

Statistical Blood Spectra Analysis using AI

Hamze Bagheri^{1*}, Mehrzad Saremi²

¹Master of Philosophy of Science. Bachelor of Physics, Master of artificial intelligence.

²Bachelor of Computer Science. Amir-Kabir University of Technology, Tehran, Iran

Abstract: *a statistical method in order to decode the blood spectrum has been presented. It has been suggested to use AI for decoding FTIR spectra of blood. In this method, there is no requirement to know what the blood samples contain, but only to know what the characteristics of the blood donors are. The principles are that the data is never lost, the phenomena are regular and that the blood spectra contains the donor's diagnostic information.*

Keywords: *Blood Spectrophotometry, Diagnostic, Spectrum, Artificial Intelligence, Statistical Analysis, FTIR Analysis.*

Introduction:

Many researchers and scientists are now working on the subject "Blood Spectrophotometry". Spectrophotometry is a method to measure how much a chemical substance absorbs light by measuring the intensity of light as a beam of light passes through sample solution. This measurement can also be used to measure the amount of a known chemical substance. Every chemical compound absorbs, transmits, or reflects light (electromagnetic radiation) over a certain range of wavelength. Spectrophotometry is a measurement of how much a chemical substance absorbs or transmits. Spectrophotometry is widely used for quantitative analysis in various areas (e.g., chemistry, physics, biology, biochemistry, material and chemical engineering, industrial applications, etc.). In clinical applications, it is used to examine blood or tissues for clinical diagnosis. A spectrophotometer is an instrument that measures the amount of photons (the intensity of light) absorbed after it passes through sample solution. With the spectrophotometer, the amount of a known chemical substance (concentrations) can also be determined by measuring the intensity of light detected. Depending on the range of wavelength of light source, it can be classified into two different types:

- UV-visible spectrophotometer: uses light over the ultraviolet range (185 - 400 nm) and visible range (400 - 700 nm) of electromagnetic radiation spectrum.
- IR spectrophotometer: uses light over the infrared range (700 - 15000 nm) of electromagnetic radiation spectrum. (CRAMI2R, 1968)

The result will be to offer a spectrophotometric equivalent for the diagnostic of a certain disease or detection of certain characteristics of a blood sample.

Current Method: In the current method, mostly researchers try to determine the presence and amount of a molecule or a biochemical in the blood. For instance, Sheeba Manoj Nair have worked on the estimation of Silodosin and Silodosin β -D-Glucuronide in human's plasma (Nair, 2016). The procedure is one that has been tried for years. For example, STIG SELANDER and KIM CRAMI2R in 1968 have tried to determine lead in blood by spectrophotometry (CRAMI2R, 1968) . The method was meant to ease the diagnostic procedure by detecting the presence of a certain molecule or exceedance of a molecule or atom from a

standard concentration in blood by spectrophotometry as a fast and efficient way. In this way, the project of diagnostic separated in two phases:

- 1- Investigation of the effect of the diseases on the blood, as existence of a certain molecule or atom or exceedance of a molecule or atom from a standard concentration in blood.
- 2- Detection of the effect (presence of a certain molecule or exceedance of a molecule or atom from a standard concentration in blood) on blood spectra.

The first phase is the one having been the subject of investigation independent to spectrophotometry, but the second phase is that the spectrophotometry researchers try to do. It is like there are three spaces:

- 1) The Space of Medical Tests and Diagnosis (T&D). (What the laboratory confirms as a person's health status.)
- 2) The Space of Blood natural elements. (The atoms and molecules contained in the blood and their proportion.)
- 3) The Space of Blood Spectra. (The Spectra obtained by spectrophotometer.)

The two steps have been illustrated in Figure 1:

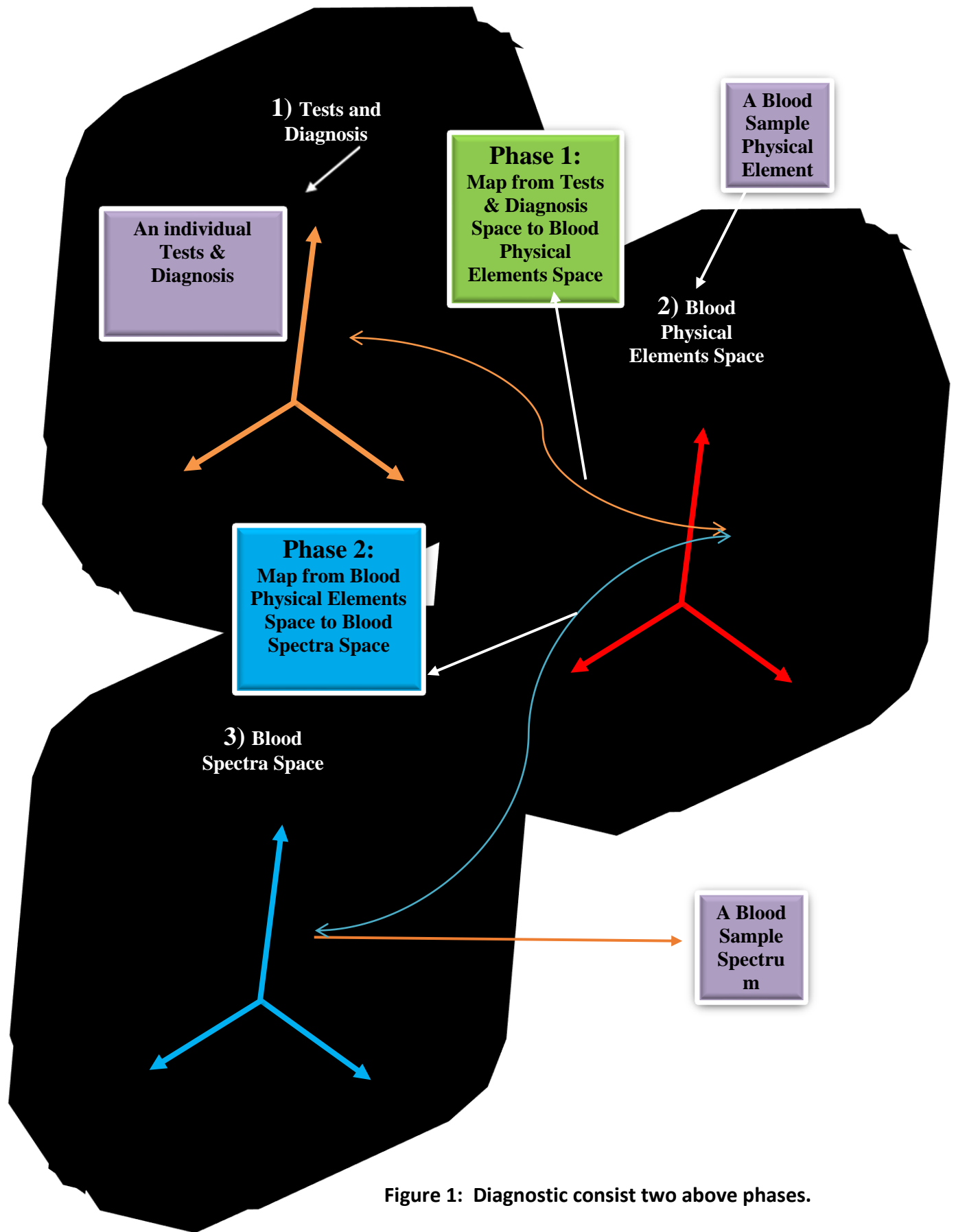


Figure 1: Diagnostic consist two above phases.

In figure 1, each point in the Test and Diagnosis Space is a possible state that a complete medical tests and diagnosis can have (For instance, age:35, disease: MS, HDL: 210 ...), and it must contain all medical tests. (the point contains information of complete medical report included: Laboratory test results, Medical images, medical history, documentation of person's diagnosis and Findings of physical and mental examinations. As more information is gathered related to the blood sample owner is better. But only verified and accurate medical tests and diagnosis result could be enough for the purpose.) Also each point in Blood Physical Elements Space, is an individual arrangement of molecules and their proportions that a blood sample can be and the space contains all possible arrangements. And each point in the Blood Spectra Space is an individual spectrum, and the space contains all possible spectra. The first phase is to find a map between the space 1 and the space 2 and the phase 2 is to find a map between the space 2 and the space 3. In this way; for instance, a research in phase 1 reveals that a special *disease* characteristic feature *is the presence of* a special element or exceedance of that in blood from an amount. Another research in phase 2 reveals that the presence of the element or the exceedance of the element from the standard degree shows itself in the blood spectra in a special range of wavelength. The first phase is to find a map between the space 1 and the space 2 and the phase 2 is to find a map between the space 2 and the space 3. In this way, for instance a research in phase 1 reveals that a special disease is equivalent with the existence of a special molecule or exceedance of a molecule in blood from an amount. Another researcher in phase 2 reveals that the existence of the molecule or the exceedance of the molecule from the standard degree shows itself in the blood spectra in a special range of wavelength or anyway they can find an equivalence in the spectra for that.

* Consequently by considering the two results, we are able to detect diseases by analyzing the spectra.

Critics to the Current Method: The method currently in use is so hard and time and effort consuming. One must first know the impact of each disease in the sample, like in case of a disease the existence of a certain molecule or exceedance of a molecule or atom from a standard concentration in blood and then, find the impact of that component on the blood spectra through spectrophotometry. In this method, diagnosis of each disease is a separate hard project and there are many diseases. If it is possible, it takes years to recognize all illnesses and health statuses in the blood spectra. We suggest a method in order to ease the process. What happens if we join the two phase and map the space 1 directly to space 3? This is the idea of using AI in blood spectrophotometry.

