Analysis of Respiratory Pressure–Volume Curves in Intensive Care Medicine Using Inductive Machine Learning

Article *in* Artificial Intelligence in Medicine · September 2002

DOI: 10.1016/S0933-3657(02)00053-2 · Source: PubMed

Steven Ganzert
Technische Universität München

Josef Guttmann
University of Freiburg

Kristian Kersting
University of Bonn

Analysis of Respiratory Pressure-Volume Curves in Intensive Care Medicine Using Inductive Machine Learning

Steven Ganzert[*/**] Josef Guttmann[*] Kristian Kersting[**]

Ralf Kuhlen[***] Christian Putensen[***] Michael Sydow[***] Stefan Kramer[**]

[*] Department of Anesthesiology and
Critical Care Medicine
Albert-Ludwigs-University
Hugstetter Str. 55
D-79106 Freiburg, Germany
guttmann@ana1.ukl.uni-freiburg.de

[**] Institute of Computer Science
Albert-Ludwigs-University Freiburg
Georges-Koehler-Allee, Building 079
D-79110 Freiburg, Germany
{ganzert, kersting, skramer}@informatik.uni-freiburg.de

[***]Scientific Board of the Clinical Multi-Center Study

Abstract: We present a case study of Machine Learning and Data Mining in intensive care medicine. In the study, we compared different methods of measuring pressure-volume curves in artificially ventilated patients suffering from the Adult Respiratory Distress Syndrome (ARDS). Our aim was to show that inductive Machine Learning can be used to gain insights into differences and similarities among these methods. We defined two tasks: The first one was to recognize the measurement method producing a given pressure-volume curve. This was defined as the task of classifying pressure-volume curves (the classes being the measurement methods). The second was to model the curves themselves, that is, to predict the volume given the pressure, the measurement method and the patient data. Clearly, this can be defined as a regression task. For these two tasks, we applied C5.0 resp. CUBIST, two inductive Machine Learning tools. Apart from medical findings regarding the characteristics of the measurement methods, we found some evidence showing the value of an abstract representation for classifying curves: Normalization and high-level descriptors from curve fitting played a crucial role in obtaining reasonably accurate methods. Another useful feature of algorithms for inductive Machine Learning is the possibility of incorporating background knowledge. In our study, the incorporation of patient data helped to improve regression results dramatically, which might open the door for the individual respiratory treatment of patients in the future.

Keywords: machine learning, data mining, classification, decision trees, regression, rule learning, classification of curves, intensive care medicine, artificial ventilation, Adult Respiratory Distress Syndrome (ARDS).

1 Introduction

Artificial ventilation is one of the key therapies in intensive care medicine. In particular if the lung is ill, it is not unproblematic, and there is always the risk of damaging it by barotrauma or volutrauma, respectively. Therefore, it is important to observe the mechanical status of the artificially ventilated lung. The two quantities most crucial for assessing the pathophysiological mechanical condition of the respiratory system (lung and thorax) are the Compliance C (the volume distensibility), and the air-flow Resistance R (both will be formally defined in the following section). There are two fundamentally different approaches to measuring Compliance C: two-point and multiple-point methods. Two-point methods obtain measurements at two different volumes, usually at the end of inspiration and at the end of expiration, whereas multiple-point methods obtain measurements during the whole breath. In order to find the respective advantages and disadvantages of various two- and multiple-point methods, the university hospitals of Aachen, Berlin, Bonn, Freiburg i. Br., Göttingen, Hamburg, Mannheim-Heidelberg and Munich are conducting a multi-center study in collaboration with the company Drägerwerk AG, Lübeck. Three multi-point methods (LOW-FLOW, SUPER-SYRINGE, SLICE) as well as two two-point methods (SCASS and PEEP-WAVE) are applied in randomized order to each patient participating in the study.

The central issue of the multi-center study is to discover systematic differences among the pressure-volume data generated by these five different measurement methods. Our aim was to show that Data Mining methods, and in particular methods from inductive Machine Learning, are generally well-suited to detect structural characteristics of measured pressure-volume curves. Another goal was to obtain interpretable results, such that physicians could inspect them and relate them to prior knowledge. This would advance the state of the art in the area, since at present pressure-volume curves are "just" analyzed visually: Important parameters like the lower inflection point *LIP* (to be defined below) and the upper inflection point *UIP*, are determined upon visual inspection of the pressure-volume curves. Both the lower and upper inflection point are critical for setting the parameters of artificial ventilatory systems. So, the application of Data Mining and Machine Learning tools might not only lead to new insights about the similarities and differences among these measurement methods, but ultimately contribute to more objective decisions.

Many different approaches can be taken to perform Data Mining. Some Data Mining techniques originate from the field of Machine Learning. One advantage of the Machine Learning methods used here is that they can flexibly handle all sorts of descriptors. For instance, in our domain it is possible to include patient data and see whether there is a connection between properties of the patients and the shape of the pressure-volume curves. In the long run, this might open the door for the individual treatment of patients under artificial ventilation.

From a Machine Learning perspective, this paper is concerned with the analysis of curves by means of classification and regression techniques. In particular, the classification of curves is an interesting topic, because surprisingly few papers in the literature deal with this type of classification task. Our experience was that normalization and curve fitting techniques dramatically improve the classification results. Apart from these findings, our Machine Learning approach may be even more widely applicable, since it yields interpretable models of the similarities and differences among curves.

The organization of this paper parallels the one in Cios *et al.* [1]: Section 2 details the medical problem domain, Section 3 presents the datasets and Section 4 the preparation of the data. Section 5 focuses on the actual Data Mining step: It describes the data modeling techniques, the training and test procedures, and the results from a quantitative point of view. Subsequently, the discovered knowledge is evaluated in Section 6. Section 7 dicusses the usage and potential usage of the discovered knowledge, before a section touching upon related work. Finally, we conclude the paper and hint at possible directions of further research.

