Data de-noise for Discriminant Analysis by using Multivariate Wavelets (Simulation with practical application)

Prof. Dr. Taha Hussein Ali, Department of Statistics, College of Administration and Economics, Salahaddin University, Erbil, Kurdistan Region, Iraq. Email: taha.ali@su.edu.krd
Asst. Prof. Dr.Nazeera Sedeek Kareem, Department of Statistics, College of Administration and Economics, Salahaddin University, Erbil, Kurdistan Region, Iraq. nazeera.kareem@su.ed.krd
Lect. Awaz shahab mohammad, Department of Mathematics, College of Education, Salahaddin University, Erbil, Kurdistan Region, Iraq. Awaz.shahab82@yahoo.com

ABSTRACT
In this paper, we suggest using the multivariate wavelet analysis in higher dimensional space (Symlet, Daubechies' least-asymmetric wavelets) with soft thresholding to de-noise of the data (Shrinkage) before use it in the Discriminant Analysis to obtain more accurate and reliable results by comparing it with the Discriminant analysis used on data before de-noise. And to know the effect of de-noise from data (proposed method) on Discriminant analysis results by simulating random data with normal distribution repeated 1000 times for different combinations of number of variables and sample sizes and real data represent leukemia patients taken from the Nanakele Hospital in Erbil. We analyzed the data depending on MATLAB Language and statistics program (SPSS). One of the most important conclusions reached by the researcher that use proposed method led to the separation between the two groups better than before de-noise (classical method) and this mean that data were classified for proposed method better than classical method.
Keywords: Discriminant analysis, Classification, multivariate wavelet, Shrinkage, Minimax, soft thresholding, Symlet Wavelets and de-noise.

تقليل ضوضائية البيانات للتحليل التمييزي باستخدام مويجات متعددة المتغيرات (محاكاة مع تطبيق عملي)

الأستاذ الدكتور طه حسين علي الزبيدي، قسم الإحصاء، كلية الإدارة والإقتصاد، جامعة صلاح الدين، أربيل، إقليم كوردستان، العراق، taha.ali@su.edu.krd

الأستاذ المساعد الدكتور نظيرة صديق كريم، قسم الإحصاء، كلية الإدارة والإقتصاد، جامعة صلاح الدين، أربيل، إقليم كوردستان، العراق، nazeera.kareem@su.ed.krd

المدرس آواز شهاب محمد، قسم الرياضيات، كلية التربية، جامعة صلاح الدين، أربيل، إقليم كوردستان، العراق، awaz.shahab8@gmail.com

الملخص :

في هذا البحث ، إقترح الباحثون استخدام تحليل المويجات متعددة المتغيرات في فضاء متعدد الأبعاد (Symlet) وهي مويجات دابشيز الأقل تماثلاً، مع قطع العتبة الناعمة لتقليل ضوضائية البيانات (التقليص المويجي) قبل استخدامها في التحليل التمييزي للحصول على نتائج أكثر دقة وموثوقية من خلال مقارنتها مع التحليل التمييزي المستخدم على البيانات قبل معالجة الضوضائية. ولمعرفة تأثير تقليل ضوضائية البيانات (الطريقة المقترحة) على نتائج التحليل التمييزي تم إستخدام محاكاة لبيانات لها توزيع طبيعي والذي تم تكراره (1000) مرة لمجموعات مختلفة من عدد المتغيرات وأحجام العينات، ومن ثم إستخدام بيانات حقيقية تمثل مرضى سرطان الدم المأخوذة من مستشفى ناناكيلي في أربيل. تم تحليل البيانات من خلال برنامج مصمم بلغة MATLABمع البرنامج الإحصائي الجاهز (SPSS) ومن أهم الاستنتاجات التي توصل إليها الباحثون هو كفاءة الطريقة المقترحة والتي أدت إلى التمييز بين مجموعتين بشكل أفضل مقارنةً مع قبل معالجة ضوضائية البيانات (الطريقة التقليدية) وهذا يعني أن البيانات تم تصنيفها للطريقة المقترحة بشكل أفضل من الطريقة التقليدية.

1: Introduction:

Multivariate wavelet denoising methodology is used to extract data of distinct characteristics, which are modeled further incorporating the multivariate framework. De-noising becomes a necessary step prior to analysis data. Multivariate wavelet shrinkage methods are particularly well suited for such de-noising works because they can yield a dispersed representation of the data. There are several reasons why multivariate

wavelet shrinkage can be used for parameter estimation. (Stanković et al., 2013), The main reasons are that wavelet shrinkage estimators are Minimax for a wide range of loss functions and for general function classes, practical and fast, adaptable to spatial and frequency in homogeneities, readily extendable to multivariate, applicable to various other problems such as density estimation and inverse problems. And which can be used to treat noise in the data.
On the other hand, Discriminant analysis is used to description of group separation in the dependent variable, in which linear functions of the variables (Discriminant functions) are used to describe or elucidate the differences between two or more groups in the dependent variable. The aims of descriptive Discriminant analysis include identifying the relative contribution of the *p* independent variables to separation of the groups in the dependent variable and finding the optimal plane on which the points can be projected to best illustrate the configuration of the groups. This analysis is also used to allocation (or prediction) of observations to groups in the dependent variable, in which linear or quadratic functions of the variables (classification functions) are employed to assign an individual sampling unit to one of the groups. The measured values in the observation vector for an individual or object are evaluated by the classification functions to find the group to which the individual most likely belongs.

2: Methodology:

This section introduces some concepts about multivariate wavelet (MW) and Discriminant Analysis:

2.1: Multivariate Wavelet Denoising (MWD):

Recently research interests have focused on using the wavelet de-noising techniques in the univariate case, but fewer attentions on the de-noising of multivariate data. The basic procedures of (MWD) method is as follows: (Kaijian et al., 2012)

1- The data into different scales using MW transform. The coefficients would include approximation coefficients as well as horizontal and vertical directions.

