Volume 4, Issue 8, August 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Algorithm for Estimation of Blood Cholesterol based on Regression Technique

Ingrid Anne P. Nazareth
Research Scholar,
Dept. of Electronics, Taleigao Plateau,
Goa University, Goa, India

Rajendra S. Gad
Associate Professor,
Dept. of Electronics, Taleigao Plateau,
Goa University, Goa, India

Sulaxana R. Vernekar
Research Scholar,
Dept. of Electronics, Taleigao Plateau,
Goa University, Goa, India

Gourish M. Naik
Professor & Head,
Dept. of Electronics, Taleigao Plateau,
Goa University, Goa, India

Abstract*—In order to lead a healthy life, it is important for a human being to know his/her blood level parameters. The human blood has over 100 constituents of which Cholesterol is an important parameter, which needs to be estimated for health purposes. It is necessary to develop an instrument to measure blood constituents non-invasively which would be user friendly, portable and reliable. This manuscript describes the algorithm to estimate Blood Cholesterol. Here the PLS and NIPALS algorithm is explained. The manuscript includes the data obtained for Cholesterol dissolved in water as per percentage found in the human blood. The data thus obtained will be fed to a multivariate system programmed in FPGA device to estimate the cholesterol concentration. The studies on other blood constituents such as Glucose, Alanine, Lactate, etc. are in progress. The results shown are in the RF absorption range of 10MHz-4GHz.

***Keywords*— Cholesterol, Multivariate System, PLS, RF Spectroscopy, NIPAL.**

I. INTRODUCTION

Cholesterol is a soft fat like waxy substance found in the bloodstream and in all the cells of the body. It is an important part of the body because it is used for building cells. A high blood cholesterol level is a major risk for coronary heart disease which can lead to heart attacks.

High cholesterol is a significant risk factor for cardiovascular disease (CVD) and specifically for coronary heart disease (CHD). CVD is the leading cause of death in the India and U.S..India spends less than 1% of its resources for health care compared to 17% in the US. Since the U.S. government spends only 17-18% of this, the out-of-pocket health expenses incurred by households is as high as 80% or more.[1][2][3] The combined inpatient and outpatient services CVD expenditure ranged from $773 in low-income to $1593 in middle income and $2917 in high income Indians, whereas the median individual monthly income was $136, $181 and $302 respectively.[4] By 2030, approximately 40% of the U.S. population is expected to have some form of CVD. As a result, direct medical costs for CVD in the U.S. are expected to triple from $273 billion currently to $818 billion in 2030.

Venipuncture is the process of obtaining intravenous access for blood sampling of venous blood. It is one of the most routinely performed invasive procedures. Health care workers use a syringe-needle technique which develops a fear among patients and could also lead to infection. The tubes in which blood/serum is transported back to the laboratory contain a variety of additives which is used in blood analysis. It is important to know which tube the individual laboratory requires for which test, as reagents vary between laboratories and may be affected by different additives. In general, whole blood needs to be mixed with Ethylene diaminete traacetic acid to prevent it from clotting.

II. METHODOLOGY

The constituents like Cholesterol whose frequency responses which are to be measured are dissolved in a known quantity of distilled water. The samples are analysed in the 10MHz-4GHz range. The Methodology for the technique of the process of preparation of the samples and the experimental setup can be referred to in my paper [5]. The multi-frequency bio-electrical impedance spectrum can be modeled through multivariate and curve-fitting statistical applications to develop summary parameters to estimate body composition like Cholesterol.

III. ALGORITHM

A Partial least squares (PLS)

PLS models are based on principal components of both the independent data **X** and the dependent data **Y**. The main idea is to calculate the principal component scores of the **X** and the **Y** data matrix and to set up a regression model between the scores. [6]

$X = T * P' + E$ ---1
$U = B * T$ ---2
$Y = U * Q' + F$ ---3

Thus the matrix **X** in equation 1 is decomposed into a matrix **T** (the score matrix) and a matrix **P'** (the loadings matrix) plus an error matrix **E.** The matrix **Y** is decomposed into **U** and **Q** and the error term **F** in equation 3. The goal of the PLS algorithm is to minimize the norm of **F** while keeping the correlation between **X** and **Y** by **U** = **BT** (equation 2) .