Statistical Analytical Method using AI: In this method, we don't need to know about the blood, but just the equivalency of blood spectrum with a specific diagnostic status. A blood spectrum would be a function of intensity in terms of wavelength. We suppose that there is a correspondence one-to-one relationship between the blood Diagnostic Characteristics-Blood Samples and Blood spectra. (Maybe a disease has no effect on blood and consequently on the blood spectrum. In this case, the disease will not be detected through spectra analysis. But for those that have effect in blood, known or unknown effect, the disease will show itself on blood spectra analysis.)

Consider figure 2:

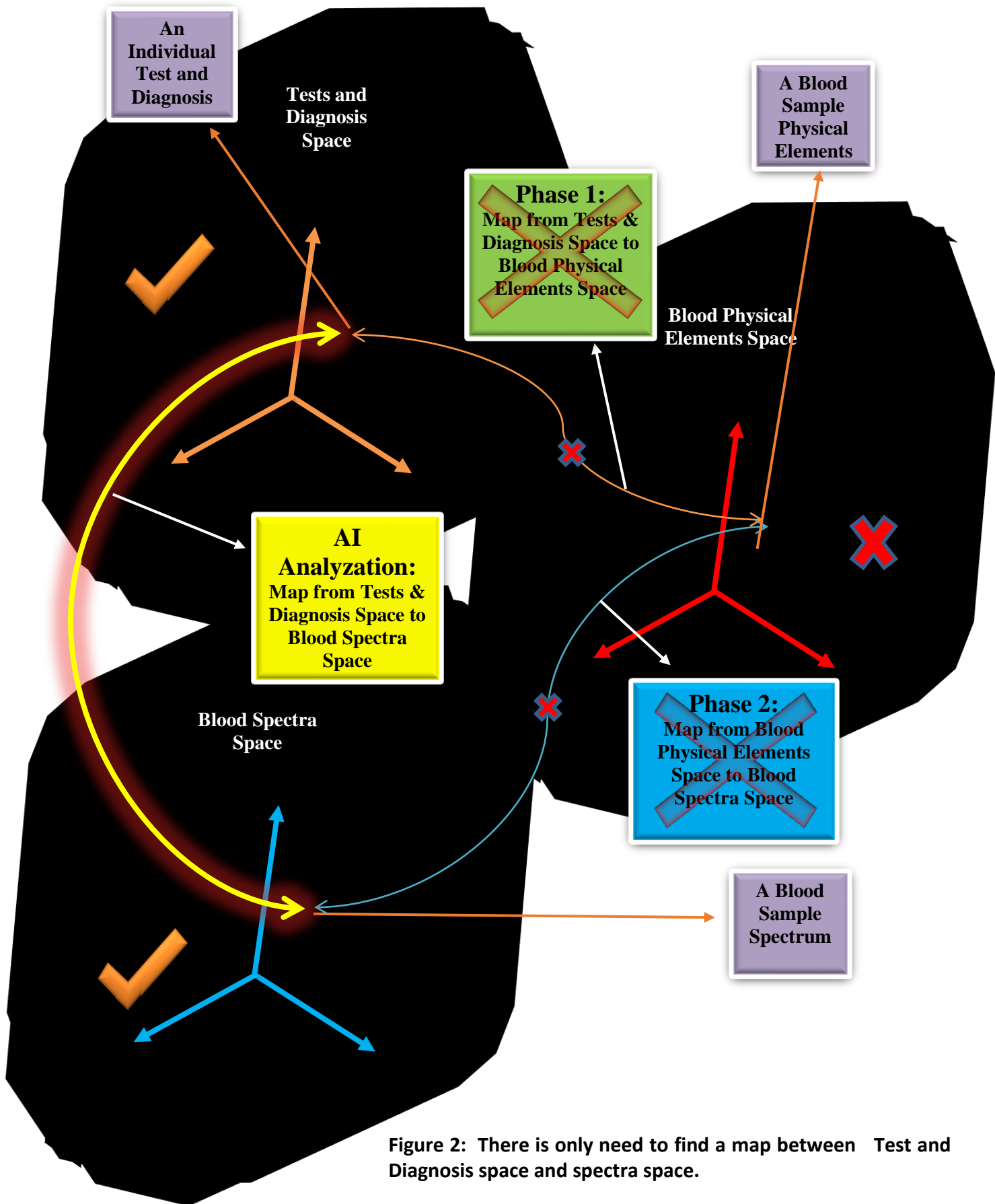


Figure 2: There is only need to find a map between Test and Diagnosis space and spectra space.

