ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

mDAE: modified Denoising AutoEncoder for missing data imputation

Mariette Dupuy[1,2], Marie CHAVENT[1] and Remi Dubois[2]

1- Univ. Bordeaux, CNRS, INRIA, Bordeaux INP, IMB, UMR 5251, F-33400 Talence, France

2- Institute of Electrophysiology and Heart Modeling (IHU Liryc) Bordeaux University, France

Abstract. This paper introduces a method based on Denoising AutoEncoder (DAE) for missing data imputation. This method, called mDAE hereafter, results from a modification of the loss function and a straightforward procedure for choosing the hyper-parameters. An ablation study shows on several UCI Machine Learning Repository datasets, the benefit of using this modified loss function and an overcomplete structure, in terms of Root Mean Squared Error (RMSE) of reconstruction. This numerical study is complemented by a comparison between mDAE and eight alternative approaches (four classical and four more recent), using the Mean Distance to the Best (MDB) criterion, which quantifies the overall performance of each method across all the datasets.

1 Introduction

Denoising autoencoders (DAE), originally proposed to recover clean outputs from noisy inputs [1], are naturally well-suited for data imputation, treating missing values as a special case of noisy input [2]. Denoising autoencoders have thus frequently been used to reconstruct missing data in previous works, typically by applying a DAE to pre-imputed data (e.g., using mean imputation) [3]. But if pre-imputation addresses the issue of loss functions that cannot handle missing values, learning to reconstruct the pre-imputed values is not meaningful. Moreover, the review paper [3] notes that most of these studies do not justify their choices of DAE structure or hyperparameters.

In this paper, we propose a method called mDAE, in which the loss function is modified to ignore pre-imputed missing values. We also propose a grid-search methodology for selecting both the hyperparameter μ (the proportion of corruption) and the DAE architecture. This method is evaluated through an ablation study to assess the contribution of its key components: the modified loss function, the hyperparameter selection via cross-validation, and the use of an overcomplete structure. It is also compared with eight other imputation methods - four based on standard machine learning and four based on deep learning and optimal transport - using a metric called Mean Distance to the Best (MDB), which measures the overall performance of a method across all datasets for a given proportion and mechanism of missing data.

The implementation of the mDAE method is available at `https://github.com/mariette-dup/mDAE.git`.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

2 The mDAE method

AutoEncoders (AE) are well-known artificial neural networks used to learn efficient representation of unlabeled data via an encoding function and to recreate the input data via a decoding function. Here, we are dealing with tabular numerical data, and we suppose that these data have been normalized so that the p features have zero mean and unit variance. Let $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ be the set of n observations which form the rows of the standardized data matrix $\mathbf{X} = (x_{ij})_{n\times p}$. The encoding function f_θ of an autoencoder (with one hidden layer) transforms an input $\mathbf{x}_i \in \mathbb{R}^p$ into a latent vector $\mathbf{y}_i \in \mathbb{R}^q$:

$$\mathbf{y}_i = f_\theta(\mathbf{x}_i) = s(\mathbf{W}\mathbf{x}_i + \mathbf{b}),$$

where $\mathbf{W} \in \mathbb{R}^{q\times p}$ is a weight matrix, $\mathbf{b} \in \mathbb{R}^p$ is a bias vector and s is an activation function (e.g., ReLU or sigmoid). The decoding function $g_{\theta'}$ then transforms the latent vector $\mathbf{y}_i \in \mathbb{R}^q$ into an output $\mathbf{z}_i \in \mathbb{R}^p$:

$$\mathbf{z}_i = g_{\theta\prime}(\mathbf{y}_i) = \mathbf{W}'\mathbf{y}_i + \mathbf{b}',$$

where $\mathbf{W}' \in \mathbb{R}^{p\times q}$ and $\mathbf{b}' \in \mathbb{R}^q$. Note that this decoding function uses no activation function since we are trying to reconstruct values in $\mathbb{R}$. Parameters θ and θ' are learned by minimizing a reconstruction loss which is here:

$$\mathcal{L}_{AE} = \sum_{i=1}^{n} \|\mathbf{x}_i - (g_{\theta'} \circ f_\theta)(\mathbf{x}_i)\|^2 = \|\mathbf{X} - \mathbf{Z}\|_F^2, \tag{1}$$

where $\mathbf{Z}$ is the reconstructed data matrix and $\|.\|_F$ is the Fronenius norm. Note that this loss function favors the reconstruction of features (columns of $\mathbf{X}$) with high variance, hence the importance of data standardization.