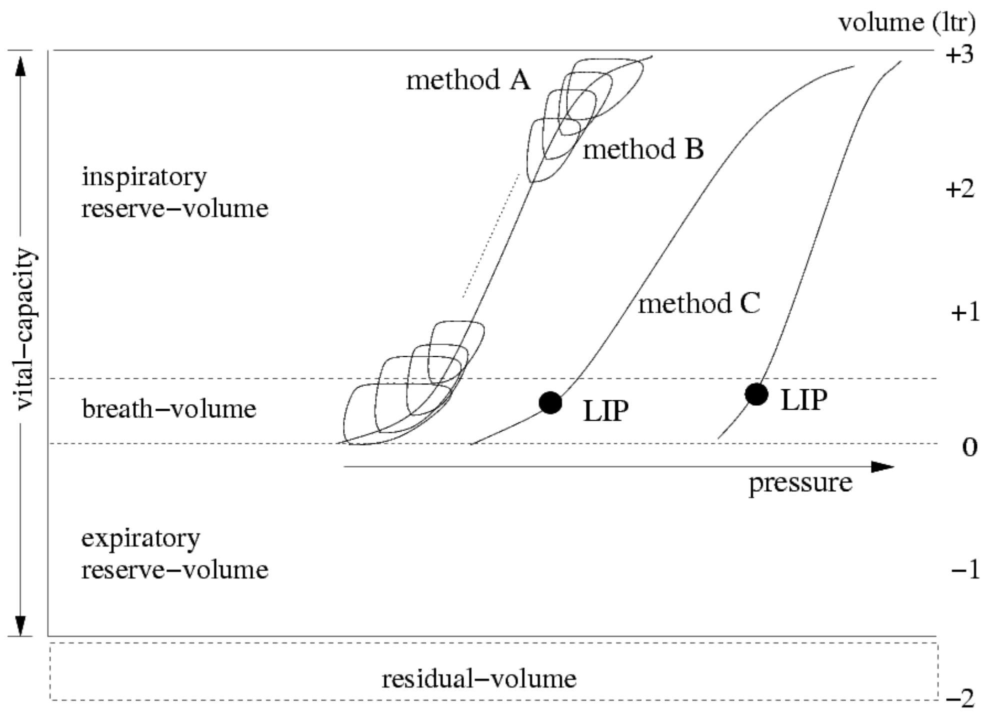

Figure 1: Schematic illustration of pressure-volume curves for various measurement methods. For the sake of clarity, the pressure-volume curves for methods *B* and *C* are shifted to higher pressure ranges.

2 Understanding Medical Problem Domain

In this section, we present the medical problem domain, the medical objectives of this study and the current solution to the addressed problem.

Figure 1 schematically shows the connection between scanning the lung volume in several breaths and the resulting pressure-volume curves. The *x*-axis represents the pressure, the *y*-axis the volume. Each of the *loops* superimposed on curve *A* describes one complete breath with its phases of inspiration and expiration. Each of the *curves* describes the pressure-volume correlation of one measurement method. For the curves *B* and *C*, the so-called lower inflection points (LIPs) are marked.

Two of the most crucial quantities for assessing the pathophysiological mechanical condition of the respiratory system (lung and thorax) are the Compliance *C* and the Resistance *R*. The Compliance describes the distensibility regarding the applied volume and can physically be described as the quotient of the difference of the end-inspiratory and the end-expiratory volume (ΔV), and the difference of end-inspiratory and end-expiratory alveolar pressure (ΔP) per breath. So, the Compliance *C* is defined as the slope of pressure-volume curves (see, e.g, Figure 1). The air-flow Resistance *R* is defined as the quotient of resistive pressure drop ΔP_{res} and the air-flow rate $\dot{V}$.

As stated in the introduction, there are two approaches to measuring the Compliance: two-point and multiple-point methods. Two-point methods measure at two different volumes, whereas multiple-point methods repeatedly perform measurements during the breath, and afterwards evaluate the Elastance E (the reciprocal value of the Compliance) and the Resistance by the equation of motion:

$$P_{rt} = E \times V + R \times \dot{V} + I \times \ddot{V} + K$$

where P_{rt} is the pressure in the respiratory tract. The inertance I describes the portion of pressure caused by inertia and can be ignored. The size K describes the pressure which has an effect on the lung already before or at the beginning of the breath and is called dynamic intrinsic PEEP (positive end-expiratory pressure).

Typical two- and multiple-point methods evaluate the Compliance as a linear quantity and give a good description of the healthy lung. However, for severely ill lungs (such as those of patients suffering from ARDS, the Adult Respiratory Distress Syndrome), the Compliance does not run linearly over the whole breath. Therefore, Guttmann et al. [2] have developed the SLICE method, where each breath is partitioned into a fixed number of six slices (parts of the applied tidal volume). For each of these slices, the compliance is evaluated separately.

2.1 Measurement Methods

In the following, we briefly introduce the four methods used to measure the pressure-volume curves, as studied in this paper. All patients investigated were mechanically ventilated using an EVITA4 ventilator by Drägerwerk AG which was modified by the manufacturer to perform the respiratory maneuvers automatically. Since recognizing PEEP-WAVE curves does not require automatic analysis due to their systematic discontinuities, we are not considering the PEEP-WAVE method in this paper.

LOW-FLOW [3]: In this method the patient's lung is inflated using extremely low air-flow rates. For that, the state of complete expiration is held for five seconds and afterwards the lung is filled with an air-flow rate of 2 l/min until P_{rt} reaches 45 mbar. Due to the low flow rate, the resistance component in the equation of motion is eliminated. The measurements are taken continuously. Under these quasi-static conditions, the recorded pressure-volume relationship can be analyzed directly.

