

3-2019

Scale-Invariance-Based Pre-Processing Drastically Improves Neural Network Learning: Case Study of Diagnosing Lung Dysfunction in Children

Nancy Avila

University of Texas at El Paso, nsavila@utep.edu

Julio Urenda

University of Texas at El Paso, jcurenda@utep.edu

Nelly Gordillo

Universidad Autónoma de Ciudad Juárez, nelly.gordillo@uacj.mx

Vladik Kreinovich

University of Texas at El Paso, vladik@utep.edu

Follow this and additional works at: https://digitalcommons.utep.edu/cs_techrep

Part of the [Applied Mathematics Commons](#), and the [Medical Sciences Commons](#)

Comments:

Technical Report: UTEP-CS-19-19

Recommended Citation

Avila, Nancy; Urenda, Julio; Gordillo, Nelly; and Kreinovich, Vladik, "Scale-Invariance-Based Pre-Processing Drastically Improves Neural Network Learning: Case Study of Diagnosing Lung Dysfunction in Children" (2019). *Departmental Technical Reports (CS)*. 1319.

https://digitalcommons.utep.edu/cs_techrep/1319

This Article is brought to you for free and open access by the Computer Science at DigitalCommons@UTEP. It has been accepted for inclusion in Departmental Technical Reports (CS) by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

Scale-Invariance-Based Pre-Processing Drastically Improves Neural Network Learning: Case Study of Diagnosing Lung Dysfunction in Children

Nancy Avila¹, Julio Urenda^{2,3},
Nelly Gordillo⁴, and Vladik Kreinovich³

¹Biomedical Engineering Program,
Department of Metallurgical, Materials and Biomedical Engineering

²Department of Mathematical Sciences

³Department of Computer Science

University of Texas at El Paso

500 W. University

El Paso, TX 79968, USA

nsavila@utep.edu, jcurenda@utep.edu, vladik@utep.edu

⁴Department of Electrical and Computer Engineering

Universidad Autónoma de Ciudad Juárez

Ave. del Charro No. 450 Norte

Ciudad Juárez, Chihuahua, México 32310

nelly.gordillo@uacj.mx

Abstract

To adequately treat different types of lung dysfunctions in children, it is important to properly diagnose the corresponding dysfunction, and this is not an easy task. Neural networks have been trained to perform this diagnosis, but they are not perfect in diagnostics: their success rate is 60%. In this paper, we show that by selecting an appropriate invariance-based pre-processing, we can drastically improve the diagnostic success, to 100% for diagnosing the presence of a lung dysfunction.

1 Formulation of the Problem

Lung dysfunctions. One of the major lung dysfunctions is asthma, a long-term inflammatory disease of the airways of the lungs. It is characterized by recurring airflow obstruction, bronchospasms, wheezing, coughing, chest tightness, and shortness of breath. These episodes may occur a few times a day or a few times per week [22].

Asthma may be preceded by Small Airway Impairment (SAI), a chronic obstructive bronchitis. If inflammation persists during SAI, it could cause asthma.

SAI, in its turn, may be preceded by a less severe condition that medical doctors classify as Possible Small Airways Impairment (PSAI).

Diagnostics of different lung dysfunctions is difficult but important.

All lung dysfunctions lead to similar symptoms like wheezing, coughing, etc. As a result, it is difficult to distinguish between these dysfunctions – and it is also difficult to distinguish these chronic dysfunctions from a common short-term respiratory disease.

However, the diagnosing of these diseases is very important, because in general, for different diseases, different treatments are efficient.

How different dysfunctions are diagnosed now. Since it is difficult to diagnose different dysfunctions solely based on the symptoms, the corresponding diagnostics involves measuring airflow in different situations. The most effective diagnostic comes from active measurements – spirometry. A patient is asked to deeply inhale, to hold their breath, and to exhale as fully as possible – and the corresponding instrument is measuring the airflow following all these instructions. Based on these measurements, symptoms, and clinical history, medical doctors come up with a diagnosis of different dysfunctions.

Children diagnostics: a serious problem. Unfortunately, the spirometry technique described above does not work with little children, especially children of pre-school age, since it not easy to make them follow the corresponding instructions; see, e.g., [16]. The same problem occurs with elderly patients and patients with certain limitations.

An additional problem is that even when children follow instructions during the spirometry testing, spirometry results are not sensitive enough to detect obstruction of small airways (2 mm or less in diameter); see, e.g., [6, 11, 12, 15, 19].

How can we diagnose children: main idea of the corresponding measurements. Since we cannot use active measuring techniques, techniques that require children’s active participation, we have to rely on passive techniques, i.e., techniques that do not require such participation. What we can do is change the incoming airflow and measure how that affects the outgoing airflow.

Passive measurements: details. The easiest way of changing the airflow is to switch a certain extra amount of airflow on or off. This is the main idea behind the Impulse Oscillometry System (IOS); see, e.g., [6]. In a usual IOS, the additional airflow is switched on and off with a period of 5 Hz, meaning that we have a 0.1 sec period with extra flow, 0.1 sec period without, then again a 0.1 sec period with extra flow, etc. The system then measures the resulting outflow $y(t)$.

In real-life clinical environment, the measurement result are affected by noise. Because of this noise, the measured values $\tilde{y}(t)$ somewhat deviate from the actual (unknown) flow results $y(t)$. The deviations $\tilde{y}(t) - y(t)$ measured at

different moments of time t are usually caused by different factors and are, thus, statistically independent. As a result of this independence, the noise is heavily oscillating – i.e., changes with high frequency. To decrease the effect of this noise, it is therefore reasonable to take a Fourier transform and ignore high-frequency components of this transform – since these components are heavily corrupted by noise.

Since the input signal is periodic, with the period of 5 Hz, we expect the output signal to also be periodic, with the same frequency. In general, when we perform Fourier transform on a signal which is periodic with frequency f , we only get components corresponding to multiples of f , i.e., to f , $2f$, $3f$, etc. In our case, this means that we will have components corresponding to 5 Hz, 10 Hz, etc. In practice, it was discovered that components above 25 Hz are too noisy to be useful – actually, the most informative values are one corresponding to 5-15 Hz range. The values corresponding to 20 and 25 Hz are also useful, but they are somewhat less informative than the 5-15 Hz values. So, the system returns the components corresponding to 5, 10, 15, 20, and 25 Hz. To be on the safe side, the system also returns the component corresponding to 35 Hz, which sometimes adds some additional information.

Another problem is related to the fact that while it is relatively easy to implement the on-off switching of the input airflow, the actual values of the on- and off-case airflows may change with time: the pressure in the system may decrease, etc. In principle, it is possible to maintain the exact airflow values, but this will make the system too complicated and thus, too expensive. Instead, the existing systems rely on the fact that while it is not easy to maintain the input airflow $a(t)$ at some pre-defined level a_0 , it is possible to accurately measure this airflow.