2- For approximation and direction coefficients at each direction, the threshold is chosen specifically at different scales for different directions and the MW coefficients are processed by either suppression or shrinkage.

3- Using the de-noised MW coefficients and the scale chosen, the processed wavelet coefficients are reconstructed into the consolidated de-noised data using MW synthesis.

MWD problems deal with models of the form:

$$X_i = A_i + e_i \qquad \cdots \quad (1)$$

Where the observation X is p-dimensional, A is the deterministic signal to be recovered, and e is a spatially correlated noise signal. This kind of model is well suited for situations for which such additive, spatially correlated noise is realistic. In the multivariate setting, given the multivariate variables X, the de-noising algorithm assumes that it consists of both deterministic data D and undesirable noises ω. Applying the MW analysis, (Ahrabian et al., 2015), this relationship is defined as in (2).

$$O \equiv \upsilon X = \upsilon D + \upsilon\omega \qquad \cdots \quad (2)$$

Where υ is an $N \times N$ orthonormal matrix, o is the N dimensional vector of MW transform coefficients $O_l : l = 0, \cdots, N-1$.

The soft thresholding is example of shrinkage rules. After you have determined your threshold, you have to decide how to apply that threshold to your data. The simplest scheme is hard thresholding. Let T denote the threshold and x your data. The soft thresholding is

$$\Gamma(X) = \begin{cases} X - T; & X > T \\ 0 \quad ; & |X| \le T \\ X + T; & X < -T \end{cases} \qquad \cdots \quad (3)$$

Minimax threshold is one of the commonly used thresholds; the minimax threshold is defined as threshold δ which minimizes the Amount (Kaijian et al., 2012):

$$\inf_{\delta} \sup_{\theta} \left\{ \frac{R_\delta(\theta)}{n^{-1} + \min(\theta^2, 1)} \right\} \qquad \cdots \quad (4)$$

Where $R_\delta(\theta) = E(\lambda_\delta(O) - \theta)^2$, $O \sim N(\theta, 1)$

On the other hand it will be explained wavelets used in the search, as follows:

Symlet Wavelets are also known as Daubechies' least-asymmetric wavelets. The symlets are more symmetric than the Extremal phase wavelets (Aminghafaria et al., 2006). N is the number of vanishing moments. These filters are also referred to in the literature by the number of filter taps, which is 2N. And this wavelet is characterized by properties near symmetric, orthogonal and biorthogonal.

Decompose data using multivariate discrete wavelet transformation (MDWT), choose MW and number of decomposition levels and compute

$$MDWT = W.X \qquad \cdots \quad (5)$$

Perform thresholding in the wavelet domain. (Matz et al., 2009), Shrink coefficients by thresholding (hard or soft) and symbolizes it $\hat{x}$, and reconstruct the data from thresholded DWT coefficients, using the following formula:

$$\hat{x} = W^{-1}.\hat{X} \qquad \cdots \quad (6)$$

2.2: Discriminant Analysis:

The objective of the discriminatory analysis is to obtain a model to predict a one qualitative variable from one or more independent variable(s). The dependent variable consists of two groups or classifications, like, normal versus abnormal blood pressure, non defaulting loan versus defaulting, patient versus not sick (for a particular disease) etc. Discriminant analysis depends on equation as linear combination of the independent variables which is discriminatory the best separation between groups in the dependent variable. This equation (linear combination) is known as the Discriminant function. The weights assigned to each independent variable are corrected for the inter-correlations among all the variables. The weights are Discriminant coefficients, (Anderson, 2003).

The Discriminant equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon \qquad \cdots \quad (7)$$

Where Y is a qualitative variable formed by the linear combination as dependent variable, $X_1,\ X_2,\ \ldots,\ X_p$ are the *p* independent variables, β_0, $\beta_1,\ \beta_2,\ \ldots,\ \beta_p$ are the weights (Discriminant coefficients) and ε is the error term (Krieng, 2012). The aim Discriminant analysis is to test if the classification of groups in a variable Y depends on at least one of the independent variables. In terms of hypothesis, it can be written as:

$H_0 : \beta_i = 0,\ for\ \ i = 1,\ 2,\ \ldots,\ p\ \ Versus\ H_1 : \beta_i \neq 0$ for at least one *i.*

Assumptions:

1. The variables $X_1,\ X_2,\ \ldots,\ X_p$ are linearity independent of each other (there is no multi-collinearity).
2. Twice the number of independent variables does not exceed the sample size.
3. Groups are mutually exclusive and group sizes are not much different.
4. The variance-covariance matrices of the independent variables are homogeneous within each group of the dependent variable.
5. Residuals (error term) are randomly distributed.
6. The variables of independent follow a multivariate normal distribution, for purposes of significance testing.

There are several purposes for Discriminant analysis:

1. To investigate differences among groups of dependent variable.
2. To determine the most parsimonious way to distinguish among groups.
3. To discard variables which are little related to distinguish the group from another.

4. To classify cases into groups.
5. To test theory by observing whether cases are classified as predicted.

Some basic concepts of Discriminant analysis:

Linear Discriminant function:

The number of computed functions equals the number of groups in dependent variable minus one. This means that two groups have one function, while three groups have two functions, and so on. Each Discriminant function is a dimension which differentiates a case into groups in the dependent variable based on its values on the independent variables. The linear discrimination function can be based on a random sample as follows (Afifi et al., 2012):

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_p X_p \quad \cdots \quad (8)$$

The Discriminant function coefficients estimated $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$ are partial coefficients that reflect the unique contribution of each independent variable to the classification of the groups in the dependent variable. We then wish to find the vector $\underline{\hat{\beta}}$ that maximizes the square standardized and we get it from the following formula:

$$\underline{\hat{\beta}} = S_{pl}^{-1}(\underline{\overline{X}}_1 - \underline{\overline{X}}_2) \quad \cdots \quad (9)$$

Where S_{pl}^{-1} represent the pooled variance matrix, $\underline{\overline{X}}_1$ and $\underline{\overline{X}}_2$ are mean vectors of the independent variables for the first and second groups, respectively.