SUPER-SYRINGE [4]: Here the ventilatory system gradually applies volume portions of 100 ml with an air-flow rate of 30 l/min to the lung. At the end of each applied volume portion**,** there is a short zero-flow maneuver to eliminate the resistive pressure component. During these breaks the measurements are taken. Afterwards a new volume portion is applied. The maneuver is repeated in volume steps of 100ml until a pressure of 45 mbar is reached.

SLICE [2]: In order to compare the SLICE method with other measurement methods, the ventilatory system applies a volume to reach an end-inspiratory pressure of 45 mbar. Normally the SLICE method does not require such a maneuver because the data is taken during a regular breath. In our case, the whole breath is similar to a normal breath but with a much larger tidal volume. Such a breath is repeated five times and the mean of these five measurements is calculated. To obtain pressure values at equidistant volumes for the phases of inspiration and expiration, the pressure measurements are linearly interpolated and the corresponding pressure values are computed. Now, the total tidal volume can be divided into six slices (volume portions) and an analysis of respiratory mechanics can be performed using a standard least-squares-fit algorithm.

SCASS [5] (Static Compliance by Automated Single Steps): In this method the applied tidal volumes have a gradual difference of 50 ml and are chosen at random. With each breath, the pressure P_{rt} rises at most up to 45 mbar. After each applied volume the respiration is stopped for 5 seconds to interrupt the air-flow. During this break the pressure-volume measurements are taken. Before the next application of a volume chosen at random, five breaths of the normal tidal volume, i.e. of the volume applied before starting the maneuver, are applied to the patient. In this way, measurements of the whole volume of the lung are taken, and the compliance can be computed under quasi-static conditions.

2.2 Medical Objectives

The central issue of the multi-center study is to discover systematic differences among pressure-volume data generated by the different measurement methods. Our aim was to show that Data Mining and Machine Learning methods are generally suitable to detect objective characteristics of the shape of the pressure-volume curves. We also wanted to obtain interpretable results in the form of trees and rules to describe these differences. In today's medical practice, the analysis of pressure-volume curves is performed upon visual.inspection. For instance, it is common practice to detect the so-called Lower-Inflection-Point (LIP) visually. The LIP is important, because the lung should be ventilated in a range of high Compliance, and it should not be ventilated at pressure values below the LIP in order to avoid alveolar collapse. In practice, a sigmoid shape of the pressure-volume curve is assumed. This sigmoid shape of the curve is required to detect the LIP in the described way (see again Figure 1). An advantage of Data Mining and Machine Learning methods is the possibility to include "background knowledge", that is, features that describe more than the mere shape of the curves. For instance, patient data or the period of artificial ventilation before the measurements could be relevant.

3 Understanding the Data

In this section, we have a closer look at the pressure-volume data that forms the basis of our study, and the size and format of the datasets.

For all four of the above measurement methods, we had the data from ten patients available. In addition, we had some incomplete datasets: one for which only the LOW-FLOW method was available, two for which the SCASS method was not applied, and one for which the LOW-FLOW method was not applied. From these 50 example curves, we had to remove one SLICE curve due to excessively many unknown values, giving only 49 measurements for the classification dataset. The examples for each measurement method were almost equally distributed: We had 13 measurements from the LOW-FLOW method, 12 from the SLICE method, 13 from SUPER-SYRINGE and 11 from SCASS.

The regression dataset contains 9,209 examples. Due to the different measurement methods, the datasets contained from 7 (SLICE method) up to several hundred (LOW-FLOW method) pressure/volume-pairs. The background data gave details about the gender, the date of birth, weight, height, the beginning of the artificial ventilation, the tube-type (tracheal or endotracheal), the internal diameter of the tube in mm (tube-ID), the calculated Compliance and the measurement method. Figure 2 shows the measurements from the four methods for one of the patients graphically.

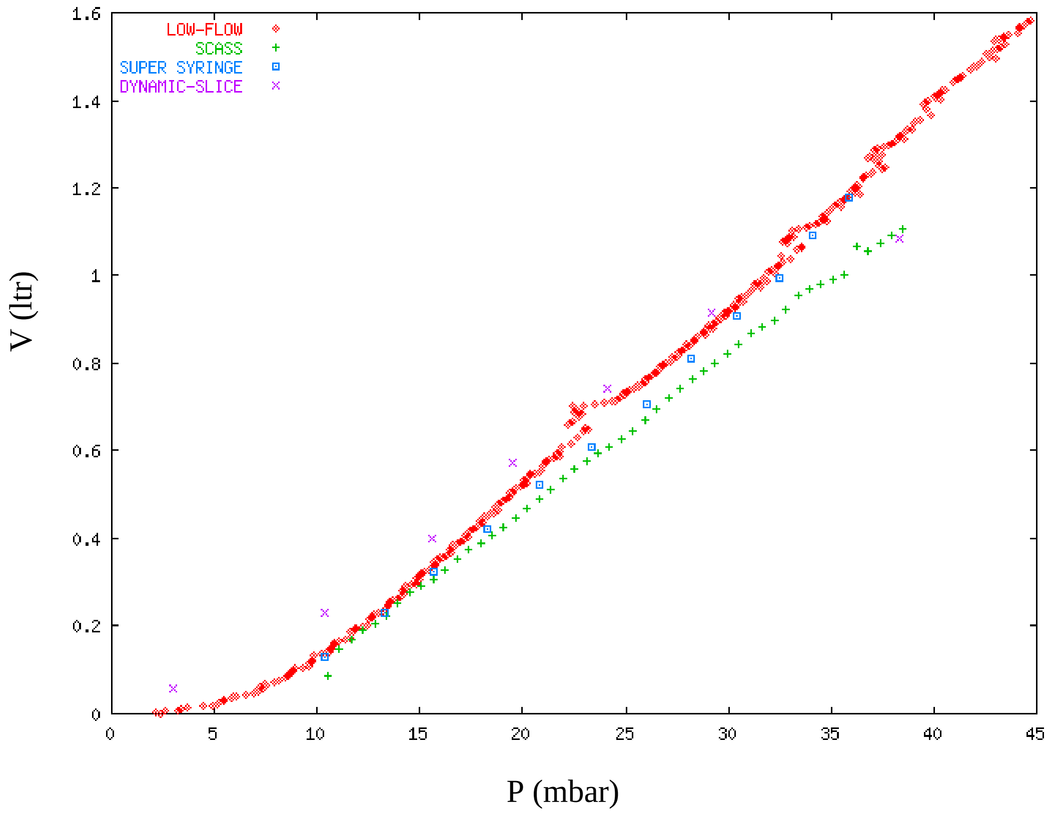

Figure 2: Pressure-volume curves obtained from four measurement methods for one patient.