The added airflow $a(t) - a_0$ is relatively small, so when estimating the reaction $y(t)$ of a human breathing system to this airflow, we can safely ignore terms which are quadratic and of higher order in terms of $a(t) - a_0$ and conclude that the dependence is linear: $y(t) = \int c(t, s) \cdot a(s) ds$ for some coefficients $c(t, s)$.

The system does not change much during the time when measurements are performed. So if we start the experiment t_0 seconds earlier, i.e., if we take $\tilde{a}(t) = a(t + t_0)$ instead of $a(t)$, then the output should change accordingly, to $\tilde{y}(t) = y(t + t_0)$. So, on the one hand, we have

$$\tilde{y}(t) = \int c(t, s) \cdot \tilde{a}(s) ds = \int c(t, s) \cdot a(s + t_0) ds,$$

which, if we introduce a new variable $\tilde{s} = s + t_0$ for which $d\tilde{s} = ds$, leads to

$$\tilde{y}(t) = \int c(t, \tilde{s} - t_0) \cdot a(\tilde{s}) d\tilde{s}.$$

On the other hand,

$$\tilde{y}(t) = y(t + t_0) = \int c(t + t_0, s) \cdot a(s) ds.$$

So, for all inputs $a(t)$, we should have

$$\int c(t + t_0, s) \cdot a(s) ds = \int c(t, s - t_0) \cdot a(s) ds.$$

Two linear functions coincide if the coefficients at all the unknown (in this case, $a(s)$) coincide, so we must have $c(t + t_0, s) = c(t, s - t_0)$. In particular, for every two values v_1 and v_2 , we can take $s = t_0 = v_2$ and $t = v_1 - v_2$ and conclude that $c(v_1, v_2) = c(v_1 - v_2, 0)$, i.e., that $c(v_1, v_2) = z(v_1 - v_2)$, where we denoted $z(a) \stackrel{\text{def}}{=} c(a, 0)$. Substituting this expression for $c(v_1, v_2)$ into the formula that describes the relation between $a(t)$ and $y(t)$, we conclude that $y(t) = \int z(t - s) \cdot a(s) ds$. This is called a *convolution* of functions $z(t)$ and $a(t)$.

It is known that the Fourier transform of the convolution is equal to the product of Fourier transforms. Thus, in this case, for the corresponding Fourier transforms $Z(f)$, $Y(f)$, and $A(f)$, we get $Y(f) = Z(f) \cdot A(f)$. We are interested in the values $Z(f)$ that do not depend on the inputs. In our case, we have computed the values $Y(f)$, so we can also compute the Fourier coefficients $A(f)$, and return the ratios $Z(f) = Y(f)/A(f)$. This is exactly what the IOS system returns: the complex numbers $Z(f) = R(f) + i \cdot X(f)$ that correspond to six frequencies $f = 5, 10, \dots, 35$. In analogy with electric circuits, the complex value $Z(f)$ is called the *impedance*, its real part $R(f)$ is called *resistance*, and its imaginary part $X(f)$ is called *reactance*.

These six complex numbers – or, equivalently, the six real parts and the six imaginary parts – are what we can use to properly diagnose the lung dysfunction.

It is not easy to make a diagnosis based on IOS data. If we plot the IOS data corresponding to patients with different diagnoses, we see that the corresponding ranges of values $R(f)$ and $X(f)$ have a huge intersection; see, e.g., Fig. 1–2. This shows that it is not easy to diagnose a patient based on IOS data.

How IOS-based diagnosis is performed now. There are no exact formulas that describe the diagnosis based on the six complex values $Z(f)$; the research about the clinical applications of the IOS parameters is still ongoing. However, we do have several patient records for which, on the one hand, we know the corresponding values $Z(f)$, and, on the other hand, we have a diagnosis provided by a skilled medical doctor. It is therefore reasonable to use machine learning and train the system to be able to diagnose a patient.

This has indeed been done: researchers have used either all or some of the 12 numbers (real parts $R(f)$ and imaginary parts $X(f)$) as input and tried to train the neural network to learn the diagnosis in the children patients.

The resulting diagnostic system is, however, not yet perfect. For adult patients, if we use spirometry results as well as IOS, we get an almost perfect separation of asthma from healthy: its accuracy is 98-99% [3, 4]. However, when we only use IOS data, the current system’s testing-data accuracy in distinguishing lung dysfunctions such as asthma, SAI, and PSAI from patients who do not have any of these diseases is only close to 60% [5].

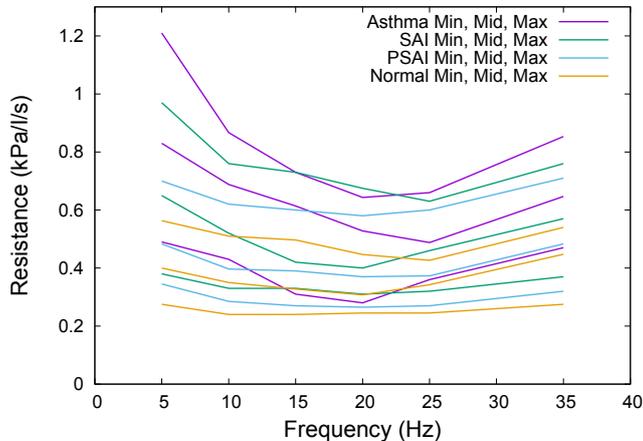


Figure 1: Resistance Maximum, Middle and Minimum Patients' Curves per Class

Comment. Similar imperfect results were obtained for a related problem: predicting asthma deterioration one week ahead. For this problem, neural networks and other machine learning techniques result, at best, in 70-75% prediction accuracy; see, e.g., [14].

Resulting problem and what we do in this paper. It is therefore desirable to come up with better diagnostic techniques. Our approach is to help a neural network by providing an appropriate pre-processing of the inputs data. It turns out that an appropriate pre-processing indeed drastically improves the diagnostic results: for diagnosing the presence of a lung dysfunction, we have 100% accuracy.

2 First Pre-Processing Stage: Scale-Invariant Smoothing

Need for further de-noising. The above-described filtering out of noisy high-frequency components eliminates some noise, but some noise remains.

Smoothing as a way to de-noise. To further decrease the noise level, it is desirable to take into account that in real life, almost all dependencies (including the dependence of the signal intensity on frequency) are smooth – in the sense that a small change in frequency leads to a small change in intensity.

How to smooth a signal. Thus, instead of considering the original (noisy) six complex numbers $Z(f)$ corresponding to $f = 5, 10, \dots$, it makes sense to approximate these values by a smooth dependence, i.e., by a function of the type $\sum_{j=1}^k c_j \cdot e_j(f)$, for some smooth functions $e_1(f), \dots, e_k(f)$.

