Analysis of Respiratory Time Series Data for Breathing Discomfort Detection Prior to Sleep Onset During APAP Therapy

by

Shelby Unger

B.S., Electrical Engineering, Texas A&M University, 2018

Submitted to the MIT Sloan School of Management and Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degrees of

Master of Business Administration

and

Master of Science in Electrical Engineering and Computer Science

in conjunction with the Leaders for Global Operations program

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Shelby Unger
MIT Sloan School of Management and
Department of Electrical Engineering and Computer Science
May 12, 2023

Certified by: Roy E. Welsch
Eastman Kodak Leaders for Global Operations Professor of Management
Thesis supervisor

Certified by: Peter Szolovits
Professor of Computer Science and Engineering
Thesis supervisor

Accepted by: Maura Herson
Assistant Dean, MBA Program, MIT Sloan School of Management

Accepted by: Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Analysis of Respiratory Time Series Data for Breathing Discomfort Detection Prior to Sleep Onset During APAP Therapy

by

Shelby Unger

Submitted to the MIT Sloan School of Management and
Department of Electrical Engineering and Computer Science
on May 12, 2023, in partial fulfillment of the
requirements for the degrees of
Master of Business Administration
and
Master of Science in Electrical Engineering and Computer Science

Abstract

Discomfort during treatment continues to be a major barrier to adherence to positive airway pressure (PAP) therapy. Thus, a key pillar of ResMed's business strategy is to deliver intelligent tools that assist healthcare providers in identifying which patients may be struggling with therapy, and why, to enable more effective interventions and personalized patient education. One potential cause of discomfort is perceived stuffiness from pressure levels that is lower than tolerable for some patient preferences. This thesis seeks to explore which patterns in the high-resolution breathing data from ResMed devices may be used to identify patients who are experiencing breathing discomfort at low pressures at the beginning of their therapy sessions. Specifically, time-series clustering is performed on sequential respiratory data to identify groups of patients with similar breathing patterns. The independence between clusters and variables pertaining to patients' demographic characteristics, therapy settings, usage habits, respiratory characteristics, and self-reported comfort levels are evaluated via statistical testing. Based on the results, features in breathing data are identified that may be meaningful indicators for whether a patient is experiencing discomfort or breathlessness. Additionally, opportunities for additional data collection that would enable further analysis and more accurate modelling are discussed.

Thesis Supervisor: Roy E. Welsch
Title: Eastman Kodak Leaders for Global Operations Professor of Management

Thesis Supervisor: Peter Szolovits
Title: Professor of Computer Science and Engineering

Acknowledgements

Before diving in to the details of this thesis, there are many people who deserve acknowledgement and gratitude.

To my colleagues at ResMed: Thank you for inviting me into the organization with open arms, providing in view into the exciting and innovative work that ResMed is conducting, and supporting this project. A special call out to my supervisor, Manish Ladha, for allowing me to explore projects that were interesting and challenging. As well, to my many mentors, Genevieve Madafiglio, Gregory Rehm, Priyanshu Gupta, Nadi Sadr, and Chinh Nguyen, who generously shared their time, knowledge, and expertise with me. Their insights and feedback were instrumental in shaping this thesis, and have helped me further develop my skills as a data scientist and product manager.

To my advisors Roy and Peter: Thank you for taking me on as an advisee, and for your invaluable guidance and encouragement throughout this project. Your knowledge, expertise, and enthusiasm was inspiring, and I am so grateful to have had the chance to learn from you both.

To my classmates in LGO: I am in awe of everything that our class has accomplished over the past two years, and the impactful work that everyone has produced during their internships. Thank you for acting as a sounding board and celebrating the wins, both big and small. I look forward to many more years of friendship and learning together.

Finally, and most importantly, to my family and partner: Thank you for your unwavering love and encouragement. Your constant support has been a source and strength and motivation as I have pursued this degree.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Acronyms

AASM American Academy of Sleep Medicine. 19, 24

AHI apnea hypopnea index. 19, 28, 49

ANOVA analysis of variance. 40

APAP automatic positive airway pressure. 16, 17, 20, 27, 28

CPAP continuous positive airway pressure. 16, 17

CVI Clustering Validation Index. 38

DTW Dynamic Time Warping. 25, 36, 37, 57

OSA obstructive sleep apnea. 15–17, 19–21, 24

PAP Positive airway pressure. 15–17, 19–24, 27, 29, 30, 52

PHI protected health information. 27

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

ResMed is a global leader in medical devices used for the treatment of obstructive sleep apnea (OSA), a common sleep-related breathing disorder characterized by repeated obstruction of the upper airway during sleep. Positive airway pressure (PAP) therapy is a common treatment for OSA, however, many patients struggle to adhere to treatment long-term, often citing discomfort as a reason for discontinuing therapy. Adjustments to device settings could help alleviate these issues, but many healthcare providers do not have access to information that could enable them to proactively intervene. Therefore, ResMed is experimenting with how the company can deliver intelligent tools to healthcare providers and patients that support earlier interventions for an improved therapy experience.

This thesis focuses specifically on the scenario where patients experience breathing discomfort at pressure levels that are too low for them. This can cause a feeling of breathlessness which may lead to changes in breathing patterns, and is often accompanied by stress or anxiety. The aim is to explore whether features in the respiratory data of patients can be used to identify if they are experiencing breathing discomfort at low pressures during the first thirty minutes of their therapy sessions. Time-series clustering is used to automatically identify groups of patients based on data from ResMed's AirSense 11 AutoSet devices. These clusters are then inspected further, revealing which features in ResMed's device data may be meaningful predictors of discomfort at low pressures and the limitations of the data in its current form.

In this chapter, introductory information about ResMed is provided. The project motivation, objectives and contributions are then discussed in detail. Finally, the structure of the remaining thesis is outlined.

1.1 Introduction to ResMed

ResMed was founded in 1989 by Dr. Peter Farrell to commercialize the treatment of OSA with a continuous positive airway pressure (CPAP) machine. Today, the company primarily provides medical devices for the treatment of sleep-disordered breathing and other respiratory disorders, which includes CPAP machines and ventilators, as well as the accompanying tubes and masks. ResMed has recently expanded to provide software as a service to out-of-hospital health care providers through the acquisitions of Brightree and MatrixCare.

Within its sleep and respiratory care product portfolio, CPAP machines for the at-home treatment of OSA are ResMed's largest product line. Modern versions of these devices are cloud-connected, allowing for the collection of data pertaining to patients' usage of the devices.

ResMed offers several cloud-based products alongside their flow generators, assisting patients and providers with disease management. AirSense 11 CPAP and APAP machines include access to myAir, a patient engagement application which allows patients to track their therapy progress, and provides them with coaching, support and educational tools. AirView is a patient management tool for healthcare professionals that includes the abilities to access patient data, and virtually adjust machine settings.

1.2 Project Motivation

Despite numerous technological advancements, adherence to PAP therapy continues to be a challenge for many patients. Long-term, physicians may observe an adherence rate between 30 to 60% across patients. [30] One of the primary reasons that patients do not adhere to PAP therapy is discomfort. There are several root causes for why patients

may be uncomfortable such as irritation caused by the mask, feeling claustrophobic, or trouble adjusting to breathing on pressure.

To support patients who may be struggling to tolerate breathing at a continuous forced pressure, ResMed has designed a ramp feature into their CPAP and APAP machines. This ramp feature allows patients to start therapy at a lower pressure relative to their prescribed pressure setting, and gradually ease into the level of pressure that they require in order to sleep without experiencing sleep-apnea symptoms. However, while a subset of OSA patients may feel like the pressure being delivered during therapy is too high without ramp, others may experience breathing discomfort when initiating therapy at a lower pressures. Patients are able to adjust ramp settings on their own to a combination that is more comfortable for them, assuming they have been educated on how to do this. Breathing discomfort can also be resolved with interventions from healthcare providers. However, providers have limited amounts of time to spend with patients and may not be aware of the problem unless it is raised by the patient. Therefore, an opportunity exists for ResMed to identify patients who may be experiencing breathing discomfort during the sleep onset period of therapy to enable more effective provider interventions and personalized patient education.

1.3 Objective and Contributions

ResMed strives to deliver intelligent tools that assist healthcare providers identify patients in need and make faster, more informed decisions, and that support patients as they navigate PAP therapy. AirSense 11 PAP devices gather data from each therapy session that a patients completes, including time series data on the pressure delivered by the flow generator and respiratory measures. This thesis aims to explore whether features in the breathing patterns of patients can be used to identify if they are experiencing breathing discomfort during period of therapy prior to sleep onset. The contributions of this work are summarized as follows:

- A pattern-driven analysis is performed by conducting time-series k-means clustering with respiratory data from the sleep onset period of patients' therapy

sessions.

- The statistical significance of differences between clusters in various demographic variables, signal characteristics, and the frequency of patient reported breathing discomfort is assessed.

- Features in respiratory signals that may be related to the presence of breathing discomfort are discussed, and general considerations regarding data management are discussed.

1.4 Thesis Outline

The remainder of this document elaborates further on the data sources and methodology used, the results of time-series clustering and statistical analysis, and recommendations for future work. Chapter 2 provides an overview of relevant background information pertaining to ResMed Devices and breathing discomfort, and a review of machine learning in sleep medicine and common techniques for the analysis of time-series data. Chapter 3 describes in detail the data that was available for this analysis, and the steps that were taken to clean and process this data into a suitable form. Chapter 4 outlines the time-series clustering procedure that was used to generate groups of patients based on similarities in their breathing patterns, and the statistical methods that were utilized to asses the significance of differences across these groups. The results of the analysis of reported in Chapter 5. Finally, Chapter 6 provides a discussion of the results and the limitations of the analysis, as well as a review of relevant data management considerations.

Chapter 2

Background

2.1 Overview of Obstructive Sleep Apnea (OSA)

OSA is a sleep disorder that affects millions of people worldwide. The condition is characterized by repeated episodes of partial or complete blockage of the upper airway during sleep, which leads to a decrease in oxygen saturation and arousal from sleep. The resulting symptoms include poor sleep quality, daytime sleepiness, and other serious health problems such as high blood pressure, cardiovascular problems, and metabolic syndrome, to name a few.

There are different levels of OSA severity, defined by a patient's apnea hypopnea index (AHI): the combined average number of apneas (the temporary cessation of breathing during sleep) and hypopneas (reductions in ventilation of at least 50% due to partial airway obstruction) that occur per hour of sleep. According to the American Academy of Sleep Medicine (AASM) OSA is categorized into mild (AHI of 5-15), moderate (AHI of 15-30), and severe (AHI greater than 30). [12] Current research shows that over 900 million adults aged 30-69 years suffer from mild to severe OSA and more than 400 million adults 30-69 have moderate to severe OSA globally. [4]

The official diagnosis of OSA involves the completing of a sleep study, either in a laboratory or at home. Several medical interventions are available for the treatment of OSA, including PAP therapy, oral appliances, and surgery. [33] PAP therapy is the most common treatment for moderate to severe OSA and involves the use of a

machine that delivers a constant flow of air through a mask worn over the nose or nose and mouth during sleep. The airflow from the machine prevents the patient's airway from collapsing during an apnea event, keeping the airway open and allowing for the patient to breathe freely. The pressure level that the machine provides is prescribed by the healthcare provider based on the patient's physiology and OSA severity. Oral appliances are also used to treat OSA, and they work by repositioning the tongue and lower jaw to keep the airway open. Additionally, surgery to enlarge or stabilize the upper airway is an option for some individuals with OSA, but it is typically reserved for cases in which other treatments have failed or are not suitable. Medical devices that stimulate the hypoglassal nerve can also be implanted. These are turned on by the patient prior to sleep and send gentle pulses in sync with the patient's breathing to prevent airway obstruction.

