# Appendix and Supplementary Materials: Impute Missing Entries with Uncertainty

## A Mathematical Derivations / Proofs

### A.1 Derivation of ELBO

We introduce a discrete uniform random variable $\alpha$ taking values on $\alpha_k = \frac{k}{K}$ for $k = 1, \cdots, K$.

$$
\begin{aligned}
& \log p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}}; \theta, \beta) \\
= \;& \log \int p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta, \beta) p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta) d\mathbf{z} \\
= \;& \log \sum_{k=1}^{K} p(\alpha_k) \int p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \alpha_k, \mathbf{x}_{\mathbf{r}}; \theta, \beta) p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta) d\mathbf{z} \\
= \;& \log \sum_{k=1}^{K} p(\alpha_k) \int p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \alpha_k, \mathbf{x}_{\mathbf{r}}; \theta, \beta) \frac{p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta)}{q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi)} q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi) d\mathbf{z} \\
\geq \;& \sum_{k=1}^{K} p(\alpha_k) \int q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi) \log \left( p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \alpha_k, \mathbf{x}_{\mathbf{r}}; \theta, \beta) \frac{p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta)}{q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi)} \right) d\mathbf{z} \quad (\because \text{Jensen's inequality}) \\
= \;& \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}};\phi)} \Big[ \log p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \alpha_k, \mathbf{x}_{\mathbf{r}}; \theta, \beta) \Big] - D_{\mathrm{KL}} \Big( q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi) \Big\| p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta) \Big) \\
= \;& \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}};\phi)} \Bigg[ \sum_{j \in I_C} \log p\big((\mathbf{x}_{1-\mathbf{r}})_j | \mathbf{z}, \alpha_k, \mathbf{x}_{\mathbf{r}}; \theta_j, \beta\big) + \sum_{j \in I_D} \log p\big((\mathbf{x}_{1-\mathbf{r}})_j | \mathbf{z}, \mathbf{x}_{\mathbf{r}}, ; \theta_j, \beta\big) \Bigg] \\
& - D_{\mathrm{KL}} \Big( q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi) \Big\| p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta) \Big) \\
= \;& \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}};\phi)} \left[ \sum_{j \in I_C} \frac{1}{K} \log \frac{\alpha_k(1-\alpha_k)}{\beta} - \rho_{\alpha_k} \left( \frac{(\mathbf{x}_{1-\mathbf{r}})_j - Q_j(\alpha_k, \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta_j)}{\beta} \right) \right] \\
& + \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}};\phi)} \left[ \sum_{j \in I_D} \sum_{c=1}^{C_j} \mathbb{I}((\mathbf{x}_{1-\mathbf{r}})_j = c) \cdot \pi_{jc}(\mathbf{z}; \theta_j) \right] - D_{\mathrm{KL}} \Big( q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi) \Big\| p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta) \Big) \\
= \;& \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}};\phi)} \left[ -\frac{1}{\beta} \cdot \sum_{j \in I_C} \frac{1}{K} \sum_{k=1}^{K} \rho_{\alpha_k} \Big( (\mathbf{x}_{1-\mathbf{r}})_j - Q_j(\alpha_k, \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta_j) \Big) \right] + \frac{|I_C|}{K} \sum_{k=1}^{K} \log \alpha_k (1-\alpha_k) - |I_C| \cdot \log \beta \\
& + \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}};\phi)} \left[ \sum_{j \in I_D} \sum_{c=1}^{C_j} \mathbb{I}((\mathbf{x}_{1-\mathbf{r}})_j = c) \pi_{jc}(\mathbf{z}; \theta_j) \right] - D_{\mathrm{KL}} \Big( q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi) \Big\| p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta) \Big) \\
= \;& -\frac{1}{\beta} \cdot \Bigg( \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}};\phi)} \left[ \sum_{j \in I_C} \frac{1}{K} \sum_{k=1}^{K} \rho_{\alpha_k} \Big( (\mathbf{x}_{1-\mathbf{r}})_j - Q_j(\alpha_k, \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta_j) \Big) \right] - \beta \cdot \frac{|I_C|}{K} \sum_{k=1}^{K} \log \alpha_k (1-\alpha_k) + \beta \cdot |I_C| \cdot \log \beta \\
& - \beta \cdot \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}};\phi)} \left[ \sum_{j \in I_D} \sum_{c=1}^{C_j} \mathbb{I}((\mathbf{x}_{1-\mathbf{r}})_j = c) \log \pi_{jc}(\mathbf{z}; \theta_j) \right] + \beta \cdot D_{\mathrm{KL}} \Big( q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi) \Big\| p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \theta) \Big) \Bigg) \\
= \;& -\frac{1}{\beta} \cdot \mathcal{L}(\mathbf{x}_{\mathbf{r}}, \mathbf{x}_{1-\mathbf{r}}; \theta, \eta, \phi)
\end{aligned}
$$

by Jensen's inequality. Note that this derivation remains valid when $\mathbf{r}$ is replaced by any masking vector $\mathbf{m} \in \{0,1\}^p$.

## A.2 Derivation of Upper Bound

$$\begin{aligned}
& D_{\mathrm{KL}}\Big(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}}) \,\|\, p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}}; \theta, \eta)\Big) \\
= \;& \mathbb{E}_{p^*(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_{\mathbf{r}})}\Big[-\log p(\mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_{\mathbf{r}}; \theta, \eta)\Big] - H\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})\right) \\
= \;& \mathbb{E}_{p^*(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_{\mathbf{r}})}\Big[-\log \int p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta) p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \eta) d\mathbf{z}\Big] - H\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})\right) \\
= \;& \mathbb{E}_{p^*(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_{\mathbf{r}})}\Big[-\log \int \frac{q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{x}_{1-\mathbf{r}}; \phi)}{q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{x}_{1-\mathbf{r}}; \phi)} p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta) p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \eta) d\mathbf{z}\Big] - H\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})\right) \\
\leq \;& \mathbb{E}_{p^*(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_{\mathbf{r}})}\Big[-\int q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{x}_{1-\mathbf{r}}; \phi) \log p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta) \frac{p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \eta)}{q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{x}_{1-\mathbf{r}}; \phi)} d\mathbf{z}\Big] - H\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})\right) \\
= \;& \mathbb{E}_{p^*(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_{\mathbf{r}})}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_{\mathbf{r}}, \mathbf{x}_{1-\mathbf{r}}; \phi)}\big[-\log p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta)\big] + D_{\mathrm{KL}}(q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{x}_{1-\mathbf{r}}; \phi) \| p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \eta))\Big] \\
& -H\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})\right) \\
= \;& \frac{1}{\beta} \cdot \mathbb{E}_{p^*(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_{\mathbf{r}})}\Big[\mathcal{L}^*(\mathbf{x}_{\mathbf{r}}, \mathbf{x}_{1-\mathbf{r}}; \theta, \eta, \phi)\Big] - \underbrace{H\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})\right)}_{\text{constant}} \quad \text{(see Appendix A.1)},
\end{aligned}$$

by Jensen's inequality, where $H(\cdot)$ is entropy.

## A.3 Proof of Proposition 1

*Proof.* First, suppose that there exist parameters $\phi$ and $\eta$ such that, for any $\varepsilon > 0$ and observation $\mathbf{x}_{\mathbf{r}}$, and its missingness pattern $\mathbf{r}$, the following holds:

$$D_{\mathrm{KL}}\left(q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi) \,\|\, p(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \eta)\right) < \varepsilon.$$

We assume that such $\phi$ and $\eta$ are given and fixed. Additionally, we define the following conditional distribution for the missing entries:

$$q(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{z}; \phi) := \frac{q(\mathbf{z} \mid \mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_{\mathbf{r}}; \phi)\, p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})}{q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi)},$$

which is a valid distribution by the definition of the aggregated posterior. Then, we have

$$\begin{aligned}
p^*(\mathbf{r}, \mathbf{x})\, q(\mathbf{z} \mid \mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_{\mathbf{r}}; \phi) &= p^*(\mathbf{r}, \mathbf{x})\, \frac{q(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{z}; \phi)\, q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi)}{p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})} \\
&= p^*(\mathbf{r} \mid \mathbf{x})\, p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})\, p^*(\mathbf{x}_{\mathbf{r}})\, \frac{q(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{z}; \phi)\, q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi)}{p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}})} \\
&= p^*(\mathbf{r} \mid \mathbf{x})\, p^*(\mathbf{x}_{\mathbf{r}})\, q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi)\, q(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{z}; \phi) \\
&= p^*(\mathbf{r}, \mathbf{x}_{\mathbf{r}})\, q(\mathbf{z} \mid \mathbf{x}_{\mathbf{r}}; \phi)\, q(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_{\mathbf{r}}, \mathbf{z}; \phi). \qquad \text{(by MAR assumption)}
\end{aligned}$$

It leads to the following objective function for imputation using U-VAE:

$$\begin{aligned}
\min_{\theta} \quad & \mathbb{E}_{p^*(\mathbf{r}, \mathbf{x})\, q(\mathbf{z}|\mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_{\mathbf{r}}; \phi)}\left[\sum_{j \in I_C : \mathbf{r}_j = 1} \int_0^1 \rho_\alpha\Big((\mathbf{x}_{1-\mathbf{r}})_j - Q_j(\alpha, \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta_j)\Big) d\alpha\right] \\
= \;& \mathbb{E}_{p^*(\mathbf{r}, \mathbf{x}_{\mathbf{r}})\, q(\mathbf{z}|\mathbf{x}_{\mathbf{r}}; \phi)\, q(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_{\mathbf{r}}, \mathbf{z}; \phi)}\left[\sum_{j \in I_C : \mathbf{r}_j = 1} \int_0^1 \rho_\alpha\Big((\mathbf{x}_{1-\mathbf{r}})_j - Q_j(\alpha, \mathbf{z}, \mathbf{x}_{\mathbf{r}}; \theta_j)\Big) d\alpha\right].
\end{aligned}$$

Under the conditional independence assumption among the variables $\mathbf{x}_{1-\mathbf{r}}$ given $\mathbf{x}_\mathbf{r}$ and $\mathbf{z}$, for $j \in I_C$ such that $\mathbf{r}_j = 0$, the solution $\hat{\theta}_j$ is defined as

$$\hat{\theta}_j \in \arg\min_{\theta_j} \mathbb{E}_{p^*(\mathbf{r},\mathbf{x}_\mathbf{r})\, q(\mathbf{z}|\mathbf{x}_\mathbf{r};\phi)} \left[ \int \int_0^1 \rho_\alpha\Big( (\mathbf{x}_{1-\mathbf{r}})_j - Q_j(\alpha, \mathbf{z}, \mathbf{x}_\mathbf{r}; \theta_j) \Big)\, d\alpha\, q\big((\mathbf{x}_{1-\mathbf{r}})_j \mid \mathbf{x}_\mathbf{r}, \mathbf{z}; \phi\big)\, d(\mathbf{x}_{1-\mathbf{r}})_j \right].$$

Since this CRPS loss is a strictly proper scoring rule relative to the conditional CDF $F_j^*\left((\mathbf{x}_{1-\mathbf{r}})_j \mid \mathbf{z}, \mathbf{x}_\mathbf{r}; \phi\right)$ for all $\mathbf{x}_\mathbf{r} \in \mathbb{R}$ and $\mathbf{z} \in \mathbb{R}^d$, we have

$$\begin{aligned}
Q_j^{-1}\big((\mathbf{x}_{1-\mathbf{r}})_j, \mathbf{z}, \mathbf{x}_\mathbf{r}; \hat{\theta}(\mathbf{x}_\mathbf{r}, \mathbf{r})_j\big) &= F_j^*\big((\mathbf{x}_{1-\mathbf{r}})_j \mid \mathbf{z}, \mathbf{x}_\mathbf{r}; \phi\big) \\
\frac{d}{d(\mathbf{x}_{1-\mathbf{r}})_j} Q_j^{-1}\big((\mathbf{x}_{1-\mathbf{r}})_j, \mathbf{z}, \mathbf{x}_\mathbf{m}; \hat{\theta}_j\big) &= q_j\big((\mathbf{x}_{1-\mathbf{r}})_j \mid \mathbf{z}, \mathbf{x}_\mathbf{r}; \phi\big), \qquad \text{(by Assumption 1)}
\end{aligned}$$

where $q_j\big((\mathbf{x}_{1-\mathbf{r}})_j \mid \mathbf{z}, \mathbf{x}_\mathbf{r}; \phi\big)$ is the marginal distribution of $q\big(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \mathbf{x}_\mathbf{r}; \phi\big)$ for the $j$th variable.

Therefore, it follows that

$$\begin{aligned}
& D_{\mathrm{KL}}\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}) \,\middle\|\, p(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}; \hat{\theta}, \eta)\right) \\
=\ & D_{\mathrm{KL}}\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}) \,\middle\|\, \int \left( \prod_{j \in I_C : \mathbf{r}_j = 0} \frac{d}{d(\mathbf{x}_{1-\mathbf{r}})_j} Q_j^{-1}\big((\mathbf{x}_{1-\mathbf{r}})_j, \mathbf{z}, \mathbf{x}_\mathbf{r}; \hat{\theta}_j\big) \right) p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)\, d\mathbf{z} \right) \\
=\ & D_{\mathrm{KL}}\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}) \,\middle\|\, \int \left( \prod_{j \in I_C : \mathbf{r}_j = 0} q_j\big((\mathbf{x}_{1-\mathbf{r}})_j \mid \mathbf{z}, \mathbf{x}_\mathbf{r}; \phi\big) \right) p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)\, d\mathbf{z} \right) \\
=\ & D_{\mathrm{KL}}\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}) \,\middle\|\, \int q(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{z}, \mathbf{x}_\mathbf{r}; \phi)\, p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)\, d\mathbf{z} \right) \qquad \text{(by Assumption 2)} \\
=\ & D_{\mathrm{KL}}\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}) \,\middle\|\, \int \frac{q(\mathbf{z} \mid \mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_\mathbf{r}; \phi)\, p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r})}{q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi)}\, p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)\, d\mathbf{z} \right) \\
=\ & D_{\mathrm{KL}}\left(p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}) \,\middle\|\, p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}) \int q(\mathbf{z} \mid \mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_\mathbf{r}; \phi) \frac{p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)}{q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi)} d\mathbf{z} \right) \\
=\ & \mathbb{E}_{p^*(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_\mathbf{r})} \left[ -\log \int q(\mathbf{z} \mid \mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_\mathbf{r}; \phi) \frac{p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)}{q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi)} d\mathbf{z} \right] \\
\leq\ & \mathbb{E}_{p^*(\mathbf{x}_{1-\mathbf{r}}|\mathbf{x}_\mathbf{r})} \left[ \int q(\mathbf{z} \mid \mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_\mathbf{r}; \phi) \log \frac{q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi)}{p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)} d\mathbf{z} \right] \\
=\ & \int \left( \int p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r})\, q(\mathbf{z} \mid \mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_\mathbf{r}; \phi) d\mathbf{x}_{1-\mathbf{r}} \right) \log \frac{q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi)}{p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)}\, d\mathbf{z} \\
=\ & \int q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi) \log \frac{q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi)}{p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta)}\, d\mathbf{z} \\
=\ & D_{\mathrm{KL}}\Big( q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi) \| p(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \eta) \Big) < \epsilon
\end{aligned}$$

where $q(\mathbf{z} \mid \mathbf{x}_\mathbf{r}; \phi) = \int p^*(\mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r})\, q(\mathbf{z} \mid \mathbf{x}_{1-\mathbf{r}}, \mathbf{x}_\mathbf{r}; \phi) d\mathbf{x}_{1-\mathbf{r}})$ is aggregated posterior.

The above equations are derived by Jensen's inequality and the Fubini-Tonelli theorem. The proof is complete. □

## A.4 Proof of Proposition 2

*Proof.* We begin by noting that the MAR assumption implies $p(\mathbf{r} \mid \mathbf{x}) = p(\mathbf{r} \mid \mathbf{x}_\mathbf{r})$. Using $\mathbf{r} \perp\!\!\!\perp \mathbf{r} \mid \mathbf{x}$ and $\mathbf{m} \perp\!\!\!\perp \mathbf{x}$, we can write

$$p(\mathbf{r}, \mathbf{m} \mid \mathbf{x}) = p(\mathbf{r} \mid \mathbf{x}) \cdot p(\mathbf{m} \mid \mathbf{x}) = p(\mathbf{r} \mid \mathbf{x}) \cdot p(\mathbf{m}) = p(\mathbf{r} \mid \mathbf{x}_\mathbf{r}) \cdot p(\mathbf{m}) = p(\mathbf{m}, \mathbf{r} \mid \mathbf{x}_\mathbf{r}).$$

Therefore, it follows that $(\mathbf{r}, \mathbf{m}) \perp\!\!\!\perp \mathbf{x}_{1-\mathbf{r}} \mid \mathbf{x}_\mathbf{r}$. The proof is complete. □

### A.5 Quantile function

Similar to [7], for $j \in I_C$, we parameterize the function $Q_j$, the location parameter of ALD, by *linear spline quantile function (LSQF)* as follows:

$$\begin{aligned} & Q_j(\alpha, \mathbf{z}, \mathbf{x_m}; \theta) \\ = \quad & \gamma_j(\mathbf{z}, \mathbf{x_m}) + \sum_{l=0}^{L} \kappa_j^{(l)}(\mathbf{z}, \mathbf{x_m})(\alpha - \delta^{(l)}) \cdot \mathbb{I}((\alpha - \delta^{(l)}) > 0) \quad \text{s.t.} \quad \sum_{l=0}^{k} \kappa_j^{(l)}(\mathbf{z}, \mathbf{x_m}) \geq 0, \quad k = 1, \cdots, L, \end{aligned}$$

where $\gamma_j(\mathbf{z}, \mathbf{x_m}) \in \mathbb{R}$, $\kappa_j(\mathbf{z}, \mathbf{x_m}) = (\kappa_j^{(1)}(\mathbf{z}, \mathbf{x_m}), \cdots, \kappa_j^{(L)}(\mathbf{z}, \mathbf{x_m})) \in \mathbb{R}^{L+1}$, $\delta = (\delta_0, \cdots, \delta_L) \in [0,1]^{L+1}$, and $0 = \delta_0 < \cdots < \delta_L = 1$.

### A.6 Closed-form Objective

$$\begin{aligned} & 2 \int_0^1 \rho_\alpha\Big((\mathbf{x_{1-m}})_j - Q_j(\alpha, \mathbf{z}, \mathbf{x_m} \theta_j)\Big) d\alpha \\ = \quad & (2\tilde{\alpha}_j - 1)(\mathbf{x_{1-m}})_j + (1 - 2\tilde{\alpha}_j)\gamma_j(\mathbf{z}, \mathbf{x_m}) \\ + \quad & \sum_{l=1}^{L} \kappa_j^{(l)}(\mathbf{z}, \mathbf{x_m}) \left( \frac{1 - (\delta^{(l)})^3}{3} - \delta^{(l)} - \max(\tilde{\alpha}_j, \delta^{(l)})^2 + 2 \max(\tilde{\alpha}_j, \delta^{(l)}) \delta^{(l)} \right), \end{aligned}$$

where $Q_j(\tilde{\alpha}_j, \mathbf{z}, \mathbf{x_m}; \theta_j) = (\mathbf{x_{1-m}})_j$, $\tilde{\alpha}_j = \frac{(\mathbf{x_{1-m}})_j - \gamma_j(\mathbf{z}, \mathbf{x_m}) + \sum_{l'=0}^{l_0} \kappa_j^{(l')}(\mathbf{z}, \mathbf{x_m}) \delta^{(l')}}{\sum_{l'=0}^{l_0} \kappa_j^{(l')}(\mathbf{z}, \mathbf{x_m})}$, and $Q_j(\delta^{(l_0)}, \mathbf{z}, \mathbf{x_m}; \theta_j) \leq (\mathbf{x_{1-m}})_j \leq Q_j(\delta^{(l_0+1)}, \mathbf{z}, \mathbf{x_m}; \theta_j)$.

### A.7 Posterior and Prior distributions

We assume that the posterior and prior distributions are

$$\begin{aligned} q(\mathbf{z} \mid \mathbf{x_r}; \phi) &= \mathrm{N}\left(\mu(\mathbf{x_r}; \phi), \mathrm{diag}(\sigma^2(\mathbf{x_r}; \phi))\right) \\ p(\mathbf{z} \mid \mathbf{x_m}; \theta) &= \mathrm{N}\left(\mu(\mathbf{x_m}; \theta), \mathrm{diag}(\sigma^2(\mathbf{x_m}; \theta))\right), \end{aligned}$$

$\mu(\cdot, \phi) : \mathbb{R}^p \mapsto \mathbb{R}^d$, $\sigma^2(\cdot, \phi) : \mathbb{R}^p \mapsto \mathbb{R}^d_+$, $\mu(\cdot, \theta) : \mathbb{R}^p \mapsto \mathbb{R}^d$, $\sigma^2(\cdot, \theta) : \mathbb{R}^p \mapsto \mathbb{R}^d_+$ are neural networks parameterized with $\phi$ and $\theta$, and $\mathrm{diag}(a), a \in \mathbb{R}^d$ denotes a diagonal matrix with diagonal elements $a$.

## B Comparison with Gaussian VAE

In conventional Gaussian VAE, the decoder $p(\mathbf{x_{1-m}} \mid \mathbf{z}, \mathbf{x_m}; \theta)$ is commonly modeled as a Gaussian distribution, primarily because the corresponding reconstruction loss reduces to the *mean squared error (MSE)*, one of the most widely used loss functions in optimization theory [12, 8, 17, 5, 1]. The MSE formulation has also been adopted in prior imputation methods, including [10, 6]. This parametric decoder assumption not only simplifies computation but also yields a closed-form reconstruction loss.

**Assumption 1** (Conditional independence)**.** *The re-masked features* $\{\mathbf{x}_j : \mathbf{m}_j = 0, \mathbf{r}_j = 1\}$ *are conditionally independent given the latent variable* $\mathbf{z}$ *and the un-masked features* $\{\mathbf{x}_j : \mathbf{m}_j = 1, \mathbf{r}_j = 1\}$.

Under Assumption 1, the decoder distribution is assumed to be a multivariate Gaussian with a diagonal covariance matrix:

$$p(\mathbf{x_{1-m}} \mid \mathbf{z}, \mathbf{x_m}; \theta) = \mathrm{N}\left(D(\mathbf{z}, \mathbf{x_m}; \theta), \beta \cdot \mathbf{I}_p\right),$$

where $D(\cdot, \cdot; \theta) : \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}^p$ denotes the mean function of the decoder, which is parameterized by a neural network with parameter $\theta$. Here, $\mathbf{I}_p$ is the $p \times p$ identity matrix, and $\beta > 0$ is a fixed (non-trainable) observation noise scalar [17].

Then, the objective function of the Gaussian VAE can be written as

$$\min_{\theta, \eta, \phi} \mathbb{E}_{q(\mathbf{z}|\mathbf{x_r}; \phi)} \left[ \frac{1}{2} \left\| \mathbf{x_{1-m}} - D(\mathbf{z}, \mathbf{x_m}; \theta) \right\|^2 \right] + \beta \cdot \mathcal{KL}\left(q(\mathbf{z} \mid \mathbf{x_r}; \phi) \| p(\mathbf{z} \mid \mathbf{x_m}; \eta)\right), \tag{1}$$

and since $(\mathbf{x}_{1-\mathbf{m}})_j = 0$ for all $j$ such that $\mathbf{m}_j = 1$, the loss in (1) can be equivalently expressed as:

$$\min_{\theta,\eta,\phi} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_\mathbf{r};\phi)} \left[ \frac{1}{2} \sum_{\substack{j:\mathbf{m}_j=0,\\ \mathbf{r}_j=1}} \Big( (\mathbf{x}_{1-\mathbf{m}})_j - D_j(\mathbf{z}, \mathbf{x}_\mathbf{m}; \theta) \Big)^2 \right] + \beta \cdot D_{\mathrm{KL}} \left( q(\mathbf{z} \mid \mathbf{x}_\mathbf{m}, \mathbf{x}_{1-\mathbf{m}}; \phi) \| p(\mathbf{z} \mid \mathbf{x}_\mathbf{m}; \eta) \right). \quad (2)$$

However, the objective in (2) cannot adequately capture the uncertainty associated with deciding which value to impute, as it merely estimates the conditional expectation of the missing entries. To address this limitation and better capture the uncertainty inherent in imputation, we shift from point estimation to distributional learning.

That is, we aim to estimate the full conditional distribution of the missing values given the observed ones, rather than a single conditional expectation. From a statistical perspective, imputation should not be treated as a prediction task. Whereas prediction aims to estimate the most likely value, imputation seeks to generate plausible values that preserve the joint distribution of the data and reflect the inherent uncertainty of the missingness mechanism [25, 20].

# C Benchmark Datasets

We provide information about the real-world tabular datasets from UCI repository[1], and Kaggle[2] used in our experiment.

