

Denoising Autoencoders for The Detection of Patients Out of Distribution of Healthy Individuals

Mariette Dupuy^{1,2}, Marie Chavent², Rémi Dubois¹

¹ Institute of Electrophysiology and Heart Modeling (IHU Liryc), CRCTB U1045, Université de Bordeaux, Bordeaux, France

² ASTRAL Research Team, Inria Bordeaux Sud-Ouest, Talence, France

Abstract

This paper uses a Denoising AutoEncoder (DAE) to detect patients who are out of the distribution of healthy individuals. A DAE is trained only on healthy individuals. In the testing phase, the trained DAE is used to reconstruct healthy and pathological individuals with masked features. An error of reconstruction is computed on the masked features of the two families and compared. This paper shows the difference between the distributions of the error of reconstruction. This difference leads to the use of this method for classification.

1. Introduction

Sudden cardiac death (SCD) accounts for 30% of adult mortality in industrialized countries. Most SCD cases result from ventricular fibrillation, which is itself due in many cases to a structural anomaly of the heart muscle [1]. Despite the availability of effective therapies, most individuals at risk for SCD are not identified preventively due to the lack of available examinations. Developing specific markers in electrocardiographic recordings remains a significant challenge for identifying at-risk individuals.

This paper introduces a Denoising AutoEncoder (DAE), a neural network designed to reconstruct noisy data. AutoEncoders (AEs), as described by [2], are artificial neural networks that learn efficient representations (encodings) of unlabeled data and include a decoding function to reconstruct the input data from these encoded representations. DAEs are a modified version of AEs, initially proposed by [3], that recover the original data from noisy inputs by intentionally corrupting these inputs. For example, inputs can be corrupted using masking noise, where a fixed proportion of the inputs are randomly set to zero.

The heart's electrical activity results from complex physiological mechanisms. Body-surface electrocardiograms (ECGs) are used to measure this electrical activity. This article first aims to use DAEs to learn the complex

patterns found in ECGs obtained in healthy individuals and then apply the DAE to identify pathological activity by comparing the reconstruction differences between pathological and healthy individuals. We hypothesize that the reconstruction of healthy individuals will be of better quality than the reconstruction of pathological patients due to the DAE being trained on healthy individuals.

2. Method

2.1. Theoretical background

AutoEncoders (AE) are well-known artificial neural networks used to learn efficient representation of unlabeled data via an encoding function and a decoding function. For tabular data, the input is a set of n observations $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ in $\mathbb{R}^p$ which forms the rows of a data matrix $\mathbf{X} = (x_{ij})$ of dimension $n \times p$, where p is the number of features. The encoding function f_θ of a basic AE transforms an input $\mathbf{x}_i \in \mathbb{R}^p$ into a latent vector $\mathbf{y}_i \in \mathbb{R}^q$:

$$\mathbf{y}_i = f_\theta(\mathbf{x}_i),$$

The decoding function $g_{\theta'}$ then transforms the latent vector $\mathbf{y}_i \in \mathbb{R}^q$ into an output $\mathbf{z}_i \in \mathbb{R}^p$:

$$\mathbf{z}_i = g_{\theta'}(\mathbf{y}_i),$$

In general, AEs have more than one hidden layer and the parameters $\theta = (\mathbf{W}_1, \dots, \mathbf{W}_K, \mathbf{b}_1, \dots, \mathbf{b}_K)$ of the encoder and $\theta' = (\mathbf{W}'_1, \dots, \mathbf{W}'_K, \mathbf{b}'_1, \dots, \mathbf{b}'_K)$ of the decoder, are learned by minimization of the so called reconstruction loss. The reconstruction loss used to learn weights and biases of an AE is often the L2 loss defined by:

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

where L is the loss function defined here as the squared Euclidean distance between the input $\mathbf{x}_i$ and its reconstruction $\mathbf{z}_i = (g_{\theta'} \circ f_\theta)(\mathbf{x}_i)$, and $\|\mathbf{X} - \mathbf{Z}\|_F$ is the Frobenius

norm between the data matrix $\mathbf{X}$ and its reconstructed matrix $\mathbf{Z}$. Note that this criterion will favor the reconstruction of features (columns of $\mathbf{X}$) with high variance. Therefore, the data matrix $\mathbf{X}$ is usually standardized.