2.2 Overview of ResMed AirSense 11 Devices

ResMed's AirSense 11 device is the company's latest version of its flagship PAP machine. Each device includes features designed to make therapy a comfortable experience for patients, as well as access to myAir, ResMed's therapy tracking and support platform for patients. AirSense 11 was launched in 2020 both globally and in the United States, to date thousands of patients rely on the AirSense 11 devices for treatment of their OSA symptoms. ResMed's AirSense 11 device, a ResMed tube and mask, and myAir are shown in Figure 2-1.

2.2.1 Categories of Devices

Three models of the AirSense 11 are currently available: AutoSet, CPAP, and Elite. CPAP and Elite devices deliver one continuous pressure level of air throughout the night as prescribed by a healthcare provider. In contrast, AutoSet machines leverage APAP technology to automatically adjust the pressure level that the device delivers throughout the night to meet the user's unique breathing needs as they change on a breath-by-breath basis. For OSA patients using an AutoSet device, healthcare

Figure 2-1: ResMed's AirSense 11 machine, mask, tube, and myAir mobile application.

providers will prescribe a minimum and maximum pressure level. This is the range of pressures that the AutoSet algorithm is able to output. On all ResMed AirSense 11 devices, the range of possible pressure levels is 4.0 cmH2O to 20.0 cmH2O.

The AutoSet pressure level is determined by analyzing the state of the user's upper airway on a breath-by-breath level. When the device senses an airway event such as an air-flow limitation, an apnea, or snoring, then the amount of pressure generated is adjusted automatically. In AutoSet mode, the amount of pressure provided by the device is only what is required to keep the upper airway open. [29] The scope of this study is limited to ResMed's AirSense 11 AutoSet devices which utilize APAP Technologies. For the duration of this paper we will refer to them as "AutoSet Devices."

2.2.2 Ramp Comfort Setting

Modern PAP devices include comfort settings that can be modified by patients to ease the process of adjusting to PAP therapy. Examples include climate control features such as a heated humidifier and expiratory pressure relief which reduces the pressure level during exhalations to reduce resistance. Many PAP devices also include a ramp feature which is designed to gradually increase pressure from a starting pressure to the prescribed pressure level over a set period of time as the patient falls asleep.

All ResMed AirSense 11 devices include an AutoRamp feature, which starts patients at a low air pressure while they are awake and automatically increases pressure up to the prescribed minimum level to ensure effective treatment of OSA.

Patients using AutoSet devices are able to configure their ramp settings in several different ways. Ramp can be disabled, meaning that patients will start therapy at their prescribed minimum pressure level as soon as their devices are turned on. Alternatively, patients can select a start pressure and an amount of time over which ramp will occur. The ramp time can be set to a value between five minutes and forty-five minutes, at five minute increments. In this scenario, ramp will occur linearly from the start pressure to the minimum prescribed pressure over the defined interval of time. Finally, ramp time can be set to 'AUTO', in which case the AutoRamp algorithm increases the pressure level from the start pressure to the prescribed minimum pressure as described prior at a rate of 1.0 cmH2O per minute.

2.3 Low Pressure During Sleep Onset and Potential Breathing Discomfort

Ramp is a feature that has been designed to assist patients who are struggling to adjust to breathing at higher pressures while on PAP therapy. While effective in many cases, some patients may find the experience of spending prolonged periods of time at a low pressure level uncomfortable. If the start pressure is set at a level that is too low for the patient during AutoRamp, the patient may spend up to 30 minutes breathing on a pressure level that is uncomfortable for them. As well, if the minimum pressure prescribed by the healthcare provider is the default 4.0 cmH2O and ramp is turned on, or if the start and minimum pressures are set to the same value, there will be no increase in pressure during the sleep onset period. Instead the patient will spend time on the same low pressure that is potentially causing them discomfort until the sleep onset period has ended and the AutoSet algorithm turns on.

Interventions from healthcare providers and adjustments to the device settings can help patients find the most suitable and comfortable settings for them. However, many providers have a limited amount of time that they can spend with each patient, and they rely on patients to proactively share about the challenges that they are

experiencing on PAP therapy.

It is hypothesized that patient's respiratory characteristics may change when they experience breathing discomfort at low pressures. In this study, we focus on discomfort during the sleep onset period when patients when patients may be held at low pressures for prolonged periods of time depending on their device settings.

2.3.1 Impact on Patient Compliance

Evaluating a random sample of 40,819 ResMed patients revealed that the time spent on low pressures did have an impact on compliance rates for patients who used their AutoSet devices for four or more hours per night on average during the first week of therapy. A patient is considered to have successfully achieved compliance if they use PAP therapy for at least four hours each night for 70% of the nights in a 30 day window, within the first 90 days after they were setup with their device.

Compliance rates were generated for the amount of time spent at low pressures, and compliance was calculated for a variety of patient groups based on average usage buckets medium to high usage (greater than four hours of usage per night), low usage (between 45 minutes and four hours of usage per night), and very low usage (less than 45 minutes of usage per night). The 95% confidence interval for each proportion was generated. For patients who are using their devices for greater than four hours per night on average during their first week of therapy there was an observable impact to compliance (approximately 4%) based on the time that patients spent at the minimum pressure. This motivates the need for continued exploration into breathing discomfort at low pressure during this period of the therapy session.

2.4 Related Work

2.4.1 Machine Learning in Sleep Medicine

Data analytics and machine learning tools are commonly leveraged to accelerate discovery and automate existing manual tasks across a variety of industries. The large

amount of high-resolution data available from sleep monitor devices and devices such as PAP machines has provided new opportunities for the analysis and extraction of meaningful information from sleep, cardiac, and respiratory signals. The AASM has identified several different applications where data science and machine learning could be leveraged in sleep medicine, such as for improving the diagnosis of sleep disorders, improving sleep scoring, and predicting disease progression. [28]

In addition to supporting the diagnosis and effective treatment of sleep disorders, artificial intelligence also has the potential to streamline operations for healthcare providers, ultimately increasing the amount of time that they have to spend caring for patients. [15] However, attention must be taken to carefully integrate these tools to ensure that the highest quality of care continues to be provided.

The treatment of OSA is uniquely positioned to utilize artificial intelligence and machine learning given the cloud-connected nature of devices. Identifying challenges that patients are having with therapy could enable real-time interventions to educate patients and empower them to improve their own treatment adherence. [14]

2.4.2 Methods for Analyzing Time-Series Data

Time series data consists of a sequence of observations recorded at equally spaced points in time. Examples include stock prices, weather data, and various types of medical data. The analysis of time series data can be challenging because of its large size and complexity, especially when multiple series interact with one another. However, there are several methods for pattern discovery in time series data that can be used to extract valuable insights.

Several methods for the visualisation of time series data have been developed that enable a user to interactively explore patterns across samples in a data base. For example, TimeSearcher2 [6] enables users to query patterns by example and explore multi-dimensional data through the use of graphs and tables. However, the user must provide a pattern to search on, so they must have some pre-existing concept of the interesting patterns that exist in the data. VizTree [21] transforms the time series into a symbolic representation via symbolic aggregate approximation (SAX) and moves

a sliding window across the time series to construct a sub-sequence tree with each branch representing one SAX symbol. Frequent patterns are identified by finding the thickest branches, and are visualized. One drawback of this method is that the user must know the approximate length of the pattern that they are searching for in order to properly define the size of the sliding window.

Principal component analysis has been applied to time-series data as a feature extraction method. [20] It involves decreasing the dimensions of a data set by mapping it to a lower dimensional subspace, and is most effective when relationships in the data are linear.

Clustering has been shown to be an effective method of extracting knowledge from time series data in various scientific areas. Feature based methods for grouping time series data and approaches that perform clustering on raw time series data have both been used for identifying and grouping similar sequences with success. Feature based methods involve a data processing pipeline where features-based representations of the temporal data are extracted, dimensionality reduction is performed, and then the clustering algorithm is applied to group data points. [11] When performing clustering on raw times series data, similarity measures such as Euclidean distance and Dynamic Time Warping (DTW) are often used, and a variety of algorithms can be applied. [36] One of the most popular methods is k-means due to its simplicity and flexibility. [1] Agglomerative hierarchical clustering algorithms [25] are also commonly used. Finally, self-organizing maps have also been used for pattern discovery in temporal data. This is a neural network method for clustering analysis which projects the input onto a low-dimensional feature map. [13]

Convolutional autoencoders have more recently been applied to the problem of pattern discovery in multivariate time series data. In this approach, learned convolution filters capture and represent patterns in the input data. The inclusion of additional layers and regularization terms ensures that the filters are interpretable. [3] However, deep learning approaches require a large amount of data.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Data

The exploration of patients' respiratory characteristics in relation to the perception of breathing discomfort while using ResMed's APAP machines requires a detailed analysis of multiple data sources. This includes patient demographic data, therapy session data, machine settings data, and qualitative responses to a survey about patients' comfort levels. The majority of the data is already available is ResMed's internal databases. Once gathered, the data was cleaned and processed into a usable form.

3.1 Data Sources

Data gathered from AutoSet devices is used by physicians to support patients throughout their PAP therapy journey, and to help patients track their progress and get comfortable with therapy. This data can be also be used by ResMed for customer support purposes, machine learning, and business intelligence. Data used in this study was accessed from a database that contained de-identified data. Specifically, protected health information (PHI) including name, contact details, and geographical subdivision is not available.

3.1.1 Therapy Data

For each therapy session that a patient completes with their AutoSet device, data is recorded on the respiratory measures of the patient and measurements pertaining to the therapy provided by the device. This data is in time series format, with successive measurements recorded at fixed time intervals throughout the night. A therapy session is defined as the period of APAP usage between the time when the patient puts their mask on and the time when the patient removes their mask. Thus, it is possible for a patient to have multiple therapy sessions during a single night. Therapy data is gathered periodically throughout the night and sent to ResMed via the cloud. For this analysis we consider data for two respiratory metrics, respiratory signal 1 (RS1) and respiratory signal 2 (RS2).

Additionally, for each night following the date that a patient was set up on a ResMed AutoSet device, data is gathered on the total number of minutes that the patient used their device. In the case where a patient did not use their device on a specific night, the duration of usage is recorded as 0 minutes.

3.1.2 Patient Demographic and Diagnostic Data

Demographic and diagnostic data about each patient is gathered from ResMed's cloud-based software tools for health care providers and patients. Data is either self-reported by the patient, or manager by their healthcare provider.