Table 1: **Download links of dataset used in this paper.**

| Dataset | Download links |
|---|---|
| anuran [3] | https://archive.ics.uci.edu/dataset/406/anuran+calls+mfccs |
| banknote [15] | https://archive.ics.uci.edu/dataset/267/banknote+authentication |
| breast [27] | https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic |
| concrete [28] | https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength |
| default [29] | https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients |
| kings (CC0: Public Domain) | https://www.kaggle.com/datasets/harlfoxem/housesalesprediction |
| letter [22] | https://archive.ics.uci.edu/dataset/59/letter+recognition |
| loan (CC0: Public Domain) | https://www.kaggle.com/datasets/teertha/personal-loan-modeling |
| redwine [4] | https://archive.ics.uci.edu/dataset/186/wine+quality |
| shoppers [21] | https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset |
| whitewine [4] | https://archive.ics.uci.edu/dataset/186/wine+quality |

Table 2: **Description of datasets.** $N$ represents the number of observations. $|I_C|$ and $|I_D|$ represent the number of continuous variables and categorical (discrete) variables, respectively. The 'Classification Target' refers to the variable used as the response variable in a classification task to evaluate imputation data utility.

| Dataset | $N$ | $\lvert I_C\rvert$ | $\lvert I_D\rvert$ | Classification Target |
|---|---|---|---|---|
| anuran | 7,195 | 22 | 3 | Species |
| banknote | 1,372 | 4 | 1 | class |
| breast | 569 | 30 | 1 | Diagnosis |
| concrete | 1,030 | 8 | 1 | Age |
| default | 30,000 | 14 | 10 | default payment |
| kings | 21,613 | 11 | 7 | grade |
| letter | 20,000 | 16 | 1 | lettr |
| loan | 5,000 | 5 | 6 | Personal Loan |
| redwine | 1,599 | 11 | 1 | quality |
| shoppers | 12,330 | 10 | 8 | Revenue |
| whitewine | 4,898 | 11 | 1 | quality |

[1] https://archive.ics.uci.edu/
[2] https://www.kaggle.com/datasets/

# D Reproducibility

We present all the experimental settings for reproducibility in this paper.

## D.1 Experimental Settings

- We run experiments using NVIDIA RTX A6000 GPU, and our experimental codes are available with `PyTorch` and `scikit-learn`.

**Hyperparameters** This demonstrates the *generalizability of our proposed model* across various tabular datasets. For all 11 datasets, we applied the following hyperparameters uniformly, without any additional tunings:

- batch size: 1024
- $\beta$ (scale parameter of asymmetric Laplace distribution): 0.1
- step (interval size of quantile levels): 0.1
- epochs: 1000 (with `AdamW` optimizer [16])
- learning rate: 0.002
- Activation function: `nn.ELU()` [2]

**Training procedure** The overall training procedure of U-VAE is outlined in Algorithm 1 (note that mini-batch style training is also applicable):

**Algorithm 1** Training of U-VAE

**Input**: Dataset: $\{(\mathbf{x}^{(i)}, \mathbf{r}^{(i)})\}_{i=1}^{n}$
**Parameter**: $\theta, \phi$

**for** $e = 1, 2, \cdots, epochs$ **do**
  $\mathbf{m}^{(i)} \sim_{i.i.d.} p(\mathbf{m})$ for $i = 1, \cdots, n$
  Update $\theta$ and $\phi$ by the gradient descent of $\sum_{i=1}^{n} \mathcal{L}(\mathbf{x}_{\mathbf{r}}^{(i)}, \mathbf{m}^{(i)}; \theta, \phi)$
**end for**

## D.2 Missing mechanisms

**Descriptions** Data missingness is a common challenge in research and practical analysis, categorized into three primary missing mechanisms: 1) Missing Completely at Random (MCAR), 2) Missing at Random (MAR), and 3) Missing Not at Random (MNAR).

Under the **MCAR** mechanism, the reason for missingness is unrelated to any data, whether observed or unobserved. In other words, the likelihood of data being missing is equal across all observations. The primary advantage of MCAR is that it does not introduce bias into the data analysis. However, despite this advantage, data missingness can still reduce the statistical power of the study because of the reduced sample size.

**MAR** occurs when the probability of missingness is related to the observed data but not the unobserved missing data. Essentially, even though the data is missing, the mechanism assumes that the missingness can be explained by other variables in the dataset. In other words, the missingness can be modeled and imputed using the available data, allowing for more accurate analyses despite the missing values.

Lastly, if the missingness is not specified by either MCAR or MAR, it becomes **MNAR**. MNAR is the most challenging mechanism, as it implies that the missingness is related to the unobserved data itself. In this case, the missing data is systematically different from the observed data, which introduces bias if not properly accounted for. For example, patients with severe symptoms may be less likely to report their health status, resulting in missing data. MNAR requires sophisticated statistical methods to handle, as ignoring or mishandling it can lead to biased and unreliable results.

**Implementations** Following [19, 11, 31], we generate the missing value mask for each dataset with three mechanisms in four settings, MCAR, MAR, MNARL, and MNARQ.

In the **MCAR** setting, each value is masked according to the realization of a Bernoulli random variable with a fixed parameter.

For the **MAR** setting, a fixed subset of variables that cannot have missing values is sampled in each experiment. The remaining variables are assigned missing values based on a logistic model with random weights, where the non-missing variables serve as inputs. A bias term is optimized via line search to achieve the desired proportion of missing values.

Finally, two different mechanisms are implemented in the MNAR setting. The first, **MNARL**, is identical to the previously described MAR mechanism, but the inputs of the logistic model are then masked by an MCAR mechanism. Hence, the logistic model's outcome depends on missing values. The second mechanism, **MNARQ**, samples a subset of variables whose values in the lower and upper $p$th percentiles are masked according to a Bernoulli random variable, and the values in-between are left not missing.

## D.3 Details of Implementing Baseline Models

To compare our proposed method with baseline models, we performed experiments across four missing data mechanisms and five missingness rates. Our reproduced codes for the baseline models are provided in the supplementary material. Below are the detailed implementations of the baseline methods:

- Mean [14]: We employ the `SimpleImputer` package [3], with the `strategy` parameter set to 'mean'.
- kNNI [24]: We employ the `KNNImputer` package [4] from `scikit-learn` for missing data imputation.
- MICE [26]: We employ the `IterativeImputer` package [5] from `scikit-learn`.
- missForest [23]: We employ the missForest module in `hyperimpute.Plugins` [6].
- GAIN [30]: We follow the implementations provided in the official repository[7]. As the paper does not explicitly discuss the separate handling of categorical and continuous variables, the code treats them simultaneously. Consequently, a rounding process is employed to handle categorical variables afterward.
- VAEAC [10]: We follow the implementations provided in the official repository[8]. The authors provided hyperparameters that adequately address both continuous and categorical variables, so we used these without further modification during model fitting.
- MIWAE [18]: The implemented MIWAE code in the official repository[9] was designed for continuous variables only. To accommodate heterogeneous tabular datasets, we treat the conditional distribution of categorical columns as categorical distributions and employ cross-entropy loss for reconstruction.
- not-MIWAE [9]: The implemented not-MIWAE code in the official repository[10] also focused on continuous variables exclusively. To handle categorical variables, we make the same modifications as in MIWAE.
- MIRACLE [13]: We utilize the MIRACLE module in `hyperimpute.Plugins` [11].
- ReMasker [6]: We follow the implementations provided in official repository [12].

[3] https://scikit-learn.org/1.5/modules/generated/sklearn.impute.SimpleImputer.html
[4] https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html
[5] https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html
[6] https://github.com/vanderschaarlab/hyperimpute/blob/main/src/hyperimpute/plugins/imputers/plugin_missforest.py
[7] https://github.com/jsyoon0823/GAIN/tree/master
[8] https://github.com/tigvarts/vaeac
[9] https://github.com/pamattei/miwae
[10] https://github.com/nbip/notMIWAE
[11] https://github.com/vanderschaarlab/hyperimpute/blob/main/src/hyperimpute/plugins/imputers/plugin_miracle.py
[12] https://github.com/tydusky/remasker

## D.4 Evaluation settings

Regarding **imputation data utility**, we conduct regression, classification, and feature selection tasks for post-imputation evaluation. Please refer to Table 2 for the classification target variable, and Table 3 for detailed machine learning model configuration for these tasks. The detailed evaluation procedure as follows:

| Tasks | Model | Description |
|---|---|---|
| `Regression` | Random Forest | Package: `sklearn.ensemble.RandomForestRegressor`, setting: random_state=0, defaulted values |
| `Classification` | Logistic Regression | Package: `sklearn.linear_model.LogisticRegression`, setting: tol=0.001, random_state=42, max_iter=1000, defaulted values |
| | $k$-Nearest Neighbors | Package: `sklearn.neighbors.KNeighborsClassifier`, setting: defaulted values |
| | Support Vector Machine | Package: `sklearn.svm.SVC`, setting: random_state=42 |
| | Random Forest | Package: `sklearn.ensemble.RandomForestClassifier`, setting: random_state=42, defaulted values |
| | Gradient Boosting | Package: `sklearn.ensemble.GradientBoostingClassifier`, setting: random_state=42, defaulted values |
| | Adaptive Boosting | Package: `sklearn.ensemble.AdaBoostClassifier`, setting: random_state=42, defaulted values |

Table 3: **Configuration of machine learning models (regressor and classifier) used to evaluate imputation data utility.** The names of all arguments used in the description are consistent with those defined in corresponding packages.

**Regression performance (SMAPE)**

1. Train an imputation method a given incomplete train dataset.
2. Impute the incomplete dataset using trained imputation model.
3. Train a machine learning model (Random Forest regressor in Table 3) using the imputed dataset, where each continuous variable serves as the response variable.
4. Assess regression prediction performance by averaging the SMAPE from the test dataset for each Random Forest regressor trained on the continuous variables.

**Classification performance (Accuracy)**

1. Train an imputation method using a given incomplete train dataset.
2. Impute the incomplete train dataset using trained imputation model.
3. Train 6 machine learning models in Table 3 using the imputed dataset.
4. Assess classification prediction performance by averaging the Accuracy from the complete test dataset from five different classifiers.

**Feature selection performance (Feature)**

1. Train an imputer using a given incomplete train dataset.
2. Impute the incomplete train dataset using trained imputer.
3. Train a Random Forest classifier using both the original complete train dataset and the imputed dataset.

4. Determine the rank-ordering of important features for both classifiers.
5. Assess feature selection performance by comparing the feature importance rank orderings of classifiers trained on the original complete train dataset and those trained on the imputed dataset using *Spearman's Rank Correlation.*

**Multiple imputation** We assess the effectiveness of multiple imputations by employing interval inference for the population mean, which was proposed by Rubin[20]. We report the raw bias, coverage, and confidence interval length (width), and the evaluation procedure for multiple imputations is outlined in Algorithm 2 [25].

---

**Algorithm 2** Evaluation procedure for multiple imputation (adapted from [20, 25])

---

**Input**: Complete data set $D^{*} = \{x_i\}_{i=1}^{n}$, Number of imputations $M$
**Output**: Bias, Coverage, and Average Confidence-Interval (CI) Length

**Target estimand** $Q^{*} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(x_i > \bar{x})$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
**for** $s = 1, \ldots, S$ **do**
  Generate missing values into $D^{*}$ to obtain incomplete data $D^{(s)}$
  Train the imputation model on $D^{(s)}$
  **for** $m = 1, \ldots, M$ **do** ▷ Multiple imputations
    Impute the incomplete dataset $\hat{D}_m^{(s)}$
    Compute estimate $\hat{Q}_m^{(s)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\hat{x}_{i,m}^{(s)} > \bar{\hat{x}}_m^{(s)})$
    Compute imputed data variance $\hat{U}_m^{(s)} = \frac{1}{n}\hat{Q}_m^{(s)}(1 - \hat{Q}_m^{(s)})$
  **end for**
  **Rubin's combining rules**[13]

$$\bar{Q}^{(s)} = \frac{1}{M}\sum_{m=1}^{M} \hat{Q}_m^{(s)}, \qquad \bar{U}^{(s)} = \frac{1}{M}\sum_{m=1}^{M} \hat{U}_m^{(s)}$$

$$B^{(s)} = \frac{1}{M-1}\sum_{m=1}^{M} \big(\hat{Q}_m^{(s)} - \bar{Q}^{(s)}\big)^2, \qquad T^{(s)} = \bar{U}^{(s)} + \Big(1 + \tfrac{1}{M}\Big)B^{(s)}$$

**end for**
**Performance metrics**

$$\begin{aligned} \text{Bias} &= \frac{1}{S}\sum_{s=1}^{S} \big|\bar{Q}^{(s)} - Q^{*}\big| \\ \text{Coverage} &= \frac{1}{S}\sum_{s=1}^{S} \mathbb{I}\Big(Q^{*} \in \big[\bar{Q}^{(s)} \pm 1.96\sqrt{T^{(s)}}\big]\Big) \\ \text{Width} &= \frac{1}{S}\sum_{s=1}^{S} 2 \times 1.96\sqrt{T^{(s)}} \end{aligned}$$

---

[13]When $M$ is small, one may replace 1.96 by the $t$-quantile with $\nu = (M-1)\big(1 + \frac{\bar{U}^{(s)}}{(1+1/M)B^{(s)}}\big)^2$ degrees of freedom, following [20].

# E Detailed experimental results

## E.1 Across missingness rates

**Q1: Imputation fidelity (RMSE)** We report the RMSE results to assess imputation fidelity across missingness rates of 0.1, 0.5, 0.7, and 0.9.

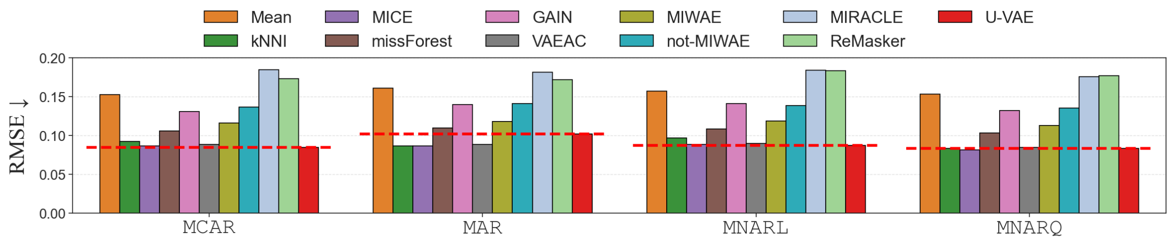


Figure 1: **Imputation fidelity (RMSE)** at 0.1 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 11 datasets and 5 random seeds are reported. ↓ denotes that lower is better.

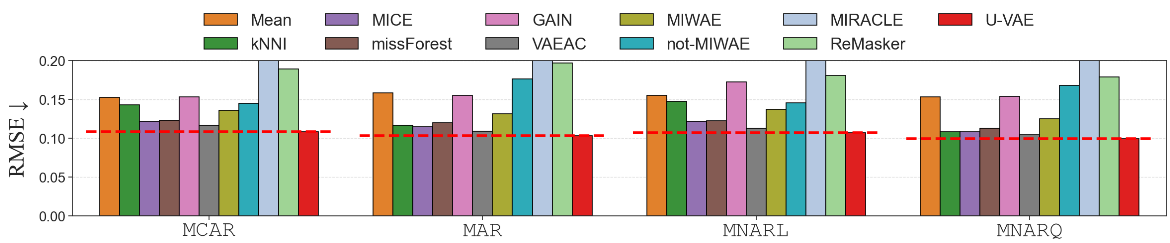


Figure 2: **Imputation fidelity (RMSE)** at 0.5 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 11 datasets and 5 random seeds are reported. ↓ denotes that lower is better.

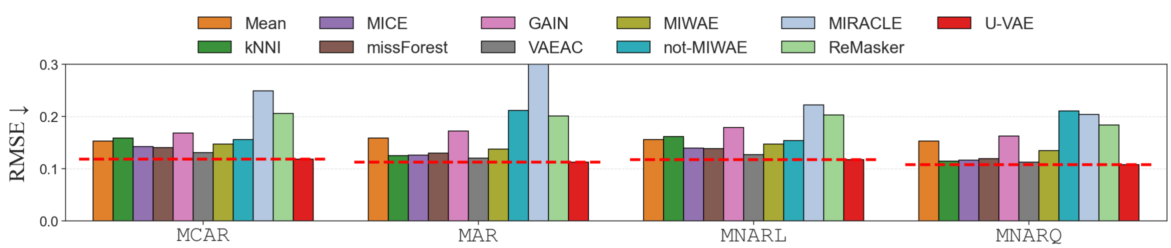


Figure 3: **Imputation fidelity (RMSE)** at 0.7 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 11 datasets and 5 random seeds are reported. ↓ denotes that lower is better.

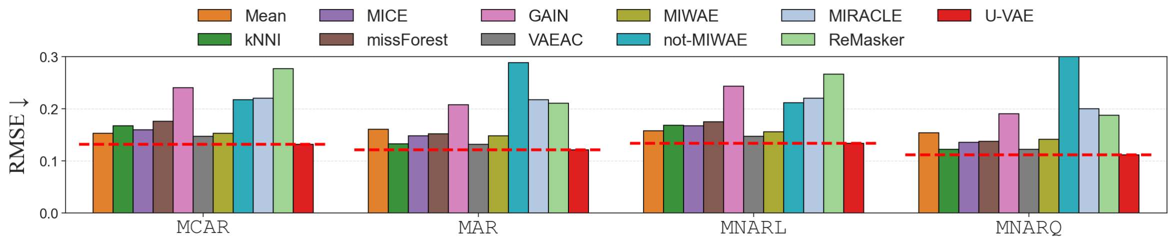


Figure 4: **Imputation fidelity (RMSE)** at 0.9 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 11 datasets and 5 random seeds are reported. ↓ denotes that lower is better.

**Q1: Imputation fidelity (WD)** We report the WD results to assess imputation fidelity across missingness rates of 0.1, 0.5, 0.7, and 0.9.

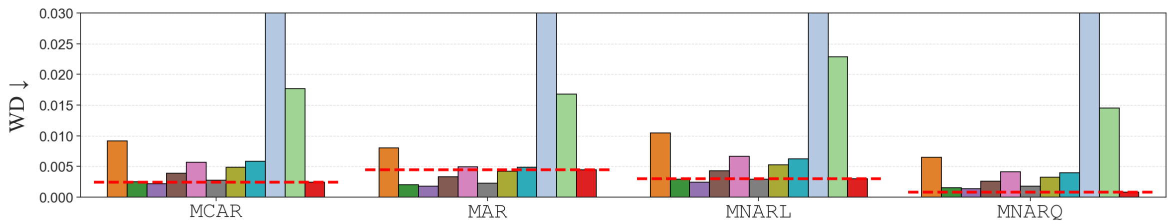


Figure 5: **Imputation fidelity (WD)** at 0.1 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 11 datasets and 5 random seeds are reported. ↓ denotes that lower is better.

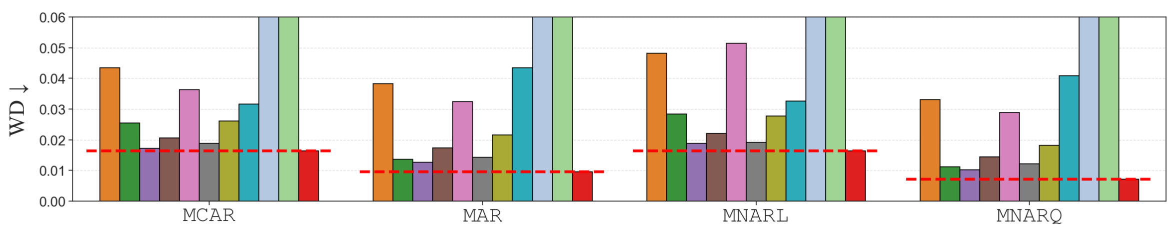


Figure 6: **Imputation fidelity (WD)** at 0.5 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 11 datasets and 5 random seeds are reported. ↓ denotes that lower is better.

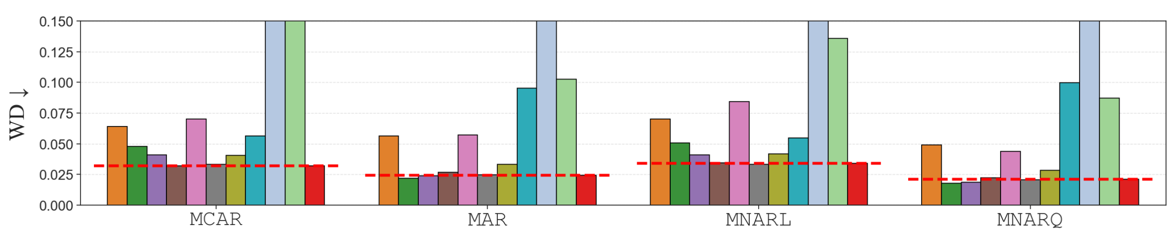


Figure 7: **Imputation fidelity (WD)** at 0.7 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 11 datasets and 5 random seeds are reported. ↓ denotes that lower is better.

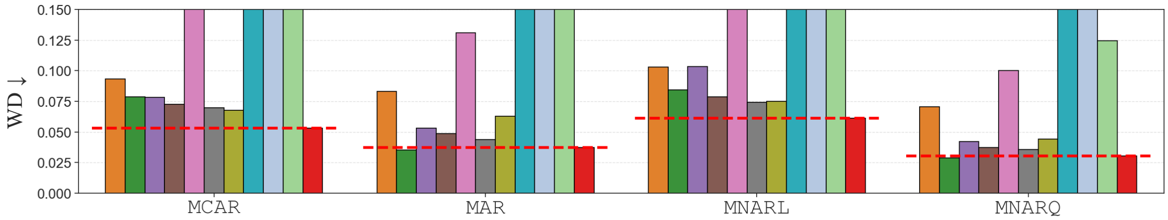


Figure 8: **Imputation fidelity (WD)** at 0.9 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 11 datasets and 5 random seeds are reported. ↓ denotes that lower is better.

**Q1: Imputation fidelity result** As the missingness rate increases, the RMSE of other baselines deteriorates, whereas U-VAE consistently maintains the lowest error in Figures 1–4, outperforming in 15 out of 16 settings (4 missingness rates × 4 missingness mechanisms). This performance gap becomes more pronounced under higher missingness rates and more challenging mechanisms (e.g., MNAR), indicating U-VAE's superior reconstruction accuracy and robustness. As shown in Figures 5–8, U-VAE again remains near the bottom in distributional divergence, indicating a closer match to the full data distribution beyond pointwise accuracy, and outperforms in 12 out of 16 settings (4 missingness rates × 4 missingness mechanisms). Moreover, its performance degrades more gradually with increasing missingness compared to other baselines.

**Q1: Imputation utility** We report the SMAPE (regression), Accuracy (classification), and Feature (classification) results to assess imputation utility across missingness rates of 0.1, 0.5, 0.7, and 0.9.

