Uncertainty-Aware Data Imputation Using Bayesian Network Guided Bayesian GANs

1st Thomás de los Santos Verrijp
Department of Advanced Computing Sciences
Maastricht University
Maastricht, Netherlands
thomasdls003@gmail.com

2nd Chang Sun
Department of Advanced Computing Sciences
Maastricht University
Maastricht, Netherlands
chang.sun@maastrichtuniversity.nl

I. Introduction

Missing data is a prevalent challenge in real-world datasets and is particularly pronounced in clinical research due to irregular and heterogeneous data collection [1]. If handled improperly, missing values can introduce bias, reduce statistical efficiency, and compromise downstream analyses [2]. Effective imputation methods should go beyond point estimates and preserve inter-feature dependencies while providing uncertainty estimates to support reliable decision-making.

Recent work has adopted deep generative models, including Generative Adversarial Networks (GANs) for missing data imputation due to their ability to learn complex joint distributions and generate realistic values for unobserved entries [3], [4]. However, most GAN-based imputers are deterministic or produce single imputed values, offering limited insight into predictive uncertainty [5]. Conversely, approaches [6] that explicitly model uncertainty often lack mechanisms to enforce conditional dependencies among variables.

This highlights an unresolved challenge: current imputation methods struggle to simultaneously achieve expressive generative modeling, uncertainty awareness, and adherence to structured variable dependencies. The impact of this limitation is amplified under non-random missingness mechanisms, where violations of dependency structure or poorly calibrated uncertainty can substantially distort downstream analyses.

To bridge this gap, this work proposes *BN-BGAN*, a Bayesian network–guided Bayesian generative adversarial framework for uncertainty-aware data imputation. BN-BGAN learns a Bayesian network from data and incorporates it into a Bayesian GAN as a structural prior, guiding the generative process while preserving stochasticity. This design enables imputations that respect learned dependencies and explicitly represent uncertainty. Experiments show that BN-BGAN improves imputation accuracy and reduces uncertainty dispersion compared to unstructured Bayesian GANs, particularly under non-random missingness mechanisms.

II. Related Works

Classical imputation techniques for tabular data include Multiple Imputation by Chained Equations (MICE) [7], MissForest [8], and related ensemble-based methods. These approaches have been applied in clinical research because of their simplicity and interpretability. However, they rely on conditional independence assumptions or iterative regression models that struggle to capture complex nonlinear dependencies and high-order interactions, particularly in high-dimensional settings [9]. Uncertainty estimates are often heuristic or limited to sampling variability, rather than reflecting model uncertainty.

Deep generative modeling such as generative adversarial networks (GAN), have been adopted for data imputation [10]. Among these, GAIN [3] introduced a conditional GAN framework that directly imputes missing values on partially observed data. However, GAN-based imputers employ deterministic generators or produce single-point estimates, limiting their ability to quantify uncertainty in the imputed values [3], [5].

Bayesian deep learning methods introduce stochasticity into neural network parameters to capture epistemic uncertainty. Bayesian GANs (BGANs) [6] place distributions over generator and discriminator weights and approximate posterior inference via sampling, enabling uncertainty-aware generation. While effective at modeling model uncertainty, existing BGAN formulations are typically unstructured and do not encode explicit relationships among input variables. As a result, generated samples may satisfy global distributional properties while violating domain-specific or conditional dependencies.

Bayesian Networks (BNs) provide a principled framework for modeling conditional dependencies through directed acyclic graphs and local conditional distributions [11], offering interpretability and explicit structural assumptions. However, structure learning and inference scale poorly with dimensionality, and standard BN parameterizations are limited in their ability to represent complex nonlinear relationships [12]. Consequently, BNs alone are often insufficient for large-scale or heterogeneous real-world datasets.

Several recent works have explored incorporating structural or relational information into deep generative models, including graph-conditioned GANs and causal generative frameworks [13], [14], [15]. These methods typically enforce structure through architectural constraints or joint optimization of graph and generator parameters. While promising, these approaches still suffer from training instability and may impose overly rigid structural assumptions that limit generative flexibility, particularly in the presence of missing data.