- **Age** - The age of the patient.
- **Gender** - The gender of the patient. Values include 'male', 'female', and 'prefer not to say'.
- **Baseline apnea hypopnea index (AHI)** - The combined average number of apneas and hypopneas per hour of sleep prior to the patient starting therapy, from myAir. According to the American Academy of Sleep Medicine these values are categorized into mild (5-15 events/hour), moderate (15-30 events/hour) and severe ($>$ 30 events/hour).

- **Mask type** - The type of mask that a patient is currently using. Values include 'Full Face', 'Nasal', and 'Nasal Pillows'. Full face masks cover both the nose and mouth, nasal mask cover only the nose, and nasal pillows are soft silicone or plastic pillows that are inserted directly into the nostrils and provide a seal around the base of the nose.

- **Setup date** - The date that the patient was set up with their ResMed device.

- **Compliance** - A value indicating whether or not the patient has successfully completed their compliance period (at least four hours of use each night for 70% of the nights in a 30 day window, within the first 90 days following setup).

3.1.3 Device Settings Data

There are many settings on ResMed's AutoSet devices that can be adjusted to meet the unique needs of individual patients. Some of these settings, such as the minimum and maximum pressure setting are prescribed by the physician and cannot be modified by the patient. Other settings related to comfort features such as ramp can be adjusted by the patient to a combination that is most suitable for them. For each night that a patient completes a therapy session, the device settings used for that therapy session are recorded. Two device settings were considered for this analysis. These will be referred to as S1 and S2 for the duration of this thesis.

3.1.4 Comfort Survey Data

As part of a broader investigation into the comfort settings available on ResMed devices, a subset of patients were surveyed on their breathing comfort. The survey was sent to ten thousand patients, and 4580 responses were received.

This survey contained a number of qualitative questions to gauge whether patient's experienced breathing discomfort during the sleep onset period of therapy. Question 1 asked patients about their perception of the quantity of air they receive from the PAP devices. Question 2 asked patients about the perceived clarity or stuffiness of

the air that they breath during PAP therapy. The distribution of responses is shown in Figure 3-1 and Figure 3-2.

Best practices were followed with regards to cleaning duplicate and incomplete responses from the data set.

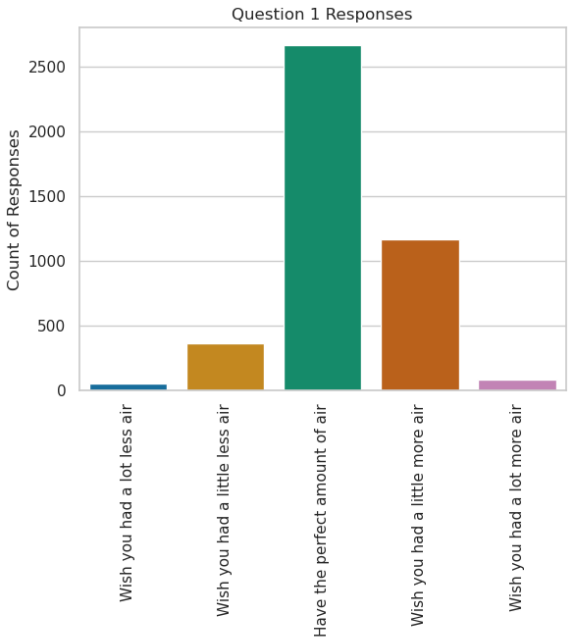

Figure 3-1: Responses to the comfort survey question pertaining to quantity of air during therapy.

3.2 Preparing the Data Set

After joining the data sets described above, several additional pre-processing steps are required to ensure that the final data set contained sufficient and valid data for the analysis. First, time series therapy data is cropped to include only the sleep onset period of the night. For this analysis the scope is limited to the period of time after a patient puts their mask on for the first time during a night.

Next, general clinical practice related to therapy sessions with high levels of leak are followed. High unintended air leaks from between the mask and a patient's nose and mouth can reduce the accuracy of measurements collected by ResMed devices.

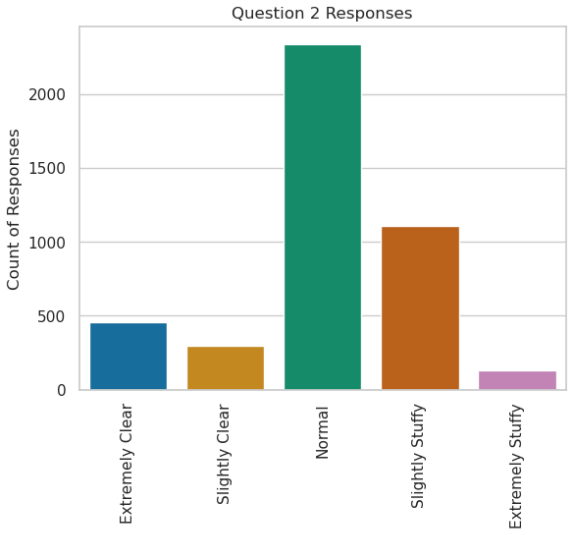

Figure 3-2: Responses to the comfort survey question pertaining to clarity of air during therapy.

These cases are excluded from the data set.

Since only one survey response was collected per patient, the assumption is made that patients are cognitively biased to the most recent therapy session that they completed. Therefore, only the last night of a patient's therapy data prior to their survey response date is considered.

To ensure that there is adequate therapy data for each patient included in the final data set, therapy sessions that are shorter than 30 minutes are removed. Finally, only therapy sessions where it is possible that the patient is being held at a low pressure for an extended period of time are considered. These cases are identified using the device setting data that is collected for each therapy session. Therapy sessions where the combination of device settings is such that the patient will not be held on low pressures for a period of time are excluded from the final data set.

3.3 Feature Engineering

Therapy data related to a patients' respiratory measures is in the form of time series data. Therefore, it is necessary to compute various characteristics of the signals

of interests. Characteristics of the first fifteen minutes of respiratory signal 1 and respiratory signal 2 are computed. This includes the mean, standard deviation, maximum, minimum, and range (the difference between the maximum and minimum values in the signal). Two features related to the device settings and pressure levels throughout the therapy session are also computed: SF1 and SF2.

3.4 Sample Population Characteristics

The final data set includes N = 1897 subjects who started therapy on ResMed's AutoSet devices between January 2022 and August 2022. Summary statistics for the sample population are described in Table 3.1. The distribution of age and gender is similar to ResMed's greater patient population. It is important to note that some demographic data is provided by the patient, so complete accuracy cannot be guaranteed.

		Missing	**Overall**
n			1897
AGE, mean (SD)		0	55.7 (12.8)
GENDER, n (%)	Female	0	923 (48.7)
	Male		967 (51.0)
	Prefer not to say		7 (0.4)
MASK TYPE, n (%)	Full face	0	727 (38.3)
	Nasal		652 (34.4)
	Nasal pillows		518 (27.3)

Table 3.1: Summary statistics for the final data set used in the analysis

Figures 3-3 and 3-4 show the distributions of two device settings used during the therapy sessions included in the final data set. Note that in each figure, patients are grouped by the value that they have selected for each of the two settings shown. The majority of patients have device setting 1 set to the same level. There is slightly more variation in the level that patients have device setting 2 set to. However, most patients fall into one of five groups — group 2, 6, 7, 10, or 11.

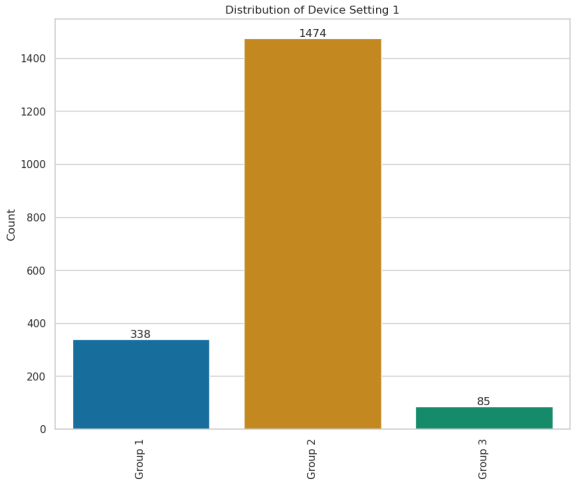

Figure 3-3: Distribution of device setting #1 in sample population

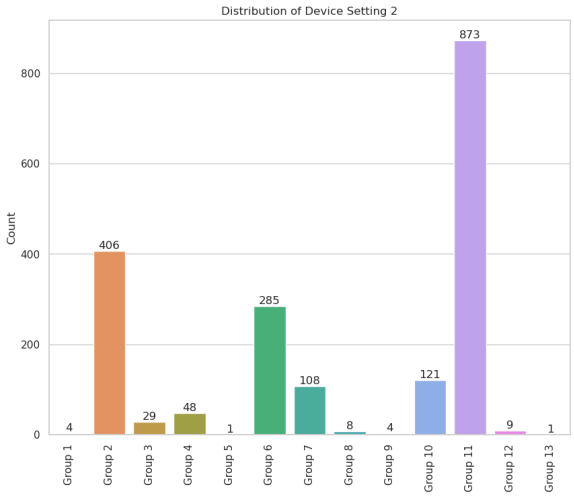

Figure 3-4: Distribution of device setting #2 in sample population

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Time Series Clustering Procedure

The goal of this analysis is to explore ResMed's therapy data using time series clustering techniques to reveal typical patterns and meaningful features in respiratory measures that may be related to breathing discomfort during the sleep onset period of therapy sessions. In this chapter the clustering procedure is described in detail, including the similarity measure selected, the clustering method, validation metrics, parameter selection, and statistical tests used to assess differences across clusters.

4.1 Similarity Measure

In clustering, the similarity measure is a way of quantifying how closely related data samples are to each other. If the scales of features used for clustering are very different, then features with the largest numeric values will be overemphasized. To prevent this outcome, especially in the multivariate case, data is normalized to a distribution between 0 and 1 prior to calculating the similarity measure. Since the goal of this analysis is to explore whether meaningful patterns exist in respiratory time series data, we want to minimize the impact that the magnitude of the signals have on the partitioning of the data. Therefore, min-max scaling is applied to the respiratory measures for each patient separately.

One widely used similarity measure is the Euclidean distance, which computes the point-wise distance between the two time series. This is the square root of the sum

of the squared distances between each point in the series. While this metric is often used to compare the similarity of two metrics, by definition it enforces a one-to-one mapping of points at each time step in two series. In other words, this metric is not invariant to time shifts. Consequentially, times series clustering tasks will break down if the vectors being compared do not align exactly along the time axis.

Distance metrics dedicated to temporal sequences can produce more meaningful results by addressing the issues described above. One such metric is Dynamic Time Warping (DTW), which stretches or compresses different regions of one times series to best match another. This makes this measure invariant to misaligned data, allowing the clustering algorithm to identify matching patterns in vectors even if the patterns occur at different time steps. [32]

First, a cost matrix containing the exhaustive combination of all pairwise Euclidean distances between points in the time series is constructed. It is then possible to trace contiguous alignment paths through the cost matrix. For a specific alignment path, the distance between the two time series is given by the sum of the costs along the path through the cost matrix. The goal of DTW is to identify the alignment path with the shortest total cost. This minimal distance is used as the similarity measure for the clustering algorithm.