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $.212_{\pm.020}$ | $.776_{\pm.024}$ | $.857_{\pm.015}$ | $.187_{\pm.017}$ | $.766_{\pm.024}$ | $.789_{\pm.017}$ | $.209_{\pm.020}$ | $.768_{\pm.025}$ | $.845_{\pm.016}$ | $.192_{\pm.017}$ | $.782_{\pm.024}$ | $.868_{\pm.014}$ |
| kNNI | $\underline{.175}_{\pm.016}$ | $.778_{\pm.024}$ | $.859_{\pm.015}$ | $\underline{.163}_{\pm.015}$ | $.767_{\pm.024}$ | $.826_{\pm.015}$ | $\mathbf{.174}_{\pm.016}$ | $.771_{\pm.024}$ | $.851_{\pm.015}$ | $\mathbf{.162}_{\pm.015}$ | $.783_{\pm.024}$ | $\underline{.892}_{\pm.013}$ |
| MICE | $.188_{\pm.019}$ | $.779_{\pm.024}$ | $\underline{.869}_{\pm.015}$ | $.171_{\pm.016}$ | $.773_{\pm.023}$ | $\underline{.868}_{\pm.014}$ | $.187_{\pm.018}$ | $.770_{\pm.024}$ | $.850_{\pm.015}$ | $.176_{\pm.016}$ | $.785_{\pm.023}$ | $.885_{\pm.012}$ |
| missForest | $.199_{\pm.019}$ | $.778_{\pm.024}$ | $.865_{\pm.015}$ | $.178_{\pm.016}$ | $.767_{\pm.024}$ | $.833_{\pm.014}$ | $.198_{\pm.019}$ | $.770_{\pm.024}$ | $.854_{\pm.015}$ | $.185_{\pm.017}$ | $.783_{\pm.024}$ | $.886_{\pm.012}$ |
| GAIN | $.196_{\pm.019}$ | $.779_{\pm.024}$ | $.811_{\pm.019}$ | $.179_{\pm.017}$ | $.777_{\pm.024}$ | $.819_{\pm.016}$ | $\underline{.183}_{\pm.017}$ | $.772_{\pm.024}$ | $.789_{\pm.015}$ | $\underline{.170}_{\pm.015}$ | $.777_{\pm.024}$ | $.839_{\pm.015}$ |
| VAEAC | $.193_{\pm.019}$ | $\underline{.787}_{\pm.023}$ | $.858_{\pm.014}$ | $.174_{\pm.016}$ | $\mathbf{.787}_{\pm.023}$ | $.860_{\pm.014}$ | $.192_{\pm.018}$ | $\mathbf{.787}_{\pm.023}$ | $\underline{.863}_{\pm.013}$ | $.180_{\pm.016}$ | $\underline{.790}_{\pm.023}$ | $.889_{\pm.012}$ |
| MIWAE | $.193_{\pm.018}$ | $.771_{\pm.023}$ | $.827_{\pm.017}$ | $.183_{\pm.017}$ | $.766_{\pm.023}$ | $.799_{\pm.016}$ | $.192_{\pm.018}$ | $.762_{\pm.023}$ | $.793_{\pm.017}$ | $.185_{\pm.017}$ | $.781_{\pm.023}$ | $.839_{\pm.014}$ |
| not-MIWAE | $.202_{\pm.019}$ | $.774_{\pm.025}$ | $.815_{\pm.016}$ | $.188_{\pm.018}$ | $.765_{\pm.024}$ | $.799_{\pm.017}$ | $.201_{\pm.019}$ | $.764_{\pm.025}$ | $.793_{\pm.016}$ | $.191_{\pm.018}$ | $.780_{\pm.024}$ | $.823_{\pm.016}$ |
| MIRACLE | $.196_{\pm.018}$ | $.771_{\pm.025}$ | $.808_{\pm.018}$ | $.177_{\pm.015}$ | $.764_{\pm.025}$ | $.805_{\pm.016}$ | $.196_{\pm.017}$ | $.765_{\pm.025}$ | $.792_{\pm.017}$ | $.183_{\pm.015}$ | $.779_{\pm.024}$ | $.831_{\pm.017}$ |
| ReMasker | $.209_{\pm.020}$ | $.769_{\pm.026}$ | $.814_{\pm.017}$ | $.188_{\pm.018}$ | $.763_{\pm.025}$ | $.793_{\pm.018}$ | $.207_{\pm.019}$ | $.762_{\pm.026}$ | $.797_{\pm.017}$ | $.195_{\pm.018}$ | $.778_{\pm.024}$ | $.832_{\pm.017}$ |
| U-VAE | $\mathbf{.174}_{\pm.019}$ | $\mathbf{.788}_{\pm.025}$ | $\mathbf{.895}_{\pm.014}$ | $\mathbf{.161}_{\pm.018}$ | $\underline{.785}_{\pm.024}$ | $\mathbf{.891}_{\pm.015}$ | $\mathbf{.174}_{\pm.019}$ | $\mathbf{.782}_{\pm.025}$ | $\mathbf{.885}_{\pm.014}$ | $.192_{\pm.018}$ | $\mathbf{.793}_{\pm.024}$ | $\mathbf{.919}_{\pm.013}$ |

Table 4: **Imputation utility** at 0.1 missingness. The means and their standard errors across 11 real-world datasets and 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better. The best value is **red bolded**, and the second best is blue underlined.

| model | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | .284$_{\pm.023}$ | .513$_{\pm.046}$ | .634$_{\pm.027}$ | .261$_{\pm.020}$ | .543$_{\pm.042}$ | .474$_{\pm.039}$ | .289$_{\pm.023}$ | .518$_{\pm.044}$ | .559$_{\pm.034}$ | .252$_{\pm.020}$ | .602$_{\pm.043}$ | .619$_{\pm.026}$ |
| kNNI | .258$_{\pm.023}$ | .548$_{\pm.040}$ | <u>.682</u>$_{\pm.033}$ | <u>.216</u>$_{\pm.019}$ | .559$_{\pm.040}$ | .584$_{\pm.030}$ | .261$_{\pm.023}$ | .550$_{\pm.039}$ | .608$_{\pm.034}$ | <u>.212</u>$_{\pm.019}$ | .615$_{\pm.042}$ | .698$_{\pm.026}$ |
| MICE | <u>.245</u>$_{\pm.022}$ | .537$_{\pm.042}$ | .664$_{\pm.029}$ | .221$_{\pm.020}$ | .562$_{\pm.039}$ | .635$_{\pm.028}$ | <u>.248</u>$_{\pm.022}$ | .548$_{\pm.041}$ | .604$_{\pm.032}$ | .218$_{\pm.019}$ | .626$_{\pm.041}$ | **.717**$_{\pm.028}$ |
| missForest | .250$_{\pm.021}$ | .513$_{\pm.045}$ | .655$_{\pm.025}$ | .230$_{\pm.019}$ | .553$_{\pm.042}$ | .575$_{\pm.031}$ | .251$_{\pm.021}$ | .521$_{\pm.044}$ | .603$_{\pm.030}$ | .227$_{\pm.019}$ | .608$_{\pm.043}$ | .686$_{\pm.027}$ |
| GAIN | .264$_{\pm.022}$ | .627$_{\pm.031}$ | .517$_{\pm.027}$ | .241$_{\pm.020}$ | .659$_{\pm.031}$ | .578$_{\pm.025}$ | .282$_{\pm.023}$ | .624$_{\pm.032}$ | .486$_{\pm.032}$ | .231$_{\pm.019}$ | .703$_{\pm.030}$ | .614$_{\pm.031}$ |
| VAEAC | **.238**$_{\pm.020}$ | **.752**$_{\pm.026}$ | .674$_{\pm.026}$ | <u>.216</u>$_{\pm.018}$ | <u>.752</u>$_{\pm.026}$ | **.708**$_{\pm.019}$ | **.240**$_{\pm.021}$ | **.746**$_{\pm.026}$ | **.650**$_{\pm.026}$ | .215$_{\pm.018}$ | <u>.764</u>$_{\pm.026}$ | <u>.713</u>$_{\pm.023}$ |
| MIWAE | .243$_{\pm.020}$ | .394$_{\pm.022}$ | .513$_{\pm.032}$ | .231$_{\pm.019}$ | .432$_{\pm.020}$ | .538$_{\pm.030}$ | .248$_{\pm.020}$ | .420$_{\pm.019}$ | .492$_{\pm.041}$ | .223$_{\pm.019}$ | .551$_{\pm.034}$ | .591$_{\pm.032}$ |
| not-MIWAE | .261$_{\pm.022}$ | .492$_{\pm.046}$ | .570$_{\pm.032}$ | .254$_{\pm.021}$ | .483$_{\pm.039}$ | .522$_{\pm.032}$ | .267$_{\pm.023}$ | .504$_{\pm.044}$ | .534$_{\pm.033}$ | .245$_{\pm.020}$ | .561$_{\pm.043}$ | .584$_{\pm.030}$ |
| MIRACLE | .281$_{\pm.022}$ | .531$_{\pm.041}$ | .592$_{\pm.036}$ | .248$_{\pm.019}$ | .562$_{\pm.039}$ | .552$_{\pm.036}$ | .282$_{\pm.022}$ | .537$_{\pm.041}$ | .528$_{\pm.036}$ | .244$_{\pm.018}$ | .614$_{\pm.042}$ | .631$_{\pm.032}$ |
| ReMasker | .276$_{\pm.022}$ | .512$_{\pm.046}$ | .574$_{\pm.033}$ | .257$_{\pm.021}$ | .539$_{\pm.044}$ | .512$_{\pm.040}$ | .281$_{\pm.023}$ | .515$_{\pm.045}$ | .531$_{\pm.037}$ | .251$_{\pm.020}$ | .595$_{\pm.044}$ | .621$_{\pm.037}$ |
| U-VAE | **.238**$_{\pm.023}$ | <u>.745</u>$_{\pm.044}$ | **.687**$_{\pm.028}$ | **.215**$_{\pm.021}$ | **.771**$_{\pm.039}$ | <u>.642</u>$_{\pm.034}$ | **.240**$_{\pm.023}$ | <u>.742</u>$_{\pm.041}$ | <u>.624</u>$_{\pm.031}$ | **.210**$_{\pm.020}$ | **.808**$_{\pm.041}$ | .690$_{\pm.030}$ |

Table 5: **Imputation utility** at 0.5 missingness. The means and their standard errors across 11 real-world datasets and 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better. The best value is **red bolded**, and the second best is <u>blue underlined</u>.

| model | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | .306$_{\pm.025}$ | .435$_{\pm.047}$ | .460$_{\pm.034}$ | .294$_{\pm.023}$ | .459$_{\pm.047}$ | .359$_{\pm.046}$ | .316$_{\pm.025}$ | .441$_{\pm.047}$ | .419$_{\pm.039}$ | .275$_{\pm.022}$ | .542$_{\pm.048}$ | .500$_{\pm.034}$ |
| kNNI | .293$_{\pm.024}$ | .492$_{\pm.044}$ | <u>.556</u>$_{\pm.041}$ | .244$_{\pm.021}$ | .481$_{\pm.045}$ | .476$_{\pm.039}$ | .299$_{\pm.024}$ | .497$_{\pm.044}$ | .475$_{\pm.042}$ | .234$_{\pm.020}$ | .552$_{\pm.047}$ | .586$_{\pm.035}$ |
| MICE | .280$_{\pm.023}$ | .449$_{\pm.044}$ | .449$_{\pm.038}$ | .253$_{\pm.022}$ | .490$_{\pm.044}$ | .529$_{\pm.034}$ | .289$_{\pm.023}$ | .470$_{\pm.045}$ | .416$_{\pm.042}$ | .238$_{\pm.020}$ | .567$_{\pm.045}$ | .590$_{\pm.037}$ |
| missForest | .269$_{\pm.021}$ | .416$_{\pm.048}$ | .492$_{\pm.034}$ | .259$_{\pm.021}$ | .470$_{\pm.047}$ | .468$_{\pm.043}$ | .274$_{\pm.022}$ | .435$_{\pm.048}$ | .451$_{\pm.036}$ | .250$_{\pm.020}$ | .548$_{\pm.048}$ | .568$_{\pm.035}$ |
| GAIN | .299$_{\pm.023}$ | .535$_{\pm.038}$ | .382$_{\pm.039}$ | .282$_{\pm.022}$ | .548$_{\pm.039}$ | .405$_{\pm.039}$ | .319$_{\pm.024}$ | .532$_{\pm.037}$ | .392$_{\pm.039}$ | .259$_{\pm.020}$ | .632$_{\pm.036}$ | .482$_{\pm.041}$ |
| VAEAC | <u>.262</u>$_{\pm.022}$ | **.719**$_{\pm.026}$ | **.588**$_{\pm.025}$ | <u>.242</u>$_{\pm.020}$ | <u>.724</u>$_{\pm.027}$ | <u>.600</u>$_{\pm.026}$ | <u>.266</u>$_{\pm.022}$ | **.710**$_{\pm.026}$ | <u>.516</u>$_{\pm.038}$ | <u>.233</u>$_{\pm.019}$ | <u>.743</u>$_{\pm.027}$ | <u>.622</u>$_{\pm.029}$ |
| MIWAE | .268$_{\pm.021}$ | .218$_{\pm.015}$ | .422$_{\pm.037}$ | .259$_{\pm.021}$ | .253$_{\pm.021}$ | .436$_{\pm.037}$ | .280$_{\pm.022}$ | .248$_{\pm.014}$ | .371$_{\pm.041}$ | .246$_{\pm.020}$ | .424$_{\pm.042}$ | .462$_{\pm.040}$ |
| not-MIWAE | .291$_{\pm.024}$ | .407$_{\pm.048}$ | .416$_{\pm.040}$ | .297$_{\pm.023}$ | .370$_{\pm.044}$ | .407$_{\pm.039}$ | .292$_{\pm.024}$ | .419$_{\pm.048}$ | .435$_{\pm.040}$ | .279$_{\pm.021}$ | .477$_{\pm.049}$ | .445$_{\pm.040}$ |
| MIRACLE | .312$_{\pm.024}$ | .459$_{\pm.043}$ | .439$_{\pm.037}$ | .286$_{\pm.021}$ | .490$_{\pm.042}$ | .453$_{\pm.040}$ | .322$_{\pm.024}$ | .460$_{\pm.043}$ | .408$_{\pm.043}$ | .272$_{\pm.020}$ | .551$_{\pm.046}$ | .532$_{\pm.040}$ |
| ReMasker | .316$_{\pm.024}$ | .427$_{\pm.047}$ | .404$_{\pm.044}$ | .292$_{\pm.022}$ | .456$_{\pm.047}$ | .426$_{\pm.047}$ | .330$_{\pm.025}$ | .433$_{\pm.047}$ | .377$_{\pm.050}$ | .281$_{\pm.022}$ | .531$_{\pm.049}$ | .514$_{\pm.041}$ |
| U-VAE | **.260**$_{\pm.025}$ | <u>.669</u>$_{\pm.049}$ | <u>.566</u>$_{\pm.043}$ | **.241**$_{\pm.024}$ | **.733**$_{\pm.046}$ | **.605**$_{\pm.041}$ | **.265**$_{\pm.025}$ | <u>.697</u>$_{\pm.048}$ | **.524**$_{\pm.049}$ | **.230**$_{\pm.022}$ | **.799**$_{\pm.049}$ | **.650**$_{\pm.038}$ |

Table 6: **Imputation utility** at 0.7 missingness. The means and their standard errors across 11 real-world datasets and 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better. The best value is **red bolded**, and the second best is <u>blue underlined</u>.

| model | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | .332$_{\pm.026}$ | .400$_{\pm.046}$ | .275$_{\pm.045}$ | .331$_{\pm.025}$ | .420$_{\pm.047}$ | .224$_{\pm.048}$ | .352$_{\pm.027}$ | .402$_{\pm.045}$ | .261$_{\pm.052}$ | .305$_{\pm.024}$ | .521$_{\pm.048}$ | .331$_{\pm.040}$ |
| kNNI | .341$_{\pm.026}$ | .436$_{\pm.046}$ | .334$_{\pm.052}$ | **.277**$_{\pm.023}$ | .425$_{\pm.047}$ | .345$_{\pm.045}$ | .348$_{\pm.026}$ | .451$_{\pm.045}$ | <u>.362</u>$_{\pm.042}$ | .260$_{\pm.022}$ | .523$_{\pm.049}$ | .428$_{\pm.042}$ |
| MICE | .339$_{\pm.024}$ | .393$_{\pm.047}$ | .272$_{\pm.048}$ | .310$_{\pm.024}$ | .427$_{\pm.046}$ | .313$_{\pm.048}$ | .357$_{\pm.026}$ | .407$_{\pm.046}$ | .249$_{\pm.052}$ | .290$_{\pm.022}$ | .515$_{\pm.046}$ | .393$_{\pm.045}$ |
| missForest | <u>.300</u>$_{\pm.022}$ | .389$_{\pm.048}$ | .217$_{\pm.053}$ | .297$_{\pm.023}$ | .421$_{\pm.047}$ | .322$_{\pm.047}$ | **.312**$_{\pm.023}$ | .393$_{\pm.047}$ | .283$_{\pm.049}$ | .281$_{\pm.022}$ | .519$_{\pm.048}$ | .406$_{\pm.045}$ |
| GAIN | .379$_{\pm.025}$ | .445$_{\pm.039}$ | .173$_{\pm.047}$ | .342$_{\pm.024}$ | .499$_{\pm.036}$ | .279$_{\pm.037}$ | .398$_{\pm.026}$ | .447$_{\pm.042}$ | .211$_{\pm.043}$ | .323$_{\pm.023}$ | .567$_{\pm.038}$ | .387$_{\pm.044}$ |
| VAEAC | .301$_{\pm.023}$ | **.631**$_{\pm.030}$ | <u>.345</u>$_{\pm.039}$ | .283$_{\pm.022}$ | <u>.658</u>$_{\pm.029}$ | <u>.401</u>$_{\pm.040}$ | <u>.313</u>$_{\pm.025}$ | <u>.610</u>$_{\pm.030}$ | .341$_{\pm.044}$ | <u>.266</u>$_{\pm.021}$ | <u>.696</u>$_{\pm.029}$ | <u>.480</u>$_{\pm.035}$ |
| MIWAE | .308$_{\pm.024}$ | .180$_{\pm.011}$ | .264$_{\pm.043}$ | .295$_{\pm.022}$ | .203$_{\pm.020}$ | .286$_{\pm.042}$ | .319$_{\pm.024}$ | .209$_{\pm.013}$ | .232$_{\pm.046}$ | .283$_{\pm.023}$ | .383$_{\pm.043}$ | .378$_{\pm.039}$ |
| not-MIWAE | .340$_{\pm.024}$ | .384$_{\pm.048}$ | .249$_{\pm.052}$ | .361$_{\pm.025}$ | .340$_{\pm.044}$ | .293$_{\pm.045}$ | .353$_{\pm.025}$ | .382$_{\pm.048}$ | .277$_{\pm.046}$ | .337$_{\pm.022}$ | .448$_{\pm.048}$ | .335$_{\pm.042}$ |
| MIRACLE | .365$_{\pm.025}$ | .407$_{\pm.046}$ | .265$_{\pm.047}$ | .350$_{\pm.024}$ | .430$_{\pm.045}$ | .324$_{\pm.043}$ | .379$_{\pm.026}$ | .405$_{\pm.046}$ | .239$_{\pm.049}$ | .323$_{\pm.022}$ | .504$_{\pm.046}$ | .346$_{\pm.044}$ |
| ReMasker | .420$_{\pm.028}$ | .381$_{\pm.047}$ | .211$_{\pm.052}$ | .344$_{\pm.024}$ | .419$_{\pm.047}$ | .277$_{\pm.049}$ | .418$_{\pm.028}$ | .393$_{\pm.047}$ | .260$_{\pm.048}$ | .327$_{\pm.024}$ | .509$_{\pm.047}$ | .395$_{\pm.041}$ |
| U-VAE | **.272**$_{\pm.026}$ | <u>.620</u>$_{\pm.050}$ | **.417**$_{\pm.046}$ | <u>.279</u>$_{\pm.025}$ | **.664**$_{\pm.049}$ | **.426**$_{\pm.047}$ | .317$_{\pm.027}$ | **.624**$_{\pm.050}$ | **.392**$_{\pm.048}$ | **.265**$_{\pm.025}$ | **.772**$_{\pm.050}$ | **.491**$_{\pm.044}$ |

Table 7: **Imputation utility** at 0.9 missingness. The means and their standard errors across 11 real-world datasets and 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better. The best value is **red bolded**, and the second best is <u>blue underlined</u>.

**Q1: Imputation utility result** As shown in Tables 4–7, U-VAE consistently ranks among the top methods in downstream utility, achieving the highest or near-highest classification accuracy and feature-selection agreement, while generally excelling in SMAPE across missingness rates, outperforming in 35 out of 48 settings (4 missingness rates × 4 missingness mechanisms × 3 metrics). These results indicate that its superior imputations translate into better downstream task performance, not just improved reconstruction metrics, with only minor exceptions (e.g., a slight SMAPE drop under MNARQ at 0.1 missingness rate).

**Q2: Multiple imputation** We report the Bias, Coverage, and Width results to assess multiple imputation across missingness rates of 0.1, 0.5, 0.7, and 0.9.

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | .005(2.0) | .867$_{\pm .024}$ | $\underline{.038}_{\pm .003}$ | .004(1.2) | .907$_{\pm .018}$ | .037$_{\pm .003}$ | .005(2.0) | .864$_{\pm .025}$ | $\underline{.038}_{\pm .003}$ | .003(1.3) | .915$_{\pm .017}$ | **.037**$_{\pm .003}$ |
| GAIN | .013(3.6) | .658$_{\pm .040}$ | .037$_{\pm .003}$ | .012(2.9) | .699$_{\pm .029}$ | .037$_{\pm .003}$ | .018(4.8) | .550$_{\pm .042}$ | $\underline{.038}_{\pm .003}$ | .008(2.2) | .786$_{\pm .029}$ | $\underline{.038}_{\pm .003}$ |
| VAEAC | .005(2.0) | .871$_{\pm .024}$ | .038$_{\pm .003}$ | .004(1.3) | .908$_{\pm .016}$ | $\underline{.038}_{\pm .003}$ | $\underline{.006}$(2.1) | .864$_{\pm .022}$ | $\underline{.038}_{\pm .003}$ | .004(1.5) | .910$_{\pm .016}$ | $\underline{.038}_{\pm .003}$ |
| MIWAE | $\underline{.005}$(1.5) | .912$_{\pm .016}$ | .041$_{\pm .003}$ | $\underline{.004}$(1.1) | $\underline{.938}_{\pm .013}$ | .040$_{\pm .003}$ | .007(1.9) | $\underline{.889}_{\pm .017}$ | .040$_{\pm .003}$ | .004(1.2) | .926$_{\pm .014}$ | .040$_{\pm .003}$ |
| not-MIWAE | .005(1.6) | **.925**$_{\pm .015}$ | .040$_{\pm .003}$ | .005(1.3) | .924$_{\pm .014}$ | .040$_{\pm .003}$ | .007(1.8) | .887$_{\pm .021}$ | .040$_{\pm .003}$ | $\underline{.003}$(1.0) | .945$_{\pm .012}$ | .039$_{\pm .003}$ |
| U-VAE | **.003**(0.9) | $\underline{.978}_{\pm .006}$ | **.037**$_{\pm .003}$ | **.004**(1.0) | **.963**$_{\pm .016}$ | **.037**$_{\pm .003}$ | **.004**(1.1) | **.977**$_{\pm .006}$ | **.037**$_{\pm .003}$ | **.002**(0.6) | .989$_{\pm .004}$ | **.037**$_{\pm .003}$ |

Table 8: **Multiple imputation performance** at 0.1 missingness rate. The means and their standard errors across 11 datasets and 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance. The best value is **red bolded**, and the second best is blue underlined.

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | .033(9.1) | .443$_{\pm .040}$ | $\underline{.038}_{\pm .003}$ | .022(5.7) | .635$_{\pm .029}$ | $\underline{.038}_{\pm .003}$ | .038(10.2) | .394$_{\pm .040}$ | $\underline{.038}$±.003 | .020(5.8) | .665$_{\pm .030}$ | $\underline{.038}_{\pm .003}$ |
| GAIN | .082(20.3) | .198$_{\pm .033}$ | **.037**$_{\pm .003}$ | .050(13.0) | .464$_{\pm .025}$ | **.037**$_{\pm .003}$ | .091(23.7) | .188$_{\pm .030}$ | **.037**$_{\pm .003}$ | .043(11.9) | .522$_{\pm .029}$ | **.037**$_{\pm .003}$ |
| VAEAC | .033(9.0) | .472$_{\pm .038}$ | .039$_{\pm .003}$ | .022(5.8) | .608$_{\pm .026}$ | $\underline{.038}_{\pm .003}$ | .033(8.9) | .458$_{\pm .035}$ | .039$_{\pm .003}$ | .020(5.7) | .652$_{\pm .029}$ | $\underline{.038}_{\pm .003}$ |
| MIWAE | $\underline{.017}$(5.5) | $\underline{.707}_{\pm .039}$ | .046$_{\pm .004}$ | $\underline{.014}$(3.9) | $\underline{.777}_{\pm .031}$ | .043$_{\pm .003}$ | $\underline{.020}$(5.9) | **.703**$_{\pm .038}$ | .045$_{\pm .004}$ | $\underline{.011}$(3.8) | $\underline{.805}_{\pm .026}$ | .042$_{\pm .004}$ |
| not-MIWAE | .024(6.7) | .572$_{\pm .039}$ | .046$_{\pm .004}$ | .026(6.2) | .626$_{\pm .025}$ | .042$_{\pm .004}$ | .026(7.0) | .523$_{\pm .038}$ | .044$_{\pm .004}$ | .022(5.8) | .656$_{\pm .025}$ | .044$_{\pm .004}$ |
| U-VAE | **.010**(3.0) | **.803**$_{\pm .030}$ | .039$_{\pm .004}$ | **.011**(3.0) | **.798**$_{\pm .024}$ | **.037**$_{\pm .004}$ | **.016**(4.4) | $\underline{.689}_{\pm .034}$ | $\underline{.038}_{\pm .004}$ | **.008**(2.2) | **.857**$_{\pm .021}$ | **.037**$_{\pm .004}$ |

Table 9: **Multiple imputation performance** at 0.5 missingness rate. The means and their standard errors across 11 datasets and 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance. The best value is **red bolded**, and the second best is blue underlined.

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | .068(16.8) | .212$_{\pm .031}$ | $\underline{.038}_{\pm .003}$ | .037(8.9) | .515$_{\pm .026}$ | .038$_{\pm .003}$ | .072(17.5) | .240$_{\pm .033}$ | $\underline{.038}_{\pm .003}$ | .033(8.3) | .567$_{\pm .027}$ | $\underline{.038}_{\pm .003}$ |
| GAIN | .139(43.1) | .135$_{\pm .030}$ | **.036**$_{\pm .003}$ | .072(20.5) | .414$_{\pm .022}$ | $\underline{.037}_{\pm .003}$ | .141(38.4) | .112$_{\pm .023}$ | **.036**$_{\pm .003}$ | .057(17.1) | .498$_{\pm .028}$ | **.037**$_{\pm .003}$ |
| VAEAC | .057(13.8) | .278$_{\pm .033}$ | .040$_{\pm .003}$ | .036(8.5) | .494$_{\pm .022}$ | .039$_{\pm .003}$ | .056(13.7) | .296$_{\pm .032}$ | .040$_{\pm .003}$ | .033(8.2) | .567$_{\pm .025}$ | .039$_{\pm .003}$ |
| MIWAE | $\underline{.023}$(6.6) | $\underline{.657}_{\pm .036}$ | .050$_{\pm .004}$ | $\underline{.018}$(5.5) | $\underline{.740}_{\pm .028}$ | .045$_{\pm .004}$ | $\underline{.029}$(7.9) | $\underline{.564}_{\pm .037}$ | .050$_{\pm .004}$ | $\underline{.016}$(5.5) | $\underline{.761}_{\pm .027}$ | .044$_{\pm .003}$ |
| not-MIWAE | .032(9.2) | .506$_{\pm .038}$ | .048$_{\pm .004}$ | .038(8.8) | .546$_{\pm .020}$ | .046$_{\pm .004}$ | .035(9.6) | .442$_{\pm .033}$ | .049$_{\pm .004}$ | .032(8.3) | .576$_{\pm .022}$ | .043$_{\pm .004}$ |
| U-VAE | **.015**(4.2) | **.729**$_{\pm .038}$ | $\underline{.038}_{\pm .004}$ | **.014**(3.7) | **.748**$_{\pm .026}$ | **.036**$_{\pm .004}$ | **.022**(6.1) | **.592**$_{\pm .034}$ | $\underline{.038}_{\pm .004}$ | **.011**(3.1) | **.802**$_{\pm .022}$ | **.037**$_{\pm .004}$ |

Table 10: **Multiple imputation performance** at 0.7 missingness rate. The means and their standard errors across 11 datasets and 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance. The best value is **red bolded**, and the second best is blue underlined.