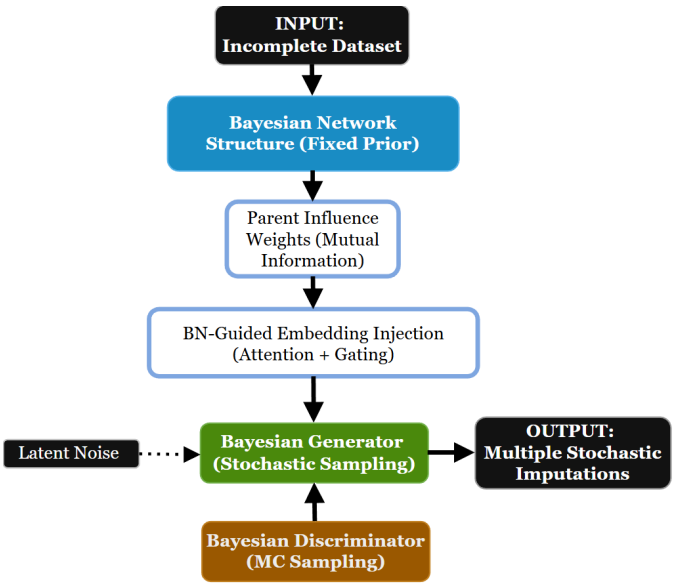

Fig. 1. Structured pipeline of BN-BGAN, where a fixed Bayesian network provides parent influence weights injected into a Bayesian generator, adversarially trained via a Bayesian discriminator (training only), to produce multiple stochastic imputations with predictive uncertainty.

III. Methodology

To address the remaining challenge from previous work, the BN-BGAN framework integrates (i) a Bayesian Network (BN) that captures conditional dependencies among features, (ii) a Bayesian GAN that models intrinsic randomness (aleatoric uncertainty) as well as model uncertainty (epistemic uncertainty) [16], and (iii) conditional generation to support structured and controllable synthesis. Refer to Fig. 1 for an overview of the pipeline.

A. Bayesian Network Structure

We learn a Bayesian Network structure using greedy hill-climbing with the Bayesian Information Criterion (BIC) score [17], [18]. This search iteratively applies local edge operations (addition, deletion, reversal) and retains modifications that improve the BIC score, balancing model fit and complexity. The resulting directed acyclic graph encodes conditional dependencies among features.

For each retained parent–child edge $(i \to j)$, we compute mutual information (MI) and normalize it across the parents of each node to obtain soft dependency weights:

$$w_{i\to j} = \frac{MI(X_i; X_j)}{\sum_{i' \in \mathrm{Pa}(j)} MI(X_{i'}; X_j)}. \tag{1}$$

These weights quantify the relative influence of parent features and define a structured prior used during generation.

1) BN-Guided Parent Context: The generator samples a latent vector $z \sim \mathcal{N}(0, 1)$ and concatenates it with a conditional vector c [19]. To inject BN structure, feature embeddings are modulated using MI-weighted parent information.

For feature j, we gather the embeddings of its parent features $E_i \mid i \in \mathrm{Pa}(j)$. Each parent embedding is first scaled by its MI-derived weight, reflecting global dependency strength. However, dependency relevance can vary across samples. We compute sample-specific attention weights from the child embedding E_j using a Multi-Layer Perceptron (MLP), producing a vector a_j whose elements lie in $(0, 1)$. The final parent weights are given by:

$$\bar{w}_{i\to j} = w_{i\to j} \cdot a_{j,i}. \tag{2}$$

These combined weights encode both global structure (MI) and local context (attention). The weighted parent embeddings are concatenated into a parent context vector:

$$\mathcal{P}'j = [\bar{w}i_1 \to jE_{i_1}; \ldots; \bar{w}_{i_m\to j}E_{i_m}]. \tag{3}$$

An MLP transforms this vector into a compact parent influence representation.

2) Gated Injection: The parent influence representation is concatenated with the child embedding and mapped through an MLP to produce a BN-informed embedding:

$$U_j = \mathrm{MLP}([E_j; \mathcal{P}'_j]). \tag{4}$$

A gating mechanism determines how strongly BN information should modify the child representation:

$$E_j \leftarrow g_j U_j + (1 - g_j)E_j, \qquad g_j \in (0, 1). \tag{5}$$

This gate allows the generator to rely on BN guidance when beneficial, while preserving its ability to override the BN when adversarial training indicates alternative structures. Finally, the BN-informed embedding is injected into the latent subvector associated with feature j:

$$z_j \leftarrow z_j + \alpha U_j, \tag{6}$$

where α controls the global strength of BN influence.