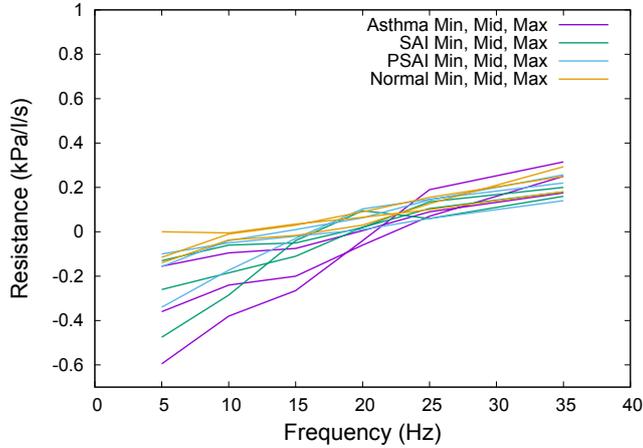


Figure 2: Reactance Maximum, Middle and Minimum Patients' Curves per Class

Which level of smoothness to choose. Usually, real-life processes are very smooth – with few exceptions like phase transitions. It is therefore desirable to select approximating functions $e_j(f)$ which are as smooth as possible.

In general, for functions, there are several different degrees of smoothness. The simplest case is when a function is one time differentiable. The next – more smooth – case is when a function is two times differentiable, etc. Then, we have functions which are infinitely many time differentiable, and finally, the smoothest of all – analytical functions, functions that can be expanded in Taylor series. Thus, to achieve maximal smoothness, we will use analytical functions $e_j(x)$.

Which analytical functions should we choose: the idea of scale-invariance. There are many different analytical functions, which ones should we choose?

A natural requirement for this choice comes from the fact that we are approximating a function from numbers (f) to numbers (Z). These numbers represent the values of the corresponding physical quantities. However, the numerical value of each physical quantity depends not only on the quantity itself, it also depends on the measuring unit that we have selected for this quantity. If we replace the original measuring unit with a new one which is λ times smaller, all the numerical values will multiply by λ . In other words, for each frequency, instead of the original numerical value f , we will have a new numerical value $\tilde{f} = \lambda \cdot f$ that describe the exact same physical quantity.

There is no physical reason why some measuring units would be preferable to others. Therefore, it makes sense to require that the selection of the resulting class C of linear combinations $\sum_{j=1}^k c_j \cdot e_j(f)$ should not change if we simply re-scale all the values by changing the measuring unit for frequencies.

The idea of scale-invariance is actively and successfully used in physics; see, e.g., [7, 20]. It is therefore reasonable to apply it to our problem as well.

In our case, scale-invariance means that the class C should be equal to the class \tilde{C} of all linear combinations $\sum_{j=1}^k c_j \cdot e_j(\tilde{f})$. In particular, this means that, for every i , the function $e_i(\lambda \cdot f)$ from the class \tilde{C} can be described as a linear combination $\sum_{j=1}^k c_{ji}(\lambda) \cdot e_j(f)$ of the original functions, with coefficients $c_{ji}(\lambda)$ depending on i and λ :

$$e_i(\lambda \cdot f) = \sum_{j=1}^k c_{ji}(\lambda) \cdot e_j(f). \quad (1)$$

Thus, we have a system of equations for the unknown functions $e_i(f)$. Let us solve this system.

Solving the corresponding system of equations. We know that the functions $e_i(f)$ are smooth. Let us show that the functions $c_{ji}(\lambda)$ are smooth as well. Indeed, for every i , by taking k different values f_1, \dots, f_k in the equation (1), we get a system of k linear equations with k unknowns $c_{1i}(\lambda), \dots, c_{ki}(\lambda)$:

$$\begin{aligned} \sum_{j=1}^k c_{ji}(\lambda) \cdot e_j(f_1) &= e_i(\lambda \cdot f_1); \\ \sum_{j=1}^k c_{ji}(\lambda) \cdot e_j(f_2) &= e_i(\lambda \cdot f_2); \\ &\dots \\ \sum_{j=1}^k c_{ji}(\lambda) \cdot e_j(f_k) &= e_i(\lambda \cdot f_k). \end{aligned}$$

It is known that, in general, the solution to a system of linear equations can be described by Cramer's rule, as a ratio of two polynomials – and thus, a smooth function – depending on the coefficients and on the right-hand sides. In our case, the coefficients $e_j(f_1), \dots, e_j(f_k)$ are constants – and thus, do not depend on λ at all, while the right-hand sides $e_i(\lambda \cdot f_1), \dots, e_i(\lambda \cdot f_k)$ are smooth functions of λ . Thus, the solutions to this system of linear equations – i.e., the coefficients $c_{ji}(\lambda)$ – are obtained by plugging smooth functions

$$e_j(f_1), \dots, e_j(f_k), e_i(\lambda \cdot f_1), \dots, e_i(\lambda \cdot f_k)$$

into a smooth expression (Cramer's rule) and are, thus, also smooth functions of λ .

Now, we can take the system of equations (1) corresponding to $i = 1, \dots, k$:

$$e_1(\lambda \cdot f) = \sum_{j=1}^k c_{j1}(\lambda) \cdot e_j(f);$$

$$\begin{aligned}
e_2(\lambda \cdot f) &= \sum_{j=1}^k c_{j2}(\lambda) \cdot e_j(f); \\
&\dots \\
e_k(\lambda \cdot f) &= \sum_{j=1}^k c_{jk}(\lambda) \cdot e_j(f).
\end{aligned}$$

All the expressions in this system are differentiable functions of λ . Thus, we can differentiate both sides of each equation by λ . As a result, we get the following system of equations:

$$\begin{aligned}
f \cdot e'_1(\lambda \cdot f) &= \sum_{j=1}^k c'_{j1}(\lambda) \cdot e_j(f); \\
f \cdot e'_2(\lambda \cdot f) &= \sum_{j=1}^k c'_{j2}(\lambda) \cdot e_j(f); \\
&\dots \\
f \cdot e'_k(\lambda \cdot f) &= \sum_{j=1}^k c'_{jk}(\lambda) \cdot e_j(f),
\end{aligned}$$

where e'_i , as usual, denotes the derivative of the function e_i . If we substitute $\lambda = 1$ into these formulas, we get the following system of ordinary differential equations:

$$\begin{aligned}
f \cdot e'_1(f) &= \sum_{j=1}^k C_{j1} \cdot e_j(f); \\
f \cdot e'_2(f) &= \sum_{j=1}^k C_{j2} \cdot e_j(f); \\
&\dots \\
f \cdot e'_k(f) &= \sum_{j=1}^k C_{jk} \cdot e_j(f),
\end{aligned}$$

where we denoted $C_{ji} \stackrel{\text{def}}{=} c'_{ji}(1)$.