Group Centroid:

Group Centroid represents the mean Discriminant scores for each group in the dependent variable for the Discriminant function. The Centroid is a one-dimensional space, one center for each group (Durvaux and Standaert, 2016). By connecting the Centroid a canonical plot can be formed depicting a Discriminant function space and has cut off-point which separate the two groups with the aim of classifying a specific item into the group to which it belongs, which is based on the following formula:

$$C.P. = 1/2(\underline{\overline{X}}_1 - \underline{\overline{X}}_2)' S_{pl}^{-1}(\underline{\overline{X}}_1 - \underline{\overline{X}}_2) \quad \cdots \quad (10)$$

Eigen value:

Eigen value is a ratio between the explained and unexplained variation of the model. For a good model the Eigen value must be greater than one. The bigger the Eigen value, the stronger is the discriminating ability of the function. It can be computed as follows:

$$\lambda = \frac{SSH(z)}{SSE(z)} \quad \cdots \quad (11)$$

Where $SSH(z)$ and $SSE(z)$ are the between and within sums of squares for z which represent Discriminant scores for each group in the dependent variable.

Canonical correlation:
The canonical correlation is a measure of the association between the groups in the dependent variable and the Discriminant function. A high value implies a high level of association between the two and vice-versa (Cagli et al., 2017).

Walk's Lambda:
The Walk's Lambda (Λ) is used to test the significance (importance) of the Discriminant function. Mathematically, it is unexplained variation or one minus the explained variation, and the value ranges from zero to one. When the value lambda for a function is small, the function is significant (unlike the F-statistics in linear regression). It can be computed as follows (Rencher, 2002):

$$\Lambda = \frac{|E|}{|E+H|} \qquad \cdots \quad (12)$$

The (p×p) matrix H has a between sum of squares on the diagonal for each of the p independent variables. Off-diagonal elements are analogous sums of products for each pair of variables. Assuming there are no multi-collinearity in the independent variables, the (p×p) matrix E has a within sum of squares (error) for each independent variable on the diagonal, with analogous sums of products off-diagonal.

Classification matrix:
The classification matrix is a cross tabulation of the observed and predicted memberships. (Rencher, 2002), the values in the diagonal must be high and the values off the diagonal must be close to zero for a good prediction. The results can be conveniently displayed in a classification table or confusion matrix.
The group-1 contains n_1 observations, n_{11} are correctly classified into group-1, and n_{12} are misclassified into group-2, where $n_1 = n_{11} + n_{12}$. Similarly, the group-2 contains n_2 observations, n_{22} are correctly classified into group-2, and n_{21} are misclassified into group-1, where $n_2 = n_{21} + n_{22}$, thus

$$\text{Apparent error rate} = \frac{n_{12} + n_{21}}{n_1 + n_2} \qquad \cdots \quad (13)$$

Similarly, we can compute

Apparent correct classification rate = $\frac{n_{11}+n_{22}}{n_1+n_2}$ $\cdots$ (14)

Box's M:

The Box's M tests the assumption of homogeneity of variance-covariance matrices in the groups. A big Box's M indicated by a small p-value indicates do not provide of this assumption. However, when the sample size is big, Box's M is usually large. The natural logarithm of the variance-covariance matrices for the groups is compared. Box's M value for two groups can be computed as follows:

$$M=\frac{\left|S_1\right|^{(n_1-1)/2}.\left|S_2\right|^{(n_2-1)/2}}{\left|S_{pl}\right|^{\sum_{i=1}^{2}(n_i-1)/2}} \quad \cdots \quad (15)$$

Where S_i is the variance-covariance matrix of the *i*th group.

2.3: Proposed Method:

The proposed method relies on de-noise of the data for all the independent variables together by using MW before performing discriminate analysis. De-noise of data depending on multivariate Symlet wavelet and get the MDWT then involve the thresholding and estimation levels it using Minimax method to obtain a modified multivariate (MMDWT), and taking its inverse as in formula (6), we get on de-noise data (IMMDW) which will be the $\hat{x}$, where $\hat{x}=[\hat{x}_1 \quad \hat{x}_2 \quad \cdots \quad \hat{x}_p]$ and by compensating in the formula (8) we get the following:

$$\hat{\mathrm{Y}}_w=\hat{\beta}_{0w}+\hat{\beta}_{1w}\hat{x}_1+\hat{\beta}_{2w}\hat{x}_2+\cdots+\hat{\beta}_{pw}\hat{x}_p \quad \cdots \quad (16)$$

The Discriminant function coefficients estimated $\hat{\beta}_{0w},\hat{\beta}_{1w},\ldots,\hat{\beta}_{pw}$ are partial coefficients that reflect the unique contribution of each independent variable after de-noise to the classification of the groups in the dependent variable. We then wish to find the vector $\underline{\hat{\beta}}_w$ that maximizes the square standardized and we get it from the following formula:

$$\underline{\hat{\beta}}_w=S_{plw}^{-1}\left(\underline{\overline{\mathrm{X}}}_{1w}-\underline{\overline{\mathrm{X}}}_{2w}\right) \quad \cdots \quad (17)$$

Where S_{plw}^{-1} represent the pooled variance matrix $\hat{x}$, $\underline{\overline{\mathrm{X}}}_{1w}$ and $\underline{\overline{\mathrm{X}}}_{2w}$ are mean vectors of the independent variables after de-noise for the first and second groups, respectively. The following diagram shows the proposed method:

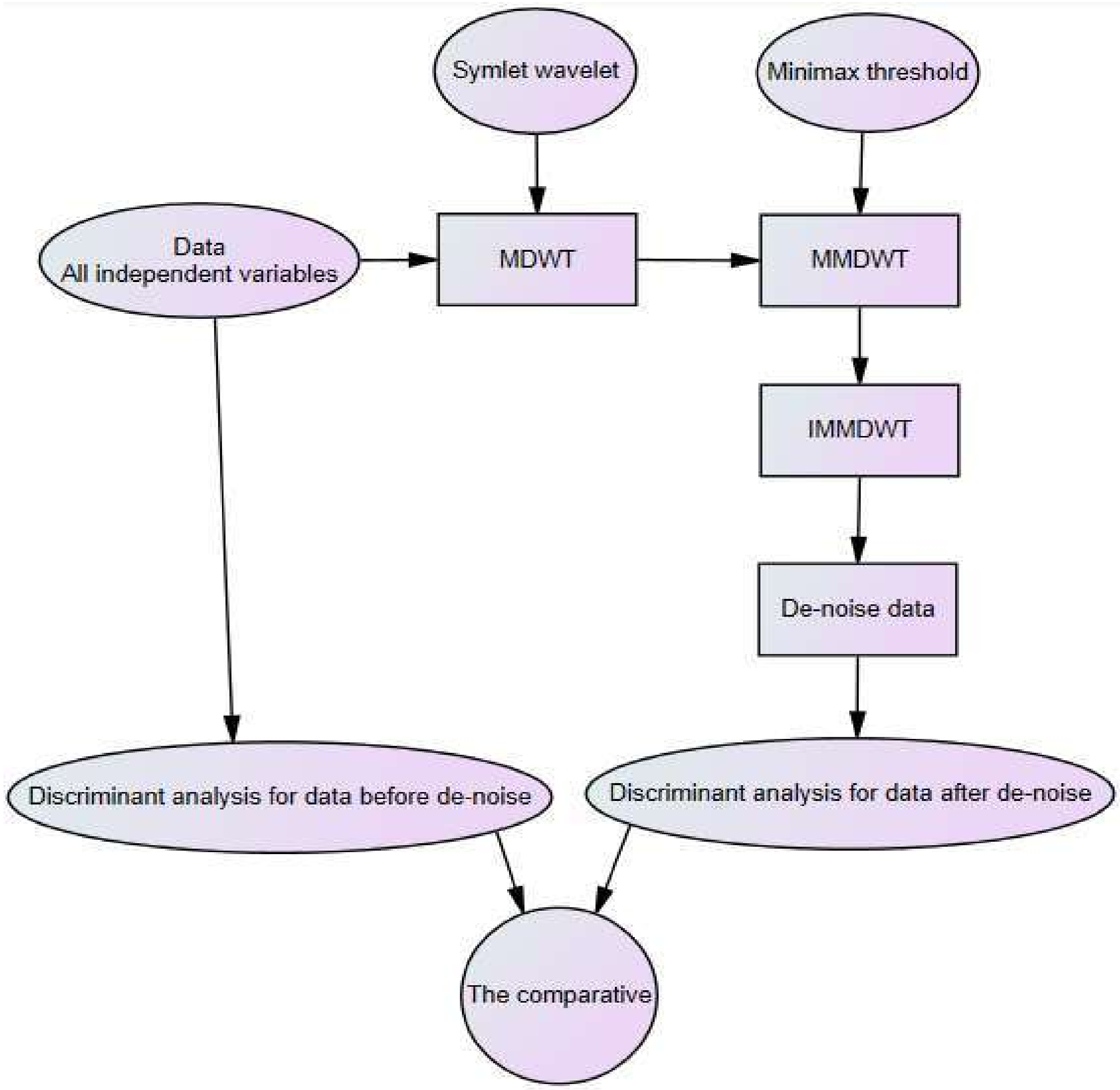

Diagram (1): Proposed method to de-noise of data with discriminate analysis

3. Simulation Study:

In this section, a comprehensive simulation study was conducted to evaluate the performance of the proposed method (depending on diagram-1), which denoted by Prop., and to compare its performance with classical method (Class.) in selection the separation between the two groups (or classification), the independent variables (p = 3, 4 and 5) for sample size (n = 30 and 50) are generated by $x = \mu + T'z$, where T represents Cholesky factorization matrix for variance covariance matrix Σ (for different selected values) and z are multivariate standard normal random numbers, x values are generated for group-1 (y = 0) and for group-2 (y = 1) in which $\mu_1 \neq \mu_2$ and $\Sigma_1 \neq \Sigma_2$, , MATLAB and SPSS were used in data analysis and the original and de-noise data for the first simulation experiment (3-variables for two groups) were illustrated in the following figure:

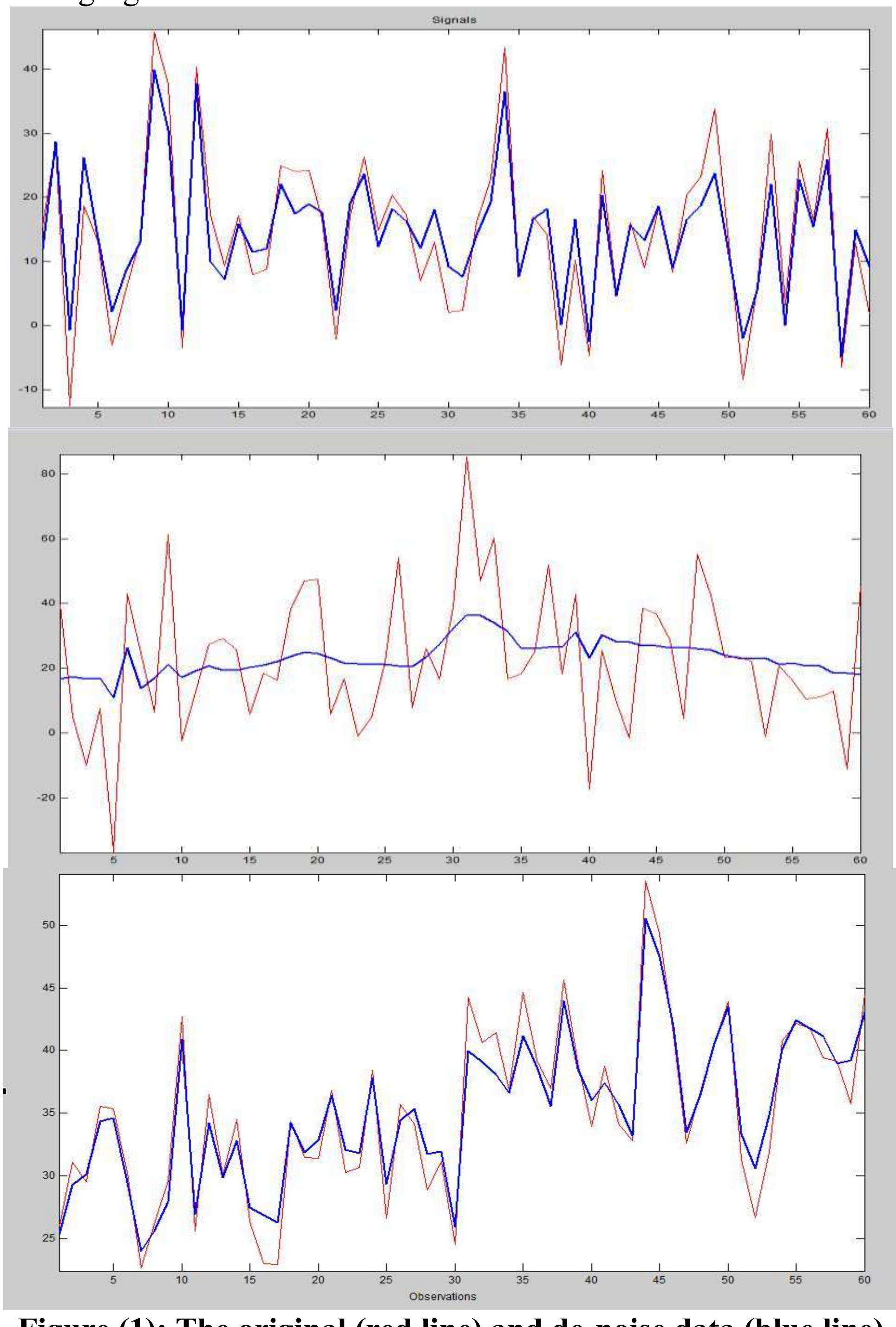

Figure (1): The original (red line) and de-noise data (blue line)

The main results for Discriminant analysis of the first simulation experiment were summarized in the following table:

Table (1): The main results for Discriminant analysis of the first simulation experiment

	No. of independent variables											
	3				4				5			
Sample size	30		50		30		50		30		50	
Method	Class.	Prop.	Class.	Prop.	Class.	Prop.	Class.	Prop.	Class.	Prop.	Class.	Prop.
No. of Sig. variables	1	3	1	1	3	3	2	3	3	3	4	4
Sig. model-F	13.895	29.028	24.95	33.34	32.25	34.74	53.61	86.04	29.05	39.22	68.34	119.04
Std. Error of the Estimate	0.392	0.324	0.383	0.357	0.286	0.278	0.284	0.239	0.274	0.245	0.240	0.190
R^2	0.427	0.609	0.438	0.510	0.701	0.716	0.693	0.784	0.729	0.784	0.784	0.864
Eigen value for function	0.744	1.555	0.780	1.042	2.345	2.526	2.257	3.623	2.690	3.631	3.635	6.332
Canonical Correlation	0.653	0.778	0.662	0.714	0.837	0.846	0.832	0.885	0.854	0.885	0.886	0.929
Wilks' Lambda	0.573	0.391	0.562	0.490	0.299	0.284	0.307	0.216	0.271	0.216	0.216	0.136
Box's M	16.43	4.325	17.55	15.040	33.20	33.17	34.27	38.67	87.73	94.35	140.8	126.07
Classification Results-Original-%	81.7	93.3	83.0	86.0	90.0	93.3	95.0	97.0	91.7	96.7	97.0	99.0
Classification Results-Cross-validated-%	80.0	90.0	82.0	84.0	90.0	93.3	95.0	96.0	90.0	93.3	95.0	98.0

Table (1) shows that the proposed method is better than the classical method of the first simulation experiment, for a combination of these different values of p and n the generated data is repeated 1000 times and calculate the average of Classification results- Original-% and Cross-validated-% as in the following table:

Table (2): Average of Classification results- Original-% and Cross-validated-%

No. of variables	Sample size	Method	Average Classification Results-Original-%	Average Classification Results- Cross-validated-%
3	30	Class.	78.60	76.34
		Prop.	86.26	83.66
	50	Class.	83.00	81.60
		Prop.	86.80	85.80
4	30	Class.	91.66	91.34
		Prop.	95.66	94.34
	50	Class.	94.40	93.20
		Prop.	97.20	96.60
5	30	Class.	96.34	95.34
		Prop.	98.66	96.66
	50	Class.	96.80	95.80
		Prop.	97.60	97.40

Table (2) shows that the average Classification Results- Original-% and Classification Results- Cross-validated-% for proposed method is better than the classical method (for different values of p and n) from generating and repeating data for 1000 times.