This procedure biases the generator toward samples that respect learned dependencies, without enforcing hard constraints. Structural information shapes the generator's internal representations rather than its outputs directly, yielding stable and flexible integration. The attention–gating mechanism is not intended to re-learn structure, but to modulate the influence of fixed BN priors when adversarial training signals contradict global dependencies.

3) Role of BN: The Bayesian network structure and MI-derived weights are learned once prior to adversarial training and remain fixed, acting as a non-differentiable structural prior while only generator, discriminator, attention, and gating parameters are optimized.

B. Bayesian GAN and Uncertainty Modelling

To enable uncertainty-aware imputation, BN-BGAN adopts a *Bayesian-inspired* generative adversarial framework in which both data uncertainty (aleatoric) and model uncertainty (epistemic) are approximated through stochastic generation and parameter sampling. Rather than performing exact Bayesian inference over network parameters, BN-BGAN employs implicit

posterior approximations that are computationally tractable and compatible with adversarial training, following prior work [6], [20].

1) Data Uncertainty: The generator produces feature-wise distributional parameters rather than point estimates. For each feature x_j, the generator outputs a mean μ_j and log-variance $\log \sigma_j^2$, defining a Gaussian likelihood:

$$p(x_j \mid z, c, \theta_G) = \mathcal{N}(\mu_j, \sigma_j^2), \tag{7}$$

where z denotes the latent noise vector, c represents the BN-guided conditional context, θ_G are the generator parameters.

Sampling is performed using the reparameterization trick:

$$x_j = \mu_j + \sigma_j \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{8}$$

which allows gradients to propagate through stochastic nodes during training. This formulation captures irreducible uncertainty arising from measurement noise and intrinsic variability in the data-generating process.

2) Epistemic Uncertainty (Model Uncertainty): Epistemic uncertainty is modeled by placing distributions over the generator and discriminator parameters. Following Bayesian GAN formulations [6], network weights are treated as random variables with implicit priors, and posterior uncertainty is approximated via stochastic sampling during training.

In practice, epistemic uncertainty is approximated using Monte Carlo (MC) sampling over generator parameters and MC dropout in the discriminator [20]. At inference time, multiple forward passes through independently sampled network instances yield a distribution of imputations, whose variance reflects uncertainty due to limited data and model capacity.

3) Training Objective: BN-BGAN is trained using a Wasserstein GAN with gradient penalty [21], augmented with uncertainty-aware regularization inspired by Bayesian GANs [6]. The generator loss is defined as

$$\mathcal{L}_G = \mathcal{L}_{\text{adv}} + \beta\, \mathcal{L}_{\text{unc}} + \lambda\, \mathcal{L}_{\text{KL}}, \tag{9}$$

where $\mathcal{L}_{\text{adv}} = -\mathbb{E}_{z\sim p(z)}[D(G(z, c))]$ forces generated samples to match real data distribution using Wasserstein distance.

To regularize aleatoric uncertainty, the uncertainty loss minimizes the negative log-likelihood of observed entries under the predicted Gaussian distributions [5], [22]:

$$\mathcal{L}_{\text{unc}} = \sum_{j \in \mathcal{O}} \left(\frac{(x_j - \mu_j)^2}{2\sigma_j^2} + \frac{1}{2} \log \sigma_j^2 \right), \tag{10}$$

where $\mathcal{O}$ denotes the set of observed features. This formulation discourages degenerate solutions in which predictive uncertainty is artificially inflated. Then, a KL divergence term is introduced to constrain the generator's predictive distributions toward a unit Gaussian prior [22]:

$$\mathcal{L}_{\text{KL}} = \sum_j \text{KL}\left(\mathcal{N}(\mu_j, \sigma_j^2) \parallel \mathcal{N}(0, 1)\right), \tag{11}$$

which admits a closed-form expression and stabilizes adversarial training by discouraging overly confident or degenerate output distributions.