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | .237(123.8) | .048$_{\pm .014}$ | $\underline{.034}_{\pm .003}$ | .095(44.6) | .396$_{\pm .021}$ | $\underline{.037}_{\pm .003}$ | .227(113.5) | .049$_{\pm .013}$ | $\underline{.034}_{\pm .003}$ | .073(32.2) | .483$_{\pm .024}$ | $\underline{.037}_{\pm .003}$ |
| GAIN | .201(81.8) | .058$_{\pm .012}$ | **.033**$_{\pm .003}$ | .099(38.0) | .392$_{\pm .019}$ | **.036**$_{\pm .003}$ | .206(72.2) | .066$_{\pm .019}$ | **.033**$_{\pm .003}$ | .078(30.9) | .419$_{\pm .019}$ | **.036**$_{\pm .003}$ |
| VAEAC | .115(25.4) | .133$_{\pm .023}$ | .043$_{\pm .004}$ | .054(11.8) | .431$_{\pm .021}$ | .039$_{\pm .003}$ | .102(23.7) | .152$_{\pm .021}$ | .041$_{\pm .004}$ | .048(10.9) | .491$_{\pm .020}$ | .039$_{\pm .003}$ |
| MIWAE | $\underline{.034}$(9.6) | $\underline{.512}_{\pm .040}$ | .050$_{\pm .004}$ | $\underline{.027}$(7.2) | $\underline{.617}_{\pm .025}$ | .048$_{\pm .004}$ | $\underline{.044}$(11.7) | $\underline{.411}_{\pm .037}$ | .055$_{\pm .004}$ | $\underline{.024}$(7.2) | $\underline{.649}_{\pm .027}$ | .048$_{\pm .004}$ |
| not-MIWAE | .056(14.6) | .308$_{\pm .026}$ | .050$_{\pm .004}$ | .047(11.0) | .466$_{\pm .018}$ | .045$_{\pm .004}$ | .058(15.0) | .261$_{\pm .020}$ | .049$_{\pm .004}$ | .047(11.1) | .503$_{\pm .020}$ | .047$_{\pm .004}$ |
| U-VAE | **.028**(7.4) | **.540**$_{\pm .038}$ | .040$_{\pm .004}$ | **.022**(5.8) | **.637**$_{\pm .027}$ | .040$_{\pm .004}$ | **.034**(9.2) | **.436**$_{\pm .036}$ | .041$_{\pm .004}$ | **.016**(4.3) | **.697**$_{\pm .027}$ | .039$_{\pm .004}$ |

Table 11: **Multiple imputation performance** at 0.9 missingness rate. The means and their standard errors across 11 datasets and 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance. The best value is **red bolded**, and the second best is blue underlined.

**Q2 result** Multiple-imputation quality is evaluated via Bias (and percent Bias), Confidence Interval Coverage (target $\approx 0.95$), and CI Width. As shown in Table 8-11, U-VAE achieves low Bias, near-target Coverage, and avoids unnecessary CI inflation, outperforming in 35 out of 48 settings (4 missingness rates $\times$ 4 missingness mechanisms $\times$ 3 metrics). Even at extreme missingness rate 0.9, it balances accuracy (Bias), reliability (Coverage), and efficiency (Width) better than others, where *many baselines show rising Bias and collapsing Coverage.*

**Q3: Sensitivity Analysis – Imputation fidelity** We report sensitivity analyses using additional evaluation metrics to assess imputation fidelity.

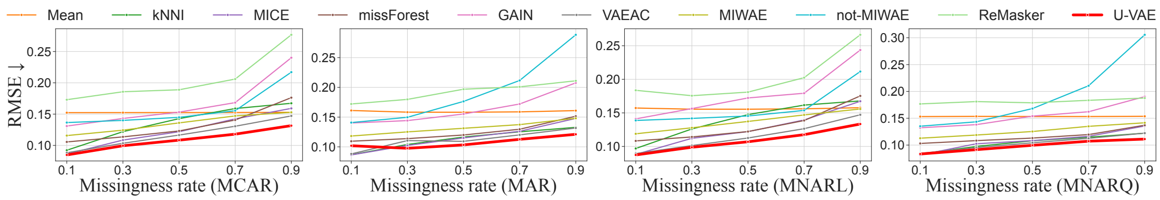


Figure 9: **Sensitivity analysis for missingness rates.** Imputation performance comparison of various imputation models under different missing data mechanisms (`MCAR`, `MAR`, `MNARL`, `MNARQ`) and missingness rates (0.1, 0.3, 0.5, 0.7, 0.9) using Bias. ↓ denotes that lower is better

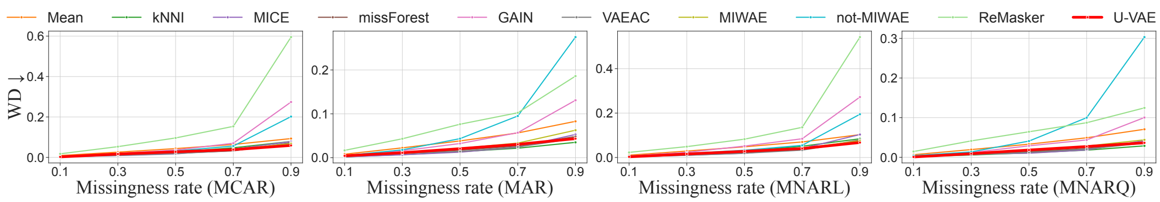


Figure 10: **Sensitivity analysis for missingness rates.** Imputation performance comparison of various imputation models under different missing data mechanisms (`MCAR`, `MAR`, `MNARL`, `MNARQ`) and missingness rates (0.1, 0.3, 0.5, 0.7, 0.9) using WD. ↓ denotes that lower is better.

**Q3: Sensitivity Analysis – Imputation utility** We report sensitivity analyses using additional evaluation metrics to assess imputation utility.

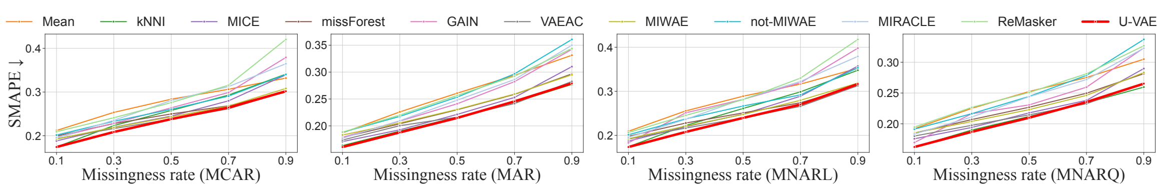


Figure 11: **Sensitivity analysis for missingness rates.** Imputation performance comparison of various imputation models under different missing data mechanisms (`MCAR`, `MAR`, `MNARL`, `MNARQ`) and missingness rates (0.1, 0.3, 0.5, 0.7, 0.9) using SMAPE (regression). ↓ denotes that lower is better.

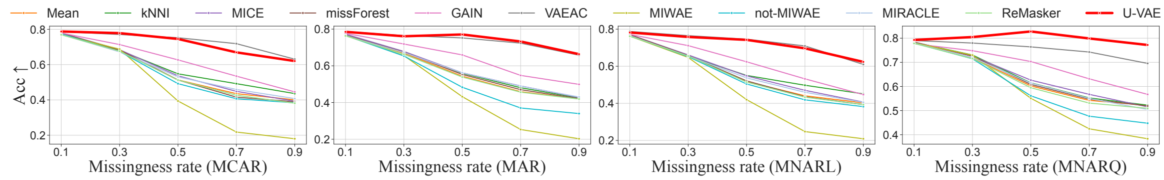


Figure 12: **Sensitivity analysis for missingness rates.** Imputation performance comparison of various imputation models under different missing data mechanisms (`MCAR`, `MAR`, `MNARL`, `MNARQ`) and missingness rates (0.1, 0.3, 0.5, 0.7, 0.9) using Accuracy (classification). ↑ denotes that higher is better.

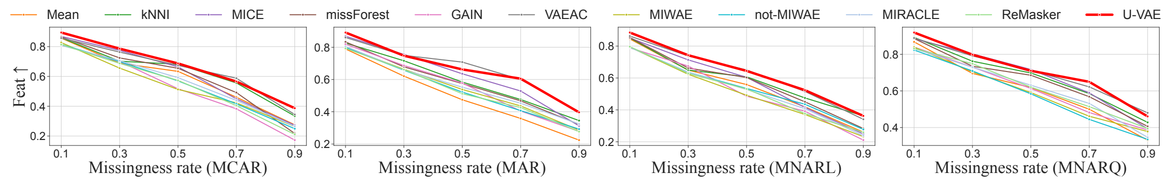


Figure 13: **Sensitivity analysis for missingness rates.** Imputation performance comparison of various imputation models under different missing data mechanisms (`MCAR`, `MAR`, `MNARL`, `MNARQ`) and missingness rates (0.1, 0.3, 0.5, 0.7, 0.9) using Feature selection. ↑ denotes that higher is better.

**Q3: Sensitivity Analysis result – Imputation fidelity & utility** As shown in Figure 9-13, U-VAE's lines remain consistently at or near the top across all metrics, with performance degrading only gradually as missingness increases. This flatness reflects strong robustness, indicating that U-VAE maintains high-quality imputations even in more challenging missingness scenarios.

**Q3: Sensitivity Analysis – Multiple imputation** We report sensitivity analyses using additional evaluation metrics to assess multiple imputation.

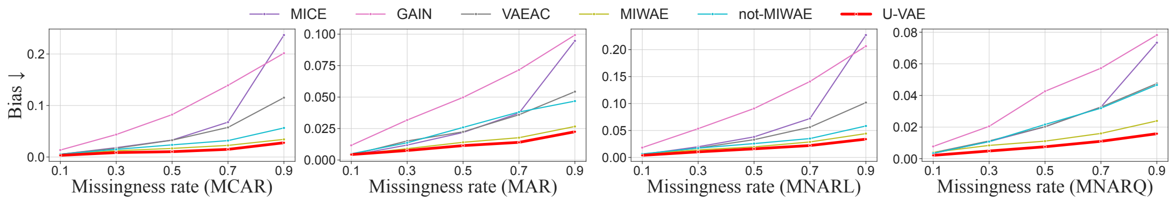


Figure 14: **Sensitivity analysis for missingness rates.** Imputation performance comparison of various imputation models under different missing data mechanisms (`MCAR`, `MAR`, `MNARL`, `MNARQ`) and missingness rates (0.1, 0.3, 0.5, 0.7, 0.9) using Width. ↓ denotes that lower is better.

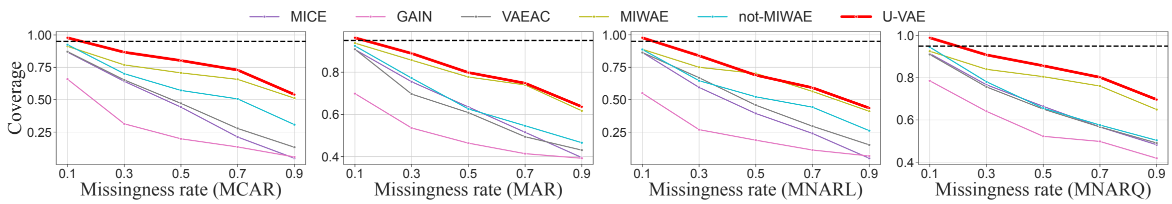


Figure 15: **Sensitivity analysis for missingness rates.** Imputation performance comparison of various imputation models under different missing data mechanisms (`MCAR`, `MAR`, `MNARL`, `MNARQ`) and missingness rates (0.1, 0.3, 0.5, 0.7, 0.9) using Coverage. Coverage close to 0.95 indicates better performance.

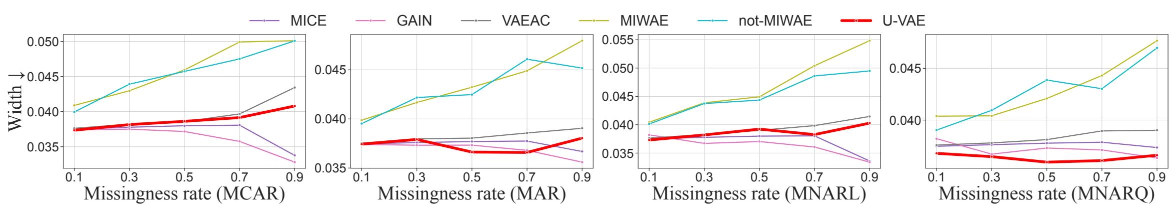


Figure 16: **Sensitivity analysis for missingness rates.** Imputation performance comparison of various imputation models under different missing data mechanisms (`MCAR`, `MAR`, `MNARL`, `MNARQ`) and missingness rates (0.1, 0.3, 0.5, 0.7, 0.9) using Width. ↓ denotes that lower is better.

**Q3: Sensitivity Analysis result – Multiple imputation** As shown in Figure 15 and 16, U-VAE maintains low Bias while keeping Coverage high without over-widening intervals. In other words, it achieves a good inference trade-off: accurate, trustworthy, and efficient uncertainty.

## E.2 Across each dataset

**Q1: Imputation fidelity (RMSE)** We report RMSE results to assess imputation fidelity for each dataset at 0.3 missingness rate.

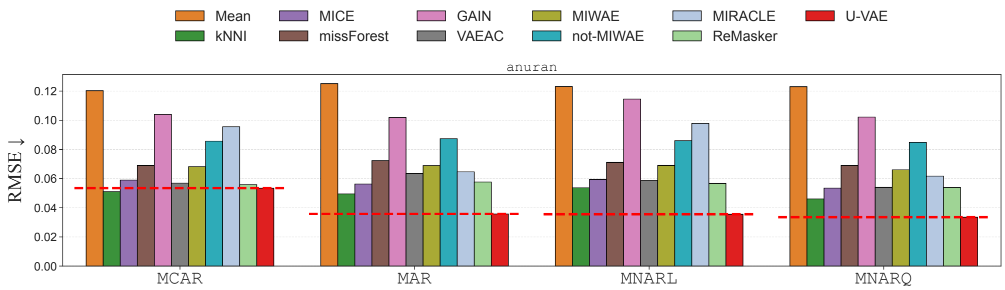


Figure 17: **Imputation fidelity (RMSE)** on the `anuran` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

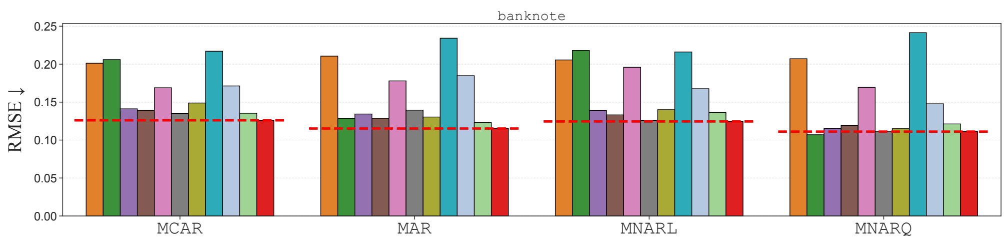


Figure 18: **Imputation fidelity (RMSE)** on the `banknote` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

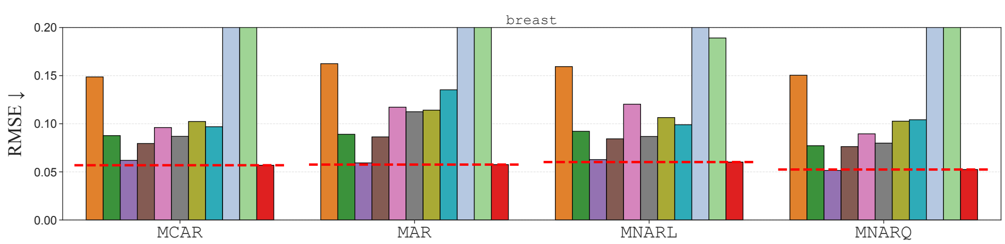


Figure 19: **Imputation fidelity (RMSE)** on the `breast` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

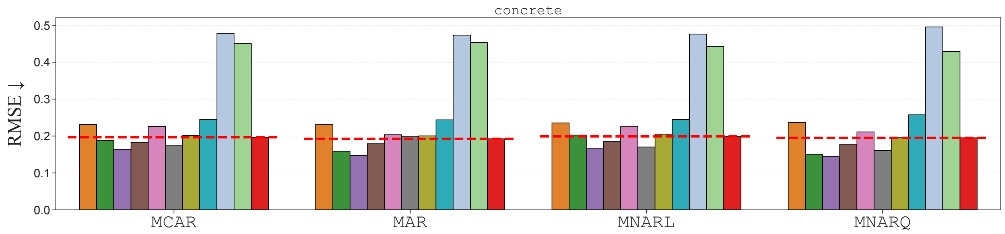


Figure 20: **Imputation fidelity (RMSE)** on the `concrete` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

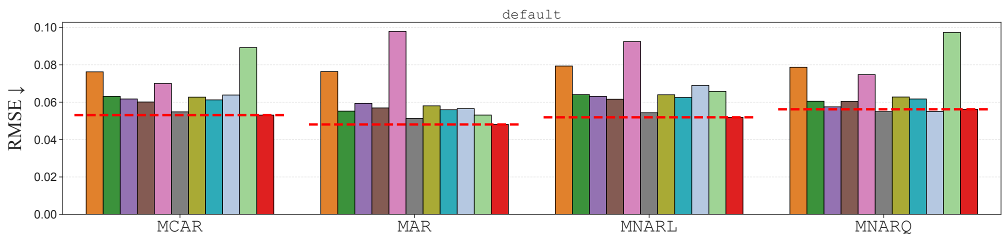


Figure 21: **Imputation fidelity (RMSE)** on the `default` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

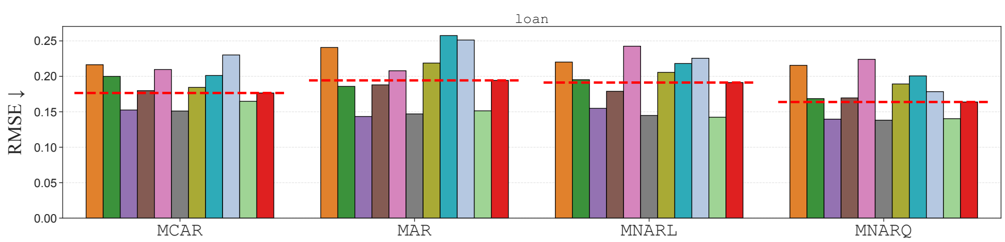


Figure 22: **Imputation fidelity (RMSE)** on the `loan` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

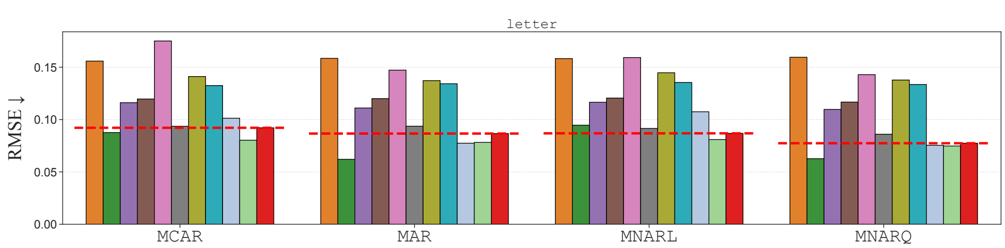


Figure 23: **Imputation fidelity (RMSE)** on the `letter` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

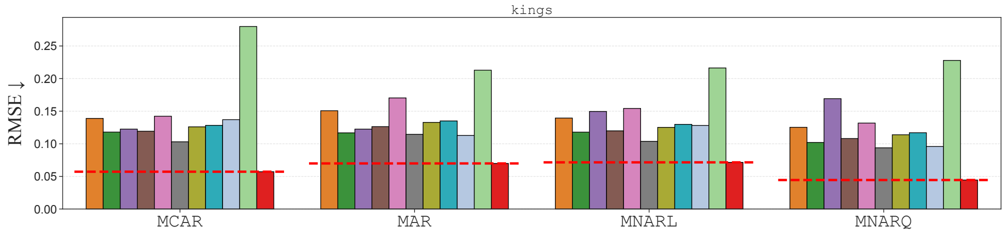


Figure 24: **Imputation fidelity (RMSE)** on the `kings` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

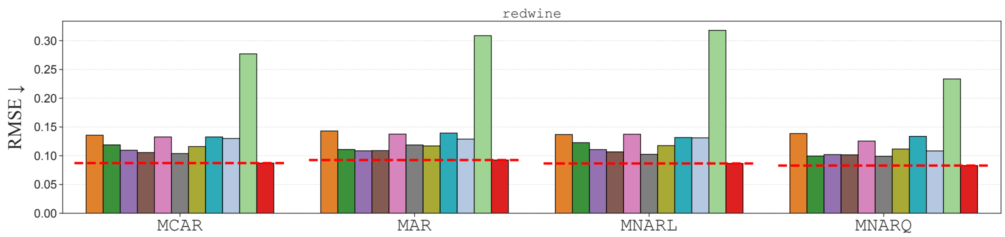


Figure 25: **Imputation fidelity (RMSE)** on the `redwine` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

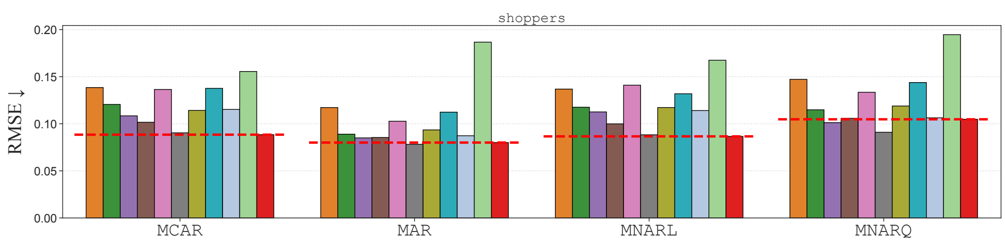


Figure 26: **Imputation fidelity (RMSE)** on the `shoppers` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

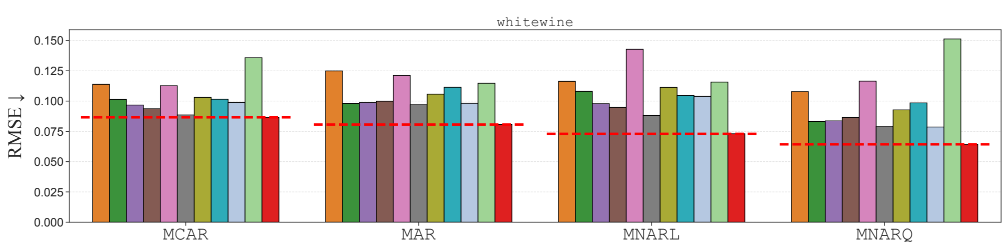


Figure 27: **Imputation fidelity (RMSE)** on the `whitewine` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

**Q1: Imputation fidelity (WD)** We report WD results to assess imputation fidelity for each dataset at 0.3 missingness rate.