Consider two temporal time sequences, x and y, with lengths n and m respectively. DTW can be formulated as the following optimization problem.

$$DTW(x, y) = \min_{\pi} \sqrt{\sum_{(i,j)\epsilon\pi} d(x_i, y_j)^2}$$

where $\pi = [\pi_0, ..., \pi_K]$ is a path that satisfies the following properties:

- It is a list of index pairs $\pi_k = (i_k, j_k)$ with $0 \leq i_k \leq n$ and $0 \leq j_k \leq m$
- $\pi_0 = (0, 0)$ and $\pi_K = (n - 1, m - 1)$
- For all $k > 0$, $\pi_k = (i_k, j_k)$ is related to $pi_{k-1} = (i_{k-1}, j_{k-1})$ as follows:
  - $i_{k-1} \leq i_k \leq i_{k-1} + 1$

- $j_{k-1} \leq j_k \leq j_{k-1} + 1$

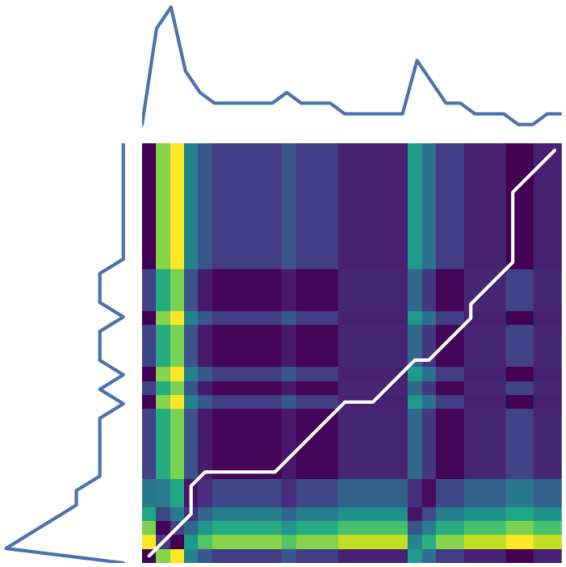

Figure 4-1: An example of DTW

Figure 4-1 provides an example of the cost matrix and optimal alignment path for two signals. In summary, DTW is calculated as the square root of the sum of squared distances between each point in x and the nearest point in y.

4.2 Clustering Method

Clustering analysis is an unsupervised machine learning method for partitioning data points into groups by minimizing intra-cluster distance between each data point and the centroid that it is assigned to. K-means is a widely used, easily interpretable and efficient algorithm for clustering. [17] Utilizing DTW as a similarity measure enables the production of more meaningful results when applying the k-means clustering algorithm to time series.

K-means uses an iterative procedure to partition data into k clusters. The value of k is an input parameter to the algorithm and is determined via the methodology outlined in Section 4.4. First, cluster centroids are initialized and each time-series is assigned to its nearest cluster using the similarity measure. Second, cluster centroids

are recomputed as the barycenters, the time series which minimizes the sum of squared distances to all of the time series in a given cluster. [24] The process of assigning time series to centroids and updating the cluster centers continues until the distance between every time series and its assigned cluster centroid is minimized.

Tslearn, a machine learning toolkit for the analysis of time series data, is utilized to perform clustering as described in this section. [35]

4.3 Clustering Validation Index

The Clustering Validation Index (CVI) is an indicator of how well clusters are formed, and is used to determine the optimal number of clusters. [2] For this analysis, two internal CVIs are used: inertia and silhouette score. Inertia is calculated as the sum of distances of each data point to their closest cluster center. The goal is to have a low inertia, as well as a low number of clusters to improve interpretability. Silhouette score was originally proposed by Rousseeuw as a method of comparing the tightness of and separation between clusters. [31] A value closer to 1 indicates that clusters are well separated from each other and clearly distinguished, while a value of -1 indicates that clusters have been assigned incorrectly.

These CVIs are both considered when selecting the optimal number of clusters for this analysis. Inertia is calculated using the tslearn library, and silhouette score is calculated with the sklearn package [23].

4.4 Parameter Selection

The k-means clustering method requires the total number of clusters, k, as an input parameter. In this study clustering is performed on data from each patient two separate times: once with time series data from a single respiratory measure, and again with time series data from two different respiratory measures. The best value of k is identified for each case by testing values from two to twenty and assessing cluster quality using the CVIs defined prior. Based on visual inspection of the results shown

in Figures 4-2 and 4-3, the optimal number of clusters was determined to be four when preforming clustering in the univariate case, and three when preforming clustering in the multivariate case.

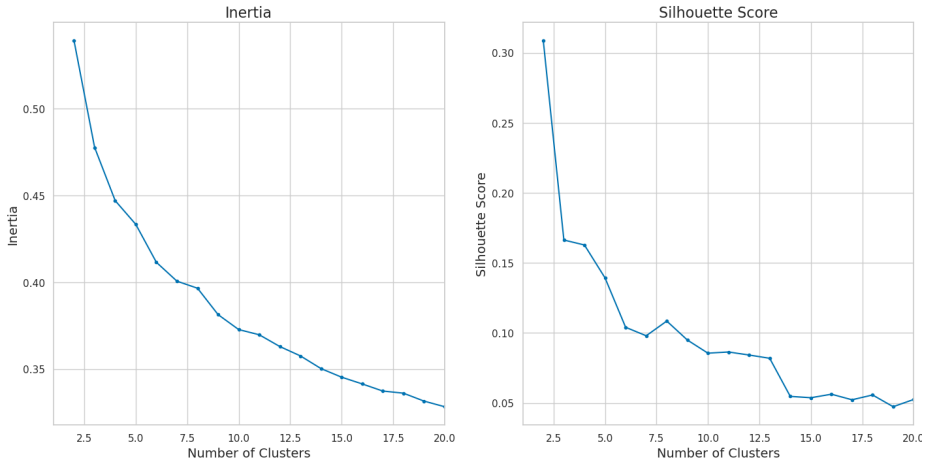

Figure 4-2: Performance evaluation for time series-clustering of one time series with different values of k.

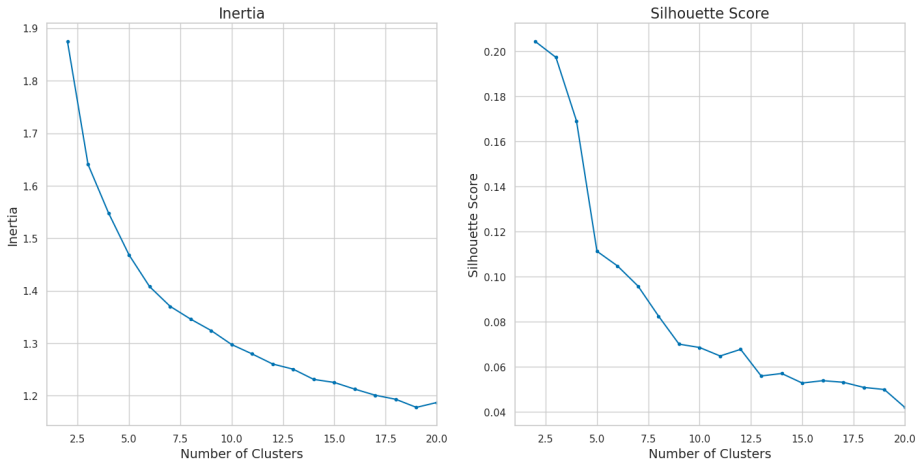

Figure 4-3: Performance evaluation for time series-clustering of multiple time series with different values of k.

4.5 Statistical Analysis Methods

Performing time-series clustering revealed several groups of patients, each with unique patterns in their respiratory measure signals. To facilitate a better understanding of what variables may be correlated with specific patterns in patients' respiratory measures during the ramp period, further statistical analysis is performed. The goal is to identify statistically significant differences between the demographic characteristics, signal features, and perception of breathing discomfort in patients within each of the clusters.

The variables being assessed are both numeric and categorical. For numeric variables the Kruskal-Wallis test, a nonparametric one-way analysis of variance (ANOVA), is applied to assess the statistical significance of differences in the median values across the clusters. [34] This test does not require that the data distributions are normal. The Pearson's chi-squared test is used to assess the statistical significance of differences between clusters for the categorical variable in the data set. [19] The significance level of all tests is fixed at 5%.

While the Kruskal-Wallis and chi-squared tests are useful for determining whether significant differences exist across clusters, they do not provide insights into exactly which clusters are different. Therefore, it is necessary to perform additional post hoc comparisons between pairs of clusters if a significant different is identified. Dunn's test is used for this purpose in this analysis. [10] The Bonferroni correction method is used to adjust p-values. [16]

Chapter 5

Results

Time series clustering was performed on the data as described in Chapter 4, once with time series data from a single respiratory measure, and again with time series data from two respiratory measures. The following sections present the resulting clusters, and the statistical analysis of the differences in selected variables between clusters.

5.1 Clustering Patients with a Single Respiratory Measure

When using only respiratory signal 1 as the input time series data to the clustering algorithm, four clusters of patients were formed. The silhouette score of 0.16, indicates that while there is likely overlap between groups, the clusters are a relatively good fit. The median, 25th percentile and 75th percentile respiratory measures for each cluster are shown in Figure 5-1. Visual inspection of the clusters reveals differences in the breathing patterns of the four groups during the sleep onset period of therapy session. Cluster A includes patients whose respiratory measure stays relatively constant throughout the time period. In contrast, patients in Clusters B, C, and D exhibit similar patterns in this respiratory measurement: the signals is at its maximum at the start of the therapy session and then decrease to a relatively stable value over time. The difference between these three clusters is in the average or the signal during

the later half of the time series. Cluster B includes patients who exhibit the largest drop in this respiratory measure from the maximum to the average. In Cluster C the difference between the peak and the average is slightly smaller, and in Cluster D this distance is less still.