4 Preparation of the Data

For the applied classification and regression algorithms (see below), the datasets had to be represented as *feature vectors of fixed size*.

4.1 Pre-Processing for Classification

The experiments were conducted with and without background information about the patients. Since we wanted to find out the differences among measurement methods, the classes to be predicted were LOW-FLOW, SCASS, SUPER SYRINGE and SLICE. Table 1 summarizes the features used for the classification task. For the classification algorithm, we represented the curves by volume values at distinct pressure values. The granularity of this representation (i.e., the number of such "pressure features") was varied. The values were generated by linear interpolation of the given values. Figure 3 shows the interpolated volume values at pressure values from 0.0 up to 50.0 mbar in steps of 5.0 mbar graphically. It depicts the raw data from Figure 2 after this pre-processing step. Pressure values below 5.0 mbar and above 40.0 mbar were not recorded at this measurement and, therefore, could not be interpolated.

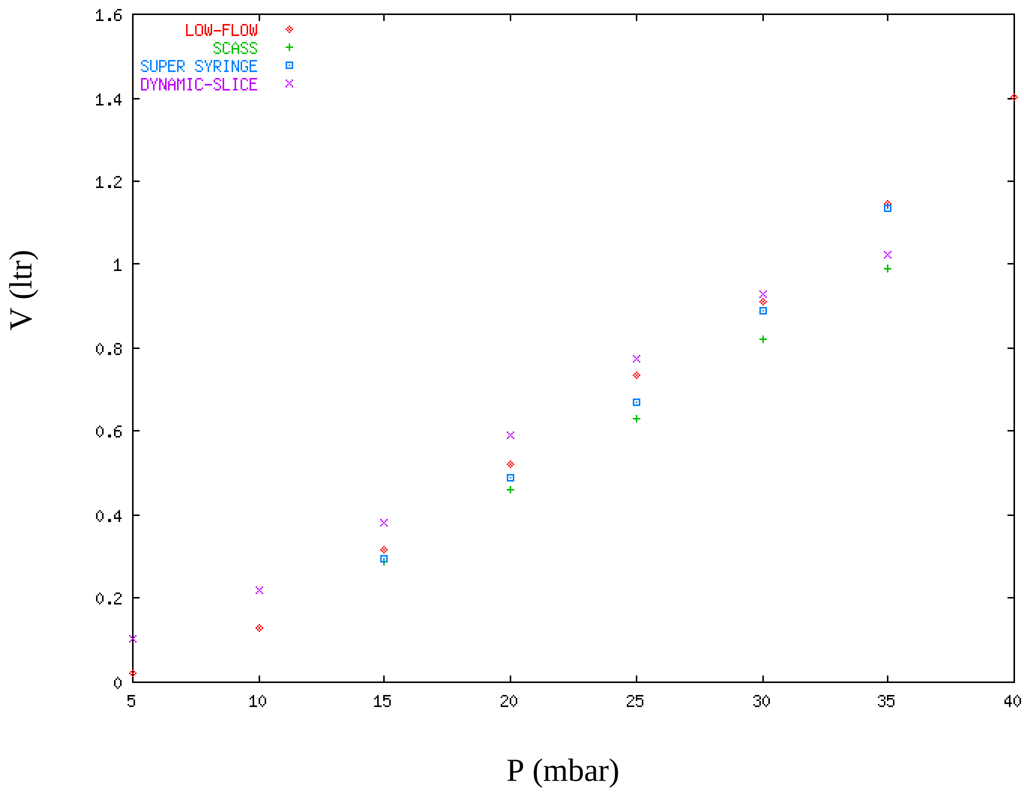

Figure 3: The four curves from the interpolated pressure/volume-pairs. Such a representation of curve data is used in the classification experiments.

As the sigmoid shape of the curve reflects the physiological basis of respiratory mechanics, we introduced a high-level descriptor representing the shape of the curve: We fitted a sigmoid function to the pressure-volume curves, obtaining parameter *b*, which indicates the sigmoidity and the slope of them.

Volume values	Volume values at pressure values of **0.0 - 50-0 mbar in steps of 5.0 mbar**.
Volume values	Volume values at pressure values of **0.0 – 50.0 mbar in steps of 1.0 mbar**.
Normalized volume values I	The first **normalization** of the pressure and volume values of each measurement by a method was calculated **over the maximal pressure resp. volume value** of this measurement. So, we generated values between 0.0 and 1.0 for the volume and pressure. We used normalized volume values between 0.0% and 100.0% of the maximum pressure **in steps of 1.0%, 2.0% and 10.0%** as features.
Normalized volume values II	The second **normalization** was calculated **over the interval of the recorded pressure resp. volume values** of a measurement. (In contrast, normalization I runs from 0.0 mbar to the maximum of a measurement.) We again generated pressure and volume values between 0.0 and 1.0. As above, we used volume values between 0.0% and 100.0% of the recorded pressure interval **in steps of 1.0%, 2.0% and 10.0%** as features.
Background features	Features **gender**, **age**, **weight**, **height**, **ventilation period** (time of respiration before the measurement), **tube type** (tracheal or endotracheal), **tube ID** (the inner diameter of the tube in mm), and the calculated **compliance**. The compliance has been derived from the curves by Drägerwerk AG.
Parameter *b*	**Parameter *b*** of **the sigmoid function** $$f(P) := min_{vol} + (max_{vol} - min_{vol})/(1 + a * e^{(-b * P)})$$ which **we obtained from fitting the pressure-volume curves**.
Dependent variable (class to be predicted)	Methods **LOW-FLOW, SCASS, SUPER SYRINGE** and **SLICE**.