Denoising AutoEncoders (DAE) are autoencoders defined to remove noise from a given input. To do this, autoencoders are trained to output the original data using corrupted data in the input. Here, each observation $\mathbf{x}_i$ is corrupted by randomly setting a proportion μ of its components to zero and the parameters θ and θ' are learned by minimizing the reconstruction loss:

$$\mathcal{L}_{DAE} = \sum_{i=1}^{n} \|\mathbf{x}_i - (g_{\theta'} \circ f_\theta)(N(\mathbf{x}_i))\|^2 = \|\mathbf{X} - \mathbf{Z}\|_F^2 \tag{2}$$

where $N(\mathbf{x}_i)$ denotes the corrupted version of $\mathbf{x}_i$. Note that μ is a hyperparameter that needs to be calibrated.

In the context of missing data imputation, denoising autoencoders have been widely used to reconstruct missing values by learning the parameters on pre-imputed data [3]. When missing values are pre-imputed with feature means (zeros here), the reconstruction loss writes:

$$\mathcal{L}_{DAE} = \sum_{i=1}^{n} \|\tilde{\mathbf{x}}_i - (g_{\theta'} \circ f_\theta)(N(\tilde{\mathbf{x}}_i))\|^2 = \|P_\Omega(\mathbf{X}) - \mathbf{Z}\|_F^2, \tag{3}$$

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

where $\tilde{\mathbf{X}} = P_\Omega(\mathbf{X})$ is the pre-imputed data matrix with

$$\tilde{\mathbf{x}}_{ij} = \begin{cases} x_{ij} & \text{if } (i,j) \in \Omega, \\ 0 & \text{if } (i,j) \notin \Omega. \end{cases} \tag{4}$$

and Ω is the set indices of non missing entries.

However, this approach learns to reconstruct zeros at the locations of the missing values, which is irrelevant. Our proposal is to modify the reconstruction loss to simply skip these locations:

$$\mathcal{L}_{mDAE} = \|P_\Omega(\mathbf{X}) - P_\Omega(\mathbf{Z})\|_F^2, \tag{5}$$

The imputed data matrix is then:

$$\hat{\mathbf{X}} = P_\Omega(\mathbf{X}) + P_{\Omega^\perp}(\mathbf{Z}), \tag{6}$$

where $\mathbf{Z}$ is the matrix reconstructed by minimization of the mDAE loss (5) and $\Omega^\perp$ is the set of indices of missing entries.

Choice of the hyper-parameter μ. The hyper-parameter μ is the proportion of zeros used to corrupt the observations $\mathbf{x}_i$. This hyper-parameter can be chosen either randomly in $[0, 1]$ or in a grid of values to minimize the error of reconstruction of the missing values. In that case, observed entries are split into a training set to learn the parameters and a validation set to estimate the error of reconstruction. This process is repeated several times, averaging the errors for more robust estimation.

Choice of the structure. Six structures (see Figure 1) are evaluated using the same procedure as for the choice of μ.

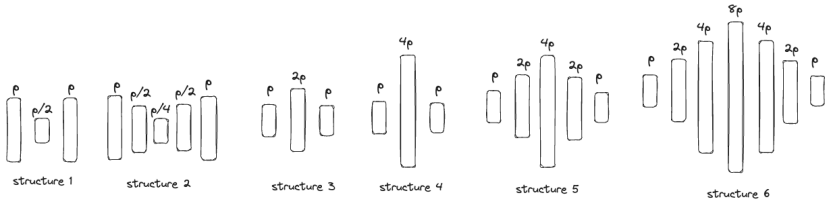

Fig. 1: Six structures where p is the number of units of the input layer.