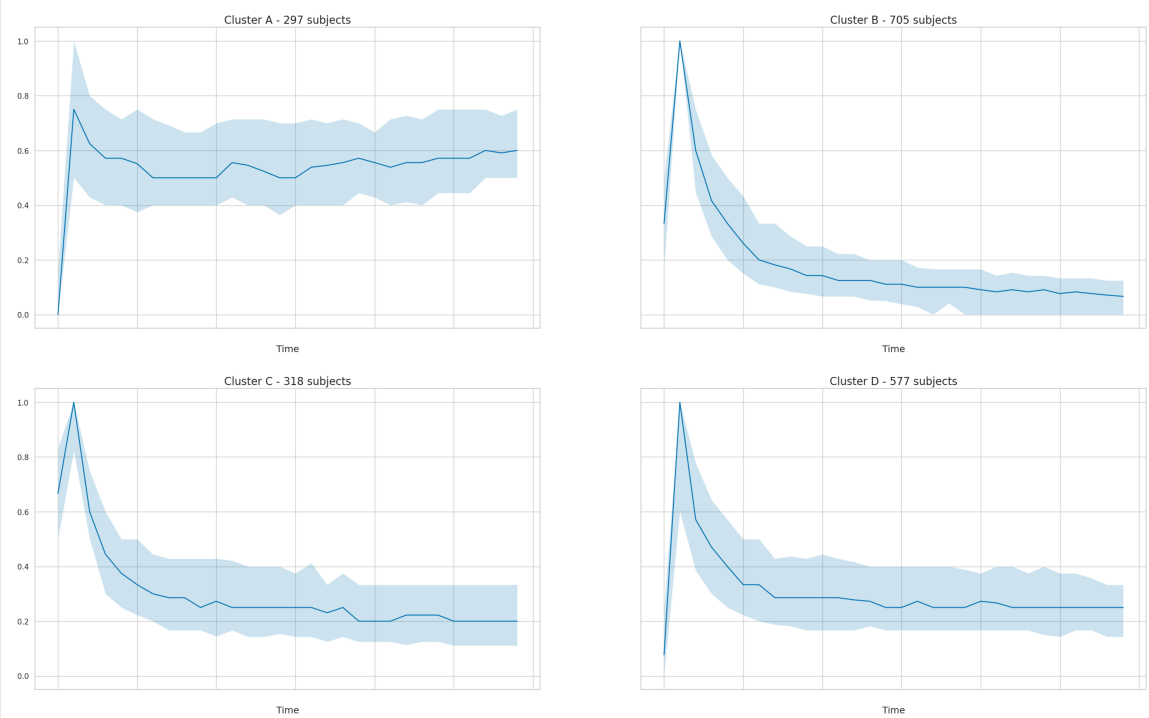

Figure 5-1: Median and 25th to 75th percentile respiratory measures for each cluster.

5.1.1 Comparison of Variables Across Clusters

Table 5.1 describes and compares the demographic characteristics, device settings, usage, survey responses, and respiratory signal characteristics of patients in each of the clusters. Note that the values of the respiratory data have been standardized. The independence between clusters and each variable is tested via the statistical tests described in Section 4.5. These tests confirm that the cluster that a patient belongs to is independent of age and gender. This indicates that the patterns in respiratory signal 1 that are used to determine which cluster a patient belongs to are not determined by either of these variables, and patients in different demographic groups will likely exhibit similar breathing patterns.

Belonging to one cluster versus another is not independent of characteristics of the patient's respiratory signal 1 data ($p < 0.001$ for all variables), and the standard deviation ($p = 0.015$), minimum ($p = 0.009$) and range ($p < 0.001$) of respiratory signal 2. Most notably, the frequency of patients that reported experiencing breathing discomfort during the start of their last therapy session is also found to be statistically significant ($p = 0.009$). Patients in Cluster B report experiencing breathing discomfort at the highest rate (31.2%), while Cluster C includes the smallest portion of patients reporting breathing discomfort (21.4%). However, there is no significant difference in the compliance rates of patients in each cluster. There is also no significant difference in the average number of minutes that patients use their device 28 days days before, 7 days before, 7 days after, 28 days after, and 60 days after completing the comfort survey. Thus, while patients may be more likely to report that they are experiencing breathing discomfort in specific clusters, there is not a meaningful difference in therapy usage over time. Finally, there are no significant differences between any of the features pertaining to the patient's device settings — S1, S2, SF1, and SF2 — across clusters.

The differences between the characteristics of respiratory signal 1 and signal 2 during the sleep onset period of therapy are visualized in Figure 5-2. When examining respiratory signal 1, Cluster B is characterized by a higher mean, standard deviation, minimum, maximum, and range than all other clusters. Cluster A has the lowest values across all respiratory signal 1 characteristics. However, there are several cases where the differences are minimal when compared to Cluster C and Cluster D, specifically when comparing standard deviation and range. There is slight variation in the median standard deviation, minimum and range of the respiratory signal 2 data of patients across clusters, although these differences are more subtle than those that exist when comparing respiratory signal 1 characteristics.

		Overall	Cluster A	Cluster B	Cluster C	Cluster D	P-Value
n		1897	297	705	318	577	
Reported Breathing Discomfort, n (%)	No	1378 (72.6)	221 (74.4)	485 (68.8)	250 (78.6)	422 (73.1)	0.009
	Yes	519 (27.4)	76 (25.6)	220 (31.2)	68 (21.4)	155 (26.9)	
Age, median [Q1,Q3]		56.0 [47.0,66.0]	56.0 [48.0,64.0]	57.0 [47.0,66.0]	55.0 [46.2,66.0]	56.0 [47.0,66.0]	0.720
Gender, n (%)	Female	923 (48.7)	146 (49.2)	351 (49.8)	141 (44.3)	285 (49.4)	0.562
	Male	967 (51.0)	151 (50.8)	350 (49.6)	176 (55.3)	290 (50.3)	
	Prefer not to say	7 (0.4)		4 (0.6)	1 (0.3)	2 (0.3)	
S1, median [Q1,Q3]		4.0 [4.0,4.0]	4.0 [4.0,4.0]	4.0 [4.0,4.0]	4.0 [4.0,4.0]	4.0 [4.0,4.0]	0.094
S2, median [Q1,Q3]		5.0 [5.0,6.0]	5.0 [5.0,6.0]	5.0 [5.0,6.0]	5.0 [5.0,6.0]	5.0 [5.0,6.0]	0.690
SF1, median [Q1,Q3]		10.0 [6.0,21.0]	10.0 [5.0,23.0]	10.0 [6.0,19.0]	9.0 [6.0,20.0]	11.0 [6.0,23.0]	0.162
SF2, median [Q1,Q3]		7.0 [4.0,15.0]	7.0 [4.0,18.0]	7.0 [3.0,14.0]	6.0 [3.2,13.0]	7.0 [3.0,15.0]	0.231
Stuffiness, n (%)	Extremely Clear	200 (10.5)	35 (11.8)	77 (10.9)	27 (8.5)	61 (10.6)	0.506
	Extremely Stuffy	51 (2.7)	12 (4.0)	17 (2.4)	7 (2.2)	15 (2.6)	
	Normal	1026 (54.1)	153 (51.5)	364 (51.6)	181 (56.9)	328 (56.8)	
	Slightly Clear	129 (6.8)	20 (6.7)	45 (6.4)	23 (7.2)	41 (7.1)	
	Slightly Stuffy	491 (25.9)	77 (25.9)	202 (28.7)	80 (25.2)	132 (22.9)	
Average Usage - 7 Days Before Survey, mean (SD)		372.2 (123.4)	357.0 (125.3)	378.6 (117.8)	377.2 (126.0)	369.4 (127.3)	0.065
Average Usage - 28 Days Before Survey, mean (SD)		367.1 (118.3)	355.2 (114.8)	373.4 (115.4)	368.3 (120.8)	364.8 (121.9)	0.152
Average Usage - 7 Days After Survey, mean (SD)		358.4 (140.1)	344.2 (138.2)	363.4 (138.3)	356.6 (144.8)	360.5 (140.5)	0.245
Average Usage - 28 Days After Survey, mean (SD)		356.8 (133.1)	345.5 (130.5)	361.8 (130.6)	355.4 (139.1)	357.4 (134.1)	0.362
Average Usage - 60 Days After Survey, mean (SD)		351.1 (133.4)	338.5 (130.1)	358.8 (129.1)	347.7 (141.6)	350.2 (135.4)	0.158
Compliance Status, n (%)	In Progress	6 (0.3)	2 (0.7)	1 (0.1)		3 (0.5)	0.081
	Compliant	1771 (93.4)	278 (93.6)	669 (94.9)	288 (90.6)	536 (92.9)	
	Not Compliant	120 (6.3)	17 (5.7)	35 (5.0)	30 (9.4)	38 (6.6)	
RS1 Mean, median [Q1,Q3]		0.3 [0.2,0.3]	0.2 [0.2,0.3]	0.3 [0.2,0.4]	0.3 [0.2,0.3]	0.3 [0.2,0.3]	<0.001
RS1 Standard Deviation, median [Q1,Q3]		0.2 [0.1,0.3]	0.1 [0.1,0.2]	0.2 [0.2,0.3]	0.1 [0.1,0.2]	0.1 [0.1,0.2]	<0.001
RS1 Maximum, median [Q1,Q3]		0.3 [0.2,0.3]	0.2 [0.1,0.2]	0.3 [0.2,0.4]	0.2 [0.2,0.3]	0.2 [0.2,0.3]	<0.001
RS1 Minimum, median [Q1,Q3]		0.4 [0.3,0.5]	0.3 [0.2,0.4]	0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.4 [0.3,0.5]	<0.001
RS1 Range, median [Q1,Q3]		0.2 [0.1,0.3]	0.1 [0.1,0.2]	0.3 [0.2,0.4]	0.2 [0.1,0.2]	0.2 [0.1,0.3]	<0.001
RS2 Mean, median [Q1,Q3]		0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.491
RS2 Standard Deviation, median [Q1,Q3]		0.2 [0.2,0.3]	0.2 [0.2,0.3]	0.2 [0.2,0.3]	0.3 [0.2,0.4]	0.2 [0.2,0.3]	0.015
RS2 Maximum, median [Q1,Q3]		0.3 [0.2,0.4]	0.3 [0.2,0.4]	0.3 [0.2,0.4]	0.3 [0.3,0.4]	0.3 [0.2,0.4]	0.452
RS2 Minimum, median [Q1,Q3]		0.5 [0.4,0.6]	0.5 [0.4,0.6]	0.5 [0.4,0.6]	0.5 [0.4,0.6]	0.5 [0.4,0.6]	0.009
RS2 Range, median [Q1,Q3]		0.2 [0.2,0.3]	0.3 [0.2,0.3]	0.2 [0.2,0.3]	0.3 [0.2,0.4]	0.2 [0.2,0.3]	<0.001

Table 5.1: Descriptive statistics for 4 clusters from performing time series clustering on respiratory signal 1 temporal data

