \qquad i = 1,\dots,n$$

Then $P^*_{Y,X} = P^*_{Y|X} P^*_X$ where X is the data and Y are the labels, and $Y_i | p_i \overset{ind}{\sim} f_{p_i} = \text{Cat}(p_i)$.

Goal: learn $P^*_{Y|X}$ and quantify the associated uncertainty.

Evidential Deep Learning (EDL)

$$\mathscr{L}^{\lambda}_{\text{EDL}}(\phi) = \mathscr{L}_{\text{data}}(\phi) + \lambda\, \mathscr{L}_{\text{reg}}(\phi),$$

Definition 2.6. *Tempered Independent Categorical-Dirichlet (ICD) Model.*

$$Y_i \mid p_i \overset{ind}{\propto} f^{\nu}_{p_i} = \prod_{k=1}^{K} p_i(k)^{\nu\, \mathbb{I}(Y_i = k)}, \quad i = 1,\dots,n,$$

$$p_i \overset{iid}{\sim} \pi = \text{Dir}(\alpha), \quad i = 1,\dots,n$$

for some $\alpha \in \mathbb{R}^K_+$ *and* $\nu > 0$.

Proposition 2.7. *Posterior distribution.*

$$\pi^{\nu}_{Y_i} = \text{Dir}(\alpha + \nu\, e_{Y_i}), \quad p_{1:n} | Y_{1:n} \sim \pi^{\nu}_{Y_{1:n}} = \prod_{i=1}^{n} \pi^{\nu}_{Y_i}$$

where e_{Y_i} *is the one-hot encoding of* Y_i, $i = 1,\dots,n$.

Definition 2.8. *Amortized VI for the Tempered ICD Model.*

$$p_{1:n} \mid Y_{1:n} \approx q^{\widehat{\phi}_n}_{X_{1:n}} = \underset{q^{\phi}_{X_{1:n}}}{\text{argmin}} \left\{ \text{KL}\left(q^{\phi}_{X_{1:n}} \,\|\, \pi^{\nu}_{Y_{1:n}} \right) \right\}$$

where

$$q^{\phi}_{X_{1:n}} = \prod_{i=1}^{n} q^{\phi}_{X_i} \quad \text{and} \quad q^{\phi}_{X_i} = \text{Dir}\left(\alpha + \text{NN}^{\phi}(X_i)\right), \quad i = 1,\dots,n,$$

$\text{NN}^{\phi} : \mathbb{R}^d \to \mathbb{R}^K_+$ *is a neural network with* $\phi \in \Phi$, *and* Φ *is the neural network class, with parameters* $\widehat{\phi}_n = \underset{\phi \in \Phi}{\text{argmin}} \left\{ \text{KL}\left(q^{\phi}_{X_{1:n}} \,\|\, \pi^{\nu}_{Y_{1:n}} \right) \right\}$.

Theorem 2.9. *Amortized VI objective is equivalent to EDL training:*

$$\widehat{\phi}_n = \underset{\phi \in \Phi}{\text{argmin}} \left\{ \mathscr{L}^{\frac{1}{\nu}}_{\text{EDL}}\left(\phi; Y_{1:n}, X_{1:n}\right) \right\}$$

Proposition 2.12. *Vacuity (under certain conditions):*

$$u(\alpha + \text{NN}^{\hat{\phi}_n}_{X_i}) = \frac{K}{\alpha_0 + \nu}, \quad i = 1,\dots,n$$

Implications:

$\Longrightarrow$ EDL is **governed by a user-defined** ν, independently of the data fit.
$\Longrightarrow$ EDL's uncertainty persists even as the sample size grows.
$\Longrightarrow$ **EDL cannot distinguish epistemic from aleatoric uncertainty.**
$\Longrightarrow$ EDL is overconfident on ID and OOD data.

Solution: DIP-EDL

We propose:

$$q^{\psi,\phi}_{X_i} = \text{Dir}\left(\alpha + n\, \text{DE}^{\psi}_{X_i}\, \text{NN}^{\phi}_{X_i}\right), \quad i = 1,\dots,n,$$

where $\text{DE}^{\psi} : \mathbb{R}^d \to \mathbb{R}_+$ is a density estimator for P^*_X with parameters ψ and $\text{NN}^{\phi} : \mathbb{R}^d \to \mathbb{R}^K_+$ is a neural network estimator for $P^*_{Y|X}$ with parameters ϕ.