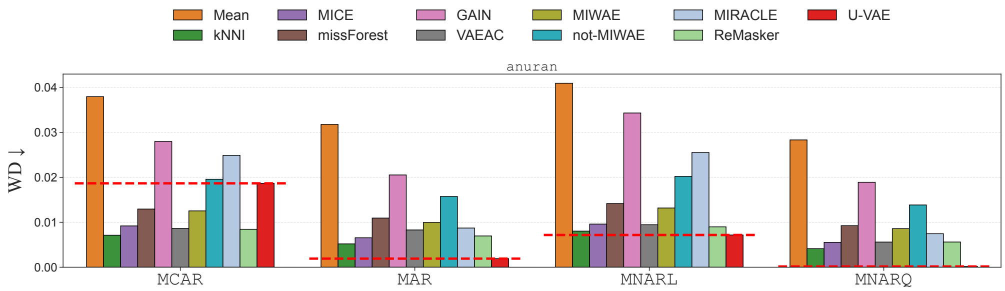


Figure 28: **Imputation fidelity (WD)** on the `anuran` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

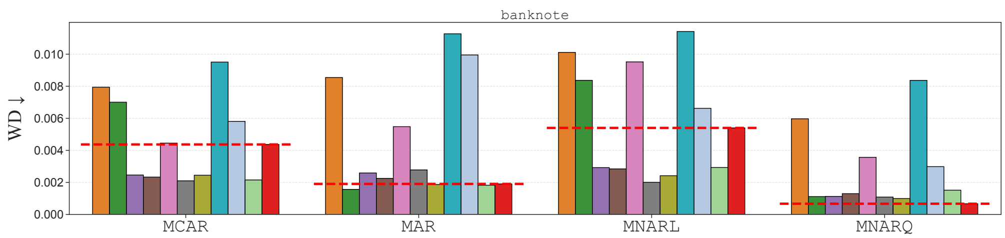


Figure 29: **Imputation fidelity (WD)** on the `banknote` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

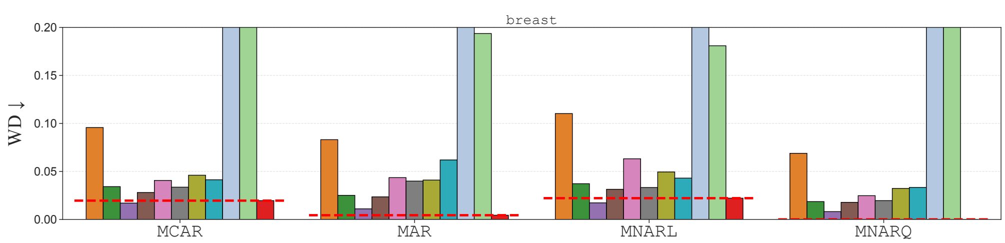


Figure 30: **Imputation fidelity (WD)** on the `breast` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

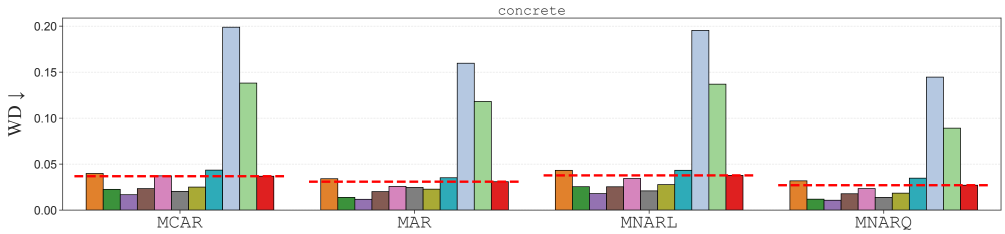


Figure 31: **Imputation fidelity (WD)** on the `concrete` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

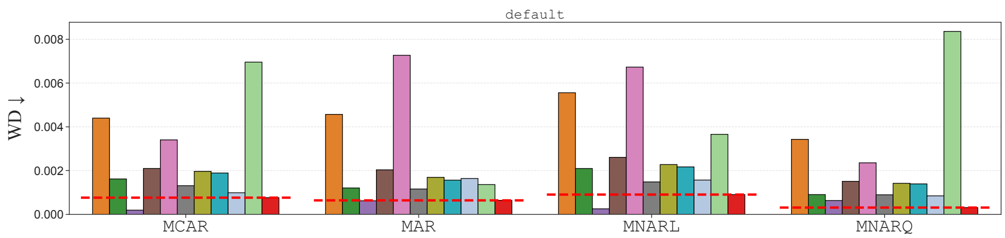


Figure 32: **Imputation fidelity (WD)** on the `default` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

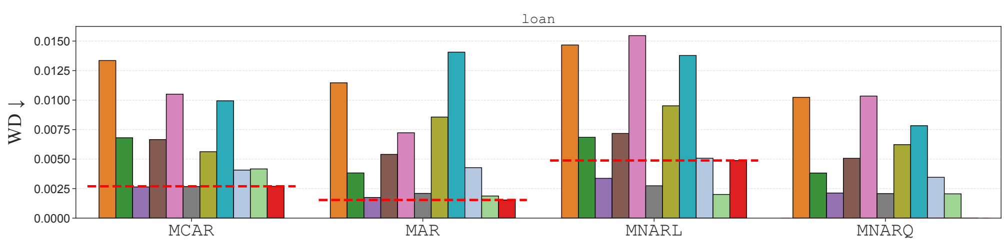


Figure 33: **Imputation fidelity (WD)** on the `loan` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

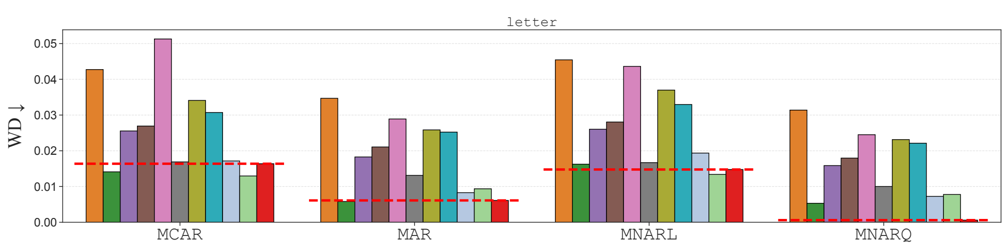


Figure 34: **Imputation fidelity (WD)** on the `letter` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

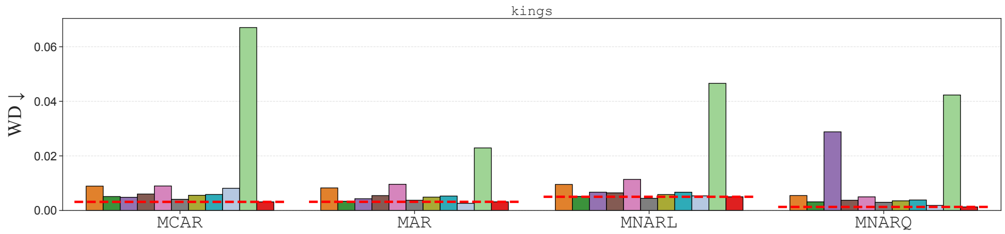


Figure 35: **Imputation fidelity (WD)** on the `kings` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

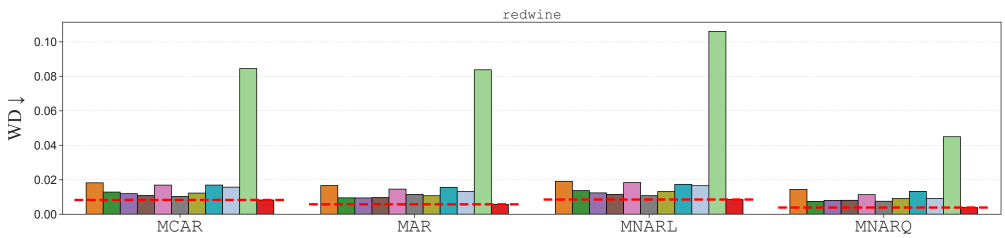


Figure 36: **Imputation fidelity (WD)** on the `redwine` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

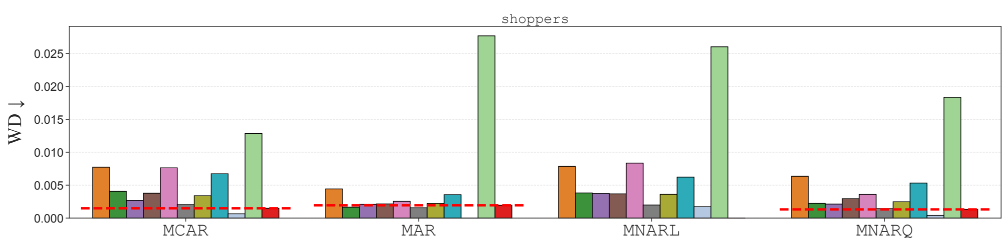


Figure 37: **Imputation fidelity (WD)** on the `shoppers` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

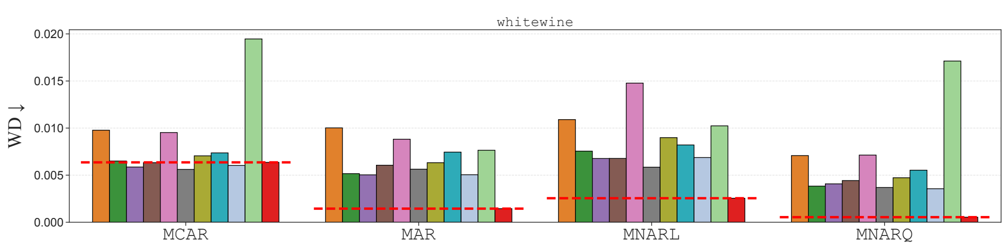


Figure 38: **Imputation fidelity (WD)** on the `whitewine` dataset at 0.3 missingness rate. The corresponding missingness mechanism is indicated below the figure. The red dashed line indicates the performance of U-VAE under each missingness mechanism. The means and their standard errors across 5 random seeds are reported. ↓ denotes that lower is better.

**Q1: Imputation utility** We report SMAPE (regression), Accuracy (Acc, classification), feature selection performance (Feat) results to assess imputation utility for each dataset at 0.3 missingness rate.

anuran

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.273_{\pm0.002}$ | $0.879_{\pm0.020}$ | $0.464_{\pm0.024}$ | $0.259_{\pm0.001}$ | $0.779_{\pm0.027}$ | $0.355_{\pm0.034}$ | $0.280_{\pm0.001}$ | $0.805_{\pm0.025}$ | $0.389_{\pm0.038}$ | $0.252_{\pm0.002}$ | $0.918_{\pm0.025}$ | $0.464_{\pm0.065}$ |
| kNNI | $0.208_{\pm0.001}$ | $0.850_{\pm0.018}$ | $0.376_{\pm0.027}$ | $0.204_{\pm0.001}$ | $0.765_{\pm0.025}$ | $0.447_{\pm0.024}$ | $0.209_{\pm0.001}$ | $0.769_{\pm0.015}$ | $0.340_{\pm0.068}$ | $0.202_{\pm0.002}$ | $0.940_{\pm0.027}$ | $0.495_{\pm0.050}$ |
| MICE | $0.201_{\pm0.001}$ | $0.872_{\pm0.012}$ | $0.424_{\pm0.018}$ | $0.199_{\pm0.001}$ | $0.739_{\pm0.026}$ | $0.497_{\pm0.031}$ | $0.203_{\pm0.001}$ | $0.787_{\pm0.017}$ | $0.403_{\pm0.040}$ | $0.196_{\pm0.001}$ | $0.909_{\pm0.039}$ | $0.595_{\pm0.059}$ |
| missForest | $0.224_{\pm0.001}$ | $0.849_{\pm0.012}$ | $0.411_{\pm0.019}$ | $0.222_{\pm0.001}$ | $0.782_{\pm0.026}$ | $0.398_{\pm0.025}$ | $0.230_{\pm0.001}$ | $0.799_{\pm0.019}$ | $0.321_{\pm0.053}$ | $0.217_{\pm0.001}$ | $0.937_{\pm0.024}$ | $0.523_{\pm0.066}$ |
| gain | $0.242_{\pm0.002}$ | $0.904_{\pm0.019}$ | $0.453_{\pm0.047}$ | $0.231_{\pm0.001}$ | $0.854_{\pm0.026}$ | $0.475_{\pm0.020}$ | $0.258_{\pm0.006}$ | $0.877_{\pm0.030}$ | $0.429_{\pm0.021}$ | $0.224_{\pm0.002}$ | $0.941_{\pm0.012}$ | $0.538_{\pm0.024}$ |
| VAEAC | $0.212_{\pm0.001}$ | $0.935_{\pm0.019}$ | $0.722_{\pm0.034}$ | $0.213_{\pm0.001}$ | $0.933_{\pm0.007}$ | $0.706_{\pm0.041}$ | $0.214_{\pm0.001}$ | $0.928_{\pm0.014}$ | $0.692_{\pm0.023}$ | $0.206_{\pm0.001}$ | $0.951_{\pm0.007}$ | $0.750_{\pm0.034}$ |
| MIWAE | $0.220_{\pm0.001}$ | $0.885_{\pm0.005}$ | $0.361_{\pm0.032}$ | $0.216_{\pm0.001}$ | $0.787_{\pm0.027}$ | $0.457_{\pm0.015}$ | $0.224_{\pm0.001}$ | $0.793_{\pm0.012}$ | $0.360_{\pm0.022}$ | $0.213_{\pm0.001}$ | $0.922_{\pm0.023}$ | $0.462_{\pm0.082}$ |
| not-MIWAE | $0.231_{\pm0.001}$ | $0.892_{\pm0.014}$ | $0.423_{\pm0.032}$ | $0.227_{\pm0.001}$ | $0.769_{\pm0.026}$ | $0.400_{\pm0.014}$ | $0.239_{\pm0.001}$ | $0.817_{\pm0.017}$ | $0.348_{\pm0.043}$ | $0.221_{\pm0.002}$ | $0.927_{\pm0.030}$ | $0.533_{\pm0.056}$ |
| MIRACLE | $0.209_{\pm0.003}$ | $0.834_{\pm0.014}$ | $0.403_{\pm0.030}$ | $0.201_{\pm0.002}$ | $0.755_{\pm0.026}$ | $0.432_{\pm0.023}$ | $0.213_{\pm0.004}$ | $0.763_{\pm0.011}$ | $0.354_{\pm0.039}$ | $0.198_{\pm0.003}$ | $0.927_{\pm0.031}$ | $0.554_{\pm0.059}$ |
| ReMasker | $0.217_{\pm0.002}$ | $0.859_{\pm0.007}$ | $0.334_{\pm0.020}$ | $0.214_{\pm0.001}$ | $0.765_{\pm0.028}$ | $0.409_{\pm0.021}$ | $0.219_{\pm0.002}$ | $0.773_{\pm0.021}$ | $0.332_{\pm0.042}$ | $0.211_{\pm0.002}$ | $0.925_{\pm0.031}$ | $0.495_{\pm0.082}$ |
| U-VAE | $0.313_{\pm0.007}$ | $0.950_{\pm0.010}$ | $0.894_{\pm0.040}$ | $0.260_{\pm0.014}$ | $0.931_{\pm0.022}$ | $0.693_{\pm0.085}$ | $0.300_{\pm0.010}$ | $0.929_{\pm0.017}$ | $0.933_{\pm0.059}$ | $0.232_{\pm0.013}$ | $0.990_{\pm0.008}$ | $0.733_{\pm0.076}$ |

Table 12: **Imputation utility** on the `anuran` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

banknote

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.378_{\pm0.006}$ | $0.910_{\pm0.013}$ | $0.863_{\pm0.058}$ | $0.353_{\pm0.004}$ | $0.899_{\pm0.021}$ | $0.644_{\pm0.119}$ | $0.388_{\pm0.005}$ | $0.886_{\pm0.023}$ | $0.717_{\pm0.100}$ | $0.322_{\pm0.006}$ | $0.946_{\pm0.013}$ | $0.577_{\pm0.141}$ |
| kNNI | $0.399_{\pm0.010}$ | $0.896_{\pm0.016}$ | $0.653_{\pm0.085}$ | $0.278_{\pm0.009}$ | $0.911_{\pm0.024}$ | $0.758_{\pm0.115}$ | $0.387_{\pm0.008}$ | $0.842_{\pm0.013}$ | $0.580_{\pm0.061}$ | $0.263_{\pm0.008}$ | $0.957_{\pm0.008}$ | $0.909_{\pm0.033}$ |
| MICE | $0.312_{\pm0.006}$ | $0.962_{\pm0.006}$ | $0.960_{\pm0.016}$ | $0.283_{\pm0.005}$ | $0.923_{\pm0.021}$ | $0.907_{\pm0.045}$ | $0.297_{\pm0.009}$ | $0.928_{\pm0.023}$ | $0.893_{\pm0.045}$ | $0.265_{\pm0.010}$ | $0.971_{\pm0.009}$ | $1.000_{\pm0.000}$ |
| missForest | $0.318_{\pm0.007}$ | $0.943_{\pm0.008}$ | $0.914_{\pm0.038}$ | $0.290_{\pm0.005}$ | $0.921_{\pm0.021}$ | $0.775_{\pm0.045}$ | $0.309_{\pm0.004}$ | $0.917_{\pm0.022}$ | $0.743_{\pm0.053}$ | $0.274_{\pm0.005}$ | $0.964_{\pm0.010}$ | $0.760_{\pm0.111}$ |
| gain | $0.339_{\pm0.008}$ | $0.942_{\pm0.013}$ | $0.867_{\pm0.030}$ | $0.315_{\pm0.006}$ | $0.963_{\pm0.011}$ | $0.751_{\pm0.061}$ | $0.367_{\pm0.018}$ | $0.929_{\pm0.008}$ | $0.804_{\pm0.046}$ | $0.288_{\pm0.004}$ | $0.977_{\pm0.005}$ | $0.861_{\pm0.045}$ |
| VAEAC | $0.303_{\pm0.009}$ | $0.979_{\pm0.004}$ | $0.945_{\pm0.022}$ | $0.295_{\pm0.006}$ | $0.966_{\pm0.007}$ | $0.880_{\pm0.039}$ | $0.299_{\pm0.007}$ | $0.969_{\pm0.005}$ | $0.917_{\pm0.037}$ | $0.261_{\pm0.005}$ | $0.984_{\pm0.005}$ | $0.947_{\pm0.039}$ |
| MIWAE | $0.322_{\pm0.011}$ | $0.864_{\pm0.013}$ | $0.748_{\pm0.057}$ | $0.293_{\pm0.010}$ | $0.800_{\pm0.020}$ | $0.806_{\pm0.075}$ | $0.317_{\pm0.005}$ | $0.810_{\pm0.010}$ | $0.695_{\pm0.075}$ | $0.272_{\pm0.003}$ | $0.929_{\pm0.026}$ | $0.855_{\pm0.083}$ |
| not-MIWAE | $0.404_{\pm0.007}$ | $0.884_{\pm0.005}$ | $0.858_{\pm0.026}$ | $0.367_{\pm0.010}$ | $0.887_{\pm0.025}$ | $0.714_{\pm0.059}$ | $0.410_{\pm0.011}$ | $0.888_{\pm0.019}$ | $0.714_{\pm0.049}$ | $0.339_{\pm0.011}$ | $0.939_{\pm0.017}$ | $0.617_{\pm0.101}$ |
| MIRACLE | $0.350_{\pm0.023}$ | $0.954_{\pm0.011}$ | $0.871_{\pm0.045}$ | $0.322_{\pm0.031}$ | $0.926_{\pm0.019}$ | $0.748_{\pm0.134}$ | $0.338_{\pm0.027}$ | $0.925_{\pm0.021}$ | $0.762_{\pm0.076}$ | $0.288_{\pm0.014}$ | $0.970_{\pm0.008}$ | $0.903_{\pm0.044}$ |
| ReMasker | $0.302_{\pm0.008}$ | $0.956_{\pm0.014}$ | $0.938_{\pm0.043}$ | $0.290_{\pm0.003}$ | $0.926_{\pm0.020}$ | $0.793_{\pm0.060}$ | $0.308_{\pm0.009}$ | $0.925_{\pm0.018}$ | $0.747_{\pm0.086}$ | $0.272_{\pm0.005}$ | $0.967_{\pm0.010}$ | $0.890_{\pm0.090}$ |
| U-VAE | $0.313_{\pm0.007}$ | $0.950_{\pm0.010}$ | $0.894_{\pm0.040}$ | $0.260_{\pm0.014}$ | $0.931_{\pm0.022}$ | $0.693_{\pm0.085}$ | $0.300_{\pm0.010}$ | $0.929_{\pm0.017}$ | $0.933_{\pm0.059}$ | $0.232_{\pm0.013}$ | $0.990_{\pm0.008}$ | $0.733_{\pm0.076}$ |

Table 13: **Imputation utility** on the `banknote` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

breast

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.089_{\pm0.001}$ | $0.915_{\pm0.008}$ | $0.400_{\pm0.051}$ | $0.084_{\pm0.002}$ | $0.917_{\pm0.022}$ | $0.446_{\pm0.047}$ | $0.092_{\pm0.001}$ | $0.922_{\pm0.014}$ | $0.510_{\pm0.044}$ | $0.083_{\pm0.001}$ | $0.932_{\pm0.012}$ | $0.460_{\pm0.085}$ |
| kNNI | $0.061_{\pm0.001}$ | $0.891_{\pm0.006}$ | $0.496_{\pm0.052}$ | $0.060_{\pm0.001}$ | $0.899_{\pm0.036}$ | $0.507_{\pm0.058}$ | $0.062_{\pm0.001}$ | $0.868_{\pm0.038}$ | $0.454_{\pm0.060}$ | $0.059_{\pm0.001}$ | $0.946_{\pm0.008}$ | $0.638_{\pm0.049}$ |
| MICE | $0.056_{\pm0.001}$ | $0.908_{\pm0.006}$ | $0.529_{\pm0.050}$ | $0.054_{\pm0.001}$ | $0.911_{\pm0.036}$ | $0.545_{\pm0.038}$ | $0.056_{\pm0.001}$ | $0.898_{\pm0.035}$ | $0.429_{\pm0.031}$ | $0.054_{\pm0.001}$ | $0.942_{\pm0.008}$ | $0.675_{\pm0.062}$ |
| GAIN | $0.062_{\pm0.002}$ | $0.923_{\pm0.007}$ | $0.485_{\pm0.015}$ | $0.062_{\pm0.002}$ | $0.932_{\pm0.012}$ | $0.565_{\pm0.041}$ | $0.065_{\pm0.002}$ | $0.920_{\pm0.021}$ | $0.522_{\pm0.029}$ | $0.059_{\pm0.001}$ | $0.943_{\pm0.007}$ | $0.631_{\pm0.060}$ |
| VAEAC | $0.064_{\pm0.001}$ | $0.952_{\pm0.003}$ | $0.552_{\pm0.043}$ | $0.068_{\pm0.001}$ | $0.950_{\pm0.003}$ | $0.584_{\pm0.040}$ | $0.064_{\pm0.001}$ | $0.949_{\pm0.004}$ | $0.646_{\pm0.030}$ | $0.061_{\pm0.001}$ | $0.955_{\pm0.002}$ | $0.589_{\pm0.033}$ |
| MIWAE | $0.067_{\pm0.001}$ | $0.818_{\pm0.011}$ | $0.474_{\pm0.017}$ | $0.065_{\pm0.001}$ | $0.798_{\pm0.040}$ | $0.508_{\pm0.064}$ | $0.067_{\pm0.001}$ | $0.769_{\pm0.014}$ | $0.468_{\pm0.026}$ | $0.063_{\pm0.001}$ | $0.904_{\pm0.030}$ | $0.615_{\pm0.030}$ |
| not-MIWAE | $0.066_{\pm0.001}$ | $0.899_{\pm0.010}$ | $0.523_{\pm0.047}$ | $0.066_{\pm0.001}$ | $0.919_{\pm0.019}$ | $0.549_{\pm0.056}$ | $0.065_{\pm0.001}$ | $0.891_{\pm0.033}$ | $0.443_{\pm0.055}$ | $0.063_{\pm0.001}$ | $0.935_{\pm0.010}$ | $0.591_{\pm0.063}$ |
| MIRACLE | $0.220_{\pm0.042}$ | $0.828_{\pm0.030}$ | $0.447_{\pm0.032}$ | $0.205_{\pm0.038}$ | $0.892_{\pm0.028}$ | $0.393_{\pm0.048}$ | $0.226_{\pm0.046}$ | $0.857_{\pm0.022}$ | $0.431_{\pm0.032}$ | $0.211_{\pm0.037}$ | $0.910_{\pm0.009}$ | $0.382_{\pm0.075}$ |
| ReMasker | $0.090_{\pm0.008}$ | $0.921_{\pm0.009}$ | $0.366_{\pm0.045}$ | $0.088_{\pm0.010}$ | $0.885_{\pm0.035}$ | $0.442_{\pm0.047}$ | $0.090_{\pm0.010}$ | $0.928_{\pm0.007}$ | $0.423_{\pm0.039}$ | $0.095_{\pm0.005}$ | $0.915_{\pm0.014}$ | $0.449_{\pm0.068}$ |
| U-VAE | $0.060_{\pm0.001}$ | $0.958_{\pm0.005}$ | $0.550_{\pm0.037}$ | $0.057_{\pm0.001}$ | $0.946_{\pm0.035}$ | $0.623_{\pm0.047}$ | $0.061_{\pm0.001}$ | $0.950_{\pm0.023}$ | $0.576_{\pm0.013}$ | $0.057_{\pm0.001}$ | $0.977_{\pm0.006}$ | $0.643_{\pm0.068}$ |

Table 14: **Imputation utility** on the `breast` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

`concrete`

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.233_{\pm0.003}$ | $0.305_{\pm0.015}$ | $0.599_{\pm0.044}$ | $0.201_{\pm0.006}$ | $0.332_{\pm0.030}$ | $0.445_{\pm0.074}$ | $0.235_{\pm0.003}$ | $0.281_{\pm0.020}$ | $0.452_{\pm0.096}$ | $0.217_{\pm0.005}$ | $0.369_{\pm0.049}$ | $0.724_{\pm0.043}$ |
| kNNI | $0.215_{\pm0.004}$ | $0.322_{\pm0.010}$ | $0.786_{\pm0.064}$ | $0.158_{\pm0.002}$ | $0.372_{\pm0.038}$ | $0.719_{\pm0.053}$ | $0.215_{\pm0.004}$ | $0.316_{\pm0.014}$ | $0.467_{\pm0.071}$ | $0.174_{\pm0.005}$ | $0.418_{\pm0.044}$ | $0.788_{\pm0.040}$ |
| MICE | $0.214_{\pm0.003}$ | $0.346_{\pm0.013}$ | $0.820_{\pm0.009}$ | $0.165_{\pm0.005}$ | $0.388_{\pm0.040}$ | $0.730_{\pm0.047}$ | $0.213_{\pm0.003}$ | $0.337_{\pm0.015}$ | $0.643_{\pm0.092}$ | $0.196_{\pm0.006}$ | $0.416_{\pm0.048}$ | $0.743_{\pm0.048}$ |
| missForest | $0.225_{\pm0.003}$ | $0.302_{\pm0.012}$ | $0.722_{\pm0.055}$ | $0.178_{\pm0.005}$ | $0.355_{\pm0.033}$ | $0.545_{\pm0.060}$ | $0.225_{\pm0.004}$ | $0.297_{\pm0.019}$ | $0.573_{\pm0.052}$ | $0.208_{\pm0.005}$ | $0.377_{\pm0.052}$ | $0.706_{\pm0.050}$ |
| GAIN | $0.230_{\pm0.003}$ | $0.423_{\pm0.017}$ | $0.724_{\pm0.044}$ | $0.191_{\pm0.008}$ | $0.443_{\pm0.021}$ | $0.736_{\pm0.041}$ | $0.227_{\pm0.003}$ | $0.427_{\pm0.016}$ | $0.643_{\pm0.062}$ | $0.210_{\pm0.004}$ | $0.462_{\pm0.024}$ | $0.657_{\pm0.075}$ |
| VAEAC | $0.216_{\pm0.003}$ | $0.454_{\pm0.010}$ | $0.691_{\pm0.042}$ | $0.193_{\pm0.005}$ | $0.428_{\pm0.011}$ | $0.587_{\pm0.019}$ | $0.217_{\pm0.004}$ | $0.434_{\pm0.011}$ | $0.595_{\pm0.053}$ | $0.201_{\pm0.006}$ | $0.495_{\pm0.013}$ | $0.693_{\pm0.077}$ |
| MIWAE | $0.226_{\pm0.003}$ | $0.313_{\pm0.013}$ | $0.649_{\pm0.061}$ | $0.211_{\pm0.004}$ | $0.367_{\pm0.031}$ | $0.556_{\pm0.089}$ | $0.228_{\pm0.003}$ | $0.320_{\pm0.015}$ | $0.422_{\pm0.085}$ | $0.215_{\pm0.004}$ | $0.397_{\pm0.035}$ | $0.746_{\pm0.054}$ |
| not-MIWAE | $0.235_{\pm0.003}$ | $0.272_{\pm0.020}$ | $0.617_{\pm0.049}$ | $0.225_{\pm0.002}$ | $0.328_{\pm0.037}$ | $0.502_{\pm0.037}$ | $0.238_{\pm0.004}$ | $0.275_{\pm0.009}$ | $0.422_{\pm0.087}$ | $0.224_{\pm0.004}$ | $0.355_{\pm0.051}$ | $0.625_{\pm0.086}$ |
| MIRACLE | $0.256_{\pm0.004}$ | $0.271_{\pm0.012}$ | $0.701_{\pm0.078}$ | $0.222_{\pm0.013}$ | $0.317_{\pm0.034}$ | $0.533_{\pm0.072}$ | $0.258_{\pm0.005}$ | $0.259_{\pm0.020}$ | $0.419_{\pm0.082}$ | $0.235_{\pm0.002}$ | $0.350_{\pm0.043}$ | $0.646_{\pm0.073}$ |
| ReMasker | $0.279_{\pm0.004}$ | $0.260_{\pm0.027}$ | $0.578_{\pm0.073}$ | $0.251_{\pm0.003}$ | $0.312_{\pm0.042}$ | $0.445_{\pm0.052}$ | $0.281_{\pm0.003}$ | $0.270_{\pm0.024}$ | $0.568_{\pm0.087}$ | $0.249_{\pm0.005}$ | $0.341_{\pm0.051}$ | $0.717_{\pm0.031}$ |
| U-VAE | $0.205_{\pm0.002}$ | $0.386_{\pm0.021}$ | $0.769_{\pm0.046}$ | $0.177_{\pm0.005}$ | $0.404_{\pm0.030}$ | $0.693_{\pm0.104}$ | $0.204_{\pm0.002}$ | $0.376_{\pm0.013}$ | $0.545_{\pm0.056}$ | $0.189_{\pm0.005}$ | $0.451_{\pm0.034}$ | $0.722_{\pm0.046}$ |