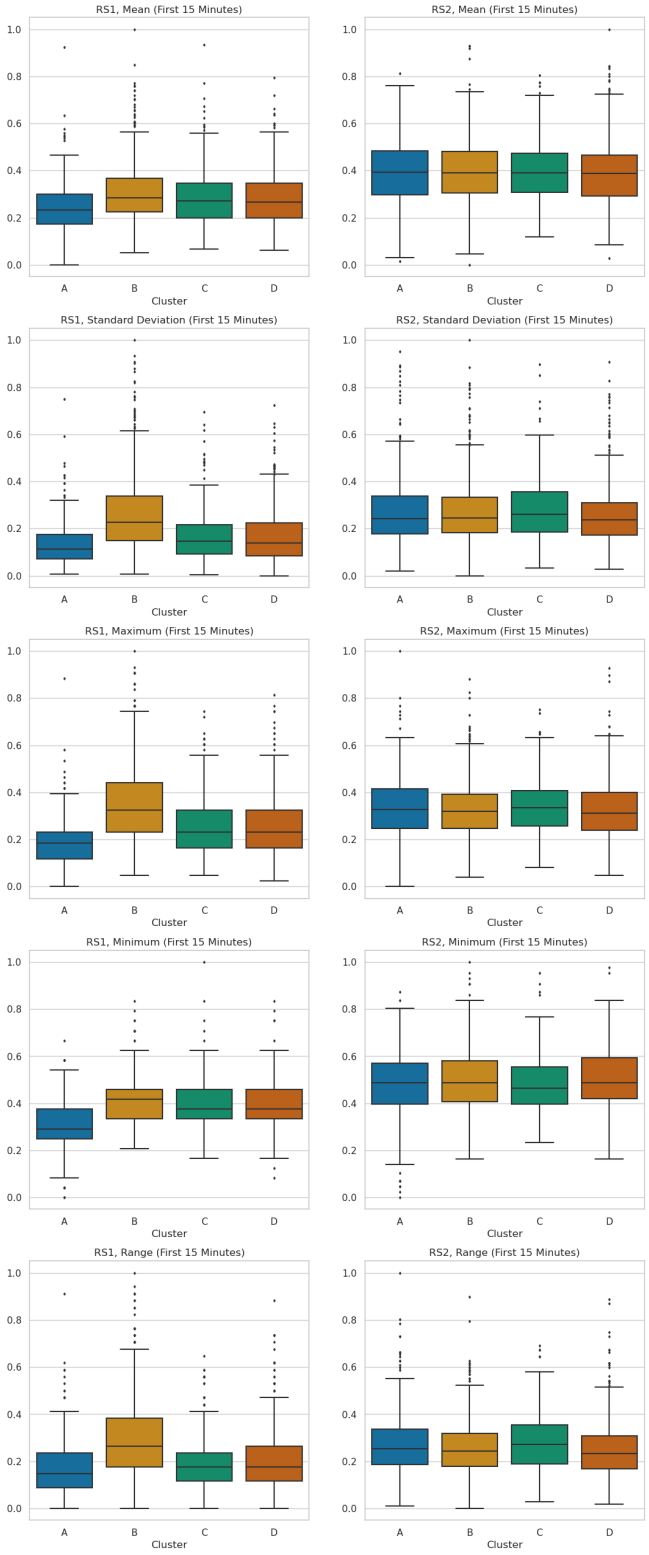

Figure 5-2: Comparison of breathing characteristics across the four clusters formed based on patient's respiratory signal 1 data

5.1.2 Post Hoc Analysis

Post hoc tests are conducted to further understand the differences between pairs of clusters, with results shown in Table 5.2. All pairs of clusters except for Clusters C and D are statistically different with regards to respiratory signal 1 mean, standard deviation, and maximum. The mean respiratory signal 1 minimum of Cluster A is significantly different from all other clusters, and there is also a significant difference in this variable between Clusters B and D. The respiratory signal 1 range is only significantly different between Cluster B and all other clusters.

Cluster #1	**A**	**A**	**A**	**B**	**B**	**C**
Cluster #2	**B**	**C**	**D**	**C**	**D**	**D**
Reported Breathing Discomfort	0.412	1	1	0.006	0.497	0.471
RS1, Mean	<0.001	<0.001	<0.001	<0.001	<0.001	1
RS1, Standard Deviation	<0.001	<0.001	<0.001	0.025	0.002	1
RS1, Maximum	<0.001	<0.001	<0.001	<0.001	<0.001	1
RS1, Minimum	<0.001	<0.001	<0.001	0.318	<0.001	0.170
RS1, Range	<0.001	0.268	0.010	<0.001	<0.001	1
RS2, Standard Deviation	1	1	0.415	0.706	0.276	0.011
RS2, Minimum	1	1	0.346	0.155	0.980	0.006
RS2, Range	0.699	1	0.021	0.012	0.444	<0.001

Table 5.2: p-values from post hoc analysis of variables identified as significantly different across clusters when analyzing respiratory signal 1

The differences in respiratory signal 2 standard deviation are not found to be statistically significant between any pairs of clusters when compared individually. The mean respiratory signal 2 minimum is only significantly different between Clusters C and D ($p = 0.006$). Significant differences in the respiratory signal 2 range are found between Clusters D and A ($p = 0.021$), D and C ($p = <0.001$), and B and C ($p = 0.012$). These results indicates that while patterns in respiratory signal 1 may be similar within the groups of patients identified by the clustering algorithm, the

relationship between respiratory signal 1 and respiratory signal 2 over time is likely different across individuals in each group due to natural biological variability.

When comparing the difference in the frequency of patients who reported experiencing breathing discomfort across clusters, there is only a statistically significant difference between Clusters B and C ($p = 0.006$). There are fewer meaningful differences in the respiratory signal characteristics between patients in Cluster B and Cluster C, however there may be additional characteristics that were not explored that are able to better capture the differences in the respiratory signal 1 patterns of patients in each group.

5.2 Clustering Patients Based on Multiple Respiratory Signals

Time series clustering was performed again on both the respiratory signal 1 and respiratory signal 2 signals from each patient, forming three clusters with a silhouette score of 0.173. The median and interquartile range of each signal are shown for each group in Figure 5-3. Cluster C is characterized by a median respiratory signal 1 that is relatively high compared to other clusters, and a steady respiratory signal 2 throughout the sleep onset period. Cluster B includes patients whose respiratory signal 1 drops off dramatically after the first one to two time steps after turning on the machine, while the median respiratory signal 2 increases slightly to a rate that is again fairly constant throughout the sleep onset period. Finally, Cluster A is characterized by patients who exhibit a decrease in their respiratory signal 2 as their respiratory signal 1 decreases throughout the first portion of the sleep onset period.

5.2.1 Comparison of Variables Across Clusters

The demographic characteristics, device settings, usage, survey responses, and respiratory measures of the clusters are again compared using the Kruskal-Wallis and chi-squared tests. Results are summarized in Table 5.3. In this case, statistically

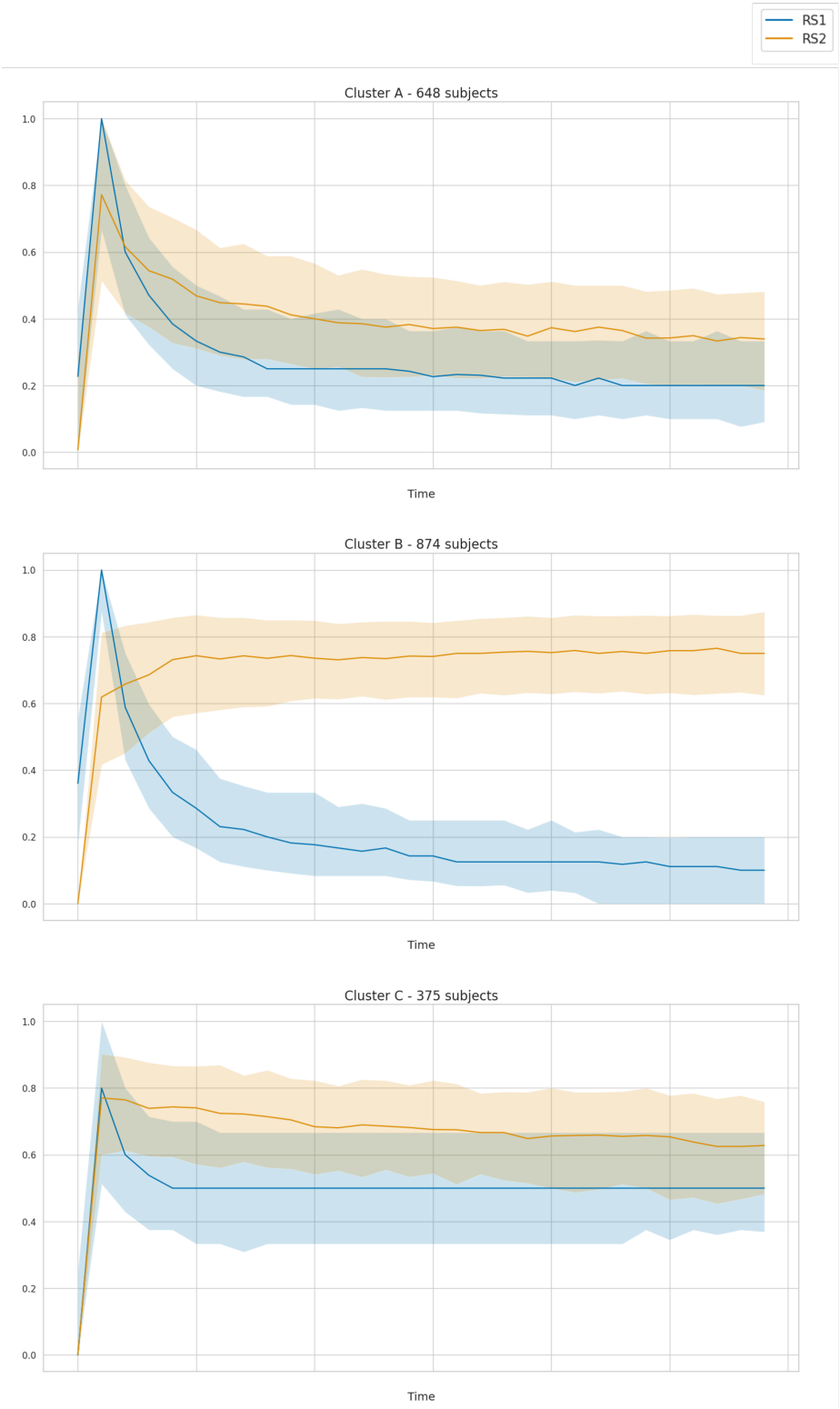

Figure 5-3: Median and 25th to 75th percentile range respiratory signal 1 and respiratory signal 2 for each cluster

significant difference between clusters are seen with regards to all respiratory signal 2 and respiratory signal 1 variables ($p = <0.001$ for all variables) except for the mean respiratory signal 2 of subjects ($p = 0.821$). One feature derived from the device settings data, SF1, was found to be significantly different across clusters ($p = 0.016$), along with the settings features S1 and S2 ($p = 0.004$ and $p = 0.001$ respectively).

It is important to note here that the median age of subjects in each cluster is significantly different with a high confidence level ($p = <0.001$), and gender is significantly different across clusters ($p = 0.010$). This may indicate that the relationship between respiratory signals changes as a patient ages, and this is the trend in the data that is influencing the formation of groups.

In this case, this is no meaningful difference between clusters in the number of subjects who report wanting more air at the start of their therapy sessions ($p = 0.191$). There are also no significant difference in usage variables, patient reports of stuffiness, compliance, or baseline AHI. Since there is no significant difference in the outcome variable that we are interested in, an ad hoc analysis is not performed.