3 Numerical study

All comparisons are made using seven datasets (see Table 1) chosen among 23 datasets of the UCI Machine Learning Repository recently used in [4] to compare several imputation methods. These 7 datasets were chosen to be all

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

numerical, of different sizes, and not too numerous, in order to avoid time-consuming experiments. To evaluate an imputation method, a certain proportion of each dataset is randomly replaced by missing values using the MCAR (Missing Completely At Random) process as implemented in the GitHub repository of [4]. Then, for a given mask $\Omega^{\perp}$ of artificial missing values, the performance of the method is evaluated by:

$$RMSE = \sqrt{\frac{1}{|\Omega^{\perp}|}|P_{\Omega^{\perp}}(\mathbf{X}) - P_{\Omega^{\perp}}(\mathbf{Z})|_F^2}. \tag{7}$$

Name	Abreviation	Rows	Columns
Breast cancer diagnostic	breast	569	30
Connectionist bench sonar	sonar	208	60
Ionosphere	iono	351	34
Blood transfusion	blood	748	4
Seeds	seeds	210	7
Climate model crashes	climate	540	18
Wine quality red	wine	1599	10

Table 1: UCI Machine Learning Repository datasets

3.1 Ablation study

An ablation study is a methodology used to evaluate the importance of different components of an algorithm, by comparing the results obtained with and without this component. Here, three components of the mDAE method are studied: the use of the modified reconstruction loss (5) rather than the standard loss (3); the use of an optimized value of the hyper-parameter μ rather than a value chosen randomly in $[0, 1]$; the use of an overcomplete structure (the 5th structure in Figure 1) rather than an undercomplete structure (the 2nd structure in Figure 1). The results of this ablation study (see Table 2) show that using the modified loss function clearly improves the reconstruction of the missing data for all datasets except the climate dataset. One can also observe that choosing the hyperparameter μ randomly in $[0, 1]$ does not significantly deteriorate the reconstruction quality, while saving computation time compared to selecting μ through grid optimization. Finally, using an overcomplete structure rather than an undercomplete one systematically improves the results.

3.2 Comparison with other methods

The mDAE method is now compared with eight other methods (see Table 3) using their respective optimal hyperparameter settings whenever possible.
Figure 2 shows that SoftImpute, missForest and mDAE methods perform reasonably well across all datasets, although no single method consistently outperforms the others. To assess overall performance across all datasets, a new metric called

[1] https://scikit-learn.org/stable/api/sklearn.impute.html

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Method	breast	climate	sonar	iono	seeds	wine	blood
mDAE	**0.466 ± 0.016**	1.007 ± 0.007	**0.656 ± 0.007**	**0.776 ± 0.018**	**0.496 ± 0.022**	**0.790 ± 0.030**	**0.701 ± 0.059**
mDAE w/o modified loss	0.685 ± 0.036 (46.996%)	**1.005 ± 0.008** (-0.199%)	0.988 ± 0.013 (50.610%)	0.808 ± 0.020 (4.124%)	0.587 ± 0.028 (18.347%)	0.828 ± 0.034 (4.810%)	0.755 ± 0.058 (7.703%)
mDAE w/o optimal μ	0.501 ± 0.043 (7.511%)	1.030 ± 0.013 (2.284%)	0.682 ± 0.049 (3.963%)	0.802 ± 0.039 (3.351%)	0.514 ± 0.054 (3.629%)	0.853 ± 0.033 (7.975%)	0.710 ± 0.055 (1.284%)
mDAE w/o overcomplete	0.500 ± 0.011 (7.296%)	1.147 ± 0.013 (13.903%)	0.699 ± 0.008 (6.555%)	0.808 ± 0.025 (4.124%)	0.671 ± 0.209 (35.282%)	0.932 ± 0.045 (17.975%)	0.960 ± 0.140 (36.947%)

Table 2: Mean RMSE of reconstruction (± the standard deviation) for $B = 8$ random draws of 20% of MCAR artificial missing values. First row : results of the mDAE method (with the modified loss, the optimal choice of the hyper-parameter μ and an overcomplete structure). Second row : results with the standard L_2 loss. Third row : results with a random choice of μ. Fourth row : results with an undercomplete structure. The results in brackets are the growth rate of mean RMSE (e.g. $46.99\% = \frac{0.685-0.466}{0.466}$).