C. Conditional Generation and Data Processing

Both generator and discriminator are conditioned using one-hot encoded categorical vectors [19]. Continuous features are modeled via Gaussian mixture representations [23]. Discrete features are handled using the Gumbel-Softmax relaxation for differentiable sampling [24].

D. Imputation Procedure

A binary mask identifies missing entries. During inference, missing values are replaced with samples generated by BN-BGAN, while observed values are retained. The imputation process is repeated multiple times to obtain multiple stochastic imputations, enabling uncertainty-aware downstream analysis.

IV. Experiments and Results

The proposed BN-BGAN is compared with state-of-the-art methods in literature, while quantifying the importance of Bayesian elements and the BN prior. All experimented GAN-based models use the same structure and parameters, with AdamW for 50 epochs. Baseline methods use default hyperparameters from their original implementations. Loss-function ablation experiments are repeated over 10 independent training runs. Imputation performance and synthetic data quality evaluations are repeated over 5 independent runs. For uncertainty analysis, each trained model generates multiple stochastic imputations under fixed random seeds.

A. Datasets

Experiments were conducted on four OpenML tabular datasets commonly used in imputation benchmarks: (1) **Breast Cancer** [25]: 699 samples with 10 numerical features, providing a clean low-dimensional benchmark; (2) **Hepatitis** [26]: 615 samples with 13 continuous biochemical markers and 2 categorical features, exhibiting correlated clinical measurements; (3) **Heart Disease** [25]: 296 samples with clinical, physiological, and categorical features, showing nonlinear dependencies; (4) **Blood Transfusion** [27]: 748 instances with four correlated temporal and behavioral features.

B. Loss Ablation

We first assess the contribution of the KL divergence and uncertainty-aware loss through a loss ablation study on the Breast Cancer dataset (Table I). To determine whether observed performance differences exceed stochastic run-to-run variability, we employ the Kruskal–Wallis test, a non-parametric alternative to one-way ANOVA that does not assume normality [28], alongside Cliff's δ [29] to quantify effect magnitude independently of sample size.

Across all configurations, differences in MMD and diversity are small relative to their variability, and Kruskal–Wallis tests fail to detect statistically significant differences (p ¿ 0.3). Effect sizes (Cliff's δ) are consistently negligible to small ($|\delta| \leq 0.18$), indicating that neither component alone produces a strong isolated effect. Minor trends are observable: KL-only yields slightly higher diversity, CL-only achieves marginally better MMD, and the adversarial baseline attains

the highest indistinguishability. However, these differences are not practically significant.

C. Uncertainty Analysis

Uncertainty calibration is assessed using Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW), which measure reliability of predictive intervals. PICP is the proportion of true values within 95% prediction intervals, while MPIW is their average width. Both are needed: high coverage can be achieved with overly wide intervals, whereas narrow intervals may indicate overconfidence.

To address the severe undercoverage observed in BN-BGAN, we apply split conformal prediction as a post-hoc calibration step. This procedure guarantees marginal coverage without modifying the underlying model.

Table III reports results after calibration. BN-BGAN achieves near-nominal coverage across all missingness rates (PICP $\approx$ 0.94–0.95), representing a substantial improvement over the uncalibrated setting (PICP $\approx$ 0.2–0.3). Calibration errors are consistently low (CE $\leq$ 0.025), indicating that conformal adjustment effectively corrects overconfident predictions. BGAN also attains valid coverage but tends to be over-conservative, with PICP values exceeding the nominal level in most cases. This results in slightly higher calibration errors compared to BN-BGAN.

D. Bayesian Network Ablation Study

To isolate the effect of BN-guided generation on epistemic uncertainty, we perform a paired uncertainty ablation comparing BGAN and BN-BGAN under identical training conditions. All architectural choices, optimization settings, and training epochs are held constant across methods. The only difference between the two models is the presence of BN-guided conditional sampling in the generator. Each model performed 30 stochastic imputations over 50 training epochs, with a 20% test split. From these imputations, the standard deviation across imputations is used as a measure of uncertainty. We report uncertainty ablation on the Breast Cancer dataset to isolate BN effects without confounding missingness patterns. Uncertainty comparisons are performed pairwise at the cell level. For each originally masked cell, we obtain a paired uncertainty estimate from BGAN and BN-BGAN under identical masking and sampling conditions. This paired design controls for data-dependent variability and allows direct attribution of uncertainty differences to the modeling choice rather than stochastic effects.