4: Application of real data:

The proposed and classical method will be used based on the proposed diagram (1) and the comparison of their efficiency by applying on sample of leukemia patients taken from the Nanakeli Hospital in Erbil. Leukemia is a type of cancer that occurs in tissues responsible for the production of blood cells, including bone marrow and lymphatic system. This type of cancer usually begins in white blood cells. The study sample included two groups where the group-1 consisted of 24 patients with leukemia, the group-2 consisted of 24 patients with non-leukemia patient, which represents the dependent variable (leukemia=0 and non-leukemia=1).The independent variables: Age represents the age of the patient, and Mch represents the Mean corpuscular hemoglobin, Mcv represents the Mean corpuscular volume and Esr represents the Erythrocyte sedimentation rate test. The noise of the data (4-independent variables) will be reduced based on the MW with the soft threshold and estimation levels it using minimax method.

Investigating Multivariate Normality:

The independent variables will be tested whether they have a normal multivariate distribution or not through the use of non parametric test (Kolmogorov-Smirnov) and parametric test χ^2 under the significance level (1%) by using Easy Fit program. The test results for classical method are summarized in the following table:

Table (3): Test of Normality

Variables of study	Classical method						Result
	K.S.			Chi-Squared			
	Statistic	p value	Critical Value	Statistic	p value	Critical Value	
Age	0.1496	0.2108	0.2306	6.3667	0.2722	15.086	Normal
Mch	0.0885	0.8146	0.2306	2.0028	0.5718	11.345	Normal
Mcv	0.1046	0.6317	0.2306	0.7299	0.6942	9.2103	Normal
Esr	0.1336	0.3288	0.2306	2.7132	0.4380	11.345	Normal
Proposed method							
Age	0.0660	0.9759	0.2306	1.0283	0.9055	13.277	Normal
Mch	0.1275	0.3839	0.2306	9.6350	0.0471	13.277	Normal
Mcv	0.1242	0.4154	0.2306	1.5336	0.6746	11.345	Normal
Esr	0.1454	0.2378	0.2306	1.4551	0.6927	11.345	Normal

Through a table (4) note that all study variables for classical and proposed method have a normal distribution because p-values for both tests are less than 1%.

To detect outlier values in the following study model:

$$Y = \beta_0 + \beta_1(Age) + \beta_2(Mch) + \beta_3(Mcv) + \beta_4(Esr) + \varepsilon$$

The values of Mahalanobis Distance were calculated in table (I) and table (II), (in appendix) for classical and proposed method respectively. For 4 degree of freedom and $\alpha = 0.01$, $\chi^2 = 13.28$ there is one value Mahalanobis Distance for proposed method greater than 13.28, so it is considered outlier.

Multiple linear regression:

To test if the classification of groups in a variable Y depends on at least one of the independent variables, we must to test significantly the impact hypothesis, and estimate parameters of multiple linear regression with the some criteria (for classical method) as in the following table:

Table (4): Multiple linear regression for classical method

Model	Coefficients of Regression	t	Sig.	VIF	F	Std. Error of the Estimate
(Constant)	1.878	3.071	0.004		22.302	0.3013
Age	-0.013	-3.937	0.000	1.005	Sig.	R^2
Mch	-0.017	-1.450	0.154	1.073	0.000	0.675
Mcv	0.001	0.147	0.884	1.075		
Esr	-0.016	-8.295	0.000	1.019		

For classical method, the F-test is highly significant (p-value equal to zero), thus we can assume that the model explains a significant amount of the variance with the R^2 = 0 .675, this means that the multiple linear regression explains 67.5% of the variance in the data. Also, the "Std. Error of the Estimate" is the standard deviation of the residuals and measures the efficiency of the estimated model and whenever it's small it indicates the efficiency of the estimated model and it is equal to 0.3013. The information in the table above also allows us to check for multi-collinearity, VIF less than 3 for all variables; this means that there is no multi-collinearity problem. We can also see that only Age and Esr variables have highly significant impact on the dependent variable (because the p-values less than 0.01).

Table (5): Multiple linear regression for proposed method

Model	Coefficients of Regression	t	Sig.	VIF	F	Std. Error of the Estimate
(Constant)	1.626	2.306	0.026		28.955	0.2749
Age	-0.006	-1.568	0.124	1.247	Sig.	R^2
Mch	-0.013	-0.871	0.389	1.135	0.000	0.729
Mcv	0.002	0.222	0.825	1.162		
Esr	-0.022	-8.851	0.000	1.225		

For proposed method, the F-test is highly significant (p-value equal to zero), thus we can assume that the model explains a significant amount of the variance with the R^2 = 0 .729, this means that the multiple linear regression explains 72.9% of the variance in the data and it's greater than 67.5% for classical method. Also the standard deviation of the residuals equal to 0.2749 and it's less than 0.3013 for classical method, this means that this model is better than before. VIF less than 3 for all variables; this means that there is no multi-collinearity problem. We can also see that only Esr variable has highly significant impact on the dependent variable (because the p-values less than 0.01).

Tests of Equality of Group Means:

Table (6): Tests of Equality of Group Means

	Classical method			Proposed method		
	Wilks' Lambda	F	Sig.	Wilks' Lambda	F	Sig.
age	.886	5.939	.019	.776	13.275	.001
Mch	.978	1.040	.313	.988	.558	.459
Mcv	.997	.153	.697	.994	.269	.606
Esr	.453	55.529	.000	.292	111.597	.000

Test of equality of means (classical method), the p-value for Esr variable is less than 0.01. Thus there is a significant difference in Esr between the group-1 and 2, but there is no a significant difference in Age, Mch and Mcv variables between the group-1 and 2.

For proposed method the p-values for both Age and Esr variables are less than 0.01. Thus, there are a significant differences in Age and Esr between the group-1 and 2, but there is no a significant difference in Mch and Mcv variables between the group-1 and 2.

