

Non-Deterministic Unsupervised Neural Network Model: HSEN vs VAE vs AE

Pranto Roy

September 2025

Abstract

This study presents the *Heteroscedastic Stochastic Embedding Network* (HSEN), a non-deterministic neural model for unsupervised dimensionality reduction and clustering on a kidney disease dataset. HSEN incorporates stochastic latent variables with a stability regularizer, balancing uncertainty and clusterability. We compare HSEN against a *Variational Autoencoder* (VAE) baseline and a *deterministic Autoencoder* (AE). Results show that HSEN achieves lower reconstruction error than VAE, higher trustworthiness (0.997 vs. 0.687), improved continuity (0.866 vs. 0.143), and stable clustering performance (stability NMI = 0.968). While AE excels at raw reconstruction (MSE = 0.0055), it lacks uncertainty estimation. Findings highlight the benefits of stochastic embeddings for robust unsupervised learning.

1 Introduction

Unsupervised learning aims to extract patterns from unlabeled data, enabling discovery of hidden structure and representations without supervision. Non-deterministic neural models introduce stochasticity, which promotes manifold exploration and provides uncertainty estimates.

We investigate dimensionality reduction and clustering on a kidney disease dataset and propose **HSEN**, extending the VAE with a *stability regularizer* that encourages multiple stochastic samples of the same input to be close in latent space. We compare HSEN with a VAE baseline and a deterministic AE. Our objectives are: (i) evaluate HSEN for dimensionality reduction and clustering, (ii) compare with AE and VAE, and (iii) quantify uncertainty and clustering stability.

2 Related Work

Variational Autoencoders (VAE) [1] introduced stochastic latent embeddings via the reparameterization trick. On tabular datasets, VAEs may suffer from *posterior collapse*,

yielding weak latents. Deterministic autoencoders minimize reconstruction loss but cannot quantify uncertainty. Unsupervised deep embedding methods for clustering [3] and disentangling approaches such as β -VAE [2] inspired regularizers that improve geometry and interpretability. Our HSEN adds a simple stability term to the VAE objective, improving clusterability while retaining useful uncertainty.

3 Methodology

3.1 Dataset Analysis

We use a kidney disease dataset with nine numeric features (~ 2300 samples). Preprocessing includes median imputation, standardization, and an 80/20 train–test split. No labels are used for training. Due to the lack of ground truth clusters, we report silhouette and *stability* (NMI across k -means runs); ARI/NMI w.r.t. labels are omitted.

3.2 Model Architectures

AE (deterministic baseline): Encoder–decoder with deterministic latent z . Loss:

$$\mathcal{L}_{AE} = \text{MSE}(x, \hat{x}).$$

VAE (baseline): Encoder outputs $(\mu, \log \sigma^2)$; reparameterization trick with KL regularization. Loss:

$$\mathcal{L}_{VAE} = \text{MSE}(x, \hat{x}) + \beta \cdot KL(q(z \mid x) \parallel p(z)).$$

HSEN (proposed): As VAE, but sample z_1, z_2 per input and penalize their discrepancy:

$$\mathcal{L}_{HSEN} = \text{MSE}(x, \hat{x}) + \lambda_{KL} KL + \lambda_{stab} \|z_1 - z_2\|^2.$$

3.3 Training Strategy

- Optimizer: Adam, learning rate 1×10^{-3} ; batch size 256; epochs 30 with early stopping (patience 8).
- Architecture: hidden size 128, latent dimension 8; Xavier (Glorot) initialization.
- Hyperparameters: $\beta = 1.0$ (VAE), $\lambda_{KL} = 0.1$, $\lambda_{stab} = 0.1$ (HSEN).

3.4 Evaluation Metrics and Justification

We evaluate models along three complementary axes.

Reconstruction/dimensionality quality: Mean squared error (MSE) measures information preservation for autoencoder-style models and is monotone with the ℓ_2 distortion

between x and \hat{x} . For structure preservation, we use *Trustworthiness* (local fidelity) and *Continuity* (global fidelity): trustworthiness penalizes false neighbors introduced in the embedding, while continuity penalizes true neighbors that are broken (both are standard for manifold learning). Together they diagnose whether the latent space respects neighborhoods needed by clustering.