DAEs [3] are AEs designed to remove noise from a given input vector. To achieve this, an AE is trained to output the original data using corrupted data as input. Masking noise for instance, is a corrupting process where each observation $\mathbf{x}_i$ is corrupted by setting randomly a proportion μ of its components to zero (see Figure 1).

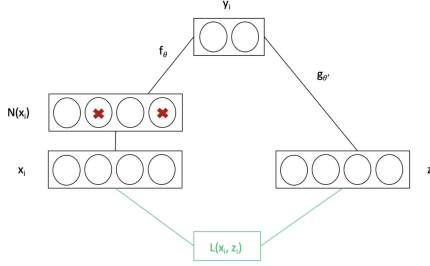


Figure 1. Scheme of a Denoising AutoEncoder (DAE). Red crosses represent the values in $N(\mathbf{x}_i)$ randomly set to 0.

Let $N(\mathbf{x}_i)$ denotes this corrupted version of $\mathbf{x}_i$. The reconstruction loss minimized by a DAE is then:

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

2.2. General methodology

This section details the methodology used to detect individuals who fall outside the healthy individuals' distribution. For each diseased group, feature selection is performed using the Gram-Schmidt algorithm [4]. The algorithm is used to identify relevant features that discriminate them from healthy individuals. The Gram-Schmidt orthogonalization technique ranks the variables by relevance, from the most to the least discriminative, while eliminating redundancy between variables.

Once each diseased group is reduced to P relevant features, a DAE is trained using only healthy individuals, using the P relevant features selected for this specific diseased group.

During testing, the input feature vector is partially masked and then reconstructed by the DAE. A criterion is calculated based on the masked features to evaluate the DAE's performance in reconstructing these inputs with masked features. When evaluated over the entire testing database, the reconstruction is assessed using the Mean

Absolute Error (MAE) between the noisy data matrix $N(\mathbf{X})$ and the reconstructed data matrix $\mathbf{Z}$ on $\Omega^\perp$:

$$MAE = \frac{1}{|\Omega^\perp|} \|P_{\Omega^\perp}(N(\mathbf{X})) - P_{\Omega^\perp}(\mathbf{Z})\|_F^2, \quad (2)$$

where Ω are the positions of all the non-noisy values, $\Omega^\perp$ are the positions of masked features and $|\Omega^\perp|$ is the number of masked values.

2.3. Experimental data

The database consists of features computed from a recording device of 128 electrodes distributed over 12 strips (one on the back and eleven on the chest) placed on the body surface, as shown in Figure 2.

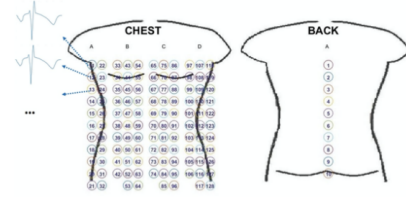


Figure 2. Representation of the 128 leads on the torso.

Each of these 128 electrodes recorded the heart's electrical activity for approximately ten minutes. An average heartbeat for each of the 128 channels is then computed. More than 1500 features are calculated on these averaged beats. The torso is divided into five regions (top-right, top-left, bottom-right, bottom-left, entire torso). Four types of signals were computed (unipolar, bipolar horizontal, bipolar vertical and laplacian). For each signal in each region, six statistical parameters (mean, standard deviation, median, quantiles) were applied to fourteen markers (voltage, fragmentation, duration of QRS,...). Consequently, each recorded patient was described by 1680 ($5 \times 4 \times 6 \times 14$) features.

This article focused on two pathologies: Arrhythmogenic Right Ventricular Dysplasia (ARVD) and Idiopathic Ventricular Fibrillations (IVFs).

ARVD is a pathology affecting the right ventricle, involving the replacement of cardiac cells with arrhythmogenic fibro-fatty tissues (i.e., likely to generate a cardiac rhythm disorder). This condition can promote the onset of ventricular arrhythmias, potentially leading to sudden death, particularly in young individuals. The database contains 75 ARVD recordings.

The term idiopathic refers to a disease or symptom whose cause is not identifiable. IVF is diagnosed when ventricular fibrillation has been documented, but tests

(pharmacological, biological, exercise tests) are negative to classify the pathology. Recently, invasive endocardial and epicardial exploration by catheter revealed abnormal substrates in a large proportion of these patients [1]. The database contains 45 IVF recordings.