Table 1: Summary of the features used for classification. Only subsets of these possible features are used in the actual experiments described below.

4.2 Pre-Processing for Regression

For the regression task, the algorithm should learn from the measured pressure/volume-pairs together with the background features. Here, the task was to predict the volume from the pressure, the background features (optionally) and the measurement method. Table 2 summarizes the features used for the regression task.

Pressure values	For the different experiments, we represented the pressure values **not normalized, normalized by method I, normalized by method II (normalization methods as for classification)**
Background features	Features **gender**, **age**, **weight**, **height**, **ventilation period, tube type,** and the **tube ID.** Note that the **compliance** is not included, since it would implicitly describe what we are trying to predict.
Measurement method	Methods **LOW-FLOW, SCASS, SUPER SYRINGE** and **SLICE**.
Dependent variable (real number to be predicted)	Volume values **not normalized, normalized by method I, normalized by method II**

Table 2: Summary of the features used for regression. Only subsets of these possible features are used in the actual experiments described below.

5 Data Mining

In this section, we present and motivate our selection of Data Mining techniques, and give a quantitative assessment of the Data Mining results. In general, inductive Machine Learning techniques are only one possible option for the Data Mining step in the Knowledge Discovery process. Other possibilities include clustering techniques and techniques for the discovery of frequently occurring patterns. The choice clearly depends on the problem at hand.

The given problem is to find similarities and differences in a sample of curves. We address it by applying classification and regression techniques from inductive Machine Learning. There are two reasons for doing so:

(1) In contrast to "black-box" approaches such as artificial neural networks, classification and regression techniques from inductive Machine Learning provide models that lend themselves to inspection and interpretation by human experts.

(2) Splitting the original problem into the two subproblems classification and regression parallels our two subproblems of finding differences and similarities. The classification problem is tackled by learning a decision tree. The attributes tested in the internal nodes of the trees represent the differences among measurement methods. The regression problem is tackled by a regression rule learner. The measurement methods, i.e. the classes of the classification problem are now part of the input (see below): The algorithm has to predict volume values given a current pressure value and a measurement method. Therefore, the learner has the possibility to generalize over different measurement methods at a given pressure value. The regression rules can indicate similarities among the measurement methods.

In the following, we describe the classification and regression techniques used in this study.

5.1 Classification

As classification learner we used C5.0 [6**]**, which is the commercial descendant of ID3 [7] and C4.5 [8]. As we wanted to gain insights into the differences and similarities among the measurement methods, we needed predictive models in a symbolic representation. C5.0 generates decision-trees and can also extract rules from them. In principle, trees and rules are models that lend themselves to interpretation by domain experts. C5.0 models are usually quite comprehensible due to the bias of the algorithm. The algorithm chooses the most "informative" features near the root of the trees, and has a built-in preference for small trees.

5.2 Regression

For the regression task, we used CUBIST [6**]**, a system that learns regression rules from examples. It is known to produce accurate rule sets for regression. In the consequent of the rules, CUBIST does not use a constant for prediction, but a linear regression model. If several rules apply to an example, CUBIST merges the predictions made by all matching rules into a single overall prediction. With each rule, the learner outputs the number of cases covered, the mean and range of the volume values of the given training instances as well as the estimated error of the prediction by the rule. This enables the validation of individual rules.

5.3 Assessment of Classification Results

In Table 3, we summarize the quantitative results from our classification experiments. The table gives the mean error rates (the rates of misclassified examples) from *ten runs of ten-fold cross-validation*. On the left-hand side of the table, we state the features used for describing the pressure-volume curves, and on the right-hand side we give the error rates. The table also represents the "history" of our experiments: We started off *without* normalized data, and varied the step-size parameters in absolute terms (mbar). Note that the number of features used for describing a curve is quite similar to the number of features from the experiments with normalized data. Using a step-size of 1 mbar, we obtain 50 features, using a step-size of 5 mbar, we obtain 10 features. Another goal of the experimentation was to find out whether the inclusion of the background features would improve the classification results. So, we systematically conducted experiments with and without background features. As it turned out, all initial results (without normalization) were quite discouraging: They were approximately as bad as always predicting the majority class.

The next step was to introduce normalizations, as described in the previous section. The rationale for these normalizations was that the absolute values of individual patients do not matter. What matters from a medical point of view is the shape of the curves alone. In both normalization I and normalization II, each pressure-volume curve is normalized individually. Whereas normalization I runs from 0 mbar to the maximal measured value, normalization II runs from the minimal measured value to the maximal measured value. The rationale behind normalization II was that in many cases measurements for lower pressure values are not available, and, thus, should not affect the representation of the curves. As in the previous set of experiments, we systematically varied the step-size parameters (i.e., the granularity of the description of the curves) and the inclusion resp. exclusion of background features. As one can see from Table 3, normalization I significantly improved over the initial results, and normalization II significantly improved over normalization I. All significant differences reported here are so strong that they can be found using all conceivable statistical tests (McNemar, t-test,...). Overall, one could observe that the results improved dramatically, but still were not satisfactory.

Finally, we wanted to enhance the rather low-level representation of our curves with a high-level description. For this purpose, we applied curve fitting to the measured pressure-volume data. Since physicians mostly assume a sigmoid shape of such curves, we fitted a sigmoid function to the measured data. The resulting parameter b (cf. Table 1) was added to the other features in the experiments. As can be seen in Table 3, parameter b again helped to improve the results drastically, and finally we obtained reasonably accurate models.