		Overall	Cluster A	Cluster B	Cluster C	P-Value
n		1897	648	874	375	
Reported Breathing Discomfort, n (%)	No	1378 (72.6)	485 (74.8)	618 (70.7)	275 (73.3)	0.191
	Yes	519 (27.4)	163 (25.2)	256 (29.3)	100 (26.7)	
Age, median [Q1,Q3]		56.0 [47.0,66.0]	59.0 [50.0,67.0]	55.0 [45.0,65.0]	56.0 [47.0,65.0]	<0.001
Gender, n (%)	Female	923 (48.7)	343 (52.9)	394 (45.1)	186 (49.6)	0.010
	Male	967 (51.0)	304 (46.9)	474 (54.2)	189 (50.4)	
	Prefer not to say	7 (0.4)	1 (0.2)	6 (0.7)		
S1, median [Q1,Q3]		4.0 [4.0,4.0]	4.0 [4.0,4.0]	4.0 [4.0,4.0]	4.0 [4.0,4.0]	0.004
S2, median [Q1,Q3]		5.0 [5.0,6.0]	5.0 [5.0,6.0]	5.0 [5.0,6.0]	5.0 [4.5,6.0]	0.001
SF1, median [Q1,Q3]		10.0 [6.0,21.0]	11.0 [6.0,21.0]	10.0 [6.0,20.0]	9.0 [5.0,20.0]	0.016
SF2, median [Q1,Q3]		7.0 [4.0,15.0]	7.0 [3.0,16.2]	7.0 [3.0,13.0]	7.0 [4.0,16.0]	0.252
Stuffiness, n (%)	Extremely Clear	200 (10.5)	71 (11.0)	90 (10.3)	39 (10.4)	0.844
	Extremely Stuffy	51 (2.7)	20 (3.1)	18 (2.1)	13 (3.5)	
	Normal	1026 (54.1)	349 (53.9)	484 (55.4)	193 (51.5)	
	Slightly Clear	129 (6.8)	45 (6.9)	58 (6.6)	26 (6.9)	
	Slightly Stuffy	491 (25.9)	163 (25.2)	224 (25.6)	104 (27.7)	
Average Usage - 7 Days Before Survey, mean (SD)		372.2 (123.4)	377.0 (126.1)	371.3 (121.2)	365.8 (124.0)	0.362
Average Usage - 28 Days Before Survey, mean (SD)		367.1 (118.3)	375.1 (119.1)	363.1 (118.2)	362.6 (116.9)	0.106
Average Usage - 7 Days After Survey, mean (SD)		358.4 (140.1)	366.4 (142.0)	354.9 (139.0)	352.7 (139.4)	0.196
Average Usage - 28 Days After, mean (SD)		356.8 (133.1)	365.3 (133.3)	352.7 (133.7)	351.8 (130.9)	0.137
Average Usage - 60 Days After, mean (SD)		351.1 (133.4)	358.8 (133.9)	347.8 (133.6)	345.8 (132.0)	0.193
Compliance Status, n (%)	In Progress	6 (0.3)	2 (0.3)	2 (0.2)	2 (0.5)	0.888
	Compliant	1771 (93.4)	608 (93.8)	815 (93.2)	348 (92.8)	
	Not Compliant	120 (6.3)	38 (5.9)	57 (6.5)	25 (6.7)	
RS1 Mean, median [Q1,Q3]		0.3 [0.2,0.3]	0.3 [0.2,0.4]	0.3 [0.2,0.4]	0.2 [0.2,0.3]	<0.001
RS1 Standard Deviation, median [Q1,Q3]		0.2 [0.1,0.3]	0.2 [0.1,0.3]	0.2 [0.1,0.3]	0.1 [0.1,0.2]	<0.001
RS1 Maximum, median [Q1,Q3]		0.3 [0.2,0.3]	0.3 [0.2,0.3]	0.3 [0.2,0.4]	0.2 [0.1,0.2]	<0.001
RS1 Minimum, median [Q1,Q3]		0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.3 [0.2,0.4]	<0.001
RS1 Range, median [Q1,Q3]		0.2 [0.1,0.3]	0.2 [0.1,0.3]	0.2 [0.1,0.4]	0.1 [0.1,0.2]	<0.001
RS2 Mean, median [Q1,Q3]		0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.4 [0.3,0.5]	0.821
RS2 Standard Deviation, median [Q1,Q3]		0.2 [0.2,0.3]	0.2 [0.2,0.3]	0.3 [0.2,0.3]	0.2 [0.2,0.3]	<0.001
RS2 Maximum, median [Q1,Q3]		0.3 [0.2,0.4]	0.4 [0.3,0.4]	0.3 [0.2,0.4]	0.3 [0.2,0.4]	<0.001
RS2 Minimum, median [Q1,Q3]		0.5 [0.4,0.6]	0.6 [0.5,0.6]	0.4 [0.4,0.5]	0.5 [0.4,0.6]	<0.001
RS2 Range, median [Q1,Q3]		0.2 [0.2,0.3]	0.2 [0.1,0.3]	0.3 [0.2,0.3]	0.3 [0.2,0.3]	<0.001

Table 5.3: Descriptive statistics for 3 clusters from performing time series clustering on respiratory signals 1 and 2 temporal data

Chapter 6

Discussion and Future Work

6.1 Discussion of Results

This study applies a time-series clustering approach to categorize respiratory data gathered throughout the night by ResMed's AirSense 11 machines with the goal of understanding how different patterns may relate to patient reports of breathing discomfort. The previous section presents the outputs of two clustering attempts applied to different combinations of respiratory data. Through the investigation of various variables pertaining to patient demographics, therapy usage over time, device settings, breathing characteristics, and self-reported outcomes we see that patterns in respiratory signal 1 are more useful for grouping patients who are more likely to be experiencing breathing discomfort than respiratory signal 1 and respiratory signal 2 together. When performing clustering on respiratory signal 1 data, a significant difference in the frequency of patients who were reported experiencing breathing discomfort was observed ($p = 0.009$). In contrast, no significant differences were found between clusters formed by grouping patients using a combination of respiratory signals ($p = 0.123$).

This finding indicates that patterns in respiratory signal 1 from a patient during the sleep onset period of the therapy session may be a meaningful indicator of whether they are experiencing breathing discomfort, independent of demographic features including age and gender. There appears to be some relationship between the magnitude and

standard deviation of respiratory signal 1 at the beginning of therapy sessions and the frequency of breathing discomfort reports. However, it is important to note that although Clusters B and C show significant differences in the reported experiences of patients, there are no meaningful differences between any other pairs of clusters. A significant amount of overlap exists in the respiratory patterns of patients who report experiencing breathing discomfort and those who do not, so it is likely that these features will not be useful to identify breathing discomfort with a high level of accuracy. It is possible that the data used in this study are obscuring features that are unique to the breathing discomfort phenomenon that could be visible in a breath-by-breath waveform.

When considering the cost of patients spending time at low pressure levels and breathing discomfort on adherence to therapy, this study yields inconclusive results. A meaningful difference in average nightly usage following the completion of the comfort survey was not observed between groups of patients identified through time series clustering. It is possible that this was the case because patients who were surveyed had been on PAP therapy for various lengths of time. The early experiences of patients are critical for determining whether or not they will adhere to therapy long term. [7] Since the comfort survey was conducted after patients had already completed their initial weeks, or even months, of therapy, it is possible that breathing discomfort during the sleep onset period of therapy is a side effect that they had already accepted. Patients who found breathing discomfort to be a critical enough issue that it caused them to reduce their PAP usage may not have been willing survey participants. To fully understand the impact that breathing discomfort has on patient compliance and long term adherence data, patient reports of breathing discomfort should be gathered periodically throughout the first month following the date that a patient receives their AutoSet device.

Finally, it should be noted that the subjective nature of breathing discomfort makes it difficult to map respiratory signals to anything other than patients' self-reported symptoms. It is possible that a patient could feel symptoms of breathing discomfort and not exhibit any changes in their breathing patterns. Biological variability also

means that breathing discomfort could present in the respiratory patterns of patients in ways that are unique to each individual.

6.2 Limitations

As with most studies, the findings presented in this thesis are subject to limitations. First, the data on each patient's self-reported experiences with breathing discomfort was taken from a general survey on all aspects of the patient's comfort while using ResMed devices. The data comes from a question that was designed specifically to measure whether a patient experienced breathing discomfort during the sleep onset period of their therapy sessions. However, the question asked about patients' experiences in general, rather than about a specific therapy session. Therefore, it was not possible to map reports of breathing discomfort to specific therapy sessions with absolute certainty. Rather than applying the same outcome label to all therapy session, the assumption was made that patients would weight their recent experiences most heavily when providing their responses and only the most recent therapy session prior to the completion of the survey was considered.

6.3 Data Management Considerations

Looking beyond the case study presented in this thesis, there are several aspects of data management that should be considered when utilizing machine learning technologies. The quality of data used for training machine learning models, including its completeness, target accuracy and feature accuracy, can have a significant impact on performance. [5] Therefore, proper data management is critical for enabling reliable decision-making.

6.3.1 Data Granularity

One essential consideration when gathering data for machine learning applications is data granularity — the level of detail in the data set. The granularity of the data is a

factor in determining what type of analysis can feasibly be performed, and whether the results of the analysis will lead to meaningful and accurate conclusions. If data granularity is too low, only large patterns will be visible in the data. Valuable insights may be lost as disparate trends are compressed into a single result. On the other hand, data that is too granular can introduce noise that makes it challenging for machine learning algorithms to distinguish which patterns in the data are meaningful. The resulting data set can also become extremely large and difficult to process.

In order to select the appropriate level of granularity of time series data for use in machine learning applications, domain knowledge of the problem is required. It is necessary to understand which features in the data may be useful for predicting the desired output, and the timescale at which these features occur at. This will help determine the frequency at which time series data should be sampled at. If the goal is to capture trends and patterns at a short-term scale, and it is important to detect changes in real-time then it is likely necessary to collect data at a fine level of granularity. Some general examples of applications where this may be the case include predictive maintenance or stock market analysis. In contrast, machine learning algorithms that are designed to identify longer-term trends and seasonality in data may perform better with data collected at a coarser level of granularity. For example, climate analysis where trends in weather patterns occur over weeks, months, or even years, and the analysis of seasonal trends in retail.

6.3.2 Data Labeling

A second critical component of data management is ensuring target label accuracy and precision. A well performing machine learning model requires high-quality data labels as these provide a "ground truth" for the algorithm. However, acquiring these high-quality labels can be challenging as it is time consuming and costly. [8]

Several methods exist for acquiring data labels. Data may be labeled by domain experts, which ensures a higher level of quality but comes at a higher cost and may take more time. Crowdsourcing leverages the crowd intelligence to generate labels for data points making it a faster and lower cost option. However, this approach can

yield results that are not reliable due to poor quality control or a lack of domain expertise amongst contributors. Synthetic data may also be generated that has the same attributes of real data. This requires more computational power, but is cost and time effective, reduces privacy risks, and may yield a more robust data set that incorporates additional edge cases. The choice of method for labeling data depends on the complexity of the problem and data, and time and resource constraints.

When looking specifically at clinical applications of machine learning and the use of patient reported outcomes as labels, it is important to consider the subjective nature of the outcomes, and the many factors that could influence patients' responses on the measures of interest. These factors may include the design of the instrument used to collect patient responses, the method and setting via which responses are collected, and patient level behaviors such as response style or recall bias. [9] It is important to recognize that all data collection is impacted by some level of error. Patient reported data is still a valuable input for understanding how patients perceive their health-related quality of life, symptoms and symptom burden, and health behaviors. [22] Therefore, a reasonable level of imperfection may be acceptable and should not discourage the continued collection and usage of patient reported outcomes.

