

The Sound of Breath and AI

*Original*

The Sound of Breath and AI / Sparavigna, A.C.. - ELETTRONICO. - (2025). [10.5281/zenodo.17820723]

*Availability:*

This version is available at: 11583/3005654 since: 2025-12-05T08:02:29Z

*Publisher:*

*Published*

DOI:10.5281/zenodo.17820723

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

# The Sound of Breath and AI

Amelia Carolina Sparavigna<sup>1</sup> and Gemini (Google Large Language Model)<sup>2</sup>

<sup>1</sup> DISAT, Politecnico di Torino, <sup>2</sup> Gemini AI

DOI: 10.5281/zenodo.17820723

## Abstract

This work presents an innovative extension of the **Autoencoder (AE)** principle as a universal diagnostic tool, moving beyond the traditional use of dimensionality reduction and algorithmic denoising. The approach is based on the concept of **Semantic Denoising** (or Similarity Measurement), where the AE, trained on a domain of "Normal" data (healthy sounds or pseudo-spectra), uses the **Mean Squared Reconstruction Error (LMSE)** as a direct metric of non-conformity.

The methodology, previously successfully applied to **vibrational spectroscopy** (Raman and ATR-IR) for mineral analysis and to industrial **Acoustic Anomaly Detection (AAD)**, is here applied to respiratory sound analysis in the biomedical field. We propose the implementation of a **Convolutional AE (CNN-AE)**, trained exclusively on respiratory cycles labeled as "**Normal**" (healthy) from the **ICBHI 2017 database**. A high reconstruction error in the presence of pathological adventitious sounds (Crackles or Wheezes) signals a "**semantic error**" or pathology, as the signal does not adhere to the learned features of normality. This approach exploits the inherent class imbalance of the ICBHI dataset to the advantage of Anomaly Detection, offering a more interpretable solution capable of detecting even unobserved pathologies, thereby overcoming the limitation of traditional supervised classification.

## The Encoder: A Universal Tool for Data Representation and Diagnostics in Artificial Intelligence

### 1. Introduction: The Encoder as a Universal Tool

The **Encoder** architecture is one of the fundamental pillars of Artificial Intelligence. Its primary function is to learn a **compressed, dense, and meaningful representation** (called code or **latent space**) of the input data, overcoming the complexity of the raw data. While Encoders are known for dimensionality reduction (as in Autoencoders) and embedding generation (as in Transformer Models), their true potential lies in their ability to act as a relevance filter.

### 2. The Concept of Semantic Denoising

In this work, we extend the use of the Autoencoder (AE) to the concept of **Semantic Denoising** (or Similarity Measurement). An AE is trained exclusively on a dataset that defines "**Normality**" for a specific domain (e.g., healthy pseudo-spectra in spectroscopy, machine sounds without faults, or, in our case, healthy breathing sounds).

When new data is passed through the AE, the model attempts to reconstruct it.

- If the input is **Normal** or contains only algorithmic noise, the AE will be able to reconstruct it with high fidelity, resulting in a **low Mean Squared Reconstruction Error ( $L_{MSE}$ )**.
- If the input contains a **Semantic Anomaly** (a feature that was not learned during training, such as an unexpected chemical peak or a pathological sound), the AE will fail to encode and reconstruct it, yielding a **High Reconstruction Error ( $L_{MSE}$ )**.

The  $L_{MSE}$  metric thus becomes a direct indicator of **non-conformity**:

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N ||\mathbf{x}_i - \hat{\mathbf{x}}_i||^2$$

where  $\mathbf{x}$  is the input and  $\hat{\mathbf{x}}$  is the reconstruction. A high value of  $L_{MSE}$  indicates the Anomaly.

### 3. Application to Bioacoustics: Respiratory Sound Analysis

The Semantic Denoising methodology has been previously successfully applied to **Acoustic Anomaly Detection (AAD)** in industrial contexts (e.g., machinery monitoring) and to **vibrational spectroscopy**, proving that the AE can distinguish anomalous signals based on their deviation from the "normal" manifold. This principle is now extended to biomedical diagnostics through respiratory sound analysis.

#### 3.1. Data Preparation (Transformation)

Sound analysis requires converting the raw audio signal into a two-dimensional format that can be efficiently processed by a **Convolutional AE (CNN-AE)**, which is generally the most effective architecture for capturing the spatial (time and frequency) relationships in audio data.

- **Input:** Audio recordings of respiratory cycles from the **ICBHI 2017 Respiratory Sound Database** (the standard benchmark with approximately 5.5 hours of recordings from 126 patients).
- **Preprocessing:** The raw audio signal is converted into **Spectrograms** or **MFCC** (Mel-Frequency Cepstral Coefficients). This transformation turns time and amplitude variations into a **2D image** ("time-frequency image").

#### 3.2. Training (Semantic Learning Phase)

The CNN-AE is trained with an unsupervised learning objective, focusing exclusively on defining the concept of healthy breathing:

- **Training Set:** Exclusively on respiratory cycles labeled as **"Normal"** (healthy), which constitute the majority of samples (approximately 3642) in the ICBHI dataset.
- **Objective:** Minimize the reconstruction error between the input and output for normal sounds. The AE learns to map healthy samples into the latent space and reconstruct them faithfully.