For each i , the expression $f \cdot e'_i = f \cdot \frac{de_i}{df}$ can be equivalently reformulated as $\frac{de_i}{df/f}$. Here, $df/f = d(\ln(f))$. So, if we introduce a new variable $F = \ln(f)$ for which $f = \exp(F)$, then for the new functions $E_i(F) \stackrel{\text{def}}{=} e_i(\exp(F))$, we get $f \cdot e'_i(f) = \frac{dE_i}{dF} = E'_i(F)$. Thus, in terms of the new functions, we get the following system of equations:

$$E'_1(F) = \sum_{j=1}^k C_{j1} \cdot E_j(F);$$

$$E'_2(F) = \sum_{j=1}^k C_{j2} \cdot E_j(F);$$

...

$$E'_k(F) = \sum_{j=1}^k C_{jk} \cdot E_j(F).$$

This is a system of linear differential equations with constant coefficients. It is known that a general solution to this system is a linear combination of terms $F^p \cdot \exp(\lambda \cdot F)$, where:

- λ is an eigenvalue of the matrix $\|C_{ji}\|$ (which is, in general, complex $\lambda = a + b \cdot i$), and
- the value $p \neq 0$ appears when we have a duplicate eigenvalue – in this case, p is a non-negative integer smaller than the dimension of the corresponding eigenspace.

In terms of real values, we get

$$F^p \cdot \exp(\lambda \cdot F) = F^p \cdot \exp((a + b \cdot i) \cdot F) =$$

$$F^p \cdot \exp(a \cdot F) \cdot (\cos(b \cdot F) + i \cdot \sin(b \cdot F)).$$

Substituting $F = \ln(f)$ into this expression, we conclude that the functions $e_i(f)$ are linear combinations of the expressions

$$(\ln(f))^p \cdot \exp(a \cdot \ln(f)) \cdot (\cos(b \cdot \ln(f)) + i \cdot \sin(b \cdot \ln(f))).$$

Here, $\exp(a \cdot \ln(f)) = (\exp(\ln(f)))^a = f^a$, so the above expression has the form

$$(\ln(f))^p \cdot f^a \cdot (\cos(b \cdot \ln(f)) + i \cdot \sin(b \cdot \ln(f))).$$

Let us take into account that the functions $e_i(f)$ should be analytical. Now, we can take into account that the functions $e_i(f)$ should be analytical, i.e., they should be expandable in Taylor series for $f = 0$. This requirement excludes possible logarithmic terms $(\ln(f))^p$, as well as cosines and sines of these logarithms, which leaves us with linear combinations of the powers f^a . Due to analyticity, all the powers should be natural numbers, so we conclude that all the functions $e_i(f)$ are linear combinations of expressions $f^0 = 1$, $f^1 = f$, f^2 , ... In other words, due to scale-invariance, all the functions $e_i(f)$ should be polynomials.

We want to approximate the function $Z(f)$ by a linear combination of the functions $e_i(f)$. A linear combination of polynomials is also a polynomial. Thus, we arrive at the following conclusion.

General conclusion of this section. Due to the natural requirement of scale-invariance, we should approximate the impedance function $Z(f)$ by a polynomial.

3 Which Order Polynomials Should We Use?

Formulation of the problem. We want to find the polynomial that fits the observations. Of course, if we take a polynomial of a sufficiently large degree, we can always find a polynomial that fits all observed data exactly – this is a well-known Lagrange interpolation polynomial.

However, the whole purpose of the polynomial smoothing is to de-noise the signal, and if we keep all the values intact, we will retain all the noise. Thus, we should not use polynomials of too high order.

On the other hand, if we use polynomials of too low order – e.g., constants or linear functions – we get a smoothing, but the approximation is too crude, and we lose the information contained in the original signal. How can we find the adequate degree of approximating polynomials?

Natural idea: general case. A natural idea is to take into account the general monotonicity of IOS curves – as described, e.g., in [11] – and select the higher order of approximating polynomials that preserve this monotonicity.

Let us apply this general idea to our problem. According to [11], for all three diseases (asthma, SAI, and PSAI):

- for the real part $R(f)$ of the impedance $Z(f)$, the corresponding value first decreases with frequency f , and then increases;
- for the imaginary part $X(f)$ of the impedance $Z(f)$, the corresponding value increases with frequency f .

So:

- for each degree, we use the usual Least Squares techniques (see, e.g., [18]) to find the polynomial of this degree that best approximates the observed values, and then
- we select the largest degree for which, on the corresponding interval of values of frequency, the resulting best-approximation polynomials follow the same monotonicity pattern.

It turns out that:

- for quadratic and cubic polynomials, we have this feature, but
- for 4th order polynomials, we no longer have the desired monotonicity: for example, for the resistance $R(f)$, the corresponding 4th order polynomial first decreases, then increases, but then decreases again; see, e.g., Fig. 3.

Because of this, in this paper, we approximate the functions $R(f)$ and $X(f)$ by *cubic polynomials*. Let us denote the corresponding approximating polynomials by $R^a(f)$ and $X^a(f)$.

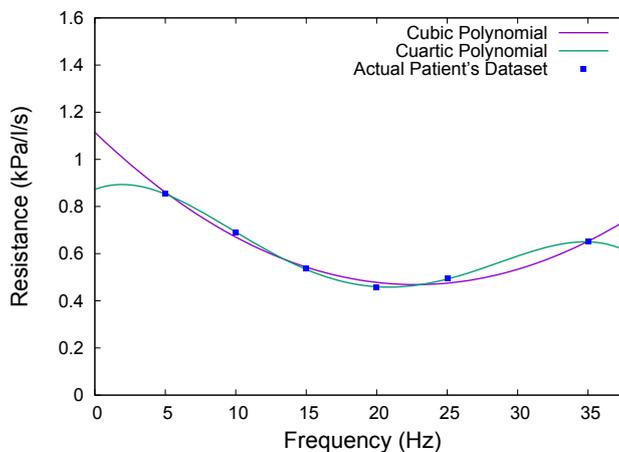


Figure 3: Cubic vs Cuartic vs a Patient's Dataset

4 Second Pre-Processing Stage: Using the Approximating Polynomials to Distinguish Between Different Diseases

Formulation of the problem. For each diagnosis d , we have several observations corresponding to patients with these diagnosis. Based on these observations, for each patient i with this diagnosis ($i \in d$), we find the smoothed functions $R_i^a(f)$ and $X_i^a(f)$.

Now, when we have a new patient with the corresponding functions $R^a(f)$ and $X^a(f)$, we would like to diagnose this patient, i.e., to classify this patient to one of the groups d . How can we do it?

How to separate different groups: general idea. How do we distinguish groups in general? How do we distinguish cats from dogs? Usually, in such situations:

- we have a mental picture of a typical cat,
- we have a mental picture of a typical dog, and
- we make our decision based on how similar the observed object is to one of these two typical ones.

A natural idea is thus:

- to form, for each group corresponding to a given diagnosis d , “typical” function $R_d(f)$ and $X_d(f)$ corresponding to this diagnosis, and then
- to base our diagnosis of new, yet-undiagnosed patients based on how similar their functions $R^a(f)$ and $X^a(f)$ are to the typical functions corresponding to each diagnosis.