Box's Test of Equality of Covariance Matrices:

Table (7): Log Determinants and Box's Test of Equality of Covariance Matrices

Classical method			Proposed method	
Group	Rank	Log Determinant	Rank	Log Determinant
Group-1	4	17.358	4	14.687
Group-2	4	15.175	4	13.680
Pooled within-groups	4	16.904	4	14.426
Test Results			Test Results	
Box's M	29.320		11.138	
F-Approx.	2.655		1.009	
Sig.	.003		0.433	

For classical method note that p-value for Box's M test is less than 0.01. Thus, inequality of variance-covariance matrix can be assumed. But for proposed method the p-value for Box's M is greater than 0.01. Thus, equality of variance-covariance matrix can be assumed. The log determinant values are quite close for both classical and proposed method. But all log determinants for proposed method less than corresponding to them classical method.

Summary of Canonical Discriminant Functions:

Table (8): Eigen values and Wilks' Lambda

Classical method				Proposed method		
Function	Eigen value	Canonical Correlation		Eigen value	Canonical Correlation	
1	2.075	0.821		2.694	0.854	
Test of Function(s)	Wilks' Lambda	Chi-square	Sig.	Wilks' Lambda	Chi-square	Sig.
1	.325	49.420	.000	0.271	57.489	0.000

There are two groups (classical method). Therefore number of function = 1. The Eigen value is 2.075 (>1). Canonical correlation, $r_c = 0.821(> 0.35)$. Wilks' Lambda = 0.325, p-value = 0.000 (<0.01). Thus, the Function 1 explains the variation well. For proposed method, the Eigen value is 2.694(>1) and its greater than (2.075) for classical method. Therefore, this function is the strongest and has a better discriminating ability than the discriminate function of classical method. Canonical correlation, $r_c = 0.854(> 0.35)$ and its greater than canonical correlation for classical method. Wilks' Lambda = 0.271, p-value = 0.000(<0.01). Thus, the Function 1 explains the variation well and it's greater than explain the variation for classical method.

The function:

Table (9): The function

Classical method			Proposed method	
Function Coefficients		Standardized Function Coefficients	Function Coefficients	Standardized Function Coefficients
age	0.053	0.690	0.028	0.276
Mch	0.073	0.276	0.058	0.165
Mcv	-0.004	-0.028	-0.008	-0.043
Esr	0.067	1.046	0.098	0.945
(Constant)	-5.758		-4.959	

Table (10) for classical and proposed method (respectively) show the Correlation between Esr, age, Mch, Mcv and Y and have the models:

$$Y = -5.758 + 0.053(age) + 0.073(Mch) - 0.004(Mcv) + 0.067(Esr)$$

$$Y = -4.959 + 0.028(age) + 0.058(Mch) - 0.089(Mcv) + 0.198(Esr)$$

Centroid Classification Statistics

Table (10): Centroid Classification Statistics

Classical method			Proposed method		
Classification Function Coefficient	Group		Group		
	Group-1	Group-2	Group-1	Group-2	
age	0.515	0.367	0.684	0.594	
Mch	1.595	1.390	1.894	1.707	
Mcv	1.648	1.660	2.716	2.741	
Esr	0.444	0.256	0.514	0.198	
(Constant)	-115.237	-98.999	-169.677	-153.741	
Functions at Group Centroids			Functions at Group Centroids		
Group	Function		Group	Function	
Group-1	1.410		Group-1	1.607	
Group-2	-1.410		Group-2	-1.607	
Classification Results			Classification Results		
	Predicted Group Membership		Predicted Group Membership		
	Group-1	Group-2	Group-1	Group-2	Total
Original Count	22	2	24	0	24
	2	22	2	22	24
%	91.7	8.3	100	.0	100
	8.3	91.7	8.3	91.7	100

Cross-validated	**Count**	**20**	**4**	**24**	**0**	**24**
		3	**21**	**4**	**20**	**24**
	%	**83.3**	**16.7**	**100**	**0**	**100**
		12.3	**87.5**	**16.7**	**83.3**	**100**

For classical method, 91.7% of original grouped cases correctly classified (8.3% incorrect), 85.4% of cross-validated grouped cases correctly classified (14.6% incorrect) as in the table-I (appendix). Functions at Group Centroids between (-1. 41) and (1. 41), the mid point is zero. For proposed method, 95.85% of original grouped cases correctly classified (4.15% incorrect), 91.65% of cross-validated grouped cases correctly classified (8.35% incorrect) as in the table-II (appendix). Functions at Group Centroids between (-1.607) and (1.607), the mid point is zero. The following figures shows:

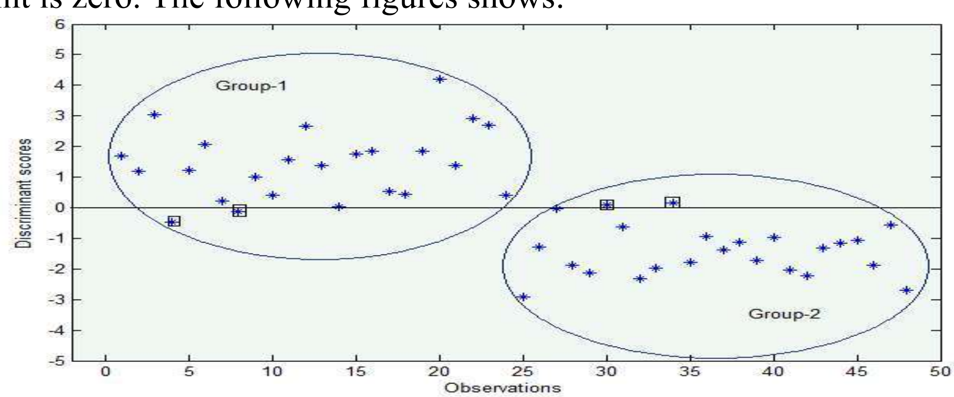

Figure (2): Discriminant Scores for classical method

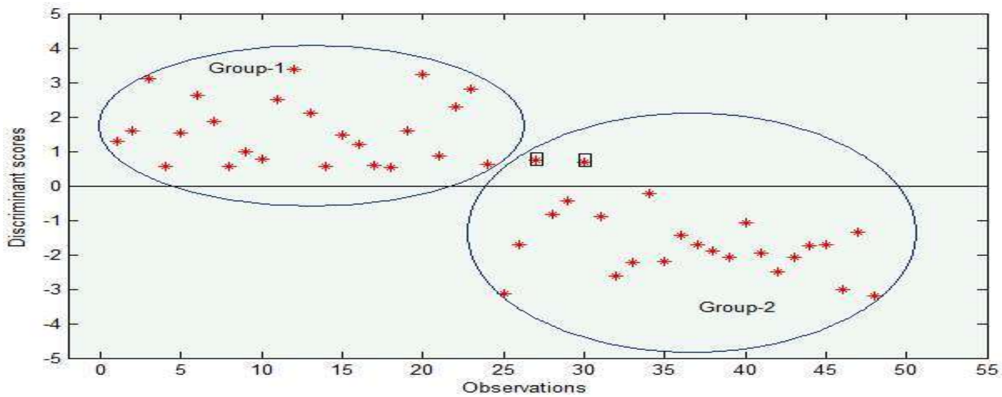

Figure (3): Discriminant Scores for proposed method

Discrimination function for proposed method, classified data is better than classical method. The following figure shows:

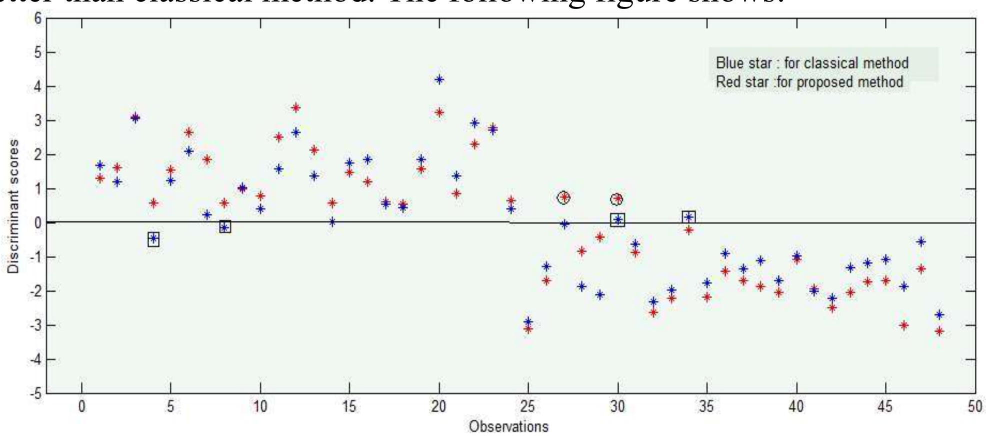

Figure (4): Discriminant Scores for data before and after de-noise

Figure (4) show that Discriminant Scores for proposed method are more separate between the two groups from classical method.

5. Conclusion:

Depending on the simulation study and real data, we conclude the following:

1. The use of multivariate wavelets in reducing the noise of data (de-noise) and then estimate the discriminate function (proposed method) led to the separation between the two groups better than before de-noise (classical method).
2. According to the first conclusion, the simulation and real data were classified for proposed method better than classical method.
3. For simulation study, there are significant differences in the no. of variables between the group-1 and 2 of the proposed method equal to or greater than classical method.
4. The variances, total and general variance of data for proposed method are less than classical method, which leads to more accurate estimates of the classification function and stronger tests.

5. For real data, study variables have a multivariate normal distribution for all data proposed and classical method.
6. For real data, there is one value Mahalanobis Distance for proposed method can be considered outlier.
7. The discriminate function for real data of the proposed method explains the variation well by 85.4% and it's greater than explain the variation for classical method (82.16%).

6. Recommendations:

1. The researcher recommends the use of multivariate wavelet in the treatment of noise data before estimating the discrimination function (proposed method).
2. Use other types of wavelets and compare them to get the best wavelet that treatments the noise of data.
3. Use of other threshold types and other methods for estimating the threshold level with wavelets and comparing them after estimating the discriminating function.

7. References:

Afifi A. A., S. May and V. A. Clark., (2012), Practical Multivariate Analysis. 5th ed. Boca Raton, FL: CRC Press.

Ahrabian A., Looney D., Stankovic L. and Mandic D., (2015), Synchrosqueezing-based time-frequency analysis of multivariate data, Signal Process., vol. 106.

Anderson, T. W. (2003), Introduction to Multivariate Statistical Analysis, 3rd ed. New York: Wiley .

Aminghafaria M., Chezea N. and Poggia J. M., (2006), Multivariate denoising using wavelets and principal component analysis, Computational Statistics & Data Analysis 50, pp. 2381-2398.

Cagli E., Dumas C. and Prouff E. ., (2017), Kernel Discriminant Analysis for Information Extraction in the Presence of Masking, pages 1–22. Springer International Publishing, Cham.

Kaijian He, Kin Keung Lai, and Guocheng Xiang, (2012), Portfolio Value at Risk Estimate for Crude Oil Markets: A Multivariate Wavelet De-noising Approach, Energies,5 , 1018-1043.

Krieng K., (2012), Comparison logistic regression and Discriminant analysis in classification groups for breast cancer, Faculty of Information Technology, Rangsit University, Thailand

Rencher A. C., (2002), Methods of Multivariate Analysis, Second edition. Wiley, New York.

Matz V., Smid R., Starman S. and Kreidl M., (2009), Signal-to-noise ratio enhancement based on wavelet filtering in ultrasonic testing, Ultrasonics 49, pp. 752-759.

Stanković L., Daković M. and Thayaparan T., (2013), Time-Frequency Signal, analysis with Applications. USA: Artech House.
Durvaux F. and Standaert F.X., (2016), From Improved Leakage Detection to the Detection of Points of Interests in Leakage Traces, pages 240–262. Springer Berlin Heidelberg, Berlin, Heidelberg.