Suppose that we have S blood samples (the number of blood samples is a matter that is needed to be discussed, but the more blood samples we have, the more reliable the result will be.)

On the other hand, we have the blood spectra of each blood sample. Consequently, we can conclude that each diagnostic characteristic is equivalent to a specific spectrum. See figure 2 for more details.

More directly, the claim is upon two postulates:

Postulate one:

- 1) For each array of diagnostic characteristic there is an individual blood sample characteristic.
- 2) For each blood sample characteristic there is an individual blood spectrum.

As a result, we have:

“For each array of diagnostic characteristic, there is an individual blood spectrum.”

Point1: That’s possible that a disease has no impact on the blood spectrum, that in the case we can’t detect the disease through spectrophotometry. But we can still use the method for other diseases.

Point2: It’s possible that a disease signature in the spectra is so small that we need that the spectrophotometry is as precise as possible.

Point3: We don’t know the exact characteristics that can change the blood spectra. But we have some knowledge about that, and we know that some illnesses or other characteristics have an equivalent impact on the blood. That’s enough, to determine the field of diagnostic characteristics, we mention all characteristics except those we are sure to have no impact on the blood. (Maybe a disease shows itself as lack of a material in the blood, but special array in different substances in the blood. We are not disappointed to detect the disease through spectrophotometry, and we appoint it to the software to discern that.)

Point4: Claim 2 can be put under question? Maybe for a sample characteristic there is more than one spectrum. In the spectrum, the area 400-1500 is called the *finger print* in spectrophotometry, that implies that for each material the spectrum in the area is unique. But majority of the researchers don’t work in the area, because of the difficulty of analyzing the area. The rest of the spectrum area, 1500-4000 also contains so much information about the material, but more known and the researchers exactly know that what the picks and different shapes in the area are representing. Altogether, we know that for each blood sample there is just an specific blood spectrum, and the rest is up to the AI to understand and decode the spectra.

To Use AI: This method doesn’t require any extra knowledge about how a disease influences the spectrum. The machine itself will learn a relation between blood spectra and diseases – hopefully, the groundtruth function. One needs to provide an adequate number of samples and let the machine train itself. The AI method is not explanatory per se, though. The term “blackbox” has been dubbed for this situation: The machine can diagnose diseases very well, but without giving the experts on why. This is not essentially a problem. The FDA has already given the green light to at least one machine learning algorithm. DeepVentricle (Automated Cardiac MRI Ventricle Segmentation using Deep Learning, 2017), which is supposed to detect heart failure accurately by segmentating the heart’s chambers and estimating the amount of blood the heart can hold and pump, has received FDA clearance in February 2017 (Food and Drug Administration: Arterys cardio dl., n.d.). This system uses Deep Neural Networks fed with CMR images to detect the contours of the heart. Using this method is as simple as setting a machine to learn from existing samples of blood samples and diagnosis pairs. The machine acts as a black box that one can use to implicitly pick up the relationship between the blood spectra and the disease. The neural networks have recently shown to work well in diagnosing disease. In (R.JackJr.a, 2008), Prashanthi Vemuri et. AI have created deep networks that can learn to diagnose the Alzheimer disease with an accuracy higher than specialized therapists. Support Vector Machines (SVMs) have also been used to detect early phases of the disease (Khedher, Ramírez, Górriz, Brahim, & Segovia, 2015). SVM’s are also machine learning systems which try to find the best separating rules between those who have Alzheimer and those who doesn’t. There are indeed three major areas in which AI is being used: Cancer, neurology and cardiology (Jiang et al., 2017). AI would learn and it can guess the next blood spectra equivalent Test and Diagnosis information, and that’s what we

want.

More Details About the Method: Suppose that we have n diagnostic characteristics. Different diseases (suppose that we have a category of d diseases), Every LDL rate (for instance, LDL can be from 0 to 1600, which is 1600 characteristics), every age (age can be presented by year or month or day. day seems more precise, but maybe that is not needed). Now we prepare n empty boxes for an individual and fill them with 0 or 1 as follow:

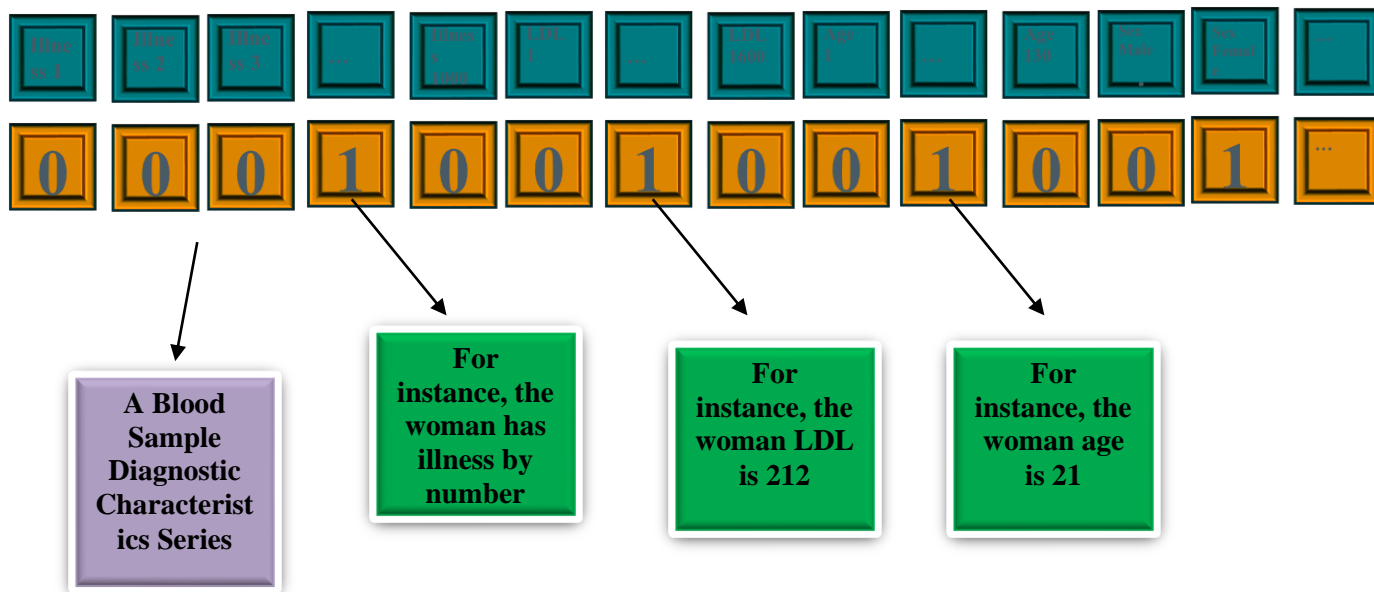


Figure 3: Coding the Test and Diagnosis Information.

As you see we have dedicated for each characteristic an empty box, and if the blood sample owner has the characteristics, we write **1** in the box, and if not we write **0**. Then for each blood sample, we will have a series of n of 0 and 1. n boxes that can be both 0 or 1, have 2^n possible arrays. And our samples series are a tiny fraction of the large number. (we suppose we have S samples). On the other hand, we have the space of blood spectra.

Suppose a spectra function:

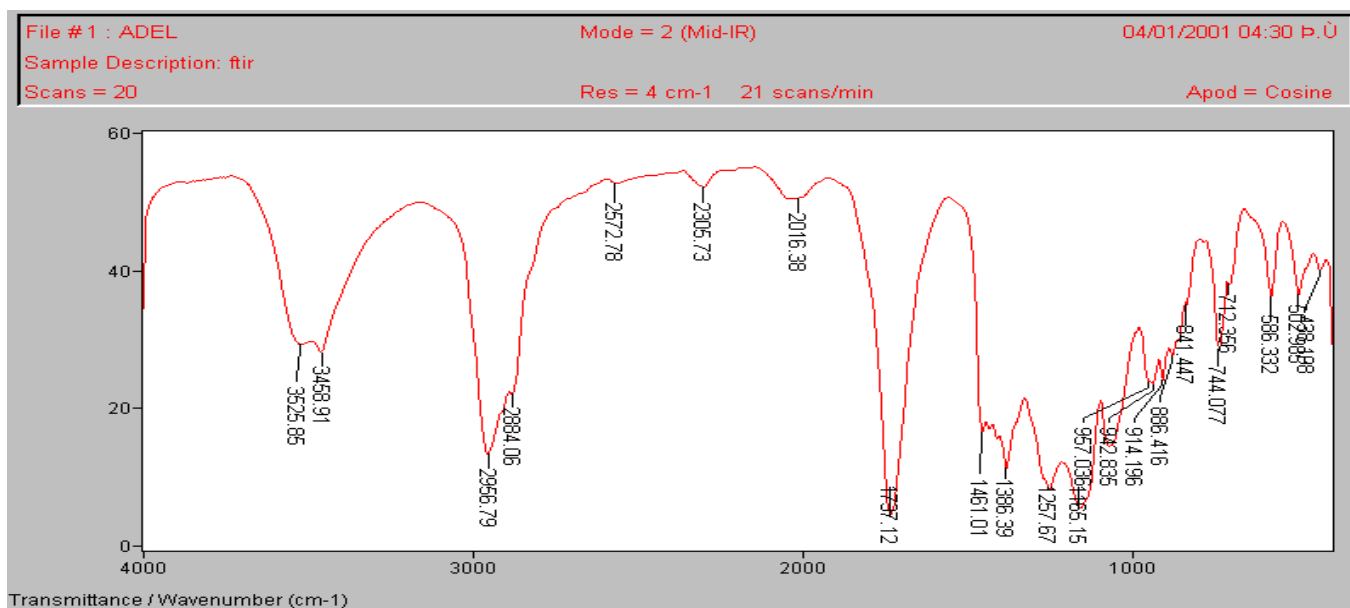


Figure 4: a spectrum.