Figure 2 shows that BN-guided generation substantially reduces epistemic uncertainty. BN-BGAN exhibits a lower median per-cell standard deviation compared to BGAN, with a tighter distribution and fewer high-uncertainty outliers. As training conditions are identical, this reduction is directly attributable to BN-guided conditional sampling.

E. Imputation Performance

To evaluate performance of advanced and baseline imputation models under different missingness mechanisms, we synthetically induce missing values [1]. Under *Missing Completely at Random* (MCAR), entries are removed uniformly at random, independent of both observed and unobserved variables. Under *Missing at Random* (MAR), missingness depends only on observed covariates, implemented by assigning higher missing probabilities to selected features conditional on other observed variables. Under *Missing Not at Random* (MNAR), missingness depends on the unobserved value itself: entries are masked with probabilities determined by the underlying (pre-missing) feature values, such that extreme or clinically significant values are more likely to be missing. This construction follows standard simulation practices for benchmarking imputation methods under MNAR [1], [30].

Epistemic uncertainty is quantified as the per-cell standard deviation across multiple stochastic imputations generated from a single trained model. For each masked cell, we compute the standard deviation over N independently sampled imputations, yielding a spatial uncertainty map over the missing entries.

Table II and Figure 3 summarize imputation performance across missingness mechanisms and datasets. BN-BGAN consistently outperforms BGAN under MAR and MNAR conditions, exhibiting lower RMSE and reduced sensitivity to missingness rate. In the Heart Disease dataset, BN-BGAN achieved the lowest RMSE under MNAR and maintained stable performance across all missingness types, whereas BGAN showed higher variability. Traditional methods, including MICE and MissForest, performed moderately, with Mean/-Mode Imputer remaining competitive. In the Hepatitis dataset,

TABLE I
LOSS ABLATION STUDY ON THE *Breast Cancer* DATASET (MEAN $\pm$ STD OVER 50 RUNS). LOWER MMD INDICATES BETTER DISTRIBUTIONAL ALIGNMENT; HIGHER DIVERSITY AND INDISTINGUISHABILITY ARE BETTER.

Config	MMD ↓	Diversity ↑	Indist. ↑
KL+CL	0.095 ± 0.012	0.052 ± 0.017	0.722 ± 0.048
KL	0.097 ± 0.009	$\mathbf{0.057 \pm 0.027}$	0.726 ± 0.053
CL	$\mathbf{0.094 \pm 0.009}$	0.0519 ± 0.022	0.726 ± 0.051
Adversarial	0.094 ± 0.010	0.051 ± 0.027	$\mathbf{0.734 \pm 0.043}$

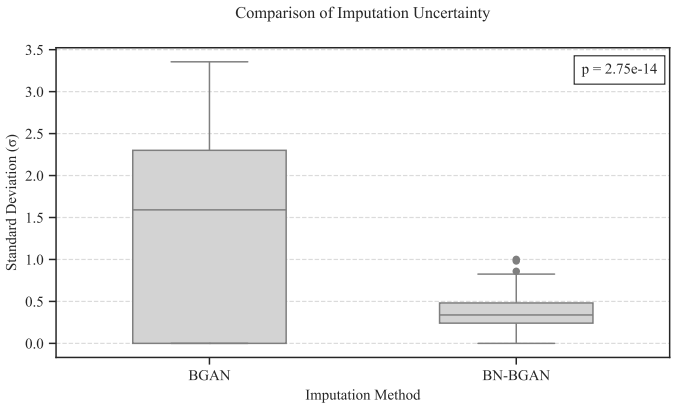

Fig. 2. Comparing the standard deviation between BN-BGAN and BGAN of 30 stochastic imputations in the Breast Cancer dataset.

TABLE II

AVERAGE IMPUTATION RMSE (MEAN ± STD) AGGREGATED OVER MISSING RATES AND TRAINING SCENARIOS. BEST RESULTS PER DATASET AND MISSINGNESS PATTERN ARE HIGHLIGHTED IN BOLD.

