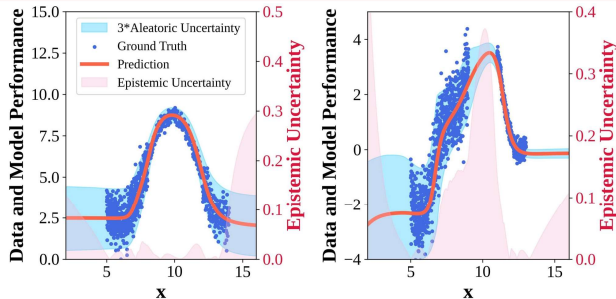


## Background

Deep models can be overconfident, so reliable prediction requires understanding **not only how uncertain a model is, but also why it is uncertain**. Uncertainty in deep learning mainly comes from two sources: aleatoric uncertainty, caused by noise or ambiguity in the data, and epistemic uncertainty, caused by limited model knowledge or insufficient training coverage. Distinguishing them is important for trustworthy decision-making. Existing methods often estimate only one uncertainty type, or they require modifying and retraining the base model. This increases computational cost and limits compatibility with pretrained systems.

### AI ← CUPID: Uncertainty

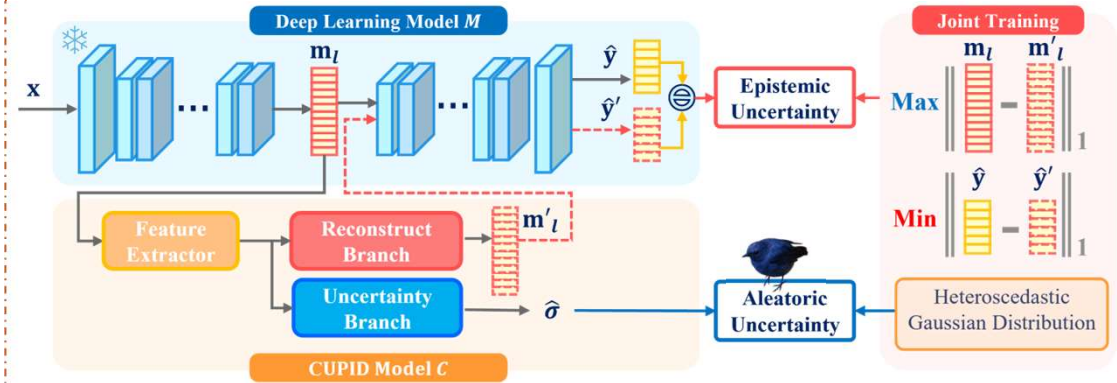


**Figure 1: CUPID uncertainty estimation on a 1D regression toy problem.** CUPID is inserted into an MLP-based predictive model. CUPID captures both aleatoric (blue) and epistemic (red) uncertainty.

**Table 1: Performance of misclassification detection (misclassified samples as positive).** CUPID Aleatoric achieved the best performance on GLV2, while CUPID Epistemic performed best on HAM10000, suggesting different dominant sources of uncertainty across datasets.

Method	GLV2			HAM10000		
	AUC (↑)	AURC (↓)	Spearman (↑)	AUC (↑)	AURC (↓)	Spearman (↑)
CUPID Alea.	<b>0.870 ± 0.002</b>	<b>0.018 ± 0.001</b>	0.941 ± 0.004	0.769 ± 0.023	0.067 ± 0.007	0.722 ± 0.014
CUPID Epis.	0.769 ± 0.015	0.034 ± 0.002	0.701 ± 0.051	<b>0.855 ± 0.006</b>	<b>0.047 ± 0.001</b>	<b>0.907 ± 0.001</b>
MC Dropout	0.768 ± 0.006	0.027 ± 0.001	0.888 ± 0.005	0.829 ± 0.001	0.076 ± 0.001	0.861 ± 0.002
Rate-in	0.815 ± 0.006	0.024 ± 0.001	0.816 ± 0.004	0.846 ± 0.001	0.048 ± 0.000	<b>0.915 ± 0.000</b>
IGRUE	0.642 ± 0.007	0.058 ± 0.002	0.199 ± 0.004	0.548 ± 0.004	0.157 ± 0.002	0.027 ± 0.018
PostNet Alea.	0.671 ± 0.006	0.182 ± 0.004	0.641 ± 0.011	0.793 ± 0.007	0.142 ± 0.003	0.764 ± 0.006
PostNet Epis.	0.559 ± 0.031	0.238 ± 0.019	0.284 ± 0.054	0.751 ± 0.017	0.158 ± 0.010	0.698 ± 0.033
BNN	0.829 ± 0.018	0.025 ± 0.003	<b>0.954 ± 0.007</b>	0.793 ± 0.006	0.096 ± 0.004	0.821 ± 0.009
DEC	0.503 ± 0.012	0.192 ± 0.006	0.803 ± 0.139	0.837 ± 0.017	0.082 ± 0.004	0.874 ± 0.007