Table 15: **Imputation utility** on the `concrete` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

`default`

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.392_{\pm0.001}$ | $0.802_{\pm0.003}$ | $0.853_{\pm0.012}$ | $0.354_{\pm0.007}$ | $0.801_{\pm0.004}$ | $0.842_{\pm0.004}$ | $0.391_{\pm0.002}$ | $0.801_{\pm0.004}$ | $0.847_{\pm0.012}$ | $0.334_{\pm0.009}$ | $0.809_{\pm0.005}$ | $0.842_{\pm0.012}$ |
| kNNI | $0.305_{\pm0.001}$ | $0.801_{\pm0.002}$ | $0.837_{\pm0.007}$ | $0.281_{\pm0.003}$ | $0.800_{\pm0.004}$ | $0.830_{\pm0.009}$ | $0.305_{\pm0.002}$ | $0.801_{\pm0.004}$ | $0.829_{\pm0.012}$ | $0.269_{\pm0.004}$ | $0.808_{\pm0.005}$ | $0.820_{\pm0.019}$ |
| MICE | $0.316_{\pm0.002}$ | $0.790_{\pm0.004}$ | $0.820_{\pm0.021}$ | $0.292_{\pm0.002}$ | $0.797_{\pm0.006}$ | $0.818_{\pm0.013}$ | $0.315_{\pm0.001}$ | $0.791_{\pm0.006}$ | $0.822_{\pm0.007}$ | $0.267_{\pm0.008}$ | $0.806_{\pm0.005}$ | $0.822_{\pm0.014}$ |
| missForest | $0.339_{\pm0.000}$ | $0.802_{\pm0.003}$ | $0.841_{\pm0.014}$ | $0.311_{\pm0.004}$ | $0.802_{\pm0.005}$ | $0.831_{\pm0.004}$ | $0.339_{\pm0.002}$ | $0.801_{\pm0.004}$ | $0.822_{\pm0.012}$ | $0.293_{\pm0.006}$ | $0.809_{\pm0.005}$ | $0.829_{\pm0.019}$ |
| GAIN | $0.342_{\pm0.007}$ | $0.799_{\pm0.002}$ | $0.756_{\pm0.025}$ | $0.314_{\pm0.007}$ | $0.799_{\pm0.004}$ | $0.792_{\pm0.016}$ | $0.363_{\pm0.010}$ | $0.799_{\pm0.005}$ | $0.773_{\pm0.016}$ | $0.289_{\pm0.009}$ | $0.805_{\pm0.005}$ | $0.783_{\pm0.024}$ |
| VAEAC | $0.308_{\pm0.003}$ | $0.819_{\pm0.003}$ | $0.836_{\pm0.010}$ | $0.290_{\pm0.005}$ | $0.820_{\pm0.002}$ | $0.819_{\pm0.020}$ | $0.312_{\pm0.002}$ | $0.819_{\pm0.003}$ | $0.831_{\pm0.014}$ | $0.275_{\pm0.007}$ | $0.817_{\pm0.002}$ | $0.842_{\pm0.009}$ |
| MIWAE | $0.307_{\pm0.001}$ | $0.767_{\pm0.003}$ | $0.830_{\pm0.017}$ | $0.297_{\pm0.001}$ | $0.709_{\pm0.029}$ | $0.826_{\pm0.006}$ | $0.307_{\pm0.002}$ | $0.713_{\pm0.026}$ | $0.813_{\pm0.007}$ | $0.289_{\pm0.003}$ | $0.796_{\pm0.013}$ | $0.836_{\pm0.013}$ |
| not-MIWAE | $0.308_{\pm0.001}$ | $0.802_{\pm0.003}$ | $0.840_{\pm0.018}$ | $0.293_{\pm0.002}$ | $0.722_{\pm0.025}$ | $0.794_{\pm0.020}$ | $0.307_{\pm0.001}$ | $0.801_{\pm0.004}$ | $0.823_{\pm0.007}$ | $0.287_{\pm0.004}$ | $0.809_{\pm0.005}$ | $0.818_{\pm0.017}$ |
| MIRACLE | $0.291_{\pm0.003}$ | $0.802_{\pm0.003}$ | $0.835_{\pm0.015}$ | $0.267_{\pm0.005}$ | $0.801_{\pm0.004}$ | $0.832_{\pm0.012}$ | $0.295_{\pm0.004}$ | $0.801_{\pm0.004}$ | $0.824_{\pm0.008}$ | $0.250_{\pm0.005}$ | $0.809_{\pm0.005}$ | $0.817_{\pm0.021}$ |
| ReMasker | $0.392_{\pm0.014}$ | $0.803_{\pm0.002}$ | $0.823_{\pm0.014}$ | $0.319_{\pm0.005}$ | $0.800_{\pm0.004}$ | $0.836_{\pm0.006}$ | $0.342_{\pm0.012}$ | $0.800_{\pm0.004}$ | $0.826_{\pm0.008}$ | $0.358_{\pm0.018}$ | $0.808_{\pm0.006}$ | $0.830_{\pm0.017}$ |
| U-VAE | $0.292_{\pm0.003}$ | $0.826_{\pm0.003}$ | $0.846_{\pm0.013}$ | $0.282_{\pm0.003}$ | $0.827_{\pm0.005}$ | $0.860_{\pm0.012}$ | $0.288_{\pm0.001}$ | $0.824_{\pm0.004}$ | $0.834_{\pm0.009}$ | $0.272_{\pm0.003}$ | $0.831_{\pm0.004}$ | $0.837_{\pm0.017}$ |

Table 16: **Imputation utility** on the `default` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

`loan`

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.298_{\pm0.002}$ | $0.930_{\pm0.001}$ | $0.839_{\pm0.023}$ | $0.291_{\pm0.004}$ | $0.929_{\pm0.002}$ | $0.853_{\pm0.015}$ | $0.301_{\pm0.004}$ | $0.927_{\pm0.002}$ | $0.823_{\pm0.021}$ | $0.288_{\pm0.002}$ | $0.936_{\pm0.002}$ | $0.896_{\pm0.018}$ |
| kNNI | $0.290_{\pm0.001}$ | $0.932_{\pm0.002}$ | $0.836_{\pm0.030}$ | $0.280_{\pm0.003}$ | $0.931_{\pm0.003}$ | $0.860_{\pm0.011}$ | $0.287_{\pm0.001}$ | $0.931_{\pm0.002}$ | $0.837_{\pm0.007}$ | $0.278_{\pm0.003}$ | $0.936_{\pm0.002}$ | $0.865_{\pm0.027}$ |
| MICE | $0.276_{\pm0.001}$ | $0.933_{\pm0.001}$ | $0.854_{\pm0.020}$ | $0.271_{\pm0.002}$ | $0.929_{\pm0.002}$ | $0.852_{\pm0.013}$ | $0.280_{\pm0.003}$ | $0.924_{\pm0.002}$ | $0.811_{\pm0.011}$ | $0.273_{\pm0.001}$ | $0.937_{\pm0.001}$ | $0.890_{\pm0.016}$ |
| missForest | $0.290_{\pm0.001}$ | $0.931_{\pm0.001}$ | $0.821_{\pm0.039}$ | $0.284_{\pm0.003}$ | $0.928_{\pm0.001}$ | $0.846_{\pm0.018}$ | $0.291_{\pm0.002}$ | $0.926_{\pm0.001}$ | $0.842_{\pm0.018}$ | $0.283_{\pm0.001}$ | $0.937_{\pm0.001}$ | $0.871_{\pm0.029}$ |
| GAIN | $0.291_{\pm0.004}$ | $0.930_{\pm0.002}$ | $0.818_{\pm0.018}$ | $0.287_{\pm0.005}$ | $0.927_{\pm0.003}$ | $0.824_{\pm0.026}$ | $0.295_{\pm0.004}$ | $0.926_{\pm0.003}$ | $0.809_{\pm0.008}$ | $0.275_{\pm0.006}$ | $0.934_{\pm0.003}$ | $0.792_{\pm0.050}$ |
| VAEAC | $0.273_{\pm0.001}$ | $0.941_{\pm0.003}$ | $0.863_{\pm0.033}$ | $0.271_{\pm0.002}$ | $0.935_{\pm0.001}$ | $0.855_{\pm0.010}$ | $0.275_{\pm0.003}$ | $0.936_{\pm0.001}$ | $0.859_{\pm0.009}$ | $0.271_{\pm0.001}$ | $0.945_{\pm0.002}$ | $0.843_{\pm0.024}$ |
| MIWAE | $0.289_{\pm0.001}$ | $0.882_{\pm0.002}$ | $0.718_{\pm0.017}$ | $0.293_{\pm0.006}$ | $0.779_{\pm0.018}$ | $0.706_{\pm0.023}$ | $0.302_{\pm0.002}$ | $0.789_{\pm0.019}$ | $0.689_{\pm0.013}$ | $0.285_{\pm0.004}$ | $0.895_{\pm0.013}$ | $0.760_{\pm0.027}$ |
| not-MIWAE | $0.292_{\pm0.001}$ | $0.931_{\pm0.001}$ | $0.855_{\pm0.018}$ | $0.294_{\pm0.005}$ | $0.898_{\pm0.031}$ | $0.815_{\pm0.038}$ | $0.299_{\pm0.005}$ | $0.928_{\pm0.002}$ | $0.833_{\pm0.020}$ | $0.290_{\pm0.004}$ | $0.901_{\pm0.016}$ | $0.814_{\pm0.036}$ |
| MIRACLE | $0.275_{\pm0.001}$ | $0.932_{\pm0.002}$ | $0.847_{\pm0.025}$ | $0.270_{\pm0.002}$ | $0.929_{\pm0.002}$ | $0.860_{\pm0.017}$ | $0.277_{\pm0.003}$ | $0.930_{\pm0.002}$ | $0.833_{\pm0.005}$ | $0.273_{\pm0.001}$ | $0.936_{\pm0.001}$ | $0.882_{\pm0.021}$ |
| ReMasker | $0.271_{\pm0.002}$ | $0.931_{\pm0.002}$ | $0.841_{\pm0.020}$ | $0.274_{\pm0.002}$ | $0.928_{\pm0.002}$ | $0.846_{\pm0.016}$ | $0.275_{\pm0.002}$ | $0.929_{\pm0.002}$ | $0.842_{\pm0.005}$ | $0.271_{\pm0.001}$ | $0.937_{\pm0.001}$ | $0.873_{\pm0.012}$ |
| U-VAE | $0.266_{\pm0.001}$ | $0.943_{\pm0.002}$ | $0.870_{\pm0.022}$ | $0.261_{\pm0.004}$ | $0.941_{\pm0.002}$ | $0.879_{\pm0.015}$ | $0.270_{\pm0.002}$ | $0.938_{\pm0.001}$ | $0.859_{\pm0.017}$ | $0.256_{\pm0.002}$ | $0.945_{\pm0.001}$ | $0.883_{\pm0.023}$ |

Table 17: **Imputation utility** on the `loan` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

`letter`

| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.103_{\pm0.000}$ | $0.578_{\pm0.008}$ | $0.710_{\pm0.028}$ | $0.092_{\pm0.002}$ | $0.591_{\pm0.040}$ | $0.650_{\pm0.019}$ | $0.106_{\pm0.000}$ | $0.548_{\pm0.008}$ | $0.713_{\pm0.035}$ | $0.089_{\pm0.001}$ | $0.624_{\pm0.040}$ | $0.709_{\pm0.023}$ |
| kNNI | $0.069_{\pm0.000}$ | $0.629_{\pm0.006}$ | $0.852_{\pm0.020}$ | $0.061_{\pm0.001}$ | $0.614_{\pm0.045}$ | $0.889_{\pm0.019}$ | $0.070_{\pm0.000}$ | $0.604_{\pm0.008}$ | $0.835_{\pm0.009}$ | $0.060_{\pm0.001}$ | $0.645_{\pm0.041}$ | $0.899_{\pm0.006}$ |
| MICE | $0.090_{\pm0.001}$ | $0.567_{\pm0.003}$ | $0.938_{\pm0.007}$ | $0.081_{\pm0.002}$ | $0.599_{\pm0.042}$ | $0.895_{\pm0.019}$ | $0.091_{\pm0.001}$ | $0.550_{\pm0.006}$ | $0.898_{\pm0.023}$ | $0.079_{\pm0.001}$ | $0.620_{\pm0.041}$ | $0.944_{\pm0.005}$ |
| missForest | $0.092_{\pm0.000}$ | $0.606_{\pm0.003}$ | $0.838_{\pm0.016}$ | $0.084_{\pm0.001}$ | $0.604_{\pm0.039}$ | $0.732_{\pm0.028}$ | $0.094_{\pm0.000}$ | $0.572_{\pm0.001}$ | $0.786_{\pm0.025}$ | $0.081_{\pm0.001}$ | $0.637_{\pm0.038}$ | $0.777_{\pm0.020}$ |
| GAIN | $0.101_{\pm0.001}$ | $0.625_{\pm0.024}$ | $0.679_{\pm0.025}$ | $0.087_{\pm0.001}$ | $0.619_{\pm0.035}$ | $0.713_{\pm0.040}$ | $0.101_{\pm0.001}$ | $0.613_{\pm0.031}$ | $0.715_{\pm0.027}$ | $0.083_{\pm0.002}$ | $0.664_{\pm0.029}$ | $0.716_{\pm0.009}$ |
| VAEAC | $0.081_{\pm0.000}$ | $0.770_{\pm0.007}$ | $0.906_{\pm0.013}$ | $0.077_{\pm0.001}$ | $0.760_{\pm0.004}$ | $0.897_{\pm0.014}$ | $0.081_{\pm0.000}$ | $0.762_{\pm0.003}$ | $0.921_{\pm0.005}$ | $0.072_{\pm0.001}$ | $0.780_{\pm0.006}$ | $0.918_{\pm0.002}$ |
| MIWAE | $0.096_{\pm0.000}$ | $0.712_{\pm0.002}$ | $0.713_{\pm0.011}$ | $0.087_{\pm0.001}$ | $0.693_{\pm0.023}$ | $0.648_{\pm0.056}$ | $0.100_{\pm0.001}$ | $0.692_{\pm0.008}$ | $0.756_{\pm0.022}$ | $0.083_{\pm0.002}$ | $0.726_{\pm0.014}$ | $0.753_{\pm0.017}$ |
| not-MIWAE | $0.094_{\pm0.000}$ | $0.592_{\pm0.003}$ | $0.770_{\pm0.023}$ | $0.085_{\pm0.001}$ | $0.597_{\pm0.040}$ | $0.719_{\pm0.020}$ | $0.096_{\pm0.001}$ | $0.564_{\pm0.005}$ | $0.718_{\pm0.010}$ | $0.082_{\pm0.001}$ | $0.631_{\pm0.043}$ | $0.766_{\pm0.024}$ |
| MIRACLE | $0.078_{\pm0.002}$ | $0.612_{\pm0.008}$ | $0.870_{\pm0.032}$ | $0.067_{\pm0.002}$ | $0.627_{\pm0.043}$ | $0.910_{\pm0.019}$ | $0.080_{\pm0.002}$ | $0.586_{\pm0.008}$ | $0.883_{\pm0.020}$ | $0.066_{\pm0.001}$ | $0.650_{\pm0.035}$ | $0.901_{\pm0.022}$ |
| ReMasker | $0.072_{\pm0.001}$ | $0.603_{\pm0.003}$ | $0.906_{\pm0.012}$ | $0.068_{\pm0.000}$ | $0.613_{\pm0.042}$ | $0.896_{\pm0.024}$ | $0.072_{\pm0.000}$ | $0.578_{\pm0.005}$ | $0.846_{\pm0.023}$ | $0.064_{\pm0.000}$ | $0.642_{\pm0.041}$ | $0.911_{\pm0.007}$ |
| U-VAE | $0.069_{\pm0.001}$ | $0.743_{\pm0.010}$ | $0.957_{\pm0.009}$ | $0.061_{\pm0.002}$ | $0.746_{\pm0.037}$ | $0.904_{\pm0.018}$ | $0.070_{\pm0.001}$ | $0.717_{\pm0.006}$ | $0.933_{\pm0.013}$ | $0.059_{\pm0.001}$ | $0.780_{\pm0.026}$ | $0.941_{\pm0.017}$ |

Table 18: **Imputation utility** on the `letter` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

| kings | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.216_{\pm 0.000}$ | $0.493_{\pm 0.004}$ | $0.857_{\pm 0.016}$ | $0.162_{\pm 0.003}$ | $0.462_{\pm 0.021}$ | $0.792_{\pm 0.016}$ | $0.215_{\pm 0.002}$ | $0.447_{\pm 0.029}$ | $0.788_{\pm 0.026}$ | $0.199_{\pm 0.008}$ | $0.549_{\pm 0.031}$ | $0.832_{\pm 0.034}$ |
| kNNI | $0.172_{\pm 0.001}$ | $0.475_{\pm 0.022}$ | $0.780_{\pm 0.038}$ | $0.130_{\pm 0.004}$ | $0.462_{\pm 0.019}$ | $0.835_{\pm 0.018}$ | $0.168_{\pm 0.002}$ | $0.458_{\pm 0.024}$ | $0.799_{\pm 0.023}$ | $0.143_{\pm 0.002}$ | $0.568_{\pm 0.029}$ | $0.833_{\pm 0.032}$ |
| MICE | $0.177_{\pm 0.006}$ | $0.491_{\pm 0.007}$ | $0.912_{\pm 0.010}$ | $0.118_{\pm 0.011}$ | $0.471_{\pm 0.020}$ | $0.901_{\pm 0.007}$ | $0.184_{\pm 0.009}$ | $0.457_{\pm 0.023}$ | $0.894_{\pm 0.018}$ | $0.150_{\pm 0.010}$ | $0.548_{\pm 0.037}$ | $0.900_{\pm 0.013}$ |
| missForest | $0.204_{\pm 0.000}$ | $0.479_{\pm 0.004}$ | $0.905_{\pm 0.008}$ | $0.155_{\pm 0.007}$ | $0.454_{\pm 0.017}$ | $0.861_{\pm 0.042}$ | $0.203_{\pm 0.002}$ | $0.445_{\pm 0.025}$ | $0.784_{\pm 0.052}$ | $0.190_{\pm 0.009}$ | $0.552_{\pm 0.031}$ | $0.864_{\pm 0.038}$ |
| GAIN | $0.151_{\pm 0.003}$ | $0.523_{\pm 0.020}$ | $0.850_{\pm 0.016}$ | $0.130_{\pm 0.008}$ | $0.535_{\pm 0.014}$ | $0.851_{\pm 0.034}$ | $0.182_{\pm 0.012}$ | $0.530_{\pm 0.013}$ | $0.828_{\pm 0.040}$ | $0.156_{\pm 0.012}$ | $0.577_{\pm 0.023}$ | $0.853_{\pm 0.047}$ |
| VAEAC | $0.191_{\pm 0.001}$ | $0.622_{\pm 0.003}$ | $0.820_{\pm 0.044}$ | $0.136_{\pm 0.009}$ | $0.621_{\pm 0.002}$ | $0.821_{\pm 0.055}$ | $0.189_{\pm 0.002}$ | $0.622_{\pm 0.002}$ | $0.794_{\pm 0.024}$ | $0.177_{\pm 0.011}$ | $0.632_{\pm 0.002}$ | $0.867_{\pm 0.036}$ |
| MIWAE | $0.150_{\pm 0.003}$ | $0.521_{\pm 0.002}$ | $0.807_{\pm 0.030}$ | $0.120_{\pm 0.007}$ | $0.525_{\pm 0.008}$ | $0.855_{\pm 0.035}$ | $0.151_{\pm 0.002}$ | $0.528_{\pm 0.007}$ | $0.809_{\pm 0.024}$ | $0.138_{\pm 0.002}$ | $0.581_{\pm 0.019}$ | $0.832_{\pm 0.025}$ |
| not-MIWAE | $0.190_{\pm 0.005}$ | $0.479_{\pm 0.006}$ | $0.838_{\pm 0.035}$ | $0.139_{\pm 0.009}$ | $0.449_{\pm 0.023}$ | $0.863_{\pm 0.005}$ | $0.197_{\pm 0.002}$ | $0.443_{\pm 0.029}$ | $0.788_{\pm 0.026}$ | $0.177_{\pm 0.010}$ | $0.556_{\pm 0.030}$ | $0.862_{\pm 0.023}$ |
| MIRACLE | $0.180_{\pm 0.001}$ | $0.483_{\pm 0.010}$ | $0.753_{\pm 0.018}$ | $0.124_{\pm 0.008}$ | $0.451_{\pm 0.017}$ | $0.800_{\pm 0.028}$ | $0.180_{\pm 0.002}$ | $0.457_{\pm 0.023}$ | $0.789_{\pm 0.030}$ | $0.164_{\pm 0.011}$ | $0.556_{\pm 0.032}$ | $0.790_{\pm 0.062}$ |
| ReMasker | $0.243_{\pm 0.023}$ | $0.451_{\pm 0.010}$ | $0.874_{\pm 0.040}$ | $0.177_{\pm 0.007}$ | $0.449_{\pm 0.032}$ | $0.838_{\pm 0.038}$ | $0.283_{\pm 0.011}$ | $0.450_{\pm 0.035}$ | $0.733_{\pm 0.046}$ | $0.232_{\pm 0.015}$ | $0.536_{\pm 0.032}$ | $0.928_{\pm 0.012}$ |
| U-VAE | $0.169_{\pm 0.001}$ | $0.638_{\pm 0.002}$ | $0.924_{\pm 0.033}$ | $0.123_{\pm 0.009}$ | $0.613_{\pm 0.011}$ | $0.850_{\pm 0.024}$ | $0.170_{\pm 0.001}$ | $0.625_{\pm 0.007}$ | $0.914_{\pm 0.037}$ | $0.155_{\pm 0.009}$ | $0.663_{\pm 0.011}$ | $0.933_{\pm 0.039}$ |