A natural way to form a typical function. A natural way to form the typical function $R_d(f)$ corresponding to a given diagnosis d is to take all the IOS values $R_i(f)$ corresponding to patients i with this diagnosis ($i \in d$), and use Least Squares to find the cubic function $R_d(f)$ that best approximate all these measurement results, i.e., for which the sum $\sum_{i \in d} \sum_f (R_d(f) - R_i(f))^2$ attains its smallest possible value (here, the summation is over the IOS frequencies $f = 5, 10, 15, 20, 25,$ and 35 Hz).

Similarly, a natural way to form the typical function $X_d(f)$ corresponding to a given diagnosis d is to take all the IOS values $X_i(f)$ corresponding to patients i with this diagnosis, and use Least Squares to find the cubic function $X_d(f)$ that best approximate all these measurement results, i.e., for which, the sum $\sum_{i \in d} \sum_f (X_d(f) - X_i(f))^2$ attains its smallest possible value.

5 Third Pre-Processing Stage: Scale-Invariant Similarity/Dissimilarity Measures

How to describe the similarity of functions: preliminary analysis. To use the above general idea for diagnosing a patient, we need to select a numerical measure of similarity/dissimilarity between the function $R^a(f)$ describing the new patient and the function $R_d(f)$ describing a typical patient with diagnosis d .

One thing to take into account is that usually the measurement record produced by the IOS device contains an initial spike, when the measured value jumps from the original 0 value to a non-zero value corresponding to the actual measurement. This initial impulse-type spike affect the Fourier transform values $R(f)$ and $X(f)$ produced by the measuring device. Since the Fourier transform of an impulse is a constant function, this means that to all the measured values $R_i(f)$ and $X_i(f)$, a constant is added that corresponds to the Fourier transform of this original impulse. Because of this added constant, the same constant gets added to the approximating cubic curves $R_i^a(f)$ and $X_i^a(f)$. For the same patient with the same disease, in different measurements, the initial impulse may be slightly different. Thus, the corresponding added constants may be different for the two measurements of the same patient. In other words, for the same patient, in two consequent measurements, we may get two functions differing by a constant.

So, to properly match, e.g., the patient's function $R^a(f)$ with the function $R_d(f)$ describing a typical patient with diagnosis d , we need to take this possible constant difference into account.

How do we estimate the corresponding constant difference? As we have mentioned earlier, the most informative part of the IOS results correspond to the 5-15 Hz range. The central point of this range is the value 10 Hz. It is therefore reasonable to take the difference $\Delta R_{id} \stackrel{\text{def}}{=} R_d(10) - R_i^a(10)$ of the values corresponding to this central frequency as an estimate for the constant

difference. Thus, we should compare the typical function $R_d(f)$ not with the actual patient's function $R_i^a(f)$, but with the "shifted" function $R_i^a(f) + \Delta R_{id}$, shifted so as to provide the best match between the two functions.

In other words, to diagnose a patient i , we need to describe the similarity/dissimilarity between the functions $R_i^a(f) + \Delta R_{id}$ and $R_d(f)$.

Similarly, for reactance, we need to describe the similarity/dissimilarity between the functions $X_i^a(f) + \Delta X_{id}$ and $X_d(f)$, where $\Delta X_{id} \stackrel{\text{def}}{=} X_d(10) - X_i^a(10)$.

How to describe the similarity/dissimilarity of functions: general idea. In general, if we have two functions $F(f)$ and $G(f)$, how can we describe their similarity/dissimilarity? For each frequency f , the larger the absolute value $|F(f) - G(f)|$ of the difference, the less similar are the corresponding values. Thus, it makes sense to assume that the degree of dissimilarity between these two values is a monotonic function of this absolute value: $m(|F(f) - G(f)|)$, for some increasing function $m(x)$.

The overall degree of dissimilarity we can then estimate by simply adding the degree corresponding to different frequencies f , i.e., by considering an integral

$$\int m(|F(f) - G(f)|) df.$$

Remaining question. Which similarity/dissimilarity measures – i.e., which functions $m(x)$ – should we use?

Main idea: use scale-invariance. To select an appropriate similarity/dissimilarity measure, let us use the same scale-invariance idea that we used to select an approximating family of functions.

Scale-invariance: from idea to formulas. If, for measuring real and imaginary components of the impedance, we select a new unit which is λ times smaller than the original one, then all the corresponding numerical values $F(f)$ and $G(f)$ get multiplied by λ :

- instead of $F(f)$, we get $\tilde{F}(f) = \lambda \cdot F(f)$, and
- instead of $G(f)$, we get $\tilde{G}(f) = \lambda \cdot G(f)$.

In this case, the absolute value of the difference $x = |F(f) - G(f)|$ also gets multiplied by λ :

$$\tilde{x} = |\tilde{F}(f) - \tilde{G}(f)| = |\lambda \cdot F(f) - \lambda \cdot G(f)| = \lambda \cdot |F(f) - G(f)| = \lambda \cdot x.$$

We want to select the function $m(x)$ (that describes degree of similarity/dissimilarity) in such a way that if we change a measuring unit for $F(f)$ and $G(f)$, the resulting value of closeness will not change – provided, of course, we appropriately change a unit for measuring dissimilarity.

Such an appropriate re-scaling is often necessary; see, e.g., [7, 20]. For example, a simple physical formula – like the fact $D = v \cdot t$ that the distance D is equal to velocity v times time t – does not change if we change the unit of time (e.g., from hours to seconds), but we need to appropriately change the related unit of velocity – from km/h to km/sec.

For each λ , let $\mu(\lambda)$ denote an appropriate re-scaling of the dissimilarity value $m(x)$. This means that if we replace x with $\tilde{x} = \lambda \cdot x$, then the resulting dissimilarity value $m(\tilde{x}) = m(\lambda \cdot x)$ differs from the original value $m(x)$ only by a re-scaling factor $\mu(\lambda)$:

$$m(\lambda \cdot x) = \mu(\lambda) \cdot m(x).$$

It is known (see, e.g., [1]) that all measurable solutions of this functional equations have the form $m(x) = C \cdot x^\alpha$.

By selecting an appropriate unit for measuring dissimilarity, we can make the coefficient C equal to 1. Thus, we arrive at the following conclusion.

Conclusion of this section. Due to scale-invariance, we measure dissimilarity between two functions as

$$\int |F(f) - G(f)|^\alpha \cdot df.$$

Remaining question. What value α should we use?

6 How to Select α : Need to Have Efficient and Robust Estimates

General idea. In the computer, we can only represent finitely many different values f_1, f_2, \dots . So, an integral, in effect, means a (weighted) sum $\sum |F_i - G_i|^\alpha$, where we denoted $F_i \stackrel{\text{def}}{=} F(f_i)$ and $G_i \stackrel{\text{def}}{=} G(f_i)$.

From this viewpoint, which value α should we choose?

Need for efficient estimates. We want to be able to have an efficient algorithm that finds the closest approximation, i.e., an approximation for which the dissimilarity degree $\sum |F_i - G_i|^\alpha$ is the smallest possible.