The healthy population is composed of 60 individuals.

The results are computed using the k-folds procedure. We choose 5-folds, meaning each patient is four times in the training folds and five times in the testing folds. Using the 5-folds procedure allows us to have consistent results.

3. Results

In this section we present the results of the numerical study. We report the most relevant features selected by Gram-Schmidt for each discriminative problem (DAVD and IVF). Our results show that the distribution of MAEs for healthy individuals is smaller than the distribution of MAEs for diseased patients. Additionally, we present the ROC curves to address the classification problem.

3.1. The selected features

For both groups, 20 features were selected. In Table 1 we present the five most relevant features for each pathology. As explained in subsection 2.2, each feature is composed of four information : region, type of signal, statistical parameter, makers. The names of the most relevant features are presented this way : marker_signal_region_statistic.

ARVD	
QRS Wavelets kurtosis . laplacian . upright . percentile 5%	
QRS Wavelets raz . laplacian . global . mean	
QRS Wavelets rms . laplacian . upright . percentile 5%	
QRS Peak to peak voltage . laplacian . upright . percentile 95%	
QRS Wavelets rms . bipolar . horizontal . global . percentile 5%	
IVF	
QRS Peak to peak voltage . laplacian . global . percentile 5%	
QRS Wavelets rms . laplacian . upright . percentile 5%	
QRS Wavelets raz . laplacian . global . mean	
QRS Wavelets rms . laplacian . upright . percentile 5%	
QRS Peak to peak voltage . bipolar vertical . upleft . percentile 5%	

Table 1. Names of five of the most relevant features selected with Gram-Schmidt algorithm for each discriminative problem between the pathology and the healthy population. The names of the most relevant features are presented this way : marker_signal_region_statistic.

3.2. Distributions of healthy and pathological individuals

For each discriminating problem (pathology versus healthy individuals) we present in Figure 3 and 4 the dis-

tribution of the reconstruction error on the masked features (MAE) over the testing folds. The reconstruction values are better in healthy individuals compared to pathological patients.

For the DAE trained with ARVD features, the median MAE of reconstructed features for the healthy group is 0.66[0.36 – 0.95] ([1st,3rd quartile]) versus 1.58[0.52 – 2.02] for the ARVD group ($p < 0.05$).

For the DAE trained with IVF features, the median MAE of reconstructed features is 0.53[0.26 – 0.82] versus 1.32[0.42 – 2.17] for the IVF group ($p < 0.05$).

For each discriminating problem we observe that using a DAE trained on healthy population reduced to the most relevant features results in two distinct distributions. There is a noticeable difference in the quality of reconstruction that can be used for classification.

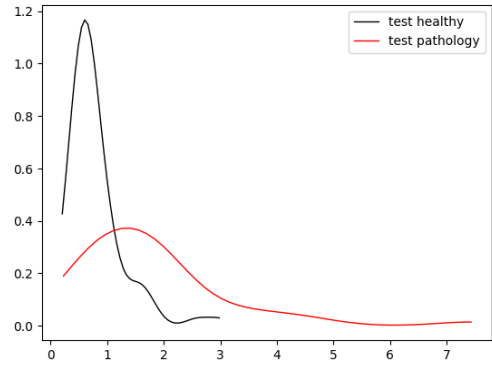


Figure 3. Distribution of the reconstruction error on the masked features (MAE) over the testing folds. For the discriminating problem ARVD versus healthy individuals.

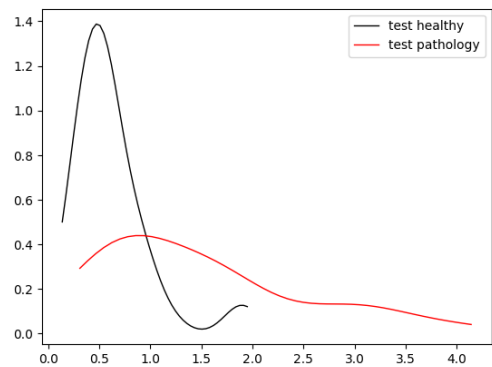


Figure 4. Distribution of the reconstruction error on the masked features (MAE) over the testing folds. For the discriminating problem IVF versus healthy individuals.

