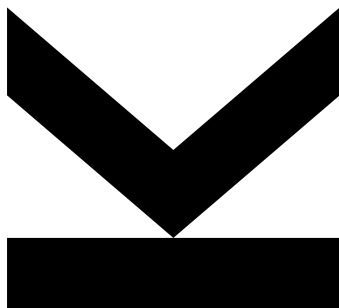


Semi-Supervised Anomaly Detection in Respiratory Sounds: A Comparative Study of Reconstruction and Density Estimation Methods



Bachelor's Thesis

to confer the academic degree of

Bachelor of Science

in the Bachelor's Program

Artificial Intelligence

Author
Lukas Selch
11941656

Submission
Institute of
Computational Perception

Thesis Supervisor
Paul Primus

November 2023

Abstract

Space for your abstract.

Contents

| | |
|--|-----------|
| Abstract | ii |
| 1 Introduction | 1 |
| 1.1 Motivation | 1 |
| 1.2 Objectives and Approach | 1 |
| 1.3 Outline | 1 |
| 2 Theoretical Background | 2 |
| 2.1 Respiratory Sounds | 2 |
| 2.2 Fundamentals of Anomaly Detection | 2 |
| 2.3 Reconstruction-Based Methods | 3 |
| 2.3.1 Essentials of Autoencoders | 4 |
| 2.4 Density Estimation Methods | 4 |
| 2.4.1 Introduction to Masked Autoencoders | 5 |
| 2.5 Evaluation Metrics for Model Comparison | 5 |
| 3 Literature Review | 6 |
| 3.1 Anomaly Detection Techniques in Respiratory Sound Analysis | 6 |
| 3.1.1 The Respiratory Sound Database | 6 |
| 3.1.2 Existing Approaches | 6 |
| 3.2 Variational Autoencoders | 6 |
| 3.3 Group Masked Autoencoders | 6 |
| 4 Methodology | 7 |
| 4.1 Dataset | 7 |
| 4.1.1 Definition | 7 |
| 4.1.2 Data Splitting | 7 |
| 4.2 Preprocessing | 7 |
| 4.3 Detailed Overview of the Models | 7 |
| 4.3.1 Implementation of the VAE | 7 |
| 4.3.2 Implementation of the G-MADE | 7 |
| 5 Experiments and Results | 8 |
| 5.1 Experimental Setup | 8 |
| 5.2 Generalization Capabilities | 8 |
| 5.3 Age and Gender Differences in Model Accuracy | 8 |
| 5.4 Model Sensitivity to Noise | 8 |
| 5.5 Model Performance on Crackles and Wheezes | 8 |
| 5.6 Impact of Hyperparameter Variations | 8 |
| 5.7 Assessment of Data Splitting at Recording Level | 8 |
| 5.8 Comparison of Feature Extraction Methods: MFCC vs. MelSpec- trogram | 8 |
| 5.9 Interpretability and Explainability of the Proposed Models | 8 |
| 5.10 Comparative Analysis and Discussion | 8 |
| 6 Conclusion and Outlook | 9 |
| 6.1 Summary of Findings | 9 |
| 6.2 Potential Applications and Practical Implications | 9 |
| 6.3 Limitations of the Proposed Approaches | 9 |

| | |
|--|-----------|
| Contents | iv |
| 6.4 Directions for Future Research | 9 |
| Bibliography | 10 |

Chapter 1

Introduction

1.1 Motivation

1.2 Objectives and Approach

1.3 Outline

Chapter 2

Theoretical Background

2.1 Respiratory Sounds

The respiratory system, comprising the airways and lungs, is responsible for the vital function of gas exchange in the human body. Respiratory sounds are generated by the airflow within this system during the inhalation and exhalation. [4] Because these sounds are known to be of great importance in the detection of respiratory pathology, listening to the sounds of breathing through the chest using a stethoscope, called *auscultation*, is a cost-effective, non-intrusive, and common part of the physical examination. [2]

The characteristics of the sounds observed during chest auscultation can be precisely defined, allowing for a clear distinction between normal and pathological sounds. Normal breathing sounds are heard throughout the inhalation phase, but only at the very beginning of the exhalation phase, and have a relatively narrow frequency band from 100 Hz to 1000 Hz. While there are a variety of different abnormal breathing sounds, we will focus on two of the most prominent and easily recognizable features.