Table 19: **Imputation utility** on the `kings` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

| redwine | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.098_{\pm 0.001}$ | $0.432_{\pm 0.019}$ | $0.680_{\pm 0.033}$ | $0.092_{\pm 0.002}$ | $0.413_{\pm 0.041}$ | $0.475_{\pm 0.093}$ | $0.099_{\pm 0.002}$ | $0.370_{\pm 0.011}$ | $0.507_{\pm 0.085}$ | $0.090_{\pm 0.001}$ | $0.498_{\pm 0.037}$ | $0.719_{\pm 0.044}$ |
| kNNI | $0.090_{\pm 0.001}$ | $0.434_{\pm 0.016}$ | $0.713_{\pm 0.055}$ | $0.084_{\pm 0.001}$ | $0.443_{\pm 0.048}$ | $0.601_{\pm 0.075}$ | $0.093_{\pm 0.001}$ | $0.396_{\pm 0.012}$ | $0.545_{\pm 0.059}$ | $0.082_{\pm 0.001}$ | $0.497_{\pm 0.038}$ | $0.723_{\pm 0.040}$ |
| MICE | $0.087_{\pm 0.001}$ | $0.438_{\pm 0.016}$ | $0.788_{\pm 0.025}$ | $0.083_{\pm 0.001}$ | $0.438_{\pm 0.044}$ | $0.659_{\pm 0.072}$ | $0.088_{\pm 0.001}$ | $0.390_{\pm 0.011}$ | $0.592_{\pm 0.075}$ | $0.084_{\pm 0.001}$ | $0.505_{\pm 0.035}$ | $0.822_{\pm 0.024}$ |
| missForest | $0.090_{\pm 0.001}$ | $0.444_{\pm 0.014}$ | $0.698_{\pm 0.048}$ | $0.087_{\pm 0.001}$ | $0.423_{\pm 0.043}$ | $0.511_{\pm 0.055}$ | $0.092_{\pm 0.001}$ | $0.380_{\pm 0.017}$ | $0.501_{\pm 0.087}$ | $0.085_{\pm 0.001}$ | $0.500_{\pm 0.031}$ | $0.762_{\pm 0.030}$ |
| GAIN | $0.095_{\pm 0.001}$ | $0.500_{\pm 0.008}$ | $0.750_{\pm 0.047}$ | $0.089_{\pm 0.001}$ | $0.515_{\pm 0.023}$ | $0.507_{\pm 0.085}$ | $0.099_{\pm 0.002}$ | $0.483_{\pm 0.010}$ | $0.625_{\pm 0.030}$ | $0.088_{\pm 0.001}$ | $0.543_{\pm 0.021}$ | $0.706_{\pm 0.047}$ |
| VAEAC | $0.087_{\pm 0.001}$ | $0.595_{\pm 0.014}$ | $0.775_{\pm 0.031}$ | $0.088_{\pm 0.001}$ | $0.587_{\pm 0.013}$ | $0.620_{\pm 0.046}$ | $0.088_{\pm 0.001}$ | $0.594_{\pm 0.007}$ | $0.568_{\pm 0.035}$ | $0.083_{\pm 0.001}$ | $0.597_{\pm 0.011}$ | $0.734_{\pm 0.033}$ |
| MIWAE | $0.091_{\pm 0.001}$ | $0.527_{\pm 0.013}$ | $0.598_{\pm 0.086}$ | $0.087_{\pm 0.000}$ | $0.509_{\pm 0.032}$ | $0.549_{\pm 0.038}$ | $0.093_{\pm 0.001}$ | $0.476_{\pm 0.017}$ | $0.493_{\pm 0.082}$ | $0.085_{\pm 0.001}$ | $0.554_{\pm 0.017}$ | $0.561_{\pm 0.067}$ |
| not-MIWAE | $0.098_{\pm 0.002}$ | $0.428_{\pm 0.014}$ | $0.582_{\pm 0.081}$ | $0.092_{\pm 0.001}$ | $0.414_{\pm 0.036}$ | $0.474_{\pm 0.096}$ | $0.099_{\pm 0.002}$ | $0.374_{\pm 0.020}$ | $0.466_{\pm 0.081}$ | $0.090_{\pm 0.001}$ | $0.486_{\pm 0.037}$ | $0.714_{\pm 0.047}$ |
| MIRACLE | $0.093_{\pm 0.004}$ | $0.445_{\pm 0.016}$ | $0.645_{\pm 0.049}$ | $0.087_{\pm 0.003}$ | $0.443_{\pm 0.035}$ | $0.483_{\pm 0.109}$ | $0.095_{\pm 0.004}$ | $0.408_{\pm 0.020}$ | $0.393_{\pm 0.098}$ | $0.086_{\pm 0.003}$ | $0.495_{\pm 0.038}$ | $0.683_{\pm 0.035}$ |
| ReMasker | $0.109_{\pm 0.003}$ | $0.419_{\pm 0.029}$ | $0.627_{\pm 0.036}$ | $0.107_{\pm 0.004}$ | $0.396_{\pm 0.050}$ | $0.393_{\pm 0.076}$ | $0.119_{\pm 0.002}$ | $0.380_{\pm 0.033}$ | $0.421_{\pm 0.099}$ | $0.099_{\pm 0.005}$ | $0.486_{\pm 0.045}$ | $0.728_{\pm 0.048}$ |
| U-VAE | $0.088_{\pm 0.002}$ | $0.605_{\pm 0.011}$ | $0.793_{\pm 0.029}$ | $0.081_{\pm 0.001}$ | $0.574_{\pm 0.037}$ | $0.664_{\pm 0.086}$ | $0.087_{\pm 0.002}$ | $0.564_{\pm 0.011}$ | $0.690_{\pm 0.068}$ | $0.079_{\pm 0.001}$ | $0.648_{\pm 0.019}$ | $0.828_{\pm 0.049}$ |

Table 20: **Imputation utility** on the `redwine` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

| shoppers | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.624_{\pm 0.006}$ | $0.874_{\pm 0.004}$ | $0.802_{\pm 0.017}$ | $0.518_{\pm 0.018}$ | $0.873_{\pm 0.004}$ | $0.763_{\pm 0.017}$ | $0.618_{\pm 0.008}$ | $0.873_{\pm 0.004}$ | $0.761_{\pm 0.030}$ | $0.528_{\pm 0.015}$ | $0.882_{\pm 0.005}$ | $0.807_{\pm 0.014}$ |
| kNNI | $0.572_{\pm 0.002}$ | $0.876_{\pm 0.004}$ | $0.755_{\pm 0.013}$ | $0.485_{\pm 0.018}$ | $0.875_{\pm 0.004}$ | $0.756_{\pm 0.009}$ | $0.570_{\pm 0.003}$ | $0.875_{\pm 0.004}$ | $0.750_{\pm 0.011}$ | $0.490_{\pm 0.015}$ | $0.882_{\pm 0.005}$ | $0.796_{\pm 0.014}$ |
| MICE | $0.616_{\pm 0.002}$ | $0.855_{\pm 0.005}$ | $0.739_{\pm 0.006}$ | $0.504_{\pm 0.020}$ | $0.879_{\pm 0.004}$ | $0.745_{\pm 0.014}$ | $0.617_{\pm 0.003}$ | $0.863_{\pm 0.006}$ | $0.737_{\pm 0.013}$ | $0.504_{\pm 0.015}$ | $0.880_{\pm 0.005}$ | $0.770_{\pm 0.010}$ |
| missForest | $0.581_{\pm 0.002}$ | $0.877_{\pm 0.004}$ | $0.734_{\pm 0.014}$ | $0.493_{\pm 0.017}$ | $0.877_{\pm 0.004}$ | $0.747_{\pm 0.014}$ | $0.581_{\pm 0.002}$ | $0.878_{\pm 0.003}$ | $0.726_{\pm 0.020}$ | $0.511_{\pm 0.016}$ | $0.883_{\pm 0.005}$ | $0.791_{\pm 0.013}$ |
| GAIN | $0.582_{\pm 0.007}$ | $0.876_{\pm 0.002}$ | $0.722_{\pm 0.017}$ | $0.513_{\pm 0.021}$ | $0.879_{\pm 0.005}$ | $0.730_{\pm 0.023}$ | $0.585_{\pm 0.005}$ | $0.871_{\pm 0.003}$ | $0.713_{\pm 0.021}$ | $0.515_{\pm 0.012}$ | $0.880_{\pm 0.004}$ | $0.779_{\pm 0.010}$ |
| VAEAC | $0.573_{\pm 0.002}$ | $0.888_{\pm 0.003}$ | $0.759_{\pm 0.014}$ | $0.499_{\pm 0.019}$ | $0.888_{\pm 0.003}$ | $0.762_{\pm 0.010}$ | $0.569_{\pm 0.003}$ | $0.886_{\pm 0.004}$ | $0.729_{\pm 0.011}$ | $0.498_{\pm 0.015}$ | $0.889_{\pm 0.003}$ | $0.799_{\pm 0.017}$ |
| MIWAE | $0.562_{\pm 0.001}$ | $0.841_{\pm 0.004}$ | $0.744_{\pm 0.021}$ | $0.517_{\pm 0.009}$ | $0.782_{\pm 0.011}$ | $0.743_{\pm 0.030}$ | $0.564_{\pm 0.003}$ | $0.783_{\pm 0.012}$ | $0.714_{\pm 0.030}$ | $0.530_{\pm 0.002}$ | $0.862_{\pm 0.013}$ | $0.760_{\pm 0.020}$ |
| not-MIWAE | $0.566_{\pm 0.005}$ | $0.874_{\pm 0.004}$ | $0.778_{\pm 0.015}$ | $0.529_{\pm 0.012}$ | $0.836_{\pm 0.022}$ | $0.720_{\pm 0.031}$ | $0.581_{\pm 0.005}$ | $0.874_{\pm 0.004}$ | $0.726_{\pm 0.022}$ | $0.535_{\pm 0.008}$ | $0.877_{\pm 0.006}$ | $0.763_{\pm 0.032}$ |
| MIRACLE | $0.563_{\pm 0.007}$ | $0.877_{\pm 0.004}$ | $0.744_{\pm 0.011}$ | $0.470_{\pm 0.015}$ | $0.878_{\pm 0.003}$ | $0.711_{\pm 0.017}$ | $0.564_{\pm 0.007}$ | $0.878_{\pm 0.004}$ | $0.729_{\pm 0.018}$ | $0.484_{\pm 0.015}$ | $0.882_{\pm 0.005}$ | $0.794_{\pm 0.016}$ |
| ReMasker | $0.596_{\pm 0.018}$ | $0.865_{\pm 0.007}$ | $0.765_{\pm 0.014}$ | $0.553_{\pm 0.018}$ | $0.865_{\pm 0.004}$ | $0.762_{\pm 0.019}$ | $0.624_{\pm 0.019}$ | $0.863_{\pm 0.003}$ | $0.757_{\pm 0.018}$ | $0.558_{\pm 0.018}$ | $0.880_{\pm 0.005}$ | $0.789_{\pm 0.009}$ |
| U-VAE | $0.562_{\pm 0.002}$ | $0.896_{\pm 0.004}$ | $0.783_{\pm 0.021}$ | $0.510_{\pm 0.019}$ | $0.892_{\pm 0.005}$ | $0.801_{\pm 0.014}$ | $0.565_{\pm 0.001}$ | $0.896_{\pm 0.004}$ | $0.734_{\pm 0.025}$ | $0.509_{\pm 0.012}$ | $0.904_{\pm 0.005}$ | $0.815_{\pm 0.013}$ |

Table 21: **Imputation utility** on the `shoppers` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

| whitewine | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ | SMAPE ↓ | Acc ↑ | Feat ↑ |
| Mean | $0.083_{\pm 0.001}$ | $0.368_{\pm 0.009}$ | $0.507_{\pm 0.065}$ | $0.081_{\pm 0.002}$ | $0.361_{\pm 0.040}$ | $0.565_{\pm 0.093}$ | $0.086_{\pm 0.001}$ | $0.319_{\pm 0.012}$ | $0.636_{\pm 0.076}$ | $0.075_{\pm 0.002}$ | $0.439_{\pm 0.038}$ | $0.622_{\pm 0.076}$ |
| kNNI | $0.078_{\pm 0.001}$ | $0.380_{\pm 0.007}$ | $0.641_{\pm 0.048}$ | $0.072_{\pm 0.002}$ | $0.381_{\pm 0.038}$ | $0.675_{\pm 0.051}$ | $0.081_{\pm 0.001}$ | $0.343_{\pm 0.015}$ | $0.709_{\pm 0.050}$ | $0.069_{\pm 0.001}$ | $0.436_{\pm 0.042}$ | $0.619_{\pm 0.105}$ |
| MICE | $0.072_{\pm 0.001}$ | $0.368_{\pm 0.012}$ | $0.736_{\pm 0.033}$ | $0.070_{\pm 0.001}$ | $0.387_{\pm 0.033}$ | $0.760_{\pm 0.034}$ | $0.073_{\pm 0.000}$ | $0.346_{\pm 0.013}$ | $0.727_{\pm 0.030}$ | $0.068_{\pm 0.001}$ | $0.445_{\pm 0.037}$ | $0.660_{\pm 0.063}$ |
| missForest | $0.079_{\pm 0.001}$ | $0.378_{\pm 0.011}$ | $0.622_{\pm 0.088}$ | $0.077_{\pm 0.001}$ | $0.368_{\pm 0.041}$ | $0.692_{\pm 0.089}$ | $0.081_{\pm 0.001}$ | $0.327_{\pm 0.014}$ | $0.721_{\pm 0.016}$ | $0.073_{\pm 0.001}$ | $0.447_{\pm 0.037}$ | $0.628_{\pm 0.071}$ |
| GAIN | $0.080_{\pm 0.002}$ | $0.403_{\pm 0.009}$ | $0.603_{\pm 0.080}$ | $0.078_{\pm 0.003}$ | $0.440_{\pm 0.027}$ | $0.611_{\pm 0.130}$ | $0.086_{\pm 0.002}$ | $0.385_{\pm 0.013}$ | $0.546_{\pm 0.084}$ | $0.074_{\pm 0.001}$ | $0.452_{\pm 0.036}$ | $0.696_{\pm 0.076}$ |
| VAEAC | $0.074_{\pm 0.001}$ | $0.521_{\pm 0.004}$ | $0.528_{\pm 0.058}$ | $0.073_{\pm 0.001}$ | $0.525_{\pm 0.004}$ | $0.715_{\pm 0.077}$ | $0.074_{\pm 0.001}$ | $0.514_{\pm 0.006}$ | $0.570_{\pm 0.065}$ | $0.069_{\pm 0.001}$ | $0.532_{\pm 0.005}$ | $0.724_{\pm 0.086}$ |
| MIWAE | $0.080_{\pm 0.001}$ | $0.435_{\pm 0.011}$ | $0.571_{\pm 0.081}$ | $0.076_{\pm 0.002}$ | $0.463_{\pm 0.013}$ | $0.647_{\pm 0.040}$ | $0.081_{\pm 0.001}$ | $0.450_{\pm 0.004}$ | $0.636_{\pm 0.049}$ | $0.072_{\pm 0.001}$ | $0.464_{\pm 0.028}$ | $0.569_{\pm 0.079}$ |
| not-MIWAE | $0.079_{\pm 0.001}$ | $0.364_{\pm 0.009}$ | $0.609_{\pm 0.088}$ | $0.077_{\pm 0.001}$ | $0.371_{\pm 0.040}$ | $0.683_{\pm 0.060}$ | $0.083_{\pm 0.001}$ | $0.314_{\pm 0.017}$ | $0.714_{\pm 0.058}$ | $0.074_{\pm 0.001}$ | $0.445_{\pm 0.037}$ | $0.694_{\pm 0.087}$ |
| MIRACLE | $0.073_{\pm 0.001}$ | $0.356_{\pm 0.007}$ | $0.650_{\pm 0.043}$ | $0.069_{\pm 0.001}$ | $0.376_{\pm 0.035}$ | $0.558_{\pm 0.061}$ | $0.076_{\pm 0.001}$ | $0.343_{\pm 0.011}$ | $0.578_{\pm 0.050}$ | $0.067_{\pm 0.001}$ | $0.434_{\pm 0.037}$ | $0.673_{\pm 0.042}$ |
| ReMasker | $0.083_{\pm 0.006}$ | $0.368_{\pm 0.013}$ | $0.510_{\pm 0.073}$ | $0.081_{\pm 0.003}$ | $0.356_{\pm 0.044}$ | $0.619_{\pm 0.049}$ | $0.087_{\pm 0.003}$ | $0.310_{\pm 0.019}$ | $0.420_{\pm 0.057}$ | $0.083_{\pm 0.008}$ | $0.434_{\pm 0.038}$ | $0.628_{\pm 0.072}$ |
| U-VAE | $0.074_{\pm 0.001}$ | $0.541_{\pm 0.011}$ | $0.718_{\pm 0.055}$ | $0.069_{\pm 0.001}$ | $0.509_{\pm 0.028}$ | $0.707_{\pm 0.063}$ | $0.073_{\pm 0.001}$ | $0.497_{\pm 0.017}$ | $0.606_{\pm 0.028}$ | $0.066_{\pm 0.001}$ | $0.570_{\pm 0.021}$ | $0.729_{\pm 0.048}$ |

Table 22: **Imputation utility** on the `whitewine` dataset at 0.3 missingness. The means and their standard errors across 5 random seeds are reported. ↑ (↓) denotes that higher (lower) is better.

**Q2: Multiple imputation** We report Bias, Coverage, and Width results to assess multiple imputation for each dataset at 0.3 missingness rate.

| anuran | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.009(1.4) | $0.955_{\pm 0.000}$ | $0.025_{\pm 0.000}$ | 0.006(1.0) | $0.936_{\pm 0.011}$ | $0.025_{\pm 0.000}$ | 0.010(1.6) | $0.900_{\pm 0.017}$ | $0.025_{\pm 0.000}$ | 0.006(0.9) | $0.955_{\pm 0.014}$ | $0.025_{\pm 0.000}$ |
| GAIN | 0.027(5.3) | $0.373_{\pm 0.030}$ | $0.025_{\pm 0.000}$ | 0.026(5.4) | $0.536_{\pm 0.039}$ | $0.025_{\pm 0.000}$ | 0.044(8.6) | $0.327_{\pm 0.072}$ | $0.025_{\pm 0.000}$ | 0.013(2.7) | $0.682_{\pm 0.043}$ | $0.025_{\pm 0.000}$ |
| VAEAC | 0.012(2.0) | $0.809_{\pm 0.022}$ | $0.025_{\pm 0.000}$ | 0.011(1.9) | $0.736_{\pm 0.017}$ | $0.025_{\pm 0.000}$ | 0.013(2.2) | $0.745_{\pm 0.023}$ | $0.025_{\pm 0.000}$ | 0.007(1.2) | $0.836_{\pm 0.018}$ | $0.025_{\pm 0.000}$ |
| MIWAE | 0.003(0.5) | $0.955_{\pm 0.000}$ | $0.026_{\pm 0.000}$ | 0.002(0.4) | $0.973_{\pm 0.011}$ | $0.026_{\pm 0.000}$ | 0.003(0.6) | $0.964_{\pm 0.009}$ | $0.026_{\pm 0.000}$ | 0.002(0.3) | $0.973_{\pm 0.011}$ | $0.026_{\pm 0.000}$ |
| not-MIWAE | 0.005(1.0) | $0.927_{\pm 0.023}$ | $0.027_{\pm 0.000}$ | 0.007(1.3) | $0.845_{\pm 0.031}$ | $0.026_{\pm 0.000}$ | 0.009(1.7) | $0.791_{\pm 0.042}$ | $0.027_{\pm 0.000}$ | 0.003(0.7) | $0.964_{\pm 0.009}$ | $0.026_{\pm 0.000}$ |
| U-VAE | 0.013(5.3) | $0.927_{\pm 0.031}$ | $0.027_{\pm 0.000}$ | 0.005(0.8) | $0.964_{\pm 0.017}$ | $0.026_{\pm 0.000}$ | 0.006(1.0) | $0.918_{\pm 0.027}$ | $0.026_{\pm 0.000}$ | 0.003(0.5) | $0.973_{\pm 0.011}$ | $0.026_{\pm 0.000}$ |

Table 23: **Multiple imputation performance** on the anuran dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| banknote | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.018(3.6) | $0.750_{\pm 0.079}$ | $0.059_{\pm 0.000}$ | 0.013(2.3) | $0.850_{\pm 0.061}$ | $0.059_{\pm 0.000}$ | 0.018(3.5) | $0.850_{\pm 0.061}$ | $0.059_{\pm 0.000}$ | 0.008(1.5) | $0.900_{\pm 0.061}$ | $0.059_{\pm 0.000}$ |
| GAIN | 0.040(8.2) | $0.450_{\pm 0.094}$ | $0.059_{\pm 0.000}$ | 0.040(7.3) | $0.550_{\pm 0.122}$ | $0.058_{\pm 0.000}$ | 0.056(12.0) | $0.350_{\pm 0.127}$ | $0.058_{\pm 0.000}$ | 0.018(3.3) | $0.800_{\pm 0.094}$ | $0.058_{\pm 0.000}$ |
| VAEAC | 0.013(2.5) | $0.950_{\pm 0.050}$ | $0.059_{\pm 0.000}$ | 0.016(3.0) | $0.850_{\pm 0.061}$ | $0.059_{\pm 0.000}$ | 0.012(2.3) | $1.000_{\pm 0.000}$ | $0.059_{\pm 0.000}$ | 0.005(1.1) | $1.000_{\pm 0.000}$ | $0.059_{\pm 0.000}$ |
| MIWAE | 0.007(1.4) | $1.000_{\pm 0.000}$ | $0.061_{\pm 0.000}$ | 0.004(0.8) | $1.000_{\pm 0.000}$ | $0.060_{\pm 0.000}$ | 0.007(1.3) | $1.000_{\pm 0.000}$ | $0.062_{\pm 0.000}$ | 0.004(0.7) | $1.000_{\pm 0.000}$ | $0.060_{\pm 0.000}$ |
| not-MIWAE | 0.020(4.0) | $0.850_{\pm 0.061}$ | $0.063_{\pm 0.000}$ | 0.013(2.5) | $0.950_{\pm 0.050}$ | $0.062_{\pm 0.000}$ | 0.020(4.0) | $0.750_{\pm 0.079}$ | $0.062_{\pm 0.000}$ | 0.015(3.0) | $0.850_{\pm 0.100}$ | $0.061_{\pm 0.001}$ |
| U-VAE | 0.007(1.5) | $1.000_{\pm 0.000}$ | $0.062_{\pm 0.000}$ | 0.007(1.4) | $1.000_{\pm 0.000}$ | $0.061_{\pm 0.000}$ | 0.010(1.9) | $1.000_{\pm 0.000}$ | $0.062_{\pm 0.000}$ | 0.004(0.8) | $1.000_{\pm 0.000}$ | $0.060_{\pm 0.000}$ |

Table 24: **Multiple imputation performance** on the banknote dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| breast | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.008(2.0) | $1.000_{\pm 0.000}$ | $0.090_{\pm 0.000}$ | 0.007(1.6) | $1.000_{\pm 0.000}$ | $0.090_{\pm 0.000}$ | 0.009(2.3) | $1.000_{\pm 0.000}$ | $0.090_{\pm 0.000}$ | 0.004(1.1) | $1.000_{\pm 0.000}$ | $0.090_{\pm 0.000}$ |
| GAIN | 0.019(4.7) | $0.887_{\pm 0.000}$ | $0.089_{\pm 0.000}$ | 0.018(4.3) | $0.900_{\pm 0.000}$ | $0.089_{\pm 0.000}$ | 0.022(5.7) | $0.867_{\pm 0.000}$ | $0.089_{\pm 0.000}$ | 0.010(2.5) | $0.947_{\pm 0.000}$ | $0.089_{\pm 0.000}$ |
| VAEAC | 0.011(2.9) | $0.987_{\pm 0.000}$ | $0.090_{\pm 0.000}$ | 0.015(3.7) | $0.947_{\pm 0.000}$ | $0.091_{\pm 0.000}$ | 0.012(3.1) | $0.993_{\pm 0.000}$ | $0.090_{\pm 0.000}$ | 0.007(1.8) | $1.000_{\pm 0.000}$ | $0.090_{\pm 0.000}$ |
| MIWAE | 0.009(2.3) | $1.000_{\pm 0.000}$ | $0.095_{\pm 0.000}$ | 0.008(2.0) | $0.987_{\pm 0.000}$ | $0.093_{\pm 0.000}$ | 0.010(2.6) | $1.000_{\pm 0.000}$ | $0.095_{\pm 0.000}$ | 0.007(1.7) | $0.993_{\pm 0.000}$ | $0.093_{\pm 0.000}$ |
| not-MIWAE | 0.009(2.3) | $1.000_{\pm 0.000}$ | $0.095_{\pm 0.000}$ | 0.013(3.1) | $0.953_{\pm 0.000}$ | $0.094_{\pm 0.000}$ | 0.009(2.3) | $0.993_{\pm 0.000}$ | $0.095_{\pm 0.000}$ | 0.007(1.7) | $0.987_{\pm 0.000}$ | $0.094_{\pm 0.000}$ |
| U-VAE | 0.007(1.8) | $1.000_{\pm 0.000}$ | $0.094_{\pm 0.000}$ | 0.005(1.3) | $1.000_{\pm 0.000}$ | $0.093_{\pm 0.000}$ | 0.007(1.8) | $1.000_{\pm 0.000}$ | $0.094_{\pm 0.000}$ | 0.005(1.2) | $1.000_{\pm 0.000}$ | $0.093_{\pm 0.000}$ |

Table 25: **Multiple imputation performance** on the breast dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| concrete | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.011(2.2%) | $1.000_{\pm 0.000}$ | $0.068_{\pm 0.000}$ | 0.008(1.7%) | $1.000_{\pm 0.000}$ | $0.068_{\pm 0.000}$ | 0.012(2.4%) | $0.975_{\pm 0.025}$ | $0.068_{\pm 0.000}$ | 0.006(1.3%) | $1.000_{\pm 0.000}$ | $0.068_{\pm 0.000}$ |
| GAIN | 0.054(11.7%) | $0.475_{\pm 0.100}$ | $0.067_{\pm 0.000}$ | 0.031(7.0%) | $0.675_{\pm 0.151}$ | $0.068_{\pm 0.000}$ | 0.032(7.2%) | $0.625_{\pm 0.056}$ | $0.068_{\pm 0.000}$ | 0.012(2.9%) | $0.950_{\pm 0.031}$ | $0.068_{\pm 0.000}$ |
| VAEAC | 0.013(2.7%) | $0.975_{\pm 0.025}$ | $0.069_{\pm 0.000}$ | 0.016(3.3%) | $0.850_{\pm 0.025}$ | $0.069_{\pm 0.000}$ | 0.014(2.9%) | $0.950_{\pm 0.031}$ | $0.069_{\pm 0.000}$ | 0.007(1.5%) | $1.000_{\pm 0.000}$ | $0.068_{\pm 0.000}$ |
| MIWAE | 0.012(2.5%) | $0.975_{\pm 0.025}$ | $0.073_{\pm 0.000}$ | 0.008(1.6%) | $1.000_{\pm 0.000}$ | $0.072_{\pm 0.000}$ | 0.011(2.2%) | $0.950_{\pm 0.031}$ | $0.073_{\pm 0.000}$ | 0.006(1.1%) | $1.000_{\pm 0.000}$ | $0.071_{\pm 0.000}$ |
| not-MIWAE | 0.017(3.5%) | $0.900_{\pm 0.047}$ | $0.075_{\pm 0.001}$ | 0.013(2.8%) | $0.925_{\pm 0.031}$ | $0.074_{\pm 0.001}$ | 0.019(4.1%) | $0.900_{\pm 0.047}$ | $0.074_{\pm 0.001}$ | 0.010(2.0%) | $0.950_{\pm 0.031}$ | $0.073_{\pm 0.001}$ |
| U-VAE | 0.010(2.2%) | $1.000_{\pm 0.000}$ | $0.073_{\pm 0.000}$ | 0.011(2.5%) | $0.925_{\pm 0.050}$ | $0.072_{\pm 0.000}$ | 0.015(3.3%) | $0.950_{\pm 0.031}$ | $0.073_{\pm 0.000}$ | 0.007(1.5%) | $1.000_{\pm 0.000}$ | $0.071_{\pm 0.000}$ |