It also may be possible to account for the uncertainty of labels when creating machine learning models. For example, performance was improved by incorporating the quantification of clinical diagnostic uncertainty when training an algorithm to detect the development of acute respiratory distress syndrome in patients. [27] The application of similar approaches could be applied to other types of data labels for machine learning application across a variety of domains.

6.3.3 Data Considerations for Deep Learning

Long short term memory networks have been applied to time series classification tasks and achieved competitive performance when compared to fully connected networks. [18] As well, transformers have shown good modeling capability for dependencies and interaction in sequential data over long ranges. This makes them well suited for time series tasks such as classification and forecasting. [37] When considering the

application of these technologies, it is important to note that these approaches typically require significantly more labeled data than traditional machine learning approaches. Leveraging pre-trained models for transfer learning may decrease the amount of training data required to achieve high levels of performance, however depending on the application this may or may not be possible. As is the case with traditional machine learning methods, it is important that the data set has an appropriate level of granularity, includes a balanced set of diverse samples, and is of high quality (i.e., data is clean, in the proper format, consistent, and has been validated). [26]

Chapter 7

Conclusion

In this thesis the application of time-series clustering analysis to trends in breathing data was discussed, specifically exploring whether patterns in time series respiratory data are correlated with patient reports of breathing discomfort. Temporal data from ResMed's AirSense 11 machines was categorized via k-means clustering using DTW as a similarity measure, and a statistical analysis was conducted to understand the differences in demographics, the number of patients reporting breathing discomfort, and features in the breathing data across clusters of patients. It was discovered that when time series clustering was performed on respiratory signal 1, a significant difference was found in the frequency of patients reporting breathing discomfort across clusters.

However, further analysis revealed that this difference was only significant between one pair of the four clusters. Therefore, in conclusion, this data is likely not useful for identifying patient who are experiencing breathing discomfort in its current form. It is possible that the biological differences across patients and the subjective nature of breathing discomfort make it such that breathing patterns in one-minute averaged data for a patient who is experiencing breathing discomfort at low pressures are not distinguishable from a patient who is comfortable and breathing normally. To further this line of research ResMed may choose to explore the use of higher fidelity data, and the collection of patient reports of breathing discomfort throughout their therapy journey. The collection of additional data could enable the application of

other supervised machine learning methods in the future, and the quantification of the cost of under-ventilation and breathing discomfort on therapy usage over time.

Bibliography

[1] Duong Tuan Anh and Le Huu Thanh. "An efficient implementation of k-means clustering for time series data with DTW distance". In: *International Journal of Business Intelligence and Data Mining* 10.3 (2015), pp. 213–232.

[2] Olatz Arbelaitz et al. "An extensive comparative study of cluster validity indices". In: *Pattern Recognition* 46.1 (2013), pp. 243–256. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2012.07.021. URL: https://www.sciencedirect.com/science/article/pii/S003132031200338X.

[3] Kevin Bascol et al. "Unsupervised Interpretable Pattern Discovery in Time Series Using Autoencoders". In: *Structural, Syntactic, and Statistical Pattern Recognition*. Ed. by Antonio Robles-Kelly et al. Cham: Springer International Publishing, 2016, pp. 427–438. ISBN: 978-3-319-49055-7.

[4] Adam V Benjafield et al. "Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis". en. In: *Lancet Respir. Med.* 7.8 (Aug. 2019), pp. 687–698.

[5] Lukas Budach et al. *The Effects of Data Quality on Machine Learning Performance*. 2022. DOI: 10.48550/ARXIV.2207.14529. URL: https://arxiv.org/abs/2207.14529.

[6] Paolo Buono et al. "Interactive pattern search in time series". In: *Visualization and Data Analysis 2005*. Ed. by Robert F. Erbacher et al. Vol. 5669. International Society for Optics and Photonics. SPIE, 2005, pp. 175–186. DOI: 10.1117/12.587537. URL: https://doi.org/10.1117/12.587537.

[7] P. G. Catcheside. "Predictors of continuous positive airway pressure adherence". In: *F1000 Med Rep* 2 (Sept. 2010).

[8] Chengliang Chai et al. "Data Management for Machine Learning: A Survey". In: *IEEE Transactions on Knowledge and Data Engineering* 35.5 (2023), pp. 4646–4667. DOI: 10.1109/TKDE.2022.3148237.

[9] Eric M Chang, Erin F Gillespie, and Narek Shaverdian. "Truthfulness in patient-reported outcomes: factors affecting patients' responses and impact on data quality". In: *Patient Related Outcome Measures* 10 (2019). PMID: 31354371, pp. 171–186. DOI: 10.2147/PROM.S178344. eprint: https://www.tandfonline.com/doi/pdf/10.2147/PROM.S178344. URL: https://www.tandfonline.com/doi/abs/10.2147/PROM.S178344.

[10] Alexis Dinno. "Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn's Test". In: *Stata Journal* 15 (Apr. 2015), pp. 292–300. DOI: 10.1177/1536867X1501500117.

[11] Jonatan Enes et al. "A pipeline architecture for feature-based unsupervised clustering using multivariate time series from HPC jobs". In: *Information Fusion* 93 (2023), pp. 1–20. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2022.12.017. URL: https://www.sciencedirect.com/science/article/pii/S1566253522002652.

[12] Lawrence J Epstein et al. "Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults". en. In: *J. Clin. Sleep Med.* 5.3 (June 2009), pp. 263–276.

[13] Tak-Chung Fu et al. "Pattern discovery from stock time series using self-organizing maps". In: *Workshop Notes of KDD2001 Workshop on Temporal Data Mining* (Jan. 2001).

[14] Cathy A Goldstein et al. "Artificial intelligence in sleep medicine: an American Academy of Sleep Medicine position statement". en. In: *J. Clin. Sleep Med.* 16.4 (Apr. 2020), pp. 605–607.

[15] Cathy A Goldstein et al. "Artificial intelligence in sleep medicine: background and implications for clinicians". en. In: *J. Clin. Sleep Med.* 16.4 (Apr. 2020), pp. 609–618.

[16] Mohieddin Jafari and Naser Ansari-Pour. "Why, when and how to adjust your P values?" en. In: *Cell J.* 20.4 (Jan. 2019), pp. 604–607.

[17] A. K. Jain, M. N. Murty, and P. J. Flynn. "Data Clustering: A Review". In: *ACM Comput. Surv.* 31.3 (Sept. 1999), pp. 264–323. ISSN: 0360-0300. DOI: 10.1145/331499.331504. URL: https://doi.org/10.1145/331499.331504.

[18] Fazle Karim et al. "LSTM Fully Convolutional Networks for Time Series Classification". In: *IEEE Access* 6 (2018), pp. 1662–1669. DOI: 10.1109/ACCESS.2017.2779939.

[19] Hae-Young Kim. "Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test." In: *Restor Dent Endod* 42.2 (2017). PMID: 28503482, pp. 152–155. DOI: 10.5395/rde.2017.42.2.152. eprint: https://doi.org/10.5395/rde.2017.42.2.152. URL: https://doi.org/10.5395/rde.2017.42.2.152.

[20] Ragnar H. Lesch, Yannick Caillé, and David Lowe. "Component analysis in financial time series". In: *Proceedings of the IEEE/IAFE 1999 Conference on Computational Intelligence for Financial Engineering (CIFEr) (IEEE Cat. No.99TH8408)* (1999), pp. 183–190.

[21] Jessica Lin et al. "VizTree: a Tool for Visually Mining and Monitoring Massive Time Series Databases." In: Dec. 2004, pp. 1269–1272. ISBN: 978-0-12-088469-8. DOI: 10.1016/B978-012088469-8/50124-8.

[22] Centers for Medicare and Medicaid Services. *Patient-Reported Outcome Measures*. Dec. 2022.

[23] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[24] François Petitjean, Alain Ketterlin, and Pierre Gançarski. "A global averaging method for dynamic time warping, with applications to clustering". In: *Pattern Recognition* 44.3 (2011), pp. 678–693. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2010.09.013. URL: https://www.sciencedirect.com/science/article/pii/S003132031000453X.

[25] Ana Radovanović et al. "Application of Agglomerative Hierarchical Clustering for Clustering of Time Series Data". In: *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. 2020, pp. 640–644. DOI: 10.1109/ISGT-Europe47291.2020.9248759.

[26] Aiswarya Raj et al. *Data Management Challenges for Deep Learning*. 2019.

[27] Narathip Reamaroon et al. "Accounting for Label Uncertainty in Machine Learning for Detection of Acute Respiratory Distress Syndrome". In: *IEEE Journal of Biomedical and Health Informatics* 23.1 (2019), pp. 407–415. DOI: 10.1109/JBHI.2018.2810820.

[28] Susan Redline and Shaun M Purcell. "Sleep and Big Data: harnessing data, technology, and analytics for monitoring sleep and improving diagnostics, prediction, and interventions-an era for Sleep-Omics?" en. In: *Sleep* 44.6 (June 2021).

[29] ResMed. *AirSense 11 Clinical Guide*.

[30] Brian W Rotenberg, Dorian Murariu, and Kenny P Pang. "Trends in CPAP adherence over twenty years of data collection: a flattened curve". en. In: *J. Otolaryngol. Head Neck Surg.* 45.1 (Aug. 2016), p. 43.

[31] Peter J. Rousseeuw. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(87)90125-7. URL: https://www.sciencedirect.com/science/article/pii/0377042787901257.

[32] H. Sakoe and S. Chiba. "Dynamic programming algorithm optimization for spoken word recognition". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.1 (1978), pp. 43–49. DOI: 10.1109/TASSP.1978.1163055.

[33] American Academy of Sleep Medicine. *Osbtructive Sleep Apnea*. Online PDF. 2008.

[34] Neil R. Smalheiser. "Chapter 12 - Nonparametric Tests". In: *Data Literacy*. Ed. by Neil R. Smalheiser. Academic Press, 2017, pp. 157–167. ISBN: 978-0-12-811306-6. DOI: https://doi.org/10.1016/B978-0-12-811306-6.00012-9. URL: https://www.sciencedirect.com/science/article/pii/B9780128113066000129.

[35] Romain Tavenard et al. "Tslearn, A Machine Learning Toolkit for Time Series Data". In: *Journal of Machine Learning Research* 21.118 (2020), pp. 1–6. URL: http://jmlr.org/papers/v21/20-091.html.

[36] T. Warren Liao. "Clustering of time series data—a survey". In: *Pattern Recognition* 38.11 (2005), pp. 1857–1874. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2005.01.025. URL: https://www.sciencedirect.com/science/article/pii/S0031320305001305.

[37] Qingsong Wen et al. *Transformers in Time Series: A Survey*. 2022. DOI: 10.48550/ARXIV.2202.07125. URL: https://arxiv.org/abs/2202.07125.