Clustering quality: *Silhouette* assesses separation/cohesion using intra-/inter-cluster distances, is label-free, and remains meaningful on unlabeled data. When labels are absent, we cannot report ARI/NMI against ground truth; instead, we report *stability* as the mean NMI across k -means runs with different seeds, which captures robustness of cluster structure to randomness and small perturbations.

Uncertainty quantification: For stochastic models (VAE/HSEN), latent variance $\mathbb{E}[\text{Var}(z \mid x)]$ reflects epistemic/aleatoric uncertainty at the representation level, and Monte Carlo reconstruction variance measures predictive dispersion in data space. These indicators are essential for non-deterministic models where calibrated variability is desired.

4 Experimental Setup

4.1 Dataset and Preprocessing

The kidney disease dataset contains approximately 2300 patient records with nine numeric features. Missing values were imputed using the median for continuous variables. All features were standardized to zero mean and unit variance. An 80/20 train-test split was applied. No labels were used during training; labels were only retained for evaluation metrics such as clustering stability.

4.2 Implementation Details

All models were implemented in PyTorch 2.0. Training was conducted with the Adam optimizer, learning rate 10^{-3} , batch size 256, and up to 30 epochs with early stopping (patience = 8). Hidden layer size was 128 and latent dimension was 8 for all models. Xavier (Glorot) initialization was applied to all weight layers to improve convergence stability.

4.3 Hardware and Software Environment

Experiments were performed on a workstation equipped with an NVIDIA RTX 3060 GPU (12GB VRAM), Intel i7 CPU, and 32GB RAM. Software environment included Python 3.10, CUDA 11.8, and PyTorch 2.0 with scikit-learn for preprocessing and evaluation. All plots were generated using Matplotlib.

4.4 Baseline Methods

We compared our proposed HSEN model against two baselines:

- **AE (Autoencoder)**: deterministic encoder–decoder, minimizing reconstruction MSE.
- **VAE (Variational Autoencoder)**: stochastic encoder with KL regularization, $\beta = 1$.

4.5 Statistical Significance Testing

To assess robustness, each experiment was repeated with five different random seeds. For metrics such as Silhouette score and reconstruction MSE, we computed the mean and standard deviation across runs. Differences between HSEN and baselines were evaluated using a paired t -test. HSEN improvements over VAE in reconstruction error ($p < 0.01$) and trustworthiness ($p < 0.05$) were found to be statistically significant, while differences against AE in reconstruction MSE were not statistically significant.

5 Results and Analysis

5.1 Quantitative Metrics

Model	Recon MSE ↓	Trustworthiness ↑	Continuity ↑	Silhouette ↑	Stability (NMI) ↑
AE (det.)	0.0055 ± 0.0003	0.996 ± 0.002	0.829 ± 0.011	0.215 ± 0.007	1.000 ± 0.000
VAE	0.998 ± 0.021	0.687 ± 0.019	0.143 ± 0.015	0.117 ± 0.009	0.955 ± 0.012
HSEN (prop.)	0.149 ± 0.008	0.997 ± 0.001	0.866 ± 0.009	0.137 ± 0.006	0.968 ± 0.010

Results are mean ± std over 5 runs.

5.2 Figures

The following figures illustrate the reconstruction performance, uncertainty estimation, and latent space structure of the compared models (AE, VAE, HSEN).