It is known (see, e.g., [13, 21]) that, in general, feasible algorithms exist for minimizing convex objective functions, while in many non-convex cases, optimization is NP-hard (i.e., crudely speaking, not feasible). Moreover, it has been proven [10] that, in general, minimization is feasible *only* for convex objective functions. Thus, it makes sense to select the value α in such a way that the objective function $\sum |F_i - G_i|^\alpha$ be convex.

In general, according to calculus, a function is convex if its second derivative is non-negative. For $x > 0$, the first derivative of the function x^α is $\alpha \cdot x^{\alpha-1}$, and the second derivative is equal to $\alpha \cdot (\alpha - 1) \cdot x^{\alpha-2}$. Here, $\alpha > 0$ – the larger the difference, the less similar are the functions, and the value $x^{\alpha-2}$ is also always

positive. Thus, the second derivative is non-negative if and only $\alpha - 1 \geq 0$, i.e., if and only if $\alpha \geq 1$.

Thus, to make sure that the corresponding optimization problems can be efficiently solved, we need to select $\alpha \geq 1$.

Need for robustness. Another important requirement for selecting α is to make sure that the resulting estimates are the least affected by noise, i.e., are the most *robust*.

It is known (see, e.g., [9]), that among all the methods based on the objective function $\sum |F_i - G_i|^\alpha$ with $\alpha \geq 1$, the most robust is the method corresponding to $\alpha = 1$. Thus, to guarantee the desired robustness, we will use $\alpha = 1$. So, we arrive at the following conclusion.

Conclusion of this section. Among all computationally efficient scale-invariant dissimilarity measures, the most robust (i.e., the most resistant to noise) is the dissimilarity measure

$$\int |F(f) - G(f)| \cdot df.$$

In our case, $F(f) = R_d(f)$ and $G(f) = R_i^a(f) + \Delta R_{id}$, so we need to use the dissimilarity measure

$$\int |R_d(f) - (R_i^a(f) + \Delta R_{id})| df.$$

When this dissimilarity measure is close to 0, this means that the function $R_i^a(f)$ corresponding to the i -th patient is very similar to the typical function $R_d(f)$ corresponding to diagnosis d . The more dissimilar these two functions, the larger the value of this dissimilarity measure.

Similarly for reactance, we use the dissimilarity measure

$$\int |X_d(f) - (X_i^a(f) + \Delta X_{id})| df.$$

7 Scale-Invariance Helps to Take into Account that Signal Informativeness Decreases with Time

For IOS, the starting part of the signal is more informative. In the previous sections, we implicitly assumed that the values of the signal $y(t)$ at different moments of time are equally informative. However, a typical IOS measurement lasts for 30-45 seconds – a reasonable time to be tied in to a strange apparatus. As a result, children’s level of stress somewhat increases as the measurement process continues. This stress level affects the breathing process – and thus, the measurement results.

So, we can conclude that values $y(t)$ corresponding to earlier time are more informative than values corresponding to later moments of time t .

How can we take this phenomenon into account. A reasonable idea of taking the above phenomenon into account is to consider not the original signals $y(t)$, but the signals weighted with some weight $w(t)$ which decreases with time. In other words, instead of the original signals $y(t)$, we consider weighted signals $w(t) \cdot y(t)$.

Which weight function should we choose: let us again apply scale-invariance. Which weight function $w(t)$ should we choose? A reasonable idea is to again use scale-invariance. In other words, we assume that if we change the unit of time to a one which is λ times smaller – which means changing all numerical values of time from t to $\tilde{t} = \lambda \cdot t$, then the formula for the weight remains the same – once we appropriately re-scale the weight function w as well, from w to $\tilde{w} = \mu(\lambda) \cdot w$, for some function $\mu(\lambda)$.

This means that if in the original units, we have $f(t) = w$, then in the new units, we will have $f(\tilde{t}) = \tilde{w}$. Substituting the expressions for \tilde{w} and \tilde{t} into this formula, we conclude that $f(\lambda \cdot t) = \mu(\lambda) \cdot w$ and thus,

$$f(\lambda \cdot t) = \mu(\lambda) \cdot f(t).$$

We have already mentioned that all measurable (in particular, all monotonic) solutions of this functional equation have the form $w = A \cdot t^\alpha$ [1]. Since we assume that the weight decreases with time, we must have $\alpha < 0$.

Which value α should we choose? Whether we use the original signal or its weighted form, what we will do next is apply Fourier transform. The original IOS device already returns the Fourier coefficients of the original signal $y(t)$. Thus, from the computational viewpoint, it is desirable to select α for which the Fourier transform of the weighted function $y(t) \cdot t^\alpha$ can be described in terms of the Fourier transform of the original function $y(t)$.

It is known that such a description is possible only for integer values α ; namely:

- the Fourier transform of $y(t)/t$ is proportional to the integral of the Fourier transform of $y(t)$;
- the Fourier transform of $y(t)/t^2$ is proportional to the second integral (integral of an integral) of the Fourier transform of $y(t)$; etc.

The simplest of these cases is the case $\alpha = -1$, which corresponds to the integral.

Thus, in addition to the original Fourier transform values, we should consider integrals of these values.

Details. Of course, when computing these integrals, we should take into account the smoothing that we have applied to the original signal. In other words, we should integrate not the original values $Z(f)$, but the corresponding smoothing polynomial approximations.

Integration should be considered over the most informative part of the spectrum – from 5 to 15 Hz. Thus, we arrive at the following conclusion.

Conclusion of this section. In addition to the smoothed signals $R^a(f)$ and $X^a(f)$, we should also consider their integrals $I_R(f) = \int_5^f R^a(x) dx$ and $X_R(f) = \int_5^f X^a(x) dx$.

8 Pre-Processing Summarized: What Information Serves as an Input to a Neural Network

Let us summarize. Let us summarize the scale-invariance-motivated pre-processing steps, and thus, describe what inputs are fed into a neural network.

This whole process consists of two stages:

- In the first, preliminary stage, we process data about known patients to find the “typical” functions $R_d(f)$ and $X_d(f)$ that correspond to each diagnosis d .
- On the working stage, we use these typical functions to diagnose a new patient.

Preliminary stage. First, for each diagnosis d , we process patients with known diagnoses d to find the typical functions $R_d(f)$ and $X_d(f)$ corresponding to each of these diagnoses. This is done as follows.

For each patient with the known diagnosis, we get the IOS values $R(f)$ and $X(f)$ corresponding to frequencies f equal to 5 Hz, 10 Hz, 15 Hz, 20 Hz, 25 Hz, and 35 Hz.

Then, we use the Least Squares techniques to find the coefficients r_0, r_1, r_2 , and r_3 of the cubic polynomial

$$R_d(f) = r_0 + r_1 \cdot f + r_2 \cdot f^2 + r_3 \cdot f^3$$

that best approximates the measured values $R_i(f)$ corresponding to the patients with this diagnosis d (of course, we only use patients from the training set, to be able to test our results of the patients from the testing set). In other words, we find the coefficients of the cubic polynomial for which the sum

$$\sum_{i \in d} \sum_f (R_d(f) - R_i(f))^2$$

is the smallest possible.