Summarizing our classification experiments, we found that

- normalization I improved over having no normalization, and normalization II improved over normalization I,
- including a parameter from curve fitting improved the results significantly,
- the inclusion of background feature harmed the classification accuracy significantly,
- and that there is no clear effect of the step-size parameter on the error rate.

Regarding the background features, we suspect that they are harmful due to overfitting. Given only 49 examples, adding 8 features of unclear relevance has the same effect as increasing the noise level. As we will see, this is not the case for the regression task, where more than 9,000 examples are available.

Classification								
Norm.	Features							Results
	Param. *b*	Step size					Back-ground	Error rate (%) Mean (stand. deviation)
		mbar		%				
		1.0	5.0	1.0	2.0	10.0		
no			X				X	75.0 (2.0)
no			X					70.5 (2.0)
no		X					X	71.2 (1.5)
no		X						68.1 (1.4)
I						X	X	69.3 (1.3)
I						X		66.2 (1.1)
I					X		X	60.1 (1.2)
I					X			53.4 (1.3)
I				X			X	59.6 (1.2)
I				X				55.1 (1.3)
II						X	X	51.2 (0.9)
II						X		44.9 (1.2)
II					X		X	42.6 (1.0)
II					X			38.5 (1.1)
II				X			X	47.1 (0.5)
II				X				38.9 (0.7)
II	X							60.7 (1.2)
II	X					X	X	32.5 (0.9)
II	X					X		33.8 (1.1)
II	X				X		X	37.5 (1.0)
II	X				X			34.1 (1.2)
II	X			X			X	39.6 (1.2)
II	X			X				33.2 (0.8)

Table 3: Classification results of C5.0 from ten runs of 10-fold cross-validation.

5.4 Assessment of Regression Results

In Table 4, we present the results from our experiments with CUBIST in terms of the average error, the relative error and the correlation coefficient. Again, the results are averaged over ten runs of ten-fold cross-validation. Here, we only systematically varied the normalizations and the inclusion resp. exclusion of background features. The most striking observation is that the inclusion of background features improves the accuracy of the models almost by one order of magnitude. Normalization is beneficial for regression as well, although normalization I and II perform more or less the same.

Regression				
Normalization	Background	Results of a 10x10 Cross-validation		
		Average error	Relative error	Correlation coefficient
no	X	0.021300	0.03	1.00
no		0.304500	0.49	0.85
I	X	0.008998	0.04	1.00
I		0.054473	0.21	0.97
II	X	0.009365	0.04	1.00
II		0.052078	0.20	0.98

Table 4: Regression results of CUBIST from ten runs of 10-fold cross-validation.

6 Evaluation of the Discovered Knowledge

6.1 Classification

In this section, we present an example tree from the classification experiments and provide an interpretation of the model. We picked the best settings from our experiments, with normalization II, including the background features and parameter *b*. This setting achieved a mean error rate of 32.5 % in ten runs of ten-fold cross-validation. Figure 4 shows a tree learned from all 49 examples using these settings.

Although the curve-fitting was not too exact, we had a good fit in the central part of the curves which is the one with the highest slope. Parameter *b* is an indicator of the sigmoidity of the curve as well as of the slope in the central part: The higher *b*, the higher the slope of the curve. For the four methods, we can conclude the following:

1. The method LOW-FLOW is classified by two tests at level 1 and 2 of the decision tree. These tests imply that the curves of LOW-FLOW measurements have a low slope and low volume values at the beginning of the curve. With these two tests, the curve is classified 100.0% correct, which indicates that these two structural characteristics are of high significance.

2. The method SCASS produces a wide range in the slope of the curves with a tendency towards lower slopes. The tests at level 2 and 3 of the tree imply that the method produces curves which end with high slopes. This is not consistent with the assumption of a sigmoid shape of the curve. Upon visual inspection of the data, one can see that this method seems to produce relatively noisy data. So, we have to be careful with our conclusions regarding SCASS.

3. Like SCASS, SUPER-SYRINGE generates curves with a wide range in the slope of the curves, but in contrast to SCASS we can observe a tendency towards higher slopes.
4. The SLICE method – in contrast to LOW-FLOW – has characteristically high slopes. Also, it has particularly high volume values at the last 30.0% of the curve.