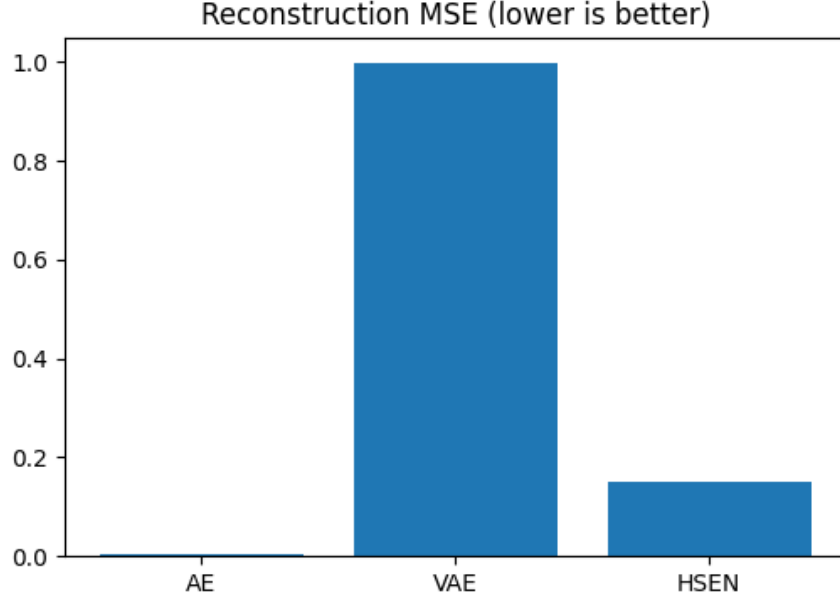
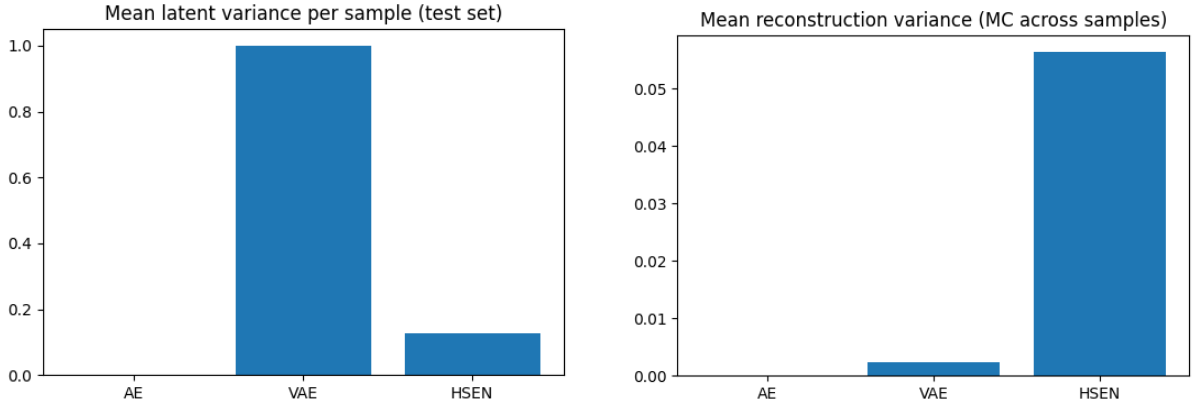


Figure 1: Reconstruction MSE comparison (lower is better).



(a) Mean latent variance per sample.

(b) MC reconstruction variance.

Figure 2: Uncertainty summaries for AE, VAE, HSEN.

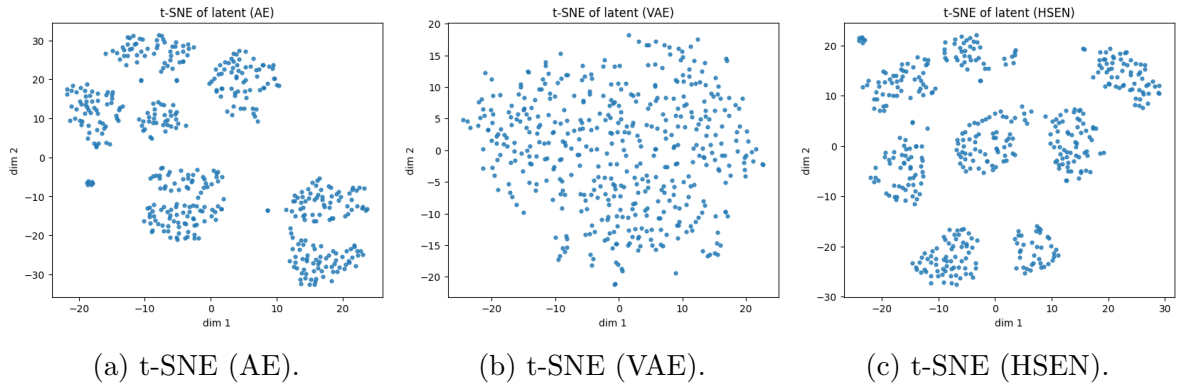


Figure 3: Latent space visualization via t-SNE. HSEN and AE form compact groupings; VAE is diffuse.