Similarly, we use the Least Squares techniques to find the coefficients x_0, x_1, x_2 , and x_3 of the cubic polynomial

$$X_d(f) = x_0 + x_1 \cdot f + x_2 \cdot f^2 + x_3 \cdot f^3$$

that best approximates the measured values $X_i(f)$ corresponding to all the patients with this diagnosis d . In other words, we find the coefficients of the cubic polynomial for which the sum $\sum_{i \in d} \sum_f (X_d(f) - X_i(f))^2$ is the smallest possible.

For each of these functions, we then compute the integrals $I_{R,d}(f) = \int_5^f R_d(x) dx$ and $I_{X,d}(f) = \int_5^f X_d(x) dx$.

Working stage. For a new patient, we get the IOS values $R(f)$ and $X(f)$ corresponding to frequencies f equal to 5 Hz, 10 Hz, 15 Hz, 20 Hz, 25 Hz, and 35 Hz. Then:

- We use the Least Squares techniques to find the coefficients r_0, r_1, r_2 , and r_3 of the cubic polynomial

$$R^a(f) = r_0 + r_1 \cdot f + r_2 \cdot f^2 + r_3 \cdot f^3$$

that best approximates the measured values $R(f)$. We also compute the integral $I_R(f) = \int_5^f R^a(x) dx$.

- After that, we use the Least Squares techniques to find the coefficients x_0, x_1, x_2 , and x_3 of the cubic polynomial

$$X^a(f) = x_0 + x_1 \cdot f + x_2 \cdot f^2 + x_3 \cdot f^3$$

that best approximates the measured values $X(f)$. We also compute the integral $I_X(f) = \int_5^f X^a(x) dx$.

- Finally, for each of the four diagnoses d , we compute the following values:

- $\int_5^{15} |R_d(f) - (R^a(f) + \Delta R_d)| df$, where $\Delta R_d \stackrel{\text{def}}{=} R_d(10) - R^a(10)$;
- $\int_5^{15} |I_{R,d}(f) - (I_R^a(f) + \Delta I_{R,d})| df$, where $\Delta I_{R,d} \stackrel{\text{def}}{=} I_{R,d}(10) - I_R^a(10)$;
- $\int_5^{15} |X_d(f) - (X^a(f) + \Delta X_d)| df$, where $\Delta X_d \stackrel{\text{def}}{=} X_d(10) - X^a(10)$;
- $\int_5^{15} |I_{X,d}(f) - (I_X^a(f) + \Delta I_{X,d})| df$, where $\Delta I_{X,d} \stackrel{\text{def}}{=} I_{X,d}(10) - I_X^a(10)$.

These four tuples of four values corresponding to four diagnoses – the total of 16 values – can then be used to train a neural networks to diagnose the patient.

Do we need all these 16 inputs? In data processing, it is known that if we use too many inputs, the prediction accuracy decreases. Indeed, if we use too many inputs, then, together with the most informative ones, we also add less informative ones. These additional inputs add noise to the result of data processing without providing us with any useful information.

Because of this, in data processing in general, it is a good idea not just to use all possible inputs, but also to check if selecting only some of these inputs will leads to more accurate results.

In our case, we tested whether we need both values corresponding to resistance R and values corresponding to reactance X . Interestingly, it turned out that the reactance-related values only decrease the prediction quality. As a result, our recommendation is to only use resistance-related values when diagnosing patients.

9 The Results of Training Neural Networks on These Pre-Processed Data

Available data. In our research, we used the data collected by our colleague Erika Meraz [2]. This data consists of 288 IOS data sets from patients with known diagnoses.

Pre-processing: first stage. First, for each data set, we used Least Square to find the coefficients of cubic polynomials $R_i^a(f)$ that best fit the observed IOS values $R_i(f)$. Then, we computed the integral $I_{R,i}(f) = \int_5^f R_i^a(x) dx$.

Neural network: general idea. We trained a neural network to distinguish patients with lung dysfunctions from patients without lung dysfunction.

A neural networks consist of neurons. Each neuron takes several inputs x_1, \dots, x_n and transforms them into the signal

$$y = s_0(w_1 \cdot x_1 + \dots + w_n \cdot x_n - w_0),$$

where w_0, w_1, \dots, w_n are coefficients that need to be determined during training, and $s_0(z)$ is a nonlinear function known as the *activation function*.

In this research, we used neural networks with sigmoid activation function $s_0(z) = 1/(1 + \exp(-z))$, the most widely used activation function. It is worth mentioning that this function can also be justified by invariance: this time, by shift-invariance (and not scale-invariance, as in the previous examples); see, e.g., [17].

Separation into training and validation data sets. Overall, we had 288 data sets, of which:

- 257 data sets correspond to patients wit lung dysfunctions, and
- 31 data sets correspond to patients without lung dysfunctions.

To train a neural networks, we separated the corresponding data set into training data set (used for training) and validation data set (used for validation). In all three cases, we used approximately 75% of the data for training and approximately 25% for validation; see, e.g., [8, 18]. Specifically:

- we selected 214 data sets for training, among which 191 corresponded to patients with lung dysfunctions and 23 corresponded to patients without lung dysfunctions, and
- the remaining 74 data sets were used for validation, among which 66 corresponded to patients with lung dysfunctions, and 8 patients without lung dysfunctions.

We have four possible diagnoses: asthma (a), SAI (s), PSAI (p), and the absence of lung dysfunctions (n). Within the training set, for each of these four diagnoses d , we applied the Least Square method to all the values $R_i(f)$

corresponding to the data sets with this diagnosis to compute the typical values $R_d(f)$ corresponding to this diagnosis. We then computed the integral $I_{R,d}(f) = \int_5^f R_d(x) dx$.