Suppose that would be something like the above figure. There are K points in the horizontal axis. Each point is a specific frequency and that means the spectrophotometer can distinguish K frequencies. And the vertical axis has J points. (for instance, in the figure above by the resolution 4, K is 35000000, and J is 10000000.)It means the spectrophotometer can distinguish J different absorbance degree for each frequency. Now we want to depict the function as binary in 0 and 1. For each point, we need e empty boxes. Each array in e bits is equivalent to an amount of absorbance. We have K points that each one can have a number between 0 to J . For each point, we need e boxes (to cover all possible arrays). Consequently, we need $e * J$ empty boxes that each one can be whether 0 or 1. Then for each possible function, there is a series in binary with 10.000.000 of 0 and 1.

All possible arrays are 2^{e*J} arrays. And as we supposed that we have S samples, we have S spectras and thus S arrays. We call these arrays the arrays of spectra.

- 1) we have S diagnostic characteristics equivalent to S spectras.
- 2) For each the diagnostic characteristic, there is a binary series of n of 0 and 1.
- 3) For each the spectra, there is a binary series of $e * J$ of 0 and 1.

Consequently:

For each Test and Diagnosis series (binary series of n of 0 and 1), there is a spectra series (binary series of $e * J$ of 0 and 1).

Then we have:

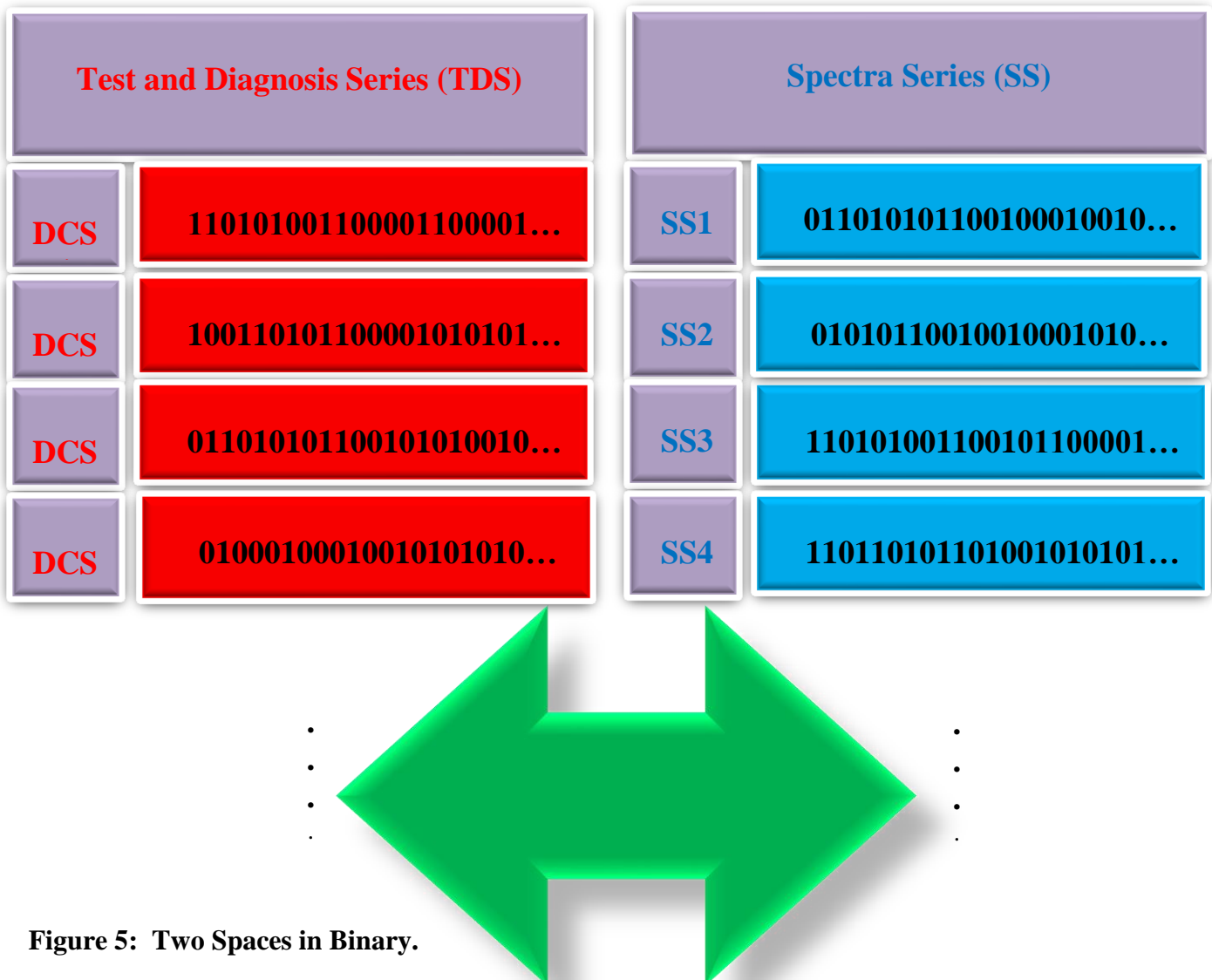


Figure 5: Two Spaces in Binary.

As you see in the above figure, there are S of TDS series and S of SS series equivalent to them. For each series in binary, there is a number. In fact, binary is a representation for numbers also. For instance, in TDS that we considered that contained n of 0 and 1, there are 2^n possible arrays. We can arrange them and correspond them to numbers in power 10. Then each series will be a number between 0 and 2^n . For SS series is also the same. Each series is a number between 0 and 2^{e*J} . Then instead of equivalence between series, we can consider the equivalence between numbers as follow:

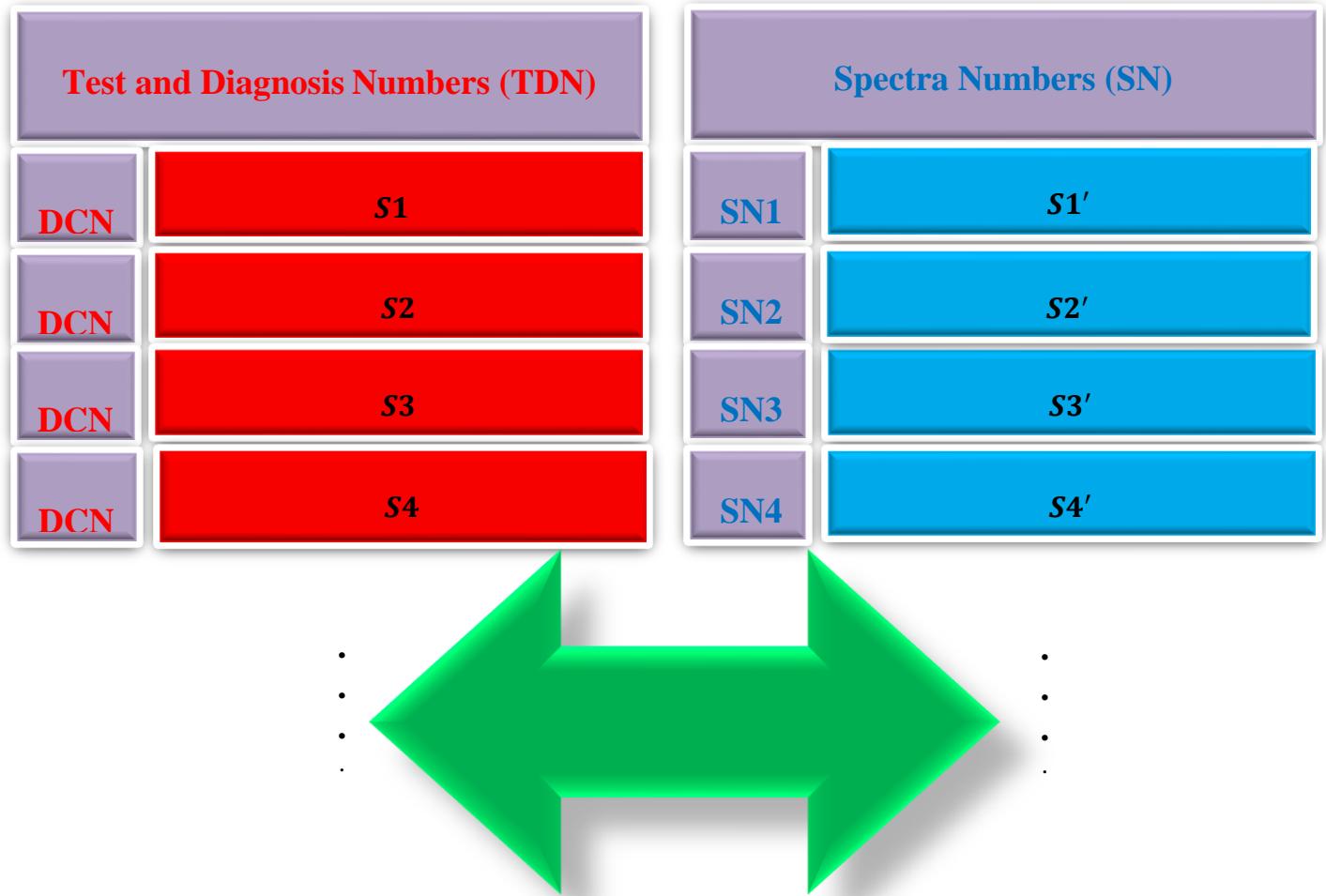


Figure 6: Two Spaces as numbers.

