ICML 2026 · ACCEPTED PAPER

Singular Bayesian Neural Networks

A low-rank variational BNN whose induced posterior has singular geometric support: theory-backed scalable uncertainty via $W = AB^T$

Mame Diarra Toure, *David A. Stephens*

McGill

TL;DR

We learn variational uncertainty over low-rank factors $W = AB^T$ and push it forward to weight space. The induced posterior q_W is supported on a measure-zero manifold, a singular Bayesian object, not just a cheaper one.

PROBLEM

Bayesian uncertainty is expensive

Full mean-field VI keeps **two variational parameters per weight**. Deep ensembles multiply the entire network. Post-hoc methods sit on a fixed-point estimate, they *cannot* change the posterior support.

$O(mn) \rightarrow O(r(m+n))$

For a 512×512 layer

r=64 → 25× · r=16 → 64× · r=8 → 128× fewer params

PRIOR LOW-RANK BAYESIAN APPROACHES

What others miss

Existing methods are not the same Bayesian object.

Approach	Posterior object	Support
Post-hoc noise	Noise around fixed W*	Full ambient $\mathbb{R}^{mn}$
Low-rank covariance approx.	Gaussian on full W	Full ambient $\mathbb{R}^{mn}$
Bayesian LoRA	Uncertainty on adapter	Rank-r adapter only
SBNN (ours)	**Pushforward q(A,B) → q_W**	**Manifold $\mathcal{R}_r \subset \mathbb{R}^{mn}$**

Same accuracy. Tighter OOD. Smaller model.

HEAD-TO-HEAD · vs SWAG

Task	SWAG	Low-Rank SBNN
SST-2 · Acc	0.808 / 208 M	**0.806 / 1.47 M**
MIMIC · AUPR-OOD / AUPR-Err	0.634 / 0.680	**0.788 / 0.540**
Beijing · PICP	narrower & costlier	**0.790 best**

KEY INSIGHT

Calibration ↔ OOD is a nuanced tradeoff. *Ensembling* 5 low-rank SBNNs closes it.

Single low-rank SBNNs trade marginal in-domain calibration for sharper OOD detection. Ensembling five low-rank SBNNs recovers, and exceeds, full-rank calibration while keeping the parameter savings.

AUROC-OOD	ECE	NLL
0.638 → 0.731	0.166 → 0.054	0.523 → 0.415

HONEST TAKEAWAY

Low-rank SBNNs shine on OOD detection, coverage, and parameter efficiency. Deep Ensembles remain strong in-domain but at 33× the cost.

EFFICIENCY · SAME GPU · SAME TASK

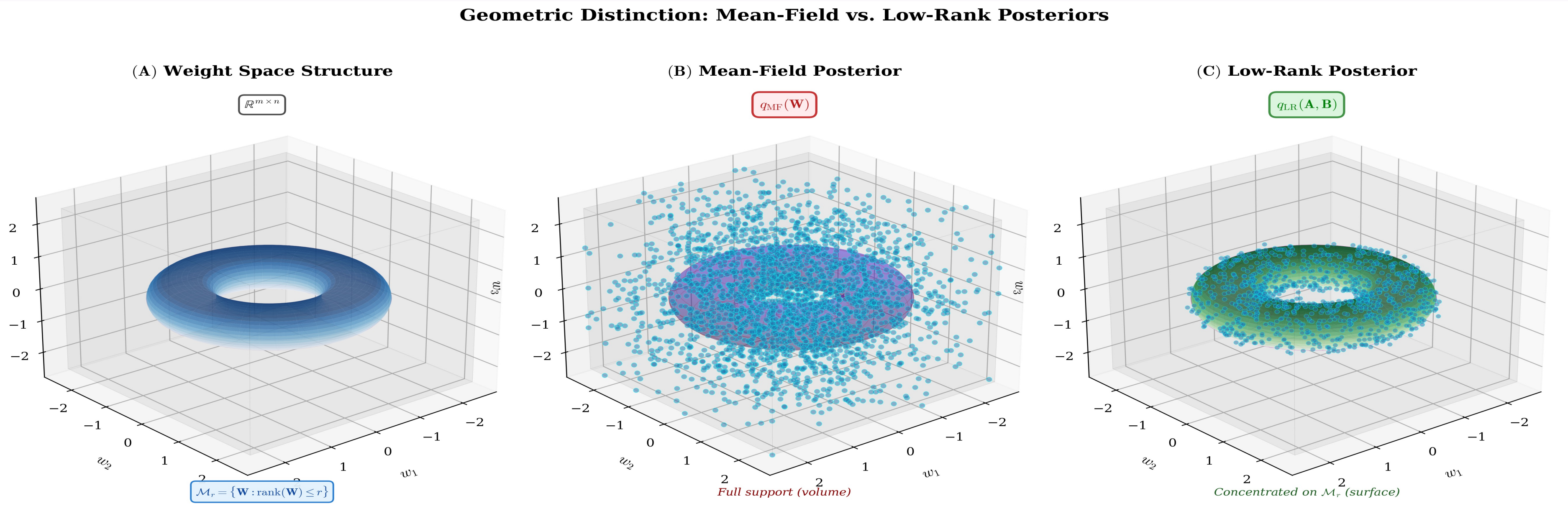
Smaller. Faster. Competitive.