Resulting typical functions $R_d(f)$. As a result of this analysis, we got the following typical functions corresponding to different diagnoses; see Fig. 4:

$$\begin{aligned} R_a(f) &= 1.152 - 7.842 \cdot 10^{-2} \cdot f + 2.686 \cdot 10^{-3} \cdot f^2 - 2.443 \cdot 10^{-5} \cdot f^3; \\ R_s(f) &= 8.960 \cdot 10^{-1} - 5.738 \cdot 10^{-2} \cdot f + 2.067 \cdot 10^{-3} \cdot f^2 - 2.024 \cdot 10^{-5} \cdot f^3; \\ R_p(f) &= 6.076 \cdot 10^{-1} - 2.717 \cdot 10^{-2} \cdot f + 8.278 \cdot 10^{-4} \cdot f^2 - 4.888 \cdot 10^{-6} \cdot f^3; \\ R_n(f) &= 4.612 \cdot 10^{-1} - 1.508 \cdot 10^{-2} \cdot f + 4.789 \cdot 10^{-4} \cdot f^2 - 2.424 \cdot 10^{-6} \cdot f^3. \end{aligned}$$

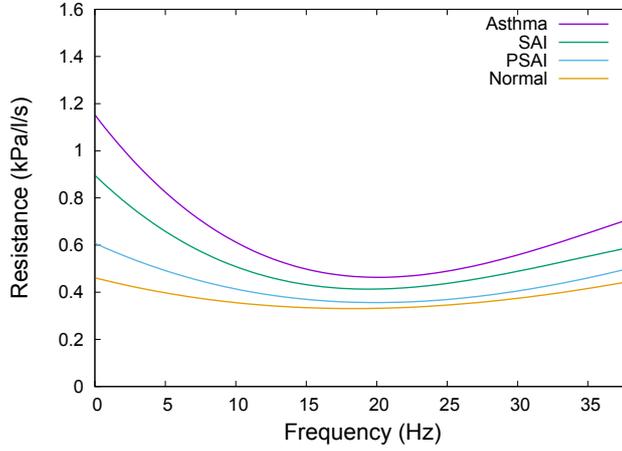


Figure 4: Cubic Resistance Functions per Class

Pre-processing: second stage. For each patient i , and for each of the four diagnoses d , we computed the following two similarity/dissimilarity measures:

- $\int |R_d(f) - (R_i^a(f) + \Delta R_{id})| df$, where $\Delta R_{id} \stackrel{\text{def}}{=} R_d(10) - R_i^a(10)$; and
- $\int |I_{R,d}(f) - (I_{R,i}^a(f) + \Delta I_{R,i,d})| df$, where $\Delta I_{R,i,d} \stackrel{\text{def}}{=} I_{R,d}(10) - I_{R,i}^a(10)$.

The resulting eight values serve as input to the neural network.

The results of training the neural network. The purpose of the neural network was to separate patients with lung dysfunctions from patients without lung dysfunction.

During the training, the network selected 50 neurons in the hidden layer. On the validation data set, the neural network achieved 100% accuracy on the validation set: all 74 cases were classified correctly.

This is much better than in the previous studies. The resulting classification accuracy is much better than the 60% accuracy achieved by neural networks without pre-processing.

Acknowledgments

This research was partially supported:

- by the National Council of Science and Technology of México (CONACYT) via a doctoral fellowship for Nancy Avila,
- by the Research Program on Migration and Health (PIMSA), a program funded by the University of California, Berkeley, Health Initiative of the Americas (HAI), and
- by the US National Science Foundation via grant HRD-1242122 (Cyber-ShARE Center of Excellence).

The data for this research were acquired by Erika Meraz as part of the NIH grant #1 S11 ES013339-01A1: “UTEP-UNM HSC ARCH Program on Border Asthma” (Pilot 1) granted to Dr. Homer Nazeran.

References

- [1] J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.
- [2] N. Avila, H. Nazeran, E. Meraz, N. Gordillo, and C. Aguilar, “Characterization of impulse oscillometric measures of respiratory small airway function in children”, *Advances in Electrical and Electronic Engineering*, to appear.
- [3] A. Badnjevic and M. Cifrek, “Classification of asthma utilizing integrated software suite”, *Proceedings of the 6th European Conference of the International Federation for Mechanical and Biological Engineering IFMBE’2015*, Dubrovnik, Croatia, 2015, pp. 415–418.
- [4] A. Badnjevic, L. Gurbeta, M. Cifrek, D. Marjanovic, “Classification of asthma using artificial neural networks”, *Proceedings of the 39th IEEE International Convention on Information and Communication Technology, Electronics, and Microelectronics MIPRO’2016*, Opatija, Croatia, 2016, pp. 387–390.
- [5] M. Barua, H. Nazeran, P. Nava, V. Granda, and B. Diong, “Classification of pulmonary diseases based on impulse oscillometric measurements of lung function using neural networks”, *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, California, September 1–5, 2004, pp. 3848–3851.

- [6] S. Bickel, J. Popler, B. Lesnick, and N. Eid, “Impulse oscillometry: interpretation and practical applications”, *Chest*, 2014, Vol. 146, No. 3, pp. 841–847.
- [7] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
- [8] A. Gholamy, V. Kreinovich, and O. Kosheleva, “Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation”, *International Journal of Intelligent Technologies and Applied Statistics*, 2018, Vol. 11, No. 2, pp. 105–111.
- [9] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, Wiley, Hoboken, New Jersey, 2009.
- [10] R. B. Kearfott and V. Kreinovich, “Beyond convex? global optimization is feasible only for convex objective functions: a theorem”, *Journal of Global Optimization*, 2005, Vol. 33, No. 4, pp. 617–624.
- [11] H. D. Komarow, I. A. Myles, A. Uzzaman, and D. D. Metcalfe, “Impulse oscillometry in the evaluation of diseases of the airways in children”, *Annals of Allergy, Asthma & Immunology*, 2011, Vol. 106, No. 3, pp. 191–199, doi 10.1016/j.anai.2010.11.011.
- [12] H. D. Komarow, J. Skinner, M. Young, D. Gaskins, C. Nelson, P. J. Gergen, and D. D. Metcalfe, “A study of the use of impulse oscillometry in the evaluation of children with asthma: Analysis of lung parameters, order effect, and utility compared with spirometry”, *Pediatric Pulmonology*, 2012, Vol. 47, No. 1, pp. 18–26.
- [13] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht, 1998.
- [14] G. Luo, B. L. Stone, C. G. Maloney, P. H. Gesteland, S. R. Yerram, and F. L. Nkoy, “Predicting asthma control deterioration in children”, *BMC Medical Informatics and Decision Making*, 2015, Vol. 15, Paper 84.
- [15] A. Marotta, M. D. Klinnert, M. R. Price, G. L. Larsen, and A. H. Liu, “Impulse oscillometry provides an effective measure of lung dysfunction in 4-year-old children at risk for persistent asthma”, *The Journal of Allergy and Clinical Immunology*, 2003, Vol. 112, No. 2, pp. 317–322, doi 10.1067/mai.2003.1627.
- [16] H. Mochizuki, K. Hirai, and H. Tabata, “Forced oscillation techniques and childhood asthma”, *Allergology International*, 2012, Vol. 61, No. 3, pp. 373–383.
- [17] H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.

- [18] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
- [19] Y. Shi, A. S. Aledia, A. V. Tatavoosian, S. Vijayalakshmi, S. P. Galant, and S. C. George, “Relating small airways to asthma control by using impulse oscillometry in children”, *The Journal of Allergy and Clinical Immunology*, 2012, Vol. 129, No. 3, pp. 671–678.
- [20] K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.
- [21] S. A. Vavasis, *Nonlinear Optimization: Complexity Issues*, Oxford University Press, New York, 1991.
- [22] World Health Organization (WHO), *Asthma*, <http://www.who.int/mediacentre/factsheets/fs307/en/>, downloaded on January 31, 2019