First, *Wheezes* are musical sounds of long duration over 100 milliseconds and of sinusoidal oscillations that can occur during expiration and inspiration caused by airway narrowing or restriction. They range from 100 Hz to 1000 Hz, with higher harmonics possible above that.

The second sound of relevance are *Crackles*. These non-musical sounds are brief and indicate occasional airway opening, possibly caused by secretions. They can be further subdivided into Fine Crackles and Coarse Crackles, which differ in frequency and duration. While Fine Crackles have a characteristic frequency of about 650 Hz and last about 5 milliseconds, Coarse Crackles are longer sounds of more than 15 milliseconds and occur at lower frequencies below 350 Hz. [2]

Knowing what constitutes physiological and pathological pulmonary sounds, it is possible to pave the way for automated systems to detect them. Electronic stethoscopes can convert the sound signals from the lung to digital signals, allowing the utilization of advanced anomaly detection algorithms for computer-aided medical diagnosis.

2.2 Fundamentals of Anomaly Detection

Anomaly detection addresses the task of identifying deviations in data from anticipated patterns, commonly termed as anomalies or outliers. [3] These anomalies often signal deviations of a system from the norm that are potentially critical and require intervention by the system user. In healthcare, as discussed in section 2.1, outliers in patient respiratory sound patterns can indicate certain conditions.

To find these outliers, anomaly detection algorithms typically define a certain concept of what is considered normal, and any data point that deviates is flagged as an anomaly. The major challenge here is that anomalies are rare and usually not available on a large scale, resulting in a huge data imbalance. In addition, the distribution of anomalies remains uncertain, and even within the set of outliers there may be substantial variation, as data can exhibit various forms of non-normality. [6] In addition, it is important to strike a balance between false positives and false negatives based on the specific application domain, as the severity of one over the other can vary significantly. Outlier detection systems must be properly calibrated to match the characteristics of the domain. Anomaly detection includes different modes depending on the available data. The simplest, but less common in practical scenarios, is *supervised anomaly detection*. In this approach, a fully labeled dataset is used and the task of the detection system is to determine the boundary between normal and abnormal data points. However, it's important to note that real-world datasets often lack comprehensive coverage of different outlier types, making *unsupervised anomaly detection* more prevalent. In unsupervised anomaly detection, only data points known to be normal are provided, and the system must autonomously learn their defining characteristics. Between these two extremes, there are intermediate modes. In our research, we will address the challenge of *weakly supervised* data, where primarily normal data points are available, but a subset of anomalies is also included to evaluate the real-world effectiveness of the learned representation within a detection system.

Outlier detection algorithms typically output a label that directly classifies the data point as normal or anomalous, or they output an anomaly score, which is a measure of the degree of deviation from normality. This score can then be used to define a threshold above which samples are considered anomalous. In addition to traditional anomaly detection methods such as k-nearest neighbor (KNN) or support vector machines (SVMs), which rely on distance metrics or decision boundaries in the feature space to identify outliers, the application of deep learning methods to anomaly detection has gained traction recently. Traditional methods such as KNN work by identifying the closest data points in a data set and flagging those that are significantly farther away as anomalies. SVMs, on the other hand, focus more on defining a boundary between classes and are particularly effective in scenarios where the separation is clear and well-defined. [3] However, while these methods have proven to be highly effective in many applications, they can be limited in a setting with complex, high-dimensional and noisy data, or data where the anomalies are very similar to the normal.

Deep learning methods, on the other hand, have the ability to learn and extract features from raw input data and have shown a remarkable ability to effectively learn patterns in complex data structures such as audio signals. In the following, we will delve into two distinct deep learning architecture families that will be employed in this research.

2.3 Reconstruction-Based Methods

In the context of anomaly detection, reconstruction-based methods use reconstruction errors as anomaly scores. The process unfolds in two main steps. First, the method reduces the dimensionality of the original data by transforming it into a latent, more compact representation. Ideally, this *latent space* captures the essential features of the data while dropping noise and unimportant information. Then follows the actual reconstruction where an attempt is made to receive back the original data from the compact representation. The chal-

lenge is to find a dimensionally reduced representation that is as compact as possible, while retaining enough essential information from the original data to allow accurate reconstruction and avoid overfitting.