PEAK MEMORY	EPOCH TIME	PARAMETERS
2.0× less than Full-Rank BBB	**3.2×** faster than Deep Ensemble	**33×** fewer than Deep Ensemble

Method	Peak GPU Memory	Time
Deep Ens (5)	670 MB	19.0 s
Full-Rank BBB	721 MB	6.5 s
Low-Rank (ours)	**358 MB**	**5.9 s**

GEOMETRIC DISTINCTION

Mean-field fills volume. Low-rank lives on a manifold.



CENTRAL THEOREM · q_W is singular w.r.t. Lebesgue measure

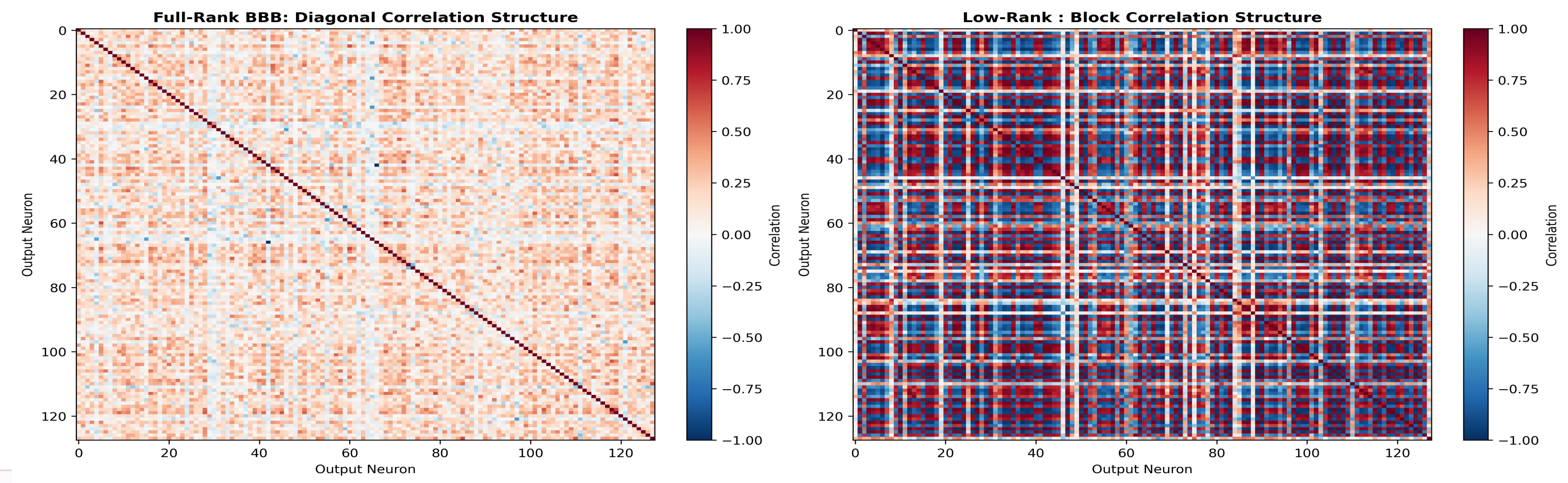
The posterior support has changed

$$\text{If } r < \min(m,n) \text{ and } W = AB^T \text{ with } A \in \mathbb{R}^{m\times r},\ B \in \mathbb{R}^{n\times r}:$$