The important point when setting up a PLS model is to make a decision for the optimum number *a* of principal components involved in the PLS model. While this can be done from variation criteria for other models, for PLS the optimum number of components has to be determined empirically by cross validation of the PLS model using an increasing number of components. A low PRESS (PRedictive Error Sum of Squares value) value indicates a good prediction model.[7][8]

B Non-linear Iterative Partial Least Squares(NIPALS)

Frequently used iterative algorithm is the NIPALS algorithm. The NIPALS Algorithm has been developed by H. Wold at first for PCA and later-on for PLS. It is the most commonly used method for calculating the principal components of a data set. It gives more numerically accurate results when compared with the SVD of the covariance matrix, but is slower to calculate.

Assuming that the data to be analysed is stored in matrix **X**, the steps to calculate the loadings **u** and scores **v** of the principal components is shown in Table I.

TABLE I: STEPS TO CALCULATE THE LOADINGS AND SCORES OF THE PRINCIPAL COMPONENTS

Step	Math	Explanation
1	$\mathbf{u} := \mathbf{x}_i$	Select a column vector $\mathbf{x}_i$ of the matrix **X** and copy it to the vector **u**
2	$\mathbf{v} := (\mathbf{X}'\mathbf{u})/(\mathbf{u}'\mathbf{u})$	Project the matrix **X** onto **u** in order to find the corresponding loading **v**
3	$\mathbf{v} := \mathbf{v}/\lvert\mathbf{v}\rvert$	Normalize the loading vector **v** to length 1
4	$\mathbf{u}_{old} := \mathbf{u}$ $\mathbf{u} := (\mathbf{Xp})/(\mathbf{v}'\mathbf{v})$	Store the score vector **u** into $\mathbf{u}_{old}$ and project the matrix **X** onto **v** in order to find corresponding score vector **u**
5	$\mathbf{d} := \mathbf{u}_{old}-\mathbf{u}$	In order to check for the convergence of the process calculate the difference vector **d** as the difference between the previous scores and the current scores. If the difference $\lvert\mathbf{d}\rvert$ is larger than a pre-defined threshold (e.g. 10^{-8}) then return to step 2.
6	$\mathbf{E} := \mathbf{X} - \mathbf{tp}'$	Remove the estimated PCA component (the product of the scores and the loadings) from **X**
7	$\mathbf{X} := \mathbf{E}$	In order to estimate the other PCA components repeat this procedure from step 1 using the matrix **E** as the new **X**

IV. RESULTS

The graphs are recorded as shown in Fig. 1 to Fig. 5. It may be noted here that even though the experiment was conducted from 10MHz to 4GHz continuously, the responses are found only in certain regions and some of them are shown in Fig. 1 to Fig. 5.