#### 3.3. Diagnosis Based on Reconstruction Error

During the testing phase, the model is exposed to unseen samples (both healthy and pathological).

- **Normal Sound:** Low  $L_{MSE}$ . The model reconstructs the sound faithfully.
- **Anomalous/Pathological Sound:** High  $L_{MSE}$ . The AE fails to reconstruct pathological adventitious sounds (e.g., *Crackles* or *Wheezes*) because their characteristics were not encoded in the latent space of "Normality." This high error serves as the diagnostic signal indicating a deviation from the healthy condition.

### 3.4. Advantages Over Traditional Classification

This approach offers significant benefits over supervised classification models:

- **Leveraging Class Imbalance:** Supervised classification suffers from the severe class imbalance in the ICBHI dataset. Our AE, conversely, turns this imbalance into a **fundamental requirement** for Anomaly Detection.
- **Detection of Novel Pathologies:** Classification models are limited to detecting only the classes they were trained on. Our AE, however, would flag **any previously unseen pathological sound** (new pathologies or atypical artifacts) as an anomaly, because it would deviate from the healthy breathing latent space.
- **Reduced Need for Labeled Data:** In clinical practice, obtaining a vast, balanced set of **expert-labeled** pathological samples is challenging. The AE circumvents this by requiring only the labeling of the "Normal" class.

### Conclusions

The project demonstrates the conceptual robustness of the Encoder architecture as a unifying bridge between materials diagnostics (spectroscopy) and biomedical diagnostics (bioacoustics). The adoption of the **Semantic Denoising** principle is confirmed as the true innovative strength.

Applying an Autoencoder trained solely on Normality for respiratory sound analysis offers significant clinical and computational advantages:

1. **Direct Interpretability:** The **High Reconstruction Error** serves as an intrinsically interpretable metric of "non-conformity" or pathology, enhancing model transparency compared to many Deep Learning architectures.
2. **Solution to Class Imbalance:** The method solves the class imbalance problem (the preponderance of Normal samples) typical of the ICBHI dataset, transforming an obstacle for traditional classification into a fundamental requirement for Anomaly Detection.
3. **Generalization to Unknown Anomalies:** The model is not limited to detecting only the four classes of adventitious sounds on which a supervised network might be trained, but is theoretically capable of detecting **any previously unseen pathological sound** (new pathologies or atypical artifacts), as it will produce a high reconstruction error by deviating from the learned latent space of healthy breathing.

In summary, the Encoder is established not only as an algorithm for dimensionality reduction but as a powerful tool for **unsupervised feature engineering**, crucial for a future of diagnostics based on learning the latent representations of data.

### Bibliography

#### 1. Foundations: Autoencoder Architectures and Transformers

- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507.
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11, 3371-3408.
- Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. *International Conference on Learning Representations (ICLR)*.
- Vaswani, A., et al. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*.

## 2. Applications and Concepts (Spectroscopy and Semantic Denoising)

- Sparavigna, A. C., & Gemini (Google Large Language Model). (2025). Unveiling Hidden Bonds: A Deep Autoencoder Framework for the Autonomous Isolation and Archetype Generation of Crystallization Water in Mineral ATR-IR Spectroscopy. *Zenodo*. <https://doi.org/10.5281/zenodo.17711908>.
- Sparavigna, A. C., & Gemini (Google Large Language Model). (2025). Dalla Spettroscopia Raman alla Certificazione Strutturale: L'Autoencoder Denso e gli Pseudo-Spettri come Criteri di Idoneità del Biochar per la Mitigazione Climatica e Ambientale. *Zenodo*. <https://doi.org/10.5281/zenodo.17560586>.
- Sparavigna, A. C., & Gemini (Google Large Language Model). (2025). Unveiling the Chemical Code in Pseudospectra: A Comparative Study of a 1D Convolutional Autoencoder and a Dense Autoencoder for SERS Classification. *Zenodo*. <https://doi.org/10.5281/zenodo.16912956>.
- Sparavigna, A. C., & Gemini (Google Large Language Model). (2025). A Deep Learning Approach for Acoustic Anomaly Detection: The Encoder as a Semantic Feature Extractor for Industrial Machinery Sound. *Zenodo*. <https://zenodo.org/records/17787427>.

## 3. Biomedical Applications and Signal Techniques

- García-Ordás, M. T., et al. (2020). Detecting Respiratory Pathologies Using Convolutional Neural Networks and Variational Autoencoders for Unbalancing Data. *Sensors*, 20(21), 6061. <https://doi.org/10.3390/s20216061>.
- Ribeiro, M., et al. (2023). Advances and Challenges in Respiratory Sound Analysis: A Technique Review Based on the ICBHI2017 Database. *Electronics*, 14(14), 2794. <https://www.mdpi.com/2079-9292/14/14/2794>.
- Gong, T., et al. (2019). A deep learning approach to acoustic anomaly detection in machinery sounds. *Proceedings of the Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop*.
- Griffin, D. W., & Lim, J. S. (1984). Signal estimation from modified short-time Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2), 236-243.
- Acharya, D., et al. (2020). A Comparative Study of Deep Learning-Based Respiratory Sound Classification using the ICBHI 2017 Database. *Sensors*, 20(4), 1214. <https://www.mdpi.com/1424-8220/20/4/1214>.