Dataset	Method	Pattern	RMSE (mean ± std)
Hepatitis	BGAN	MAR	31.35 ± 8.42
Hepatitis	BGAN	MCAR	**30.49 ± 2.28**
Hepatitis	BGAN	MNAR	25.48 ± 3.47
Hepatitis	BN-BGAN	MAR	**23.37 ± 2.02**
Hepatitis	BN-BGAN	MCAR	33.95 ± 1.37
Hepatitis	BN-BGAN	MNAR	**11.63 ± 1.02**
Heart	BGAN	MAR	26.92 ± 12.28
Heart	BGAN	MCAR	33.40 ± 6.27
Heart	BGAN	MNAR	34.92 ± 5.85
Heart	BN-BGAN	MAR	**18.70 ± 2.19**
Heart	BN-BGAN	MCAR	**20.32 ± 2.41**
Heart	BN-BGAN	MNAR	**16.78 ± 1.55**
Cancer	BGAN	MAR	5.02 ± 0.67
Cancer	BGAN	MCAR	4.76 ± 0.04
Cancer	BGAN	MNAR	4.13 ± 0.16
Cancer	BN-BGAN	MAR	**4.73 ± 0.12**
Cancer	BN-BGAN	MCAR	**3.23 ± 0.08**
Cancer	BN-BGAN	MNAR	**1.37 ± 0.05**

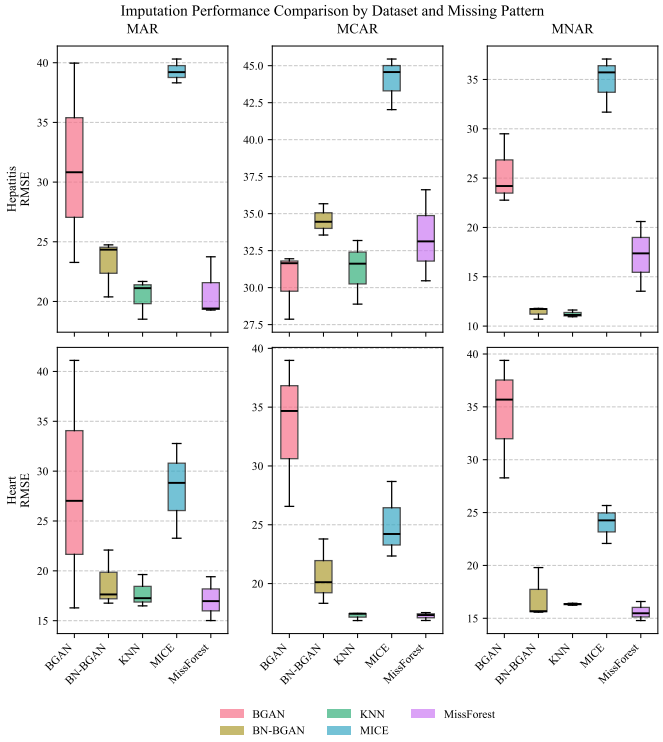

Fig. 3. Compares the RMSE between different imputation methods under different missingness patterns (aggregation of performance under missingess rates of 10%, 20%, and 30%). Compared performance between Heart Disease and Hepatitis datasets.

KNN and Mean/Mode Imputers yielded strong results, while MICE consistently recorded the highest RMSE. BN-BGAN's advantage is most pronounced under structured or non-random missingness, highlighting the importance of dependency-aware uncertainty modeling.

TABLE III

UNCERTAINTY CALIBRATION ON THE *Blood* DATASET UNDER MCAR MISSINGNESS AFTER SPLIT CONFORMAL CALIBRATION. REPORTED VALUES ARE MEAN ± STD OVER 5 TRIALS. PICP DENOTES EMPIRICAL COVERAGE (TARGET: 0.95), MPIW IS MEAN INTERVAL WIDTH, AND CE IS $|\text{PICP} - 0.95|$.