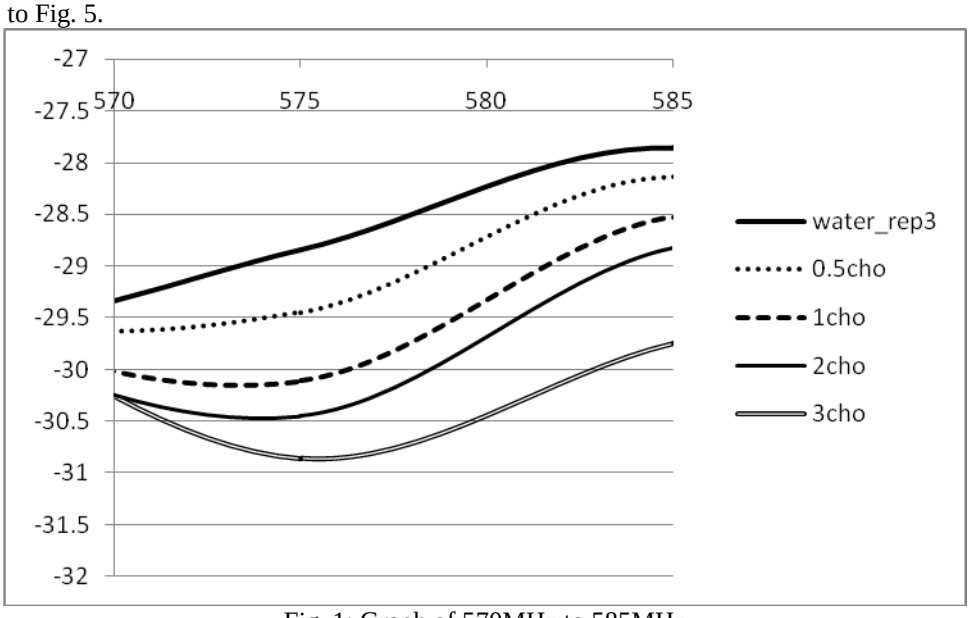

Fig. 1: Graph of 570MHz to 585MHz

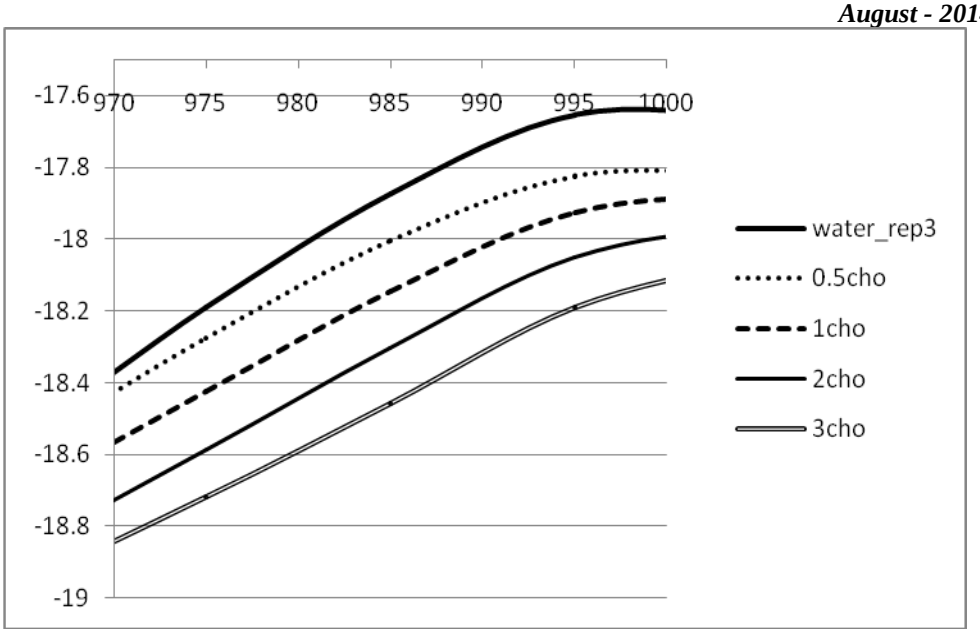

Fig. 2: Graph of 970MHz to 1000MHz

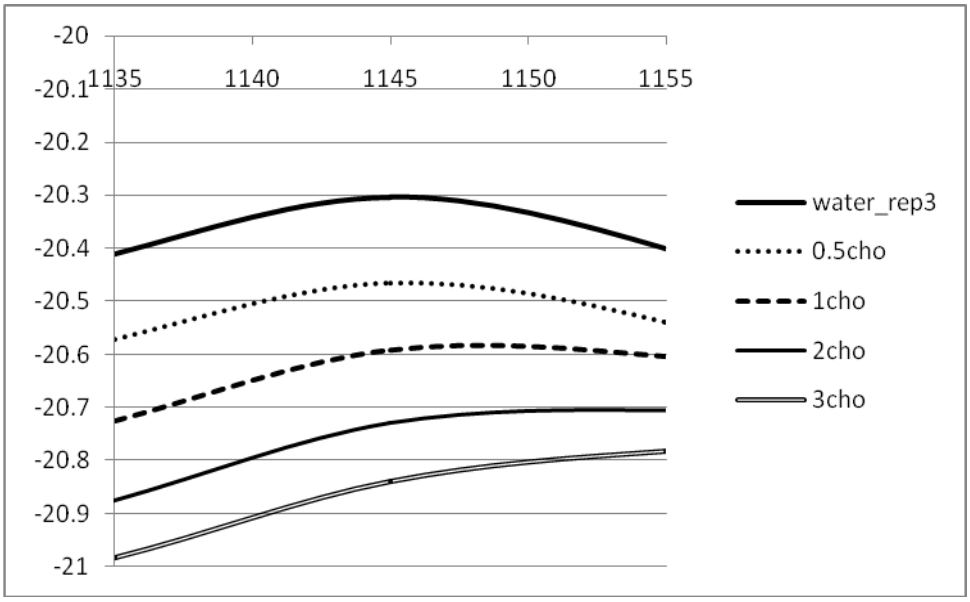

Fig. 3: Graph of 1135MHz to 1155MHz

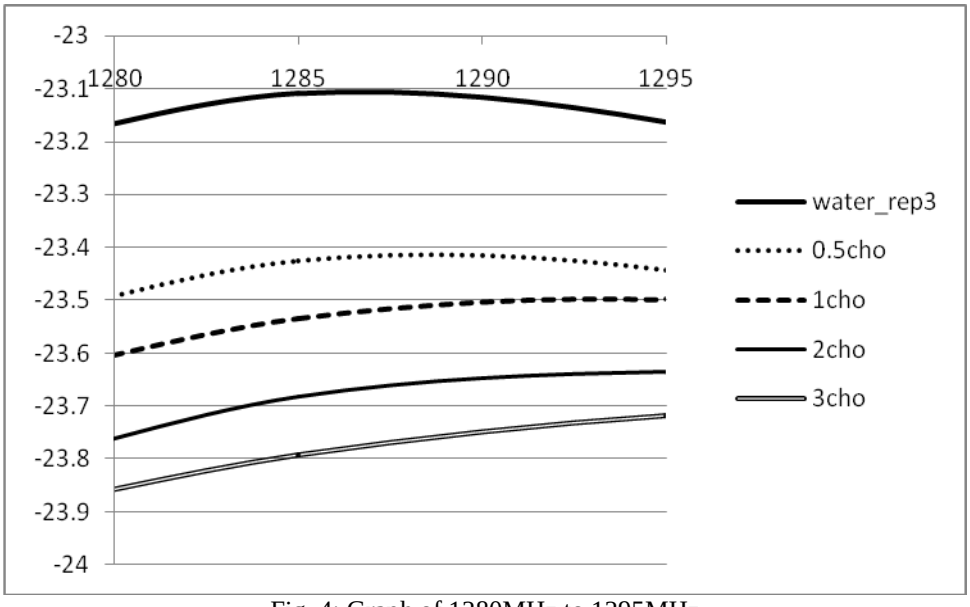

Fig. 4: Graph of 1280MHz to 1295MHz

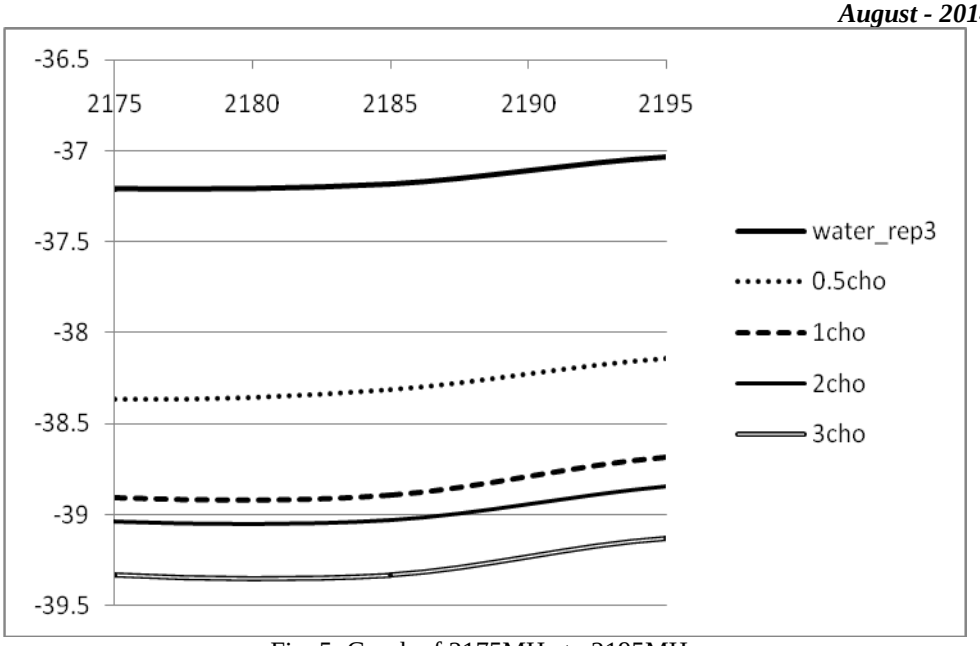

Fig. 5: Graph of 2175MHz to 2195MHz

It can be observed from the graphs shown in Fig. 1 to Fig. 5 that as the concentration of cholesterol increases, the absorption increases at 575MHz, 995MHz, 1145MHz, 1285MHz and 2185MHz

TABLE II : VARIATION OF ABSORPTION LEVEL OF CHOLESTEROL IN DB

Freq in MHz	Water	Concentration			
		0.5	1	2	3
575	-28.84	-29.44	-30.11	-30.45	-30.86
995	-17.65	-17.83	-17.93	-18.05	-18.19
1145	-20.30	-20.47	-20.59	-20.73	-20.84
1285	-23.11	-23.43	-23.53	-23.68	-23.79
2185	-37.18	-38.31	-38.89	-39.03	-39.33