We write the above as follow:

$$F(DCN1) = SN1$$

$$F(DCN2) = SN2$$

$$F(DCN3) = SN3$$

⋮
⋮
⋮

$$F(DCNs) = SNs$$



$$F(TDN) = SN$$

The problem is to find the function **F**. In 2D, you can suppose that as a shape that we have 2000000 points

of the shape, and the problem is to guess the shape. That's like this:

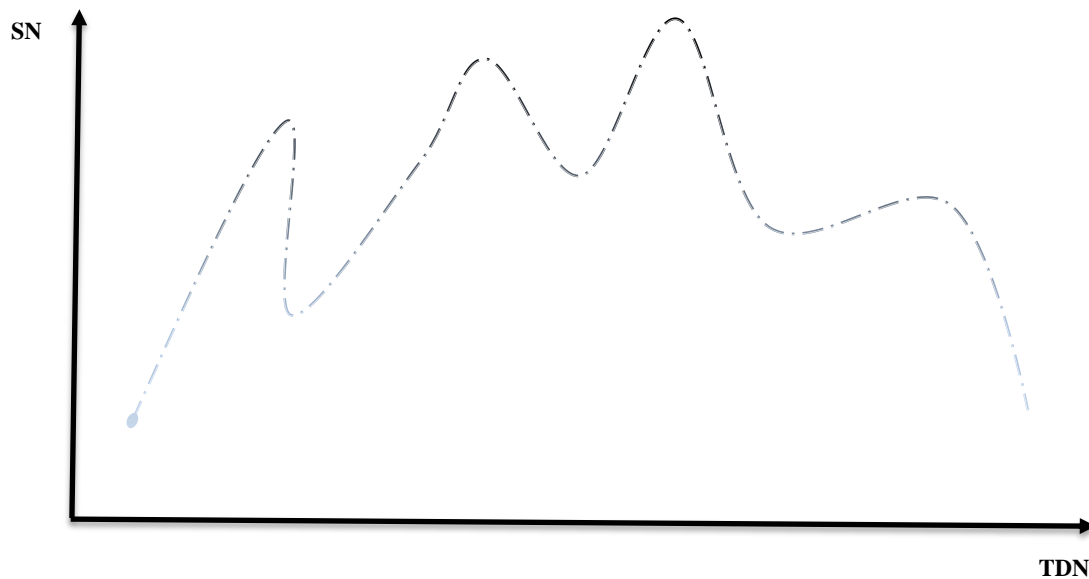


Figure 7: The Estimated Function.

Of course, the number of points we have related to the number we don't have is so few and the figure is just for illustration. But the problem is this, to guess the figure by this point, or equivalently to guess the function that maps the space of **TDN** to the space of **SN**. And this is what a black box does.

References:

- [1] Nair, S. M. (2016). Development and Validation of High Performance LCMS Methods for Estimation of Silodosin and Silodosin β -D-Glucuronide in Human Plasma. *Pharmaceutical Analytical*.
- [2] CRAMI2R, S. S. (1968). Determination of Lead in Blood by Atomic Absorption. *Brit. J. industr. Med.*, 209.
- [3] Smith, Janice Gorzynski (2011). "Chapter 13 Mass Spectrometry and Infrared Spectroscopy". *Organic chemistry(Book)* (3rd ed.). New York, NY: McGraw-Hill. pp. 463–488.
- [4] Filip Monica Sanda, M. E. (2012). SPECTROPHOTOMETRIC MEASUREMENTS TECHNIQUES FOR FERMENTATION PROCESS. *Two countries, one goal, joint success*.
- [5] *Automated Cardiac MRI Ventricle Segmentation using Deep Learning* . (2017, May 9). Retrieved from http://on-demand.gputechconf.com/gtc/2017/presentation/s7654-daniel-golden_cloud-based-deep-learning-as%20the-radiologists-best.pdf
- [6] *Food and Drug Administration: Arterys cardio dl* . (n.d.). Retrieved from http://www.accessdata.fda.gov/cdrh_docs/pdf16/K163253.pdf
- [7] R.JackJr.a, P. L. (2008). Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *Elsevier*.