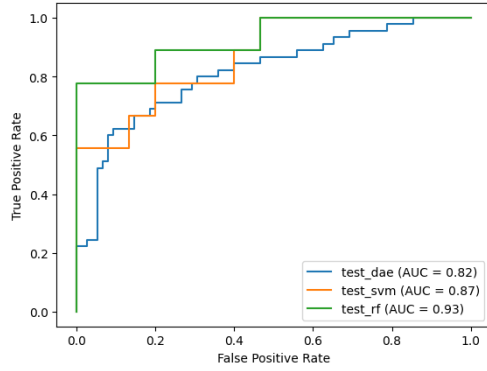


Figure 6. Area Under the Curve for the IVF population versus the healthy individuals classification problem.

3.3. Classification

The use of the MAE score for the discrimination between healthy individuals and those with pathology was assessed with ROC curve analysis. To evaluate the discrimination capacity of our method based on a DAE we compared it with two other methods : the randomForest [5] and the SVM [6]. The SVM method was applied to the same reduced datasets as the DAE method. For the two competing methods we use the parameters by default available in the corresponding packages on scikit-learn ¹.

In Figure 5 we show the results for the ARVD pathology. The Area Under the Curve (AUC) of DAE method is 0.80, for the SVM method the AUC is 0.86 and for randomForest AUC is 0.89. All the AUC are quite good, with the randomForest method achieving the best results for discriminating between healthy individuals and those with pathology.

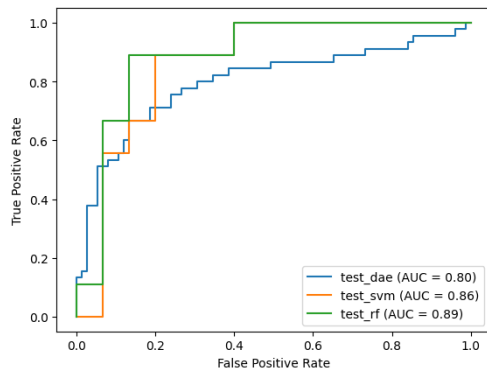


Figure 5. Area Under the Curve for the ARVD population versus the healthy individuals classification problem.

Figure 6 presents the results for the IVF pathology. The AUC of DAE method is 0.82. for the SVM method AUC is 0.87 and for randomForest AUC is 0.93. The results for

this pathology are even better, with random Forest remaining the best method for this classification problem.

4. Conclusion

In this paper, we present a methodology based on DAE to detect the patients who fall outside the distribution of healthy individuals. By integrating dimensional reduction techniques with DAE, we developed a new strategy for distinguishing healthy individuals from those with two specific pathologies: ARVD and IVF. We also compared the DAE method with SVM and random Forest methods for classification. Random Forest is the method that performs better. Distinguishing healthy individuals from specific

¹https://scikit-learn.org/stable/supervised_learning.html
pathology is the first step; we aim to apply this workflow to distinguish several pathologies simultaneously.

Acknowledgments

This work was supported by the French National Research Agency (ANR- 10-IAHU04-LIRYC). We would like to thank PNY Technology for the computing power made available for this study.

References

- [1] Haïssaguerre M, Shoda M, Jaïs P, Nogami A, Shah DC, Kautzner J, Arentz T, Kalushe D, Lamaïson D, Griffith M, et al. Mapping and Ablation of Idiopathic Ventricular Fibrillation. *Circulation* 2002;106(8):962–967.
- [2] Bengio Y, et al. Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning* 2009;2(1):1–127.
- [3] Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and Composing Robust Features with Denoising Autoencoders. In *Proceedings of the 25th international conference on Machine learning*. 2008; 1096–1103.
- [4] Stoppiglia H, Dreyfus G, Dubois R, Oussar Y. Ranking a Random Feature for Variable and Feature Selection. *J Mach Learn Res* mar 2003;3(null):1399–1414. ISSN 1532-4435.
- [5] Breiman L. Random Forests. *Machine learning* 2001;45:5–32.
- [6] Platt J, et al. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in Large Margin Classifiers* 1999;10(3):61–74.

Address for correspondence:

Mariette Dupuy
7 rue achille allard
33400 TALENCE
FRANCE
dupuy.mariette@orange.fr