The figures in Table II show the absorption of Cholesterol in dB. Here it shows that as the concentration of cholesterol increases the attenuation increases at frequencies 575MHz, 995MHz, 1145MHz, 1285MHz and 2185MHz.

From the Table shown above, the spectra of each of the blood constituents are unique and therefore the data of the spectra is fed to a PLSR model to find out unknown value of blood constituents like cholesterol, glucose, urea, etc. Development of full instrumentation is in progress.

ACKNOWLEDGMENT

Author Ingrid Anne P. Nazareth thanks Goa Univerity for providing a scholarship for the above work.

REFERENCES

[1] Vamadevan AS, Shah BR, Califf RM, Prabhakaran D. Cardiovascular research in India: a perspective. *Am Heart J.* Mar 2011;161(3):431-438.

[2] Gupta R. Coronary heart disease in India: Absolute numbers and economic burden. 2004 April 6.

[3] Prabhakaran D, Yusuf S. Cardiovascular disease in India: lessons learnt & challenges ahead. *Indian J Med Res.* Nov 2010;132(5):529-530.

[4] Huffman M D, Rao KD, Pichon-Riviere A, et al. A cross-sectional study of the microeconomic impact of cardiovascular disease hospitalization in four low- and middle-income countries. *PLoS ONE.* 2011;6(6):e20821.

[5] Ingrid Anne P. Nazareth, Sulaxana R.Vernekar, Rajendra S. Gad, Gourish M. Naik, IJIREEICE Vol 1, Issue 9, December 2013: 413-417

[6] Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. Anal Chim Acta 1986;185:1–17.

[7] Wold S, Ruhe A, Wold H, Dunn WJ. The Collinearity problem in linear regression. The partial least squares approach to generalized inverse. SIAM J Sci Stat Comput 1984;3:735–43.

[8] Wold, Svante; Esbensen, Kim (1987), "Principle Component Analysis", *Chemometrics and Intelligent Laboratory Systems* **2**: 37–52