During the training process, the input to a reconstruction-based model contains only data points that are known to be normal, making the process unsupervised. This is critical for the model to learn the typical patterns of standard data. Throughout training, a *reconstruction error* is calculated, which is a measure of how well the reconstructed data matches the original data. In other words, this error evaluates how accurately the model can recreate the input data. The basic assumption is that a model trained exclusively on normal data will be able to reconstruct the original data with minimal error. Consequently, when the trained model encounters anomalous data that deviates from the normal patterns, it will struggle to reconstruct it, resulting in a higher reconstruction error. The magnitude of the reconstruction error can be used as an anomaly score, where a higher error indicates a higher probability that a data point is anomalous and vice versa, allowing the model to be used as an anomaly detection system.

2.3.1 Essentials of Autoencoders

Autoencoders are one way to address the dimensionality reduction and are a popular choice of neural network architecture for reconstruction-based methods, as they are capable of efficiently learn a compact representation of data in an unsupervised fashion. An autoencoder consists of two key components: the Encoder (denoted as $A : \mathbb{R}^n \rightarrow \mathbb{R}^l$) and the Decoder (denoted as $B : \mathbb{R}^l \rightarrow \mathbb{R}^n$).

The Encoder is responsible for mapping high-dimensional input data of dimensionality n into a lower-dimensional latent space of dimensionality $l < n$. This represents the compression step into a efficient, compact representation of the original data. The Decoder performs the inverse operation where it takes the data from the latent space and reconstructs it back to its original dimensionality, aiming to reproduce the original input as accurately as possible. The result is an optimization problem defined as

$$\arg \min_{A,B} E[\Delta(\mathbf{x}, B \circ A(\mathbf{x}))],$$

where E is the expected value of the reconstruction loss Δ for input \mathbf{x} . [1]

2.4 Density Estimation Methods

Probability distributions are mathematical functions that describe the probability associated with each possible value of a random variable. If the random variable can take any value within a certain range, the probability distribution is continuous. It is typically described using a *probability density function* (PDF). The PDF is a fundamental concept in probability and statistics, providing a way to calculate the probability of the random variable falling within a specific interval.

For a continuous real-valued random variable X , the PDF is defined as

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

for all $a, b \in \mathbb{R}$. [5] Here, $f(x)$ represents the probability density function of X . It's important to note that the PDF itself does not give probabilities directly. Instead, the probability of X falling within the interval from a to b is given by the area under the curve of the PDF between these two points. Test

2.4.1 Introduction to Masked Autoencoders

2.5 Evaluation Metrics for Model Comparison

Chapter 3

Literature Review

3.1 Anomaly Detection Techniques in Respiratory Sound Analysis

3.1.1 The Respiratory Sound Database

3.1.2 Existing Approaches

3.2 Variational Autoencoders

3.3 Group Masked Autoencoders

Chapter 4

Methodology

4.1 Dataset

4.1.1 Definition

4.1.2 Data Splitting

4.2 Preprocessing

4.3 Detailed Overview of the Models

4.3.1 Implementation of the VAE

4.3.2 Implementation of the G-MADE

Chapter 5

Experiments and Results

5.1 Experimental Setup

5.2 Generalization Capabilities

5.3 Age and Gender Differences in Model Accuracy

5.4 Model Sensitivity to Noise

5.5 Model Performance on Crackles and Wheezes

5.6 Impact of Hyperparameter Variations

5.7 Assessment of Data Splitting at Recording Level

5.8 Comparison of Feature Extraction Methods: MFCC vs. MelSpectrogram

5.9 Interpretability and Explainability of the Proposed Models

5.10 Comparative Analysis and Discussion

Chapter 6

Conclusion and Outlook

6.1 Summary of Findings

6.2 Potential Applications and Practical Implications

6.3 Limitations of the Proposed Approaches

6.4 Directions for Future Research

Bibliography

- [1] Dor Bank, Noam Koenigstein, and Raja Giryes. 2023. Autoencoders. *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, 353–374.
- [2] Abraham Bohadana, Gabriel Izbicki, and Steve S Kraman. 2014. Fundamentals of lung auscultation. *New England Journal of Medicine*, 370, 8, 744–751.
- [3] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41, 3, 1–58.
- [4] J Earis. 1992. Lung sounds. *Thorax*, 47, 9, 671.
- [5] Charles Miller Grinstead and James Laurie Snell. 2006. *Grinstead and Snell's introduction to probability*. Chance Project.
- [6] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. 2021. Deep learning for anomaly detection: A review. *ACM computing surveys (CSUR)*, 54, 2, 1–38.