$$q_W(\mathcal{R}_r) = 1 \quad\text{and}\quad \lambda(\mathcal{R}_r) = 0 \quad\Longrightarrow\quad q_W \perp \lambda$$

The Bayesian object cannot be represented as a full-dimensional density over the weight space.

Mean-field factors → structured weight correlations: full-rank BBB stays diagonal, low-rank SBNN learns block structure.



EXPERIMENTAL DESIGN

Three architectures × three shift regimes

MLP / LSTM / Transformer trained on MIMIC-III mortality, Beijing $PM_{2.5}$, and SST-2 sentiment. Each evaluated under in-domain accuracy, OOD detection, and selective-prediction coverage, same rank-r SBNN recipe across modalities and scales.

RESULT · TASK 1 / 3

MIMIC-III mortality · MLP classifier

In-distribution AUROC matches the full-rank baseline; OOD discrimination and error-aware coverage

Method	Params	AUROC	AUC-OOD	AUPR-OOD	AUPR-Err
Deterministic	22.4 K	0.922	—	—	—
Deep Ensemble (5)	112 K	0.929	0.738	0.754	0.237
Full-Rank BBB	44.8 K	0.895	0.770	0.759	0.412
Low-Rank SBNN (ours)	**13.6 K**	**0.895**	**0.802**	**0.788**	**0.540**

70% fewer parameters vs Full-Rank BBB
13.6K vs 44.8K

0.802 AUC-OOD best of all methods
+0.032 over FR-BBB

0.540 AUPR-Err coverage gains
+0.036 over DE

RESULT · TASK 2 / 3

Beijing $PM_{2.5}$ · LSTM regression

Comparable MAE to full-rank baselines at far lower cost; calibrated coverage on noisy sequential data.

Method	Params	MAE ↓	PICP	AUC-OOD	AUPR-OOD
Deterministic	33 K	10.79	—	—	—
Deep Ensemble (10)	330 K	10.45	0.310	0.730	0.883
Full-Rank BBB	132 K	10.55	0.788	0.492	0.743
Rank-1 Multiplicative	66 K	10.80	0.449	0.580	0.751
Low-Rank SBNN (ours)	**47 K**	**10.63**	**0.790**	**0.710**	**0.861**

64% fewer parameters vs Full-Rank BBB
47K vs 132K

17.4% MAE reduction @ 80% retention
8.71 vs 9.21

0.790 PICP coverage + best AUC-OOD
0.710

RESULT · TASK 3 / 3 · SCALE TEST

SST-2 sentiment · Transformer (9.9 M)

In-domain accuracy matches Deep Ensemble at 33× fewer parameters and 8× lower wall-clock.

Method	Params	Acc ↑	AUROC-OOD ↑	Train time
Deep Ensemble (5)	49.6 M	0.825	0.657	64.7 min
Full-Rank BBB	19.8 M	0.752	0.622	23.1 min
Low-Rank SBNN (ours)	**1.47 M**	**0.806**	**0.640**	**8.2 min**

0.302 AUPR-In best of all methods
vs DE 0.267, FR 0.222

8× faster training vs Deep Ensemble
8.2 min vs 64.7 min

0.640 AUROC-OOD second to Deep Ensemble
−0.017 vs DE

RESEARCH AGENDA

Adaptive rank selection · Foundation models· Hierarchical priors over learned factors

FIND THE PAPER

arXiv:2602.00387 · arradiat.github.io/projects/singular-bnn . github.com/arradiat/SBNN · mame.toure@mail.mcgill.ca