Table 26: **Multiple imputation performance** on the `concrete` dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| default | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.023(8.8) | $0.443_{\pm 0.014}$ | $0.012_{\pm 0.000}$ | 0.017(6.2) | $0.614_{\pm 0.036}$ | $0.011_{\pm 0.000}$ | 0.025(9.0) | $0.343_{\pm 0.073}$ | $0.012_{\pm 0.000}$ | 0.012(4.5) | $0.671_{\pm 0.036}$ | $0.011_{\pm 0.000}$ |
| GAIN | 0.036(11.7) | $0.071_{\pm 0.023}$ | $0.011_{\pm 0.000}$ | 0.028(8.7) | $0.386_{\pm 0.070}$ | $0.011_{\pm 0.000}$ | 0.048(14.2) | $0.086_{\pm 0.035}$ | $0.011_{\pm 0.000}$ | 0.025(7.1) | $0.500_{\pm 0.051}$ | $0.011_{\pm 0.000}$ |
| VAEAC | 0.020(7.4) | $0.429_{\pm 0.000}$ | $0.012_{\pm 0.000}$ | 0.016(5.9) | $0.500_{\pm 0.060}$ | $0.012_{\pm 0.000}$ | 0.020(7.6) | $0.443_{\pm 0.042}$ | $0.012_{\pm 0.000}$ | 0.012(4.3) | $0.686_{\pm 0.043}$ | $0.011_{\pm 0.000}$ |
| MIWAE | 0.012(4.6) | $0.371_{\pm 0.069}$ | $0.014_{\pm 0.000}$ | 0.011(4.1) | $0.586_{\pm 0.086}$ | $0.013_{\pm 0.001}$ | 0.015(5.7) | $0.443_{\pm 0.027}$ | $0.014_{\pm 0.001}$ | 0.007(2.9) | $0.671_{\pm 0.036}$ | $0.012_{\pm 0.000}$ |
| not-MIWAE | 0.015(5.9) | $0.314_{\pm 0.043}$ | $0.013_{\pm 0.000}$ | 0.013(5.0) | $0.557_{\pm 0.042}$ | $0.014_{\pm 0.001}$ | 0.019(6.9) | $0.257_{\pm 0.080}$ | $0.013_{\pm 0.000}$ | 0.009(3.5) | $0.600_{\pm 0.036}$ | $0.012_{\pm 0.000}$ |
| U-VAE | 0.008(3.0) | $0.671_{\pm 0.077}$ | $0.016_{\pm 0.001}$ | 0.007(2.8) | $0.686_{\pm 0.058}$ | $0.014_{\pm 0.001}$ | 0.006(2.5) | $0.686_{\pm 0.048}$ | $0.015_{\pm 0.001}$ | 0.004(1.5) | $0.800_{\pm 0.042}$ | $0.012_{\pm 0.000}$ |

Table 27: **Multiple imputation performance** on the `default` dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| loan | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.016(4.1) | $0.520_{\pm 0.049}$ | $0.030_{\pm 0.000}$ | 0.011(2.5) | $0.640_{\pm 0.075}$ | $0.030_{\pm 0.000}$ | 0.023(5.7) | $0.400_{\pm 0.063}$ | $0.030_{\pm 0.000}$ | 0.013(3.5) | $0.600_{\pm 0.063}$ | $0.030_{\pm 0.000}$ |
| GAIN | 0.073(18.2) | $0.120_{\pm 0.120}$ | $0.030_{\pm 0.000}$ | 0.039(9.1) | $0.560_{\pm 0.040}$ | $0.030_{\pm 0.000}$ | 0.087(21.0) | $0.040_{\pm 0.040}$ | $0.030_{\pm 0.000}$ | 0.035(8.7) | $0.520_{\pm 0.049}$ | $0.030_{\pm 0.000}$ |
| VAEAC | 0.019(4.6) | $0.480_{\pm 0.120}$ | $0.031_{\pm 0.000}$ | 0.009(2.2) | $0.800_{\pm 0.063}$ | $0.031_{\pm 0.000}$ | 0.016(4.0) | $0.560_{\pm 0.147}$ | $0.033_{\pm 0.002}$ | 0.016(4.0) | $0.480_{\pm 0.080}$ | $0.031_{\pm 0.000}$ |
| MIWAE | 0.015(3.9) | $0.760_{\pm 0.040}$ | $0.034_{\pm 0.001}$ | 0.004(1.0) | $0.960_{\pm 0.040}$ | $0.035_{\pm 0.001}$ | 0.016(4.0) | $0.720_{\pm 0.049}$ | $0.037_{\pm 0.001}$ | 0.015(3.7) | $0.800_{\pm 0.063}$ | $0.033_{\pm 0.001}$ |
| not-MIWAE | 0.015(3.6) | $0.680_{\pm 0.080}$ | $0.039_{\pm 0.002}$ | 0.011(2.8) | $0.720_{\pm 0.080}$ | $0.035_{\pm 0.001}$ | 0.015(3.6) | $0.720_{\pm 0.136}$ | $0.039_{\pm 0.001}$ | 0.014(3.5) | $0.800_{\pm 0.063}$ | $0.036_{\pm 0.002}$ |
| U-VAE | 0.005(1.3) | $0.960_{\pm 0.040}$ | $0.034_{\pm 0.001}$ | 0.004(0.9) | $1.000_{\pm 0.000}$ | $0.035_{\pm 0.001}$ | 0.006(1.5) | $0.960_{\pm 0.040}$ | $0.036_{\pm 0.001}$ | 0.005(1.3) | $0.960_{\pm 0.040}$ | $0.034_{\pm 0.001}$ |

Table 28: **Multiple imputation performance** on the `loan` dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| letter | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.010 (2.1) | $0.400_{\pm 0.025}$ | $0.015_{\pm 0.000}$ | 0.010 (2.1) | $0.550_{\pm 0.012}$ | $0.015_{\pm 0.000}$ | 0.014 (2.9) | $0.288_{\pm 0.042}$ | $0.015_{\pm 0.000}$ | 0.006 (1.3) | $0.688_{\pm 0.044}$ | $0.015_{\pm 0.000}$ |
| GAIN | 0.062 (13.8) | $0.175_{\pm 0.023}$ | $0.015_{\pm 0.000}$ | 0.045 (10.1) | $0.450_{\pm 0.070}$ | $0.015_{\pm 0.000}$ | 0.094 (22.6) | $0.100_{\pm 0.025}$ | $0.015_{\pm 0.000}$ | 0.024 (5.4) | $0.525_{\pm 0.015}$ | $0.015_{\pm 0.000}$ |
| VAEAC | 0.010 (2.2) | $0.438_{\pm 0.048}$ | $0.017_{\pm 0.001}$ | 0.017 (3.2) | $0.538_{\pm 0.032}$ | $0.015_{\pm 0.000}$ | 0.020 (4.3) | $0.462_{\pm 0.051}$ | $0.016_{\pm 0.001}$ | 0.007 (1.6) | $0.650_{\pm 0.042}$ | $0.015_{\pm 0.000}$ |
| MIWAE | 0.032 (6.5) | $0.575_{\pm 0.091}$ | $0.039_{\pm 0.006}$ | 0.028 (6.4) | $0.650_{\pm 0.042}$ | $0.034_{\pm 0.009}$ | 0.039 (7.9) | $0.425_{\pm 0.031}$ | $0.044_{\pm 0.008}$ | 0.024 (4.7) | $0.600_{\pm 0.108}$ | $0.025_{\pm 0.004}$ |
| not-MIWAE | 0.031 (6.3) | $0.475_{\pm 0.076}$ | $0.039_{\pm 0.008}$ | 0.039 (8.8) | $0.550_{\pm 0.036}$ | $0.031_{\pm 0.009}$ | 0.043 (9.7) | $0.312_{\pm 0.040}$ | $0.033_{\pm 0.010}$ | 0.021 (4.7) | $0.638_{\pm 0.023}$ | $0.021_{\pm 0.003}$ |
| U-VAE | 0.011 (2.3) | $0.650_{\pm 0.073}$ | $0.021_{\pm 0.002}$ | 0.018 (4.2) | $0.675_{\pm 0.012}$ | $0.030_{\pm 0.010}$ | 0.024 (5.6) | $0.600_{\pm 0.064}$ | $0.022_{\pm 0.004}$ | 0.006 (1.4) | $0.762_{\pm 0.023}$ | $0.019_{\pm 0.002}$ |

Table 29: **Multiple imputation performance** on the `letter` dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| kings | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.030 (13.7) | $0.345_{\pm 0.018}$ | $0.014_{\pm 0.000}$ | 0.014 (6.2) | $0.582_{\pm 0.046}$ | $0.013_{\pm 0.000}$ | 0.030 (13.6) | $0.327_{\pm 0.055}$ | $0.014_{\pm 0.000}$ | 0.021 (10.5) | $0.582_{\pm 0.055}$ | $0.014_{\pm 0.000}$ |
| GAIN | 0.045 (15.6) | $0.036_{\pm 0.022}$ | $0.013_{\pm 0.000}$ | 0.032 (9.1) | $0.382_{\pm 0.034}$ | $0.013_{\pm 0.000}$ | 0.059 (18.9) | $0.055_{\pm 0.036}$ | $0.013_{\pm 0.000}$ | 0.028 (11.3) | $0.382_{\pm 0.078}$ | $0.013_{\pm 0.000}$ |
| VAEAC | 0.027 (12.7) | $0.436_{\pm 0.034}$ | $0.014_{\pm 0.000}$ | 0.015 (6.1) | $0.582_{\pm 0.036}$ | $0.013_{\pm 0.000}$ | 0.023 (11.6) | $0.509_{\pm 0.036}$ | $0.014_{\pm 0.000}$ | 0.022 (10.7) | $0.582_{\pm 0.022}$ | $0.014_{\pm 0.000}$ |
| MIWAE | 0.008 (5.9) | $0.618_{\pm 0.034}$ | $0.015_{\pm 0.000}$ | 0.008 (3.2) | $0.855_{\pm 0.022}$ | $0.015_{\pm 0.000}$ | 0.012 (6.8) | $0.582_{\pm 0.079}$ | $0.015_{\pm 0.000}$ | 0.005 (4.1) | $0.727_{\pm 0.029}$ | $0.014_{\pm 0.000}$ |
| not-MIWAE | 0.010 (6.1) | $0.527_{\pm 0.018}$ | $0.015_{\pm 0.000}$ | 0.010 (4.0) | $0.673_{\pm 0.046}$ | $0.014_{\pm 0.000}$ | 0.014 (6.3) | $0.455_{\pm 0.029}$ | $0.015_{\pm 0.000}$ | 0.009 (4.4) | $0.636_{\pm 0.029}$ | $0.014_{\pm 0.000}$ |
| U-VAE | 0.006 (3.7) | $0.655_{\pm 0.034}$ | $0.014_{\pm 0.000}$ | 0.007 (2.8) | $0.800_{\pm 0.045}$ | $0.014_{\pm 0.000}$ | 0.011 (5.5) | $0.600_{\pm 0.055}$ | $0.015_{\pm 0.000}$ | 0.004 (2.8) | $0.745_{\pm 0.018}$ | $0.014_{\pm 0.000}$ |

Table 30: **Multiple imputation performance** on the `kings` dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| redwine | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.021 (5.3) | $0.782_{\pm 0.022}$ | $0.054_{\pm 0.000}$ | 0.011 (2.7) | $0.909_{\pm 0.041}$ | $0.054_{\pm 0.000}$ | 0.020 (5.2) | $0.745_{\pm 0.034}$ | $0.054_{\pm 0.000}$ | 0.014 (3.8) | $0.836_{\pm 0.034}$ | $0.054_{\pm 0.000}$ |
| GAIN | 0.035 (8.1) | $0.473_{\pm 0.067}$ | $0.054_{\pm 0.000}$ | 0.028 (6.2) | $0.582_{\pm 0.036}$ | $0.053_{\pm 0.000}$ | 0.047 (11.0) | $0.291_{\pm 0.034}$ | $0.054_{\pm 0.000}$ | 0.018 (4.7) | $0.727_{\pm 0.081}$ | $0.054_{\pm 0.000}$ |
| VAEAC | 0.020 (5.2) | $0.745_{\pm 0.034}$ | $0.055_{\pm 0.000}$ | 0.016 (3.7) | $0.764_{\pm 0.074}$ | $0.054_{\pm 0.000}$ | 0.021 (5.2) | $0.745_{\pm 0.067}$ | $0.054_{\pm 0.000}$ | 0.013 (3.5) | $0.818_{\pm 0.050}$ | $0.054_{\pm 0.000}$ |
| MIWAE | 0.013 (3.5) | $0.891_{\pm 0.034}$ | $0.061_{\pm 0.001}$ | 0.008 (2.0) | $0.964_{\pm 0.022}$ | $0.058_{\pm 0.001}$ | 0.015 (4.1) | $0.855_{\pm 0.022}$ | $0.062_{\pm 0.001}$ | 0.009 (2.6) | $0.927_{\pm 0.034}$ | $0.059_{\pm 0.001}$ |
| not-MIWAE | 0.015 (3.9) | $0.927_{\pm 0.034}$ | $0.062_{\pm 0.001}$ | 0.009 (2.2) | $0.964_{\pm 0.022}$ | $0.059_{\pm 0.000}$ | 0.013 (3.4) | $0.909_{\pm 0.029}$ | $0.062_{\pm 0.000}$ | 0.010 (2.6) | $0.891_{\pm 0.018}$ | $0.059_{\pm 0.001}$ |
| U-VAE | 0.009 (2.3) | $0.982_{\pm 0.018}$ | $0.061_{\pm 0.000}$ | 0.007 (1.7) | $1.000_{\pm 0.000}$ | $0.058_{\pm 0.001}$ | 0.011 (3.0) | $0.982_{\pm 0.018}$ | $0.060_{\pm 0.001}$ | 0.007 (1.7) | $0.945_{\pm 0.022}$ | $0.057_{\pm 0.000}$ |

Table 31: **Multiple imputation performance** on the `redwine` dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| shoppers | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.041 (17.2) | $0.120_{\pm 0.037}$ | $0.017_{\pm 0.000}$ | 0.023 (9.5) | $0.420_{\pm 0.020}$ | $0.017_{\pm 0.000}$ | 0.043 (17.8) | $0.060_{\pm 0.024}$ | $0.018_{\pm 0.000}$ | 0.020 (9.0) | $0.420_{\pm 0.049}$ | $0.017_{\pm 0.000}$ |
| GAIN | 0.048 (17.2) | $0.120_{\pm 0.037}$ | $0.017_{\pm 0.000}$ | 0.028 (10.8) | $0.380_{\pm 0.037}$ | $0.017_{\pm 0.000}$ | 0.056 (21.4) | $0.014_{\pm 0.014}$ | $0.017_{\pm 0.000}$ | 0.025 (11.5) | $0.429_{\pm 0.047}$ | $0.017_{\pm 0.000}$ |
| VAEAC | 0.029 (12.9) | $0.260_{\pm 0.051}$ | $0.017_{\pm 0.000}$ | 0.023 (9.5) | $0.420_{\pm 0.049}$ | $0.017_{\pm 0.000}$ | 0.029 (12.5) | $0.320_{\pm 0.037}$ | $0.017_{\pm 0.000}$ | 0.016 (7.6) | $0.560_{\pm 0.024}$ | $0.017_{\pm 0.000}$ |
| MIWAE | 0.014 (7.6) | $0.440_{\pm 0.075}$ | $0.018_{\pm 0.000}$ | 0.010 (4.6) | $0.560_{\pm 0.075}$ | $0.018_{\pm 0.000}$ | 0.015 (7.9) | $0.420_{\pm 0.086}$ | $0.018_{\pm 0.000}$ | 0.008 (4.9) | $0.640_{\pm 0.024}$ | $0.017_{\pm 0.000}$ |
| not-MIWAE | 0.018 (7.7) | $0.220_{\pm 0.058}$ | $0.019_{\pm 0.000}$ | 0.013 (5.2) | $0.480_{\pm 0.020}$ | $0.019_{\pm 0.001}$ | 0.018 (7.3) | $0.260_{\pm 0.068}$ | $0.020_{\pm 0.001}$ | 0.013 (5.0) | $0.440_{\pm 0.040}$ | $0.018_{\pm 0.000}$ |
| U-VAE | 0.006 (3.0) | $0.760_{\pm 0.103}$ | $0.018_{\pm 0.000}$ | 0.006 (2.4) | $0.800_{\pm 0.045}$ | $0.018_{\pm 0.000}$ | 0.008 (3.5) | $0.660_{\pm 0.040}$ | $0.018_{\pm 0.000}$ | 0.004 (1.9) | $0.860_{\pm 0.024}$ | $0.018_{\pm 0.000}$ |

Table 32: **Multiple imputation performance** on the `shoppers` dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

| whitewine | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCAR | | | MAR | | | MNARL | | | MNARQ | | |
| model | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ | Bias↓ | Coverage | Width↓ |
| MICE | 0.013 (2.8) | $0.745_{\pm 0.045}$ | $0.031_{\pm 0.000}$ | 0.010 (2.2) | $0.800_{\pm 0.045}$ | $0.031_{\pm 0.000}$ | 0.014 (3.0) | $0.655_{\pm 0.060}$ | $0.031_{\pm 0.000}$ | 0.008 (1.8) | $0.782_{\pm 0.022}$ | $0.031_{\pm 0.000}$ |
| GAIN | 0.041 (8.8) | $0.273_{\pm 0.064}$ | $0.031_{\pm 0.000}$ | 0.035 (7.3) | $0.491_{\pm 0.084}$ | $0.031_{\pm 0.000}$ | 0.040 (8.5) | $0.309_{\pm 0.062}$ | $0.031_{\pm 0.000}$ | 0.014 (3.0) | $0.673_{\pm 0.102}$ | $0.031_{\pm 0.000}$ |
| VAEAC | 0.013 (3.0) | $0.673_{\pm 0.062}$ | $0.032_{\pm 0.000}$ | 0.013 (2.7) | $0.673_{\pm 0.036}$ | $0.031_{\pm 0.000}$ | 0.014 (3.1) | $0.636_{\pm 0.057}$ | $0.031_{\pm 0.000}$ | 0.010 (2.2) | $0.709_{\pm 0.045}$ | $0.031_{\pm 0.000}$ |
| MIWAE | 0.009 (2.0) | $0.873_{\pm 0.036}$ | $0.037_{\pm 0.001}$ | 0.008 (1.7) | $0.891_{\pm 0.034}$ | $0.035_{\pm 0.001}$ | 0.010 (2.2) | $0.891_{\pm 0.045}$ | $0.038_{\pm 0.001}$ | 0.006 (1.3) | $0.909_{\pm 0.029}$ | $0.034_{\pm 0.001}$ |
| not-MIWAE | 0.010 (2.2) | $0.891_{\pm 0.045}$ | $0.037_{\pm 0.001}$ | 0.009 (1.9) | $0.873_{\pm 0.062}$ | $0.035_{\pm 0.001}$ | 0.012 (2.8) | $0.764_{\pm 0.062}$ | $0.040_{\pm 0.001}$ | 0.007 (1.6) | $0.836_{\pm 0.018}$ | $0.036_{\pm 0.001}$ |
| U-VAE | 0.011 (2.4) | $0.927_{\pm 0.053}$ | $0.039_{\pm 0.001}$ | 0.006 (1.4) | $0.927_{\pm 0.018}$ | $0.035_{\pm 0.001}$ | 0.010 (2.2) | $0.873_{\pm 0.022}$ | $0.037_{\pm 0.001}$ | 0.005 (1.1) | $0.945_{\pm 0.022}$ | $0.036_{\pm 0.001}$ |

Table 33: **Multiple imputation performance** on the `whitewine` dataset at 0.3 missingness rate. The means and their standard errors across 5 random seeds are reported. The values in parentheses next to Bias represent the percent bias. ↓ denotes lower is better. Coverage close to 0.95 indicates better performance.

## References


[1] Minwoo Chae, Dongha Kim, Yongdai Kim, and Lizhen Lin. A likelihood approach to nonparametric estimation of a singular distribution using deep generative models. *Journal of Machine Learning Research*, 24(77):1–42, 2023.

[2] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). In Yoshua Bengio and Yann LeCun, editors, *ICLR (Poster)*, 2016.

[3] Juan Colonna, Eduardo Nakamura, Marco Cristo, and Marcelo Gordo. Anuran Calls (MFCCs). UCI Machine Learning Repository, 2015. DOI: https://doi.org/10.24432/C5CC9H.

[4] Paulo Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis. Wine Quality. UCI Machine Learning Repository, 2009. DOI: https://doi.org/10.24432/C56S3T.

[5] Bin Dai and David Wipf. Diagnosing and enhancing VAE models. In *International Conference on Learning Representations*, 2019.

[6] Tianyu Du, Luca Melis, and Ting Wang. Remasker: Imputing tabular data with masked autoencoding. In *The Twelfth International Conference on Learning Representations*, 2024.

[7] Jan Gasthaus, Konstantinos Benidis, Yuyang Wang, Syama Sundar Rangapuram, David Salinas, Valentin Flunkert, and Tim Januschowski. Probabilistic forecasting with spline quantile function rnns. In *The 22nd international conference on artificial intelligence and statistics*, pages 1901–1910. PMLR, 2019.

[8] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017.

[9] Niels Bruun Ipsen, Pierre-Alexandre Mattei, and Jes Frellsen. not-{miwae}: Deep generative modelling with missing not at random data. In *International Conference on Learning Representations*, 2021.

[10] Oleg Ivanov, Michael Figurnov, and Dmitry Vetrov. Variational autoencoder with arbitrary conditioning. In *International Conference on Learning Representations*, 2019.

[11] Daniel Jarrett, Bogdan C Cebere, Tennison Liu, Alicia Curth, and Mihaela van der Schaar. HyperImpute: Generalized iterative imputation with automatic model selection. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 9916–9937. PMLR, 17–23 Jul 2022.

[12] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[13] Trent Kyono, Yao Zhang, Alexis Bellot, and Mihaela van der Schaar. MIRACLE: Causally-aware imputation via learning missing data mechanisms. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.

[14] R.J.A. Little and D.B. Rubin. *Statistical analysis with missing data.* Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley, 2002.

[15] Volker Lohweg. Banknote Authentication. UCI Machine Learning Repository, 2012. DOI: https://doi.org/10.24432/C55P57.

[16] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2017.

[17] James Lucas, G. Tucker, Roger Baker Grosse, and Mohammad Norouzi. Don't blame the elbo! a linear vae perspective on posterior collapse. In *Neural Information Processing Systems*, 2019.

[18] Pierre-Alexandre Mattei and Jes Frellsen. Miwae: Deep generative modelling and imputation of incomplete data sets. In *International Conference on Machine Learning*, 2019.

[19] Boris Muzellec, Julie Josse, Claire Boyer, and Marco Cuturi. Missing data imputation using optimal transport. In *International Conference on Machine Learning*, pages 7130–7140. PMLR, 2020.

[20] Donald B. Rubin and Nathaniel Schenker. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, 81:366–374, 1986.

[21] C. Sakar and Yomi Kastro. Online Shoppers Purchasing Intention Dataset. UCI Machine Learning Repository, 2018. DOI: https://doi.org/10.24432/C5F88Q.

[22] David Slate. Letter Recognition. UCI Machine Learning Repository, 1991. DOI: https://doi.org/10.24432/C5ZP40.

[23] Daniel J Stekhoven and Peter Bühlmann. Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, 2012.

[24] Olga G. Troyanskaya, Michael N. Cantor, Gavin Sherlock, Patrick O. Brown, Trevor J. Hastie, Robert Tibshirani, David Botstein, and Russ B. Altman. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17 6:520–5, 2001.

[25] Stef van Buuren. Flexible imputation of missing data. 2012.

[26] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011.

[27] William Wolberg, Olvi Mangasarian, Nick Street, , and W. Street. Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository, 1993. DOI: https://doi.org/10.24432/C5DW2B.

[28] I-Cheng Yeh. Concrete Compressive Strength. UCI Machine Learning Repository, 1998. DOI: https://doi.org/10.24432/C5PK67.

[29] I-Cheng Yeh. Default of Credit Card Clients. UCI Machine Learning Repository, 2009. DOI: https://doi.org/10.24432/C55S3H.

[30] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. GAIN: Missing data imputation using generative adversarial nets. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5689–5698. PMLR, 10–15 Jul 2018.

[31] He Zhao, Ke Sun, Amir Dezfouli, and Edwin V. Bonilla. Transformed distribution matching for missing value imputation. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 42159–42186. PMLR, 23–29 Jul 2023.