Properties:

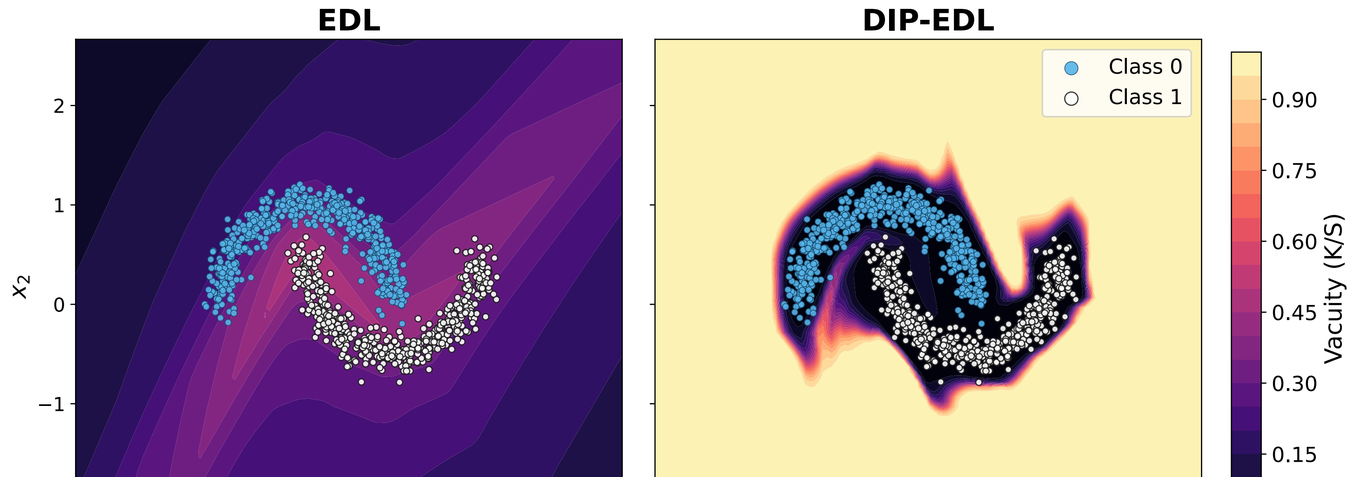
- **Distributional awareness** - uncertainty is higher in low-density regions and outside the training distribution.
- **Asymptotic consistency** - epistemic uncertainty vanishes asymptotically as the number of samples grow, while aleatoric uncertainty may persist.

$$\text{DE}^{\hat{\psi}_n}_{X_i} \xrightarrow{p} P^*_X(X_i), \quad \text{NN}^{\hat{\phi}_n}_{X_i} \xrightarrow{p} P^*_{Y|X}(\,\cdot \mid X_i) \quad \Rightarrow \quad p_i \mid X_{1:n}, Y_{1:n} \approx q^{\hat{\psi}_n, \hat{\phi}_n}_{X_i} \xrightarrow{p} P^*_{Y|X}(\,\cdot \mid X_i)\,.$$

- **Modularity** - density estimator and classifier trained separately.
- **Efficiency** - single forward pass at inference.

Applications: disaster response management, medical diagnosis, autonomous driving, fraud detection

Vacuity comparison (toy example)



Experiments and results: CIFAR-10

Model	ID Performance Metrics		OOD Performance Metrics					
	Accuracy (↑)	Brier Score (↓)	AUROC (↑)		AUPR (↑)		OOD Brier Score (↓)	
	CIFAR-10		CIFAR-100	SVHN	CIFAR-100	SVHN	CIFAR-100	SVHN
EDL	0.7535	0.3520	0.7588	0.7519	0.6900	0.8094	**0.0960**	**0.0492**
R-EDL	0.8957	0.1737	0.8484	0.8741	0.8016	0.9150	0.3603	0.3262
Re-EDL	0.8939	0.1737	0.8603	0.9102	0.8229	0.9427	0.6613	0.6303
DAEDL	0.8879	0.1760	0.8323	0.8606	0.7913	0.9088	0.4087	0.3843
PostNet	0.8169	0.2623	0.7691	0.7760	0.7254	0.8532	0.4609	0.4734
MC Dropout	***0.9534***	***0.0729***	0.8833	0.8802	0.8407	0.9053	0.5958	0.6111
Deep Ensembles	**0.9614**	**0.0587**	**0.9072**	***0.9628***	***0.8807***	***0.9802***	0.5222	0.4200
DIP-EDL	0.9503	0.0978	***0.9002***	**0.9660**	**0.8859**	**0.9830**	***0.3251***	***0.1109***