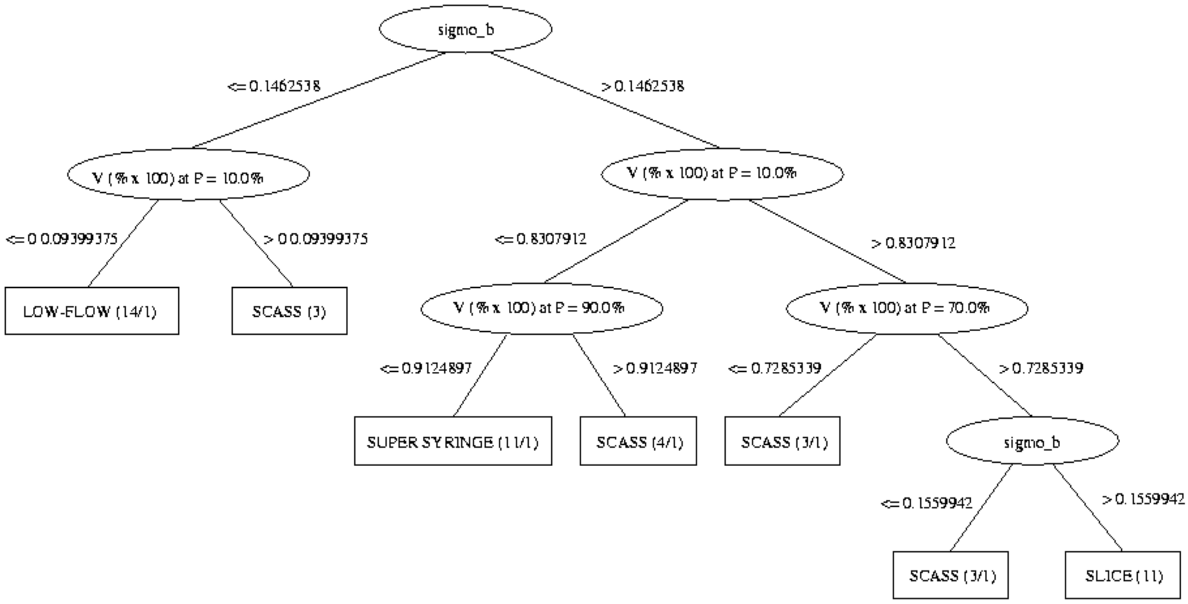

Figure 4: Decision tree learned with C5.0 from all 49 examples. In this particular example tree, the background features do not occur.

6.2 Regression

In this section, we interpret the regression rules learned by CUBIST using normalization II and without background knowledge. We chose this model, because it is small (7 rules only), and still quite accurate (see Table 5). The interpretation of the best regression models (consisting of 45 rules) cannot be included due to space limitations.

Because the data is normalized over the recorded pressure and volume interval, one can make a structural analysis of the regression rules. The seven regression rules clearly divide the curves into four parts:

1. From 0.0% up to 16.0% of the recorded pressure intervals, we can distinguish the methods as follows: SCASS and SLICE are quite similar, but different from LOW-FLOW, which in turn is different from SUPER-SYRINGE.

```
Rule 1: [4252 cases, mean 0.2533045, range -0.05183585 to 0.657424, est err 0.0645252]
if
Method = LOW-FLOW
P <= 0.5423189
then
V = -0.0141977 + 0.708 P

Rule 2: [299 cases, mean 0.2644253, range -0.02333333 to 0.7339268, est err 0.0484499]
if
Method in {SCASS, SUPER-SYRINGE, SLICE}
P <= 0.5423189
then
V = 0.0031941 + 0.951 P

Rule 3: [208 cases, mean 0.2747052, range -0.02333333 to 0.7339268, est err 0.0520135]
if
Method in {SCASS, SLICE}
P <= 0.5423189
then
V = -0.0071364 + 1.074 P

Rule 4: [3491 cases, mean 0.3022743, range 0.007869249 to 0.657424, est err 0.0652686]
if
Method in {LOW-FLOW, SUPER-SYRINGE}
P > 0.1619474
P <= 0.5423189
then
V = -0.0808896 + 1.056 P

Rule 5: [2890 cases, mean 0.6600054, range 0.4107611 to 0.9266548, est err 0.0637494]
if
Method in {LOW-FLOW, SUPER-SYRINGE}
P > 0.5423189
P <= 0.8305588
then
V = -0.105292 + 1.117 P

Rule 6: [102 cases, mean 0.7147808, range 0.5033333 to 0.8991098, est err 0.0520263]
if
Method in {SCASS, SLICE}
P > 0.5423189
P <= 0.8305588
then
V = 0.0642547 + 0.955 P

Rule 7: [1666 cases, mean 0.9191762, range 0.6978347 to 1.068571, est err 0.0225499]
If
```

Table 5: Regression rules induced from 9209 examples described by 11 attributes. Normalization II yields an average error of 0.0517693, a relative error of 0.20 and a correlation coefficient of 0.98 in ten runs of 10-fold cross-validation.

2. From 16.0% up to 54.0% of the recorded pressure intervals, we observe the same pattern. But while in the previous part of the curves the methods LOW-FLOW and SUPER SYRINGE were clearly separated, in this part of the pressure interval the two methods have a common rule. Of course they are not equal, because each of them has another rule for volume prediction in this part, but they appear to have structural similarities.
3. From 54.0% up to 83.0% of the recorded pressure intervals, we can identify two groups of methods: one includes SCASS and SLICE again, and the other LOW-FLOW and SUPER SYRINGE.

4. From 83.0% up to 100.0% of the recorded pressure intervals, we finally cannot distinguish the four measurement methods.

In summary, one can say that the most differences can be found in the first half of the recorded pressure intervals. This is well in accordance with current medical practice, since at present diagnoses are based on the lower part of pressure-volume curves. Another point is that SCASS and SLICE cannot be distinguished according to these rules. Of course we have to be careful with our interpretations, but we can clearly observe tendencies from our models.

7 Using the Discovered Knowledge

From a medical perspective, the discovered knowledge enables an objective analysis of the pressure-volume relationship. At present, in the daily routine, physicians analyze pressure-volume curves only visually, if they are measured at all. In contrast to this approach, Data Mining and Machine Learning offer the opportunity to gain a more objective and general understanding of these measurement methods.

Furthermore, both classification and regression techniques from Machine Learning allow for easy incorporation of many different sorts of features, such as patient-related features. One possible usage of the regression models would be as follows: Given the attribute values for a particular patient, one can generate typical curves for each measurement method. Based on these predicted curves, one could chose the method that is expected to minimize the invasivity for the patient. Of course, this scenario would still require a lot more research into this direction, but it hints at the potential of the approach.

8 Related Work

Inductive Machine Learning techniques have been applied to data from artificial ventilation before: Muller and co-workers [9,10] describe the development of a knowledge-based alarm system for ventilatory therapy using techniques from inductive Machine Learning. Starting with a mathematical model, the authors generated data for "simulated patients", and subsequently applied an Machine Learning algorithm to induce rules linking signal features from each simulated breath to events that occurred while the breath was recorded. The reported classification accuracy was excellent. One of the differences from our study is that we are not analyzing data from simulations, but actually measured pressure-volume curves from a multi-center study.