## Method



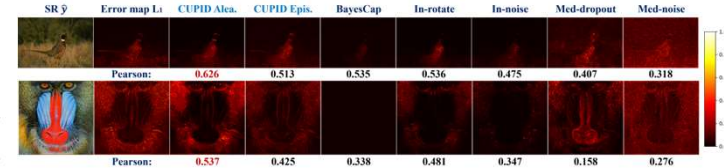
**Figure 2: The CUPID pipeline.** Aleatoric uncertainty is estimated using a dedicated Uncertainty Branch, while epistemic uncertainty is captured by measuring the variance between the original model output  $\hat{y}$  and the perturbed output  $\hat{y}'$ .

## Results

**Table 2: Performance of OOD detection (OOD samples as positive).** PAPILA and ACRIMA share the same research problem (glaucoma detection) with the ID dataset while CIFAR10 is a general classification dataset. CUPID Epistemic performs best on PAPILA and ACRIMA, suggesting that it is more sensitive to subtle distribution shifts within related clinical domains. In contrast, CUPID Aleatoric performs best on CIFAR-10, showing stronger response to extreme domain mismatch.

Method	PAPILA		ACRIMA		CIFAR10	
	AUC(↑)	AUPR(↑)	AUC(↑)	AUPR(↑)	AUC(↑)	AUPR(↑)
CUPID Alea.	0.379 ± 0.027	0.333 ± 0.007	0.717 ± 0.029	0.661 ± 0.027	<b>0.983 ± 0.005</b>	<b>0.998 ± 0.001</b>
CUPID Epis.	<b>0.877 ± 0.032</b>	<b>0.854 ± 0.027</b>	<b>0.978 ± 0.010</b>	<b>0.984 ± 0.007</b>	0.898 ± 0.054	0.991 ± 0.005
MC Dropout	0.733 ± 0.002	0.586 ± 0.007	0.869 ± 0.003	0.816 ± 0.009	0.887 ± 0.004	0.986 ± 0.001
Rate-in	0.328 ± 0.005	0.329 ± 0.008	0.363 ± 0.003	0.390 ± 0.003	0.620 ± 0.001	0.927 ± 0.002
IGRUE	0.636 ± 0.114	0.486 ± 0.097	0.941 ± 0.008	0.944 ± 0.008	0.978 ± 0.005	<b>0.998 ± 0.001</b>
PostNet Alea.	0.638 ± 0.060	0.487 ± 0.067	0.549 ± 0.040	0.487 ± 0.040	0.657 ± 0.032	0.952 ± 0.005
PostNet Epis.	0.577 ± 0.097	0.425 ± 0.088	0.685 ± 0.154	0.654 ± 0.151	0.773 ± 0.082	0.976 ± 0.011
BNN	0.707 ± 0.040	0.612 ± 0.050	0.708 ± 0.073	0.699 ± 0.042	0.643 ± 0.108	0.959 ± 0.013
DEC	0.515 ± 0.024	0.457 ± 0.024	0.680 ± 0.003	0.685 ± 0.012	0.660 ± 0.015	0.963 ± 0.003

**Figure 3: Comparison of visual results between error and uncertainty maps.** CUPID Aleatoric shows the best texture alignment and highest correlation with error maps.



**Figure 4: Performance of CUPID inserted at varying locations: misclassification detection (Left) and super-resolution (Right).**