Table 1. **ID and OOD performance when models trained on CIFAR-10.** We evaluate ID accuracy and calibration, alongside OOD detection against CIFAR-100 (near-OOD) and SVHN (far-OOD).

Components			ID Performance Metrics		OOD Performance Metrics					
			Accuracy (↑)	Brier Score (↓)	AUROC (↑)		AUPR (↑)		OOD Brier Score (↓)	
n	DE	NN	CIFAR-10		CIFAR-100	SVHN	CIFAR-100	SVHN	CIFAR-100	SVHN
✓	✓	X	0.1000	0.9000	**0.9000**	**0.9686**	**0.8864**	**0.9850**	**0.0000**	**0.0000**
✓	X	✓	**0.9496**	**0.0831**	0.5036	0.5014	0.5014	0.7224	0.6945	0.6412
X	✓	✓	**0.9496**	0.7653	**0.9000**	**0.9686**	**0.8864**	**0.9850**	0.0004	**0.0000**
✓	✓	✓	**0.9496**	0.0978	**0.9000**	**0.9686**	**0.8864**	**0.9850**	0.3303	0.1032

Table 2: **Ablation study of DIP-EDL.** We evaluate the individual and combined contributions of training set size, n, the density estimator, $\text{DE}^{\psi}_{X_i}$, and the discriminative classifier, $\text{NN}^{\phi}_{X_i}$. The results demonstrate how the model splits it's tasks across the three components and achieves the best performance when all three components are included.

Experiments and results (real data): LAMOST

Model	ID Performance Metrics		OOD Performance Metrics		
	Accuracy (↑)	Brier Score (↓)	AUROC (↑)	AUPR (↑)	OOD Brier Score (↓)
	LAMOST		Star	Star	Star
EDL	0.4955	0.7585	0.4934	0.9884	**0.3595**
R-EDL	0.7706	0.4273	0.6022	0.9933	0.0783
Re-EDL	0.7706	0.4272	0.5957	0.9932	0.0752
DAEDL	0.7706	0.2802	0.5748	0.9904	0.1418
PostNet	0.7328	0.4231	0.5315	0.9913	0.0297
MC Dropout	***0.5851***	***0.6055***	0.4974	0.9886	0.3729
Deep Ensembles	**0.8929**	**0.1497**	**0.7425**	***0.9957***	0.2394
DIP-EDL	0.8929	0.1555	***0.6789***	**0.9948**	***0.2517***

Table 3. **ID and OOD performance when models trained on LAMOST (Galaxy, Quasar).** We evaluate ID accuracy and calibration, alongside OOD detection against LAMOST Star spectra.

References

[1] Sensoy, M., Kaplan, L., and Kandemir, M. Evidential deep learning to quantify classification uncertainty. Advances in neural information processing systems, 31, 2018.
[2] Chen, M., Gao, J., and Xu, C. R-edl: Relaxing nonessential settings of evidential deep learning. In The Twelfth International Conference on Learning Representations, 2024.
[3] Mengyuan Chen, Junyu Gao, and Changsheng Xu. Revisiting essential and nonessential settings of evidential deep learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025.
[4] Yoon, T. and Kim, H. Uncertainty estimation by density aware evidential deep learning. In Proceedings of the 41st International Conference on Machine Learning, ICML'24. JMLR.org, 2024.
[5] Charpentier, B., Zuegner, D., and Guennemann, S. Posterior network: uncertainty estimation without ood samples via density-based pseudo-counts. In Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.
[6] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning, pages 1050–1059. PMLR, 2016.
[7] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in neural information processing systems, 30, 2017.