Miss.	Method	PICP	MPIW ↓	CE ↓
10%	BN-BGAN	$\mathbf{0.942 \pm 0.016}$	1646.8 ± 781.7	$\mathbf{0.011 \pm 0.014}$
	BGAN	0.987 ± 0.012	$\mathbf{1519.5 \pm 544.9}$	0.037 ± 0.012
20%	BN-BGAN	$\mathbf{0.939 \pm 0.036}$	$\mathbf{1436.8 \pm 761.5}$	$\mathbf{0.025 \pm 0.025}$
	BGAN	0.981 ± 0.015	1482.9 ± 266.6	0.031 ± 0.015
30%	BN-BGAN	$\mathbf{0.949 \pm 0.014}$	2004.4 ± 1721.4	$\mathbf{0.012 \pm 0.004}$
	BGAN	0.984 ± 0.021	$\mathbf{1923.6 \pm 646.1}$	0.034 ± 0.021
40%	BN-BGAN	$\mathbf{0.953 \pm 0.011}$	1844.9 ± 1018.4	$\mathbf{0.007 \pm 0.008}$
	BGAN	0.954 ± 0.029	$\mathbf{1314.3 \pm 457.0}$	0.025 ± 0.008

V. DISCUSSION

The experimental results indicate that incorporating explicit Bayesian structure into GAN imputation models can improve stability, robustness to missingness, and behavior under uncertainty.

The loss ablation study highlights that the effects of KL divergence and uncertainty-aware losses are data-dependent. Although no configuration achieves statistically significant dominance across metrics, observed trends suggest that KL regularization primarily influences distributional alignment, while uncertainty loss affects diversity and dispersion. These findings indicate that auxiliary loss terms shape model behavior in complementary but data-sensitive ways, and should therefore be validated per dataset rather than assumed to generalize uniformly.

The Bayesian network ablation provides direct evidence that structural priors significantly affect epistemic uncertainty. Compared to BGAN, BN-BGAN produces lower and more concentrated uncertainty estimates under identical training conditions, indicating that BN-guided sampling constrains stochastic generation in accordance with learned dependencies. This suggests that explicit structural priors can stabilize uncertainty estimates.

Imputation performance comparisons further show that BN-BGAN maintains competitive accuracy across MCAR, MAR, and MNAR mechanisms, with the largest gains observed under non-random missingness. This robustness suggests that dependency-aware generation enables more effective use of observed variables when inferring missing values. Simultaneously, strong performance by simpler baselines on certain datasets shows that generative models do not universally dominate, underscoring the importance of aligning model complexity with dataset characteristics.

The uncertainty calibration experiment reveals a key trade-off between uncertainty and coverage. Before calibration, BN-BGAN produced narrower intervals with undercoverage, while BGAN achieved higher coverage with wider intervals. After applying split conformal calibration, both models achieve near-nominal coverage, coming at the cost of increased interval width. Notably, BN-BGAN often retains competitive or smaller interval widths after calibration, indicating that structurally guided models can yield more efficient calibrated

uncertainty estimates. This highlights that while structural priors improve precision, explicit calibration mechanisms are necessary to ensure reliable probabilistic coverage. However, generalisability claims of this experimentation are limited by the scale of the evaluation; all datasets contain 750 or fewer samples and at most 15 features.

VI. Conclusion

This paper presented BN-BGAN, a Bayesian network–guided Bayesian generative adversarial framework for uncertainty-aware data imputation. BN-BGAN is able to capture inter-feature dependencies while enabling epistemic uncertainty estimation through stochastic sampling.

Experiments across multiple benchmark datasets show that BN-BGAN achieves competitive and often improved imputation accuracy compared to baseline methods, particularly under MAR and MNAR missingness. BN-guided generation improves stability as missingness increases and reduces variability in stochastic imputations, while maintaining strong distributional fidelity in synthetic data generation.

Ablation studies indicate that both structural priors and uncertainty-aware regularization influence performance in a dataset-dependent manner. We demonstrate that post-hoc conformal calibration effectively restores coverage guarantees, but at the cost of increased interval width. This highlights a key limitation: calibration methods correct empirical coverage without addressing the underlying source of miscalibration, and therefore do not eliminate the trade-off between precision and reliability.

These findings suggest that uncertainty calibration should be treated as a first-class objective in generative imputation, rather than a post-processing step. Future work should explore integrating calibration-aware objectives directly into training, experimenting with larger datasets, and extending structural priors to more flexible or learned dependency representations.

Declarations

The authors have no conflict of interest to declare that are relevant to the content of this article. Portions of this manuscript were refined assisted by ChatGPT. All content was reviewed by the authors.