Name	Abreviation
k-nearest neighbors[1]	knn
SoftImput [5]	si
missForest [6]	rf
BayesianRidge[1]	br
Generative Adversarial Imputation Network [7]	gain
Missing Data Importance Weighted Autoencoders [8]	miwae
Batch Sinkhorn Imputation [4]	skh
Transformed Distribution Matching for data imputation [9]	tdm

Table 3: Imputation methods

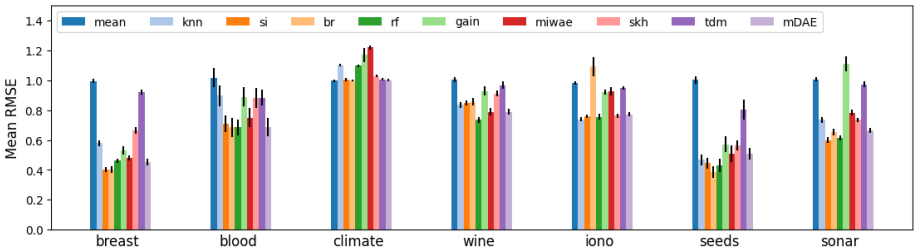

Fig. 2: Mean RMSE of reconstruction (± the standard deviation) for $B = 12$ random draws of 20% of MCAR artificial missing values.

the Mean Distance to the Best (MDB) measures how far, on average, a method is from the best one:

$$MDB(j) = \frac{1}{I}\sum_{i=1}^{I}\left(R_{ij} - \min_{\ell=1\ldots J} R_{i\ell}\right) \tag{8}$$

where R_{ij} is the RMSE obtained with the method j on the dataset i.
Figure 3 confirms the good results of the si, rf, and mDAE methods while high-

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

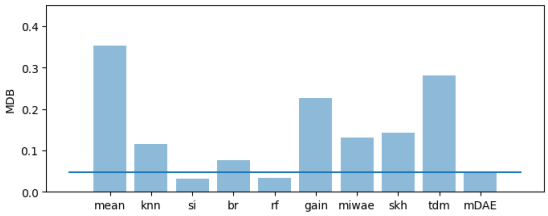

Fig. 3: Mean Distance to the Best (MDB).

lighting weaker results for the four recent gain, miwae, skh and tdm methods.

It should be noted that similar conclusions were obtained with other proportions of missing data and alternative missing data generation processes.

4 Conclusion

This article proposes a simple method for missing data imputation using denoising autoencoders, along with a procedure for selecting the hyper-parameters (the proportion of noise μ and the structure of the network). A new criterion called Mean Distance to the Best (MDB) showed the good behavior of the proposed method globally across several datasets. In perspective, the mDAE's specific features should enable block-wise missing values to be considered by imposing block-wise structuring of the masking noise.

References

[1] Vincent et al. Extracting and composing robust features with denoising autoencoders. In *Proceedings of ICML 2008*, pages 1096–1103, 2008.

[2] Abiri et al. Establishing strong imputation performance of a denoising autoencoder in a wide range of missing data problems. *Neurocomputing*, 365:137–146, 2019.

[3] Pereira et al. Reviewing autoencoders for missing data imputation. *Journal of Artificial Intelligence Research*, 67, 2020.

[4] Muzelec et al. Missing data imputation using optimal transport. In *Proceedings of ICML 2020*, volume 119, pages 7130–7140, 2020.

[5] Mazumder et al. Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11:2287–2322, 2010.

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

[7] Yoon et al. Gain: Missing data imputation using generative adversarial nets. In *Proceedings of ICML 2018*, volume 80, pages 5689–5698, 2018.

[8] Mattei and Frellsen. Miwae: Deep generative modelling and imputation of incomplete data sets. In *Proceedings of ICML 2019*, volume 97, pages 4413–4423, 2019.

[9] Zhao et al. Transformed distribution matching for missing value imputation. In *Proceedings of ICML 2023*, volume 202, pages 42159–42186, 2023.