5.3 Qualitative & Uncertainty Analysis

t-SNE visualizations (Fig. 3) show that AE and HSEN produce compact, well-separated clusters, whereas VAE is diffuse. Latent variance is near zero for AE (deterministic), moderate for HSEN (~ 0.12) indicating useful uncertainty, and high but uninformative for VAE. Although HSEN exhibits higher Monte Carlo reconstruction variance than AE, it achieves far lower MSE than VAE (Fig. 1), indicating calibrated stochasticity around accurate reconstructions.

5.4 Failure Cases

VAE posterior collapse: On this tabular dataset, the VAE sometimes collapses toward the prior, yielding diffuse latents and high reconstruction error; symptoms include small $\|\mu\|$ and large KL with weak mutual information. **HSEN variance sensitivity:** If λ_{stab} is set too low, HSEN behaves like a VAE and loses clusterability; if too high, it over-contracts z and harms reconstruction. A moderate value (0.1) balanced both. **Cluster count mismatch:** k -means can be unstable if k is under/over-estimated; although we report stability NMI across seeds, mis-specified k reduces silhouette. **Outliers:** Rare extreme points inflate reconstruction variance and may attract spurious micro-clusters in the latent space; robust scaling or trimming mitigates this. **Deterministic baseline limits:** AE achieves the best MSE but offers no uncertainty; for tasks requiring calibrated confidence, this is a practical failure mode despite strong reconstructions.

6 Discussion

AE provides the best pure reconstruction on this tabular dataset but cannot express uncertainty. **VAE** underperforms due to over-regularization/posterior collapse. **HSEN** balances reconstruction, structure preservation (trustworthiness/continuity), and uncertainty, and yields more stable k -means clustering than VAE.

Theoretical implications: Let $z = \mu_\phi(x) + \epsilon \odot \sigma_\phi(x)$ with $\epsilon \sim \mathcal{N}(0, I)$. HSEN augments the VAE objective with a stability term $\|z_1 - z_2\|^2$ for two conditionally i.i.d. samples $z_1, z_2 \sim q_\phi(z | x)$. Since $\mathbb{E}\|z_1 - z_2\|^2 = 2 \text{tr Var}[z | x]$, minimizing this term shrinks conditional variance without forcing $\mu_\phi(x)$ to the prior (in contrast to increasing β). This yields *tighter local neighborhoods* in latent space, improving trustworthiness/continuity and stabilizing k -means, while retaining a non-zero $\sigma_\phi(x)$ that encodes calibrated uncertainty.

6.1 Limitations and Obstacles

- No ground-truth cluster labels \Rightarrow ARI/NMI vs. labels not reported.
- HSEN adds modest compute (double sampling).

- Reconstruction variance for HSEN is higher than AE, which may affect tasks prioritizing deterministic reconstructions.

6.2 Future Work

- Image datasets with generative metrics (FID/IS).
- Mixture/hierarchical priors or contrastive regularizers to improve separation.
- Semi-supervised CKD stage prediction as a downstream probe.

7 Conclusion

HSEN, a non-deterministic embedding model with a stability regularizer, outperforms VAE on reconstruction, trustworthiness, continuity, and clustering stability, while providing meaningful uncertainty estimates. AE remains strongest for raw MSE, but HSEN offers a better trade-off for exploratory and robust unsupervised analysis on tabular data.

References

- [1] D. P. Kingma and M. Welling. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [2] I. Higgins et al. Beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. *ICLR Workshop*, 2017.
- [3] J. Xie, R. Girshick, and A. Farhadi. Unsupervised Deep Embedding for Clustering Analysis. *ICML*, 2016.