Data Availability. All data used in this research is publicly available through OpenML [31].

Code Availability. https://github.com/annonymous8923849/bn-bgan-impl

References

[1] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 3rd ed. Wiley, 2019.

[2] J. L. Schafer and J. W. Graham, "Missing data: Our view of the state of the art," *Psychological Methods*, vol. 7, no. 2, pp. 147–177, 2002.

[3] J. Yoon, J. Jordon, and M. van der Schaar, "GAIN: missing data imputation using generative adversarial nets," in *ICML*, ser. Proceedings of Machine Learning Research, vol. 80, 2018, pp. 5675–5684.

[4] A. Figueira and B. Vaz, "Survey on synthetic data generation, evaluation methods and gans," *Mathematics*, vol. 10, no. 15, p. 2733, 2022.

[5] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, p. 5580–5590.

[6] Y. Saatci and A. G. Wilson, "Bayesian GAN," in *NIPS*, 2017, pp. 3622–3631.

[7] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," *Journal of Statistical Software*, vol. 45, no. 3, p. 1–67, 2011.

[8] D. J. Stekhoven and P. Bühlmann, "Missforest - non-parametric missing value imputation for mixed-type data," *Bioinform.*, vol. 28, no. 1, pp. 112–118, 2012.

[9] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: What is it and how does it work?" *International Journal of Methods in Psychiatric Research*, vol. 20, no. 1, pp. 40–49, 2011.

[10] Y. Sun, J. Li, Y. Xu, T. Zhang, and X. Wang, "Deep learning versus conventional methods for missing data imputation: A review and comparative study," *Expert Syst. Appl.*, vol. 227, p. 120201, 2023.

[11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, ser. Morgan Kaufmann series in representation and reasoning. Elsevier Science, 1988.

[12] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[13] T. N. Kipf, E. Fetaya, K.-C. Wang, M. Welling, and R. Zemel, "Neural relational inference for interacting systems," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

[14] S. Ma, M. Gong, and K. Zhang, "Causal generative neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[15] Y. Yang, Y. Ding, and Y. Liu, "Causalgan: Learning causal implicit generative models with adversarial training," in *International Conference on Learning Representations (ICLR)*, 2020.

[16] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods," *Mach. Learn.*, vol. 110, no. 3, pp. 457–506, 2021.

[17] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning bayesian networks: The combination of knowledge and statistical data," *Machine Learning*, vol. 20, no. 3, pp. 197–243, 1995.

[18] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[19] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014.

[20] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *ICML*, ser. JMLR Workshop and Conference Proceedings, vol. 48, 2016, pp. 1050–1059.

[21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *NIPS*, 2017, pp. 5767–5777.

[22] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *ICLR*, 2014.

[23] D. A. Reynolds, "Gaussian mixture models," in *Encyclopedia of Biometrics*. Springer, 2009, pp. 659–663.

[24] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," in *ICLR (Poster)*, 2017.

[25] O. L. Mangasarian and W. H. Wolberg, "Cancer diagnosis via linear programming," *SIAM News*, vol. 23, no. 5, pp. 1–18, 1990.

[26] G. Hoffmann, A. Bietenbeck, R. Lichtinghagen, and F. Klawonn, "Using machine learning techniques to generate laboratory diagnostic pathways—a case study," *Journal of Laboratory and Precision Medicine*, vol. 3, no. 6, 2018.

[27] I.-C. Yeh, K.-J. Yang, and T.-M. Ting, "Knowledge discovery on rfm model using bernoulli sequence," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5866–5871, 2009.

[28] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[29] N. Cliff, "Dominance statistics: Ordinal analyses to answer ordinal questions," *Psychological Bulletin*, vol. 114, no. 3, pp. 494–509, 1993. [Online]. Available: https://doi.org/10.1037/0033-2909.114.3.494

[30] S. Van Buuren, *Flexible imputation of missing data*. New York, USA: CRC press Boca Raton, FL, 2012, vol. 10.

[31] J. Vanschoren, J. N. Van Rijn, B. Bischl, and L. Torgo, "Openml: Networked science in machine learning," *SIGKDD Explorations*, vol. 15, no. 2, pp. 49–60, 2013.