Regarding the classification of curves, surprisingly few studies are reported in the literature. Džeroski *et al.* [11] presented a case study of diterpene structure elucidation from NMR spectra using techniques from Inductive Logic Programming (ILP). Here, the task was to identify the type of skeleton of diterpenes given their ^{13}C NMR spectrum. While the task was also to classify curves, these curves have a fundamentally different meaning and the techniques applied do not easily carry over to our task.

In the pattern recognition literature, Kudo *et al.* [12] presented an approach that turns example curves with a varying number of measurements into binary feature vectors of fixed-size. The features are defined in terms of the regions that all or almost all curves of one class pass through and no other class does. One of the advantages of this method is that the resulting classification rules can easily be visualized and interpreted.

Gasser and Kneip [13] introduced a methodology for searching for structure in curve samples. In contrast to the approach by Kudo *et al.*, they are not concerned with the classification of curves, but with regression. Gasser and Kneip basically propose some kind of feature extraction: The structural features characterizing a curve can be related to extrema, inflection points, plateaus, etc. The formal basis of this work is non-parametric regression and Kernel density estimation. The method has the advantage that no prior assumptions about the functional model of the curves are made.

9 Conclusion

To the best of our knowledge, this is the first application of inductive Machine Learning techniques to real, measured respiratory data from intensive care medicine. We compared different methods of measuring pressure-volume curves of artificially ventilated patients suffering from the Adult Respiratory Distress Syndrome (ARDS). The goal was to find structural differences and similarities among the different methods. Our approach introduces an objective dimension along which the measurement methods can be analyzed: We induced symbolic classification and regression models from the data, and inspected the induced models to gain insight about the shapes of the curves. Classification was used to find the differences among the methods, and regression to find the similarities.

Apart from medical findings regarding the characteristics of the measurement methods, we found that normalization and high-level descriptors from curve fitting are important for obtaining reasonably accurate classification models. We also found that background information about the patients helps to improve the regression results considerably.

Our results hint at different possible topics of future research. More and more empirical investigations result in samples of curves following some common pattern. Therefore, it would be interesting to apply our method as a general approach to analyzing samples of curves depending on "implicit" features such as measurement methods.

Another open problem which is at the heart of intensive care medicine and, especially, of artificial ventilation is the individual ventilator setting. Given that inductive Machine Learning techniques enable us to integrate background information, we are planning to conduct more experiments with data from intensive care medicine in order to answer the question: Does incorporating patient data facilitate the individual ventilatory treatment? If so, it should, in principle, be possible to select the adequate ventilatory setting depending on patient type and to adjust the selected method. This is closely related to one of the long term goals within the multi-center study to automatically predict characteristics such as the lower-inflection point given patient data and the measurement method used.

10 Acknowledgements

We would like to thank Drägerwerk AG for supporting the multi-center study, especially Dipl.-Ing. E.-W. Schubert and Dipl.-Ing. T. Handzsuj. Special thanks also go to all our scientific collaborators in this multi-center study.

11 References

1. K.J. Cios, A. Teresinska, S. Konieczna, J. Potocka, S. Sharma, *Diagnosing Myocardial Perfusion SPECT Bull's-eye Maps - A Knowledge Discovery Approach*. IEEE Engineering in Medicine and Biology Magazine, special issue on Medical Data Mining and Knowledge Discovery, 19(4):17-25, 2000.
2. J. Guttmann, L. Eberhard, B. Fabry, G. Wolff, *Analysis of non-linear respiratory mechanics using the new SLICE method*, In Technol Health Care, 2(1):1-30, 1993.
3. L. Holzapfel, D. Robert, F. Terrin, P. L. Blanc, B. Palmier, C. Guerin, *Static Pressure-Volume Curves and Effect of Positive End-Expiratory Pressure on Gas-Exchange in Adult Respiratory Distress Syndrome.* In Crit. Care Med. 11:591-597, 1983.
4. D. Matamis, F. Lemaire, A. Harf, C. Brune-Buisson, J. C. Ansquer, G. Atlan, *Total Respiratory Pressure-Volume Curves in the Adult Respiratory Distress Syndrome*. In Chest 86: 58-66, 1984.
5. M. Sydow, H. Burchardi, J. Zinserling, H. Ische, T. A. Crozier, W. Weyland, *Improved Determination of Static Compliance by Automated Single Volume Steps in Ventilated Patients*. In Intensive Care Med. 17:108-114, 1991.
6. RuleQuest Research Pty Ltd. `http://www.rulequest.com/`
7. R. Quinlan. *Induction of Decision Trees*, Machine Learning 1(1):81-106, 1986.

8. R. Quinlan. *C4.5: Programs for Machine Learning*, San Mateo, CA, Morgan Kaufmann, 1993.
9. B. Muller, A. Hasman, J. A. Blom. *Building intelligent alarm systems by combining mathematical models and inductive machine learning techniques*. In International Journal of Biomedical Computation, 41(2):107-124, 1996.
10. B. Muller, A. Hasman, J. A. Blom. *Building intelligent alarm systems by combining mathematical models and inductive machine learning techniques Part 2--sensitivity analysis*. In International Journal of Biomedical Computation, 42(3):165-79, 1996.
11. S. Džeroski, S. Schulze-Kremer, K.R. Heidtke, K. Siems, D. Wettschereck. *Applying ILP to Diterpene Structure Elucidation from ^{13}C NMR Spectra*, In Applied Artificial Intelligence, 1998.
12. M. Kudo, J. Toyama, M. Shimbo. *Multidimensional Curve Classification Using Passing-Through Regions*, Pattern Recognition Letters, 20(11):1103-1111, 1999.
13. T. Gasser, A Kneip. *Searching for Structure in Curve Sample*. In Journal of the American Statistical Association, 90(432):1179-1188, 1995.