International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor (2018): 7.426

Classification Data of Binary Data Respond to the Diagnosis of Lymphocytic Leukemia using Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA)

Fatma A. Abdelati Hanaa T. Elgohari Ali Y .AL-Zubaidi

Department of Applied Statistics, College of Faculty of Commerce, University of Mansoura, Mansoura, Egypt

Abstract: ***In this research, the process of distinguishing or categorizing statistical data was studied by using the dual-response Support vector machine and Linear discriminating function , by based on the correct classification, The two methods were then applied to the practical side and to real data of lymphocytic leukemia in Iraq , comparison was also made between the two methods used in the study. The study found that the method or method of the supporting vector machine was the best in the classification whether using the real data in the practical application side***

Keywords: lymphocytic leukemia, classification, SVM, LDA

1. Introduction

The single classification or a new view of one of the groups under study can be done according to the method of classification of the vocabulary to their original communities according to a set of statistical methods. For the implementation of the statistical classification mechanism, there are a number of methods, including the Support Vector Machine (SVM) method and the linear differential analysis method. Both methods focus on the concept of correct classification of new observations with the least possible classification error. These are the two main methods of statistical classification, which focus on a dual-response dependent variable to express two states, or these two methods can be used with a dependent variable to respond to multiple situations**t**he problem of our current study lies **in** the difficulty of classifying the original data to their societies in the natural state and also the difficulty in determining the relative importance of the variables of the phenomenon studied.

2. Experimental

2.1 The concept of lymphocytic leukemia:- [2] [17]

Global statistics show the incidence of millions of people in various types of cancers, including lymphoid leukemia, which is one of the most serious types of lymphoma, which is described in the category of malignant diseases, but to a low degree. This disease is different from other types of tumors, because the bone marrow is the one that generates the cancerous cells that are transmitted to the bloodstream of the infected person and appear in blood directly, and these cells are missing vital activities. Medical studies indicate that people with lymphoma can resist the disease for a period of 5-15 years, accompanied by health problems that have emerged in recent years. The disease causes the accumulation of fibroblasts in the bone marrow and the lymph nodes constantly, and the transfer of these cells to the organs of the manufacture of blood leads to accelerating the death of cells.

2.2 Diagnosis of chronic lymphocytic leukemia[2][18][19]

Laboratory tests: - Through laboratory analysis of whole blood cells (CBC), which is a laboratory analysis common in medical examinations, this laboratory analysis focuses on the main components of blood, such as red blood cells, white blood cells, and platelets Hemoglobin and others, the increase or decrease in the numbers of these components indicates a specific condition.

Second: Clinical tests: To diagnose Lymphoma, a clinical examination such as an enzyme test (LDH), which has a set of features that enable it to participate in the metabolism and the release of energy in the cells of the body. This enzyme is also characterized by its presence in almost all tissues of the body. This enzyme is highly present in the cells of the heart, liver, kidneys, lungs, brain, red blood cells and skeletal muscles. This enzyme takes five forms:(LDH_1) - Mainly found in Red blood cells , Kidney, and heart muscle
(LDH_2)- Mainly found in Red blood cells , Kidney, and heart muscle.
(LDH_3)- Mainly found in Spleen, lungs, and pancreas.
(LDH_4)- Mainly found in the liver, kidneys and skeletal muscles.
(LDH_5)- Mainly found in skeletal muscle and liver .

2.3 Linear discriminating function [5][7]

The linear discrimination function is suitable for societies that follow a multiple natural distribution that has vectors from the different computational classes μ_i and the array of variance and equal contrast. The linear discrimination function can be applied in two cases of differentiation between groups. The first case distinguishes between two groups only. Two groups. But in our current study the focus will be on the first case that the vocabulary of our study community is divided into only two groups, the first group

(representing the presence of lymphocytic leukemia) and the second group (absence of lymphocytic leukemia).[5][7]

2.4 Discrimination or classification in the case of two groups:[7][10][14]

If two entities are randomly drawn from two societies that follow the natural distribution and possess different means, μ1, and μ2, and a matrix of variation and equal common variation for both communities:

$$X_1 \sim N(\mu 1, \Sigma)$$
$$X_2 \sim N(\mu 2, \Sigma)$$

Through the above data, a function with high flexibility can be structured to determine the viewing and classification of the original community to which it belongs. This idea was developed in 1921 by the researcher "Karl person", which depends on the idea of measuring the statistical distance between the two selected samples. This measure is called the coefficient of relative similarity Coefficient Racial likeness, which symbolizes It has a C.R.L which is calculated as in the formula shown below.

$$C.R.L = \frac{1}{p}\left[\sum_{i=1}^{p}\left(\frac{di^2}{\frac{S_{1i}^2}{n_1}+\frac{S_{2i}^2}{n_2}}\right)\right]$$

And it can be inserted into a set of steps to summarize the process of excellence

1) Calculate the mean of the variables of each group separately, then calculate the difference between the averages of each variable in the two groups. The average in the first group is

$$\overline{X}_{(1)} = \sum_{i=1}^{n1} Xi / n1$$

The average in the second group is

$$\overline{X}_{(2)} = \sum_{j=1}^{n2} Xj / n2$$

But the difference between the two groups is short di he is

$$di = \overline{X}_{(1)} - \overline{X}_{(2)}$$

2) Calculate the sum of the squares of each variable in each group, as well as the sum of the result of multiplying each variable within each group

$$Sii = \sum Xi^2 - (\sum Xi)^2 / n$$

$$Sij = \sum Xi.Xj - (\sum Xi)(\sum Xj) / n$$

3) Calculation of the built-in contrast and contrast (within totals)

$$Vii = \frac{Sii_{(1)} + Sii_{(2)}}{n_1 + n_2 - 2}$$

$$Vij = \frac{Sij_{(1)} + Sij_{(2)}}{n_1 + n_2 - 2}$$

4) Therefore, the mathematical model of the function of discrimination is: -

$$Y = a'X$$

whereas *, When compensation is worth In the above example we get the following formula:

$$Y = (\overline{X}_{\cdot} - \overline{X}_{\cdot})' S^{-1} X$$

2.5 Support vector machine (SVM)[1][3][6][8][13]

2.5.1Support vector machine for classification

This method is one of the most important methods of machine learning, which was suggested by the researcher (vapnik) in (1992), the summary of this method is to build a learning algorithm by supervisor or supervisor (supervised), and the basic idea in the work of this technique depends On the theory of statistical learning (Statistical Learning Theory)]. The discovery of the technology of the vector support machine is to find the best solution to the problem of pattern recognition by selecting the level of data separation, this technique is centered around the main objective is to find the level of separation and optimization of the data to be classified and separated into two categories.

2.5.2 Linear Support Vector Machine [3][11][12][16]

Suppose that we have L of the points and each input value of xi has a given dimension D (dimensionality) and that the variable yi is equal to either +1 or -1. The data is defined as training data according to the following

{xi, yi} where i = 1, ... L, yi∈ {-1, + 1}, x ∈ RD

Below I assume that the available data is separated by a linear line. A line of x1 against x2 can be plotted with the possibility of separating available data into two or two categories according to the following cases. If D = 2, the level of the separator on the graph is x1, ...xD, but where D> 2 can be found the level of the separator (Hyperplane) depending on the following formulas:

w'xi + b = 0

Points near the level of the Hyperplane are supportive vectors

Their values range from +1 to -1, if +1 is for the first group and 1 is for the second group

The support vector machines (SVM) have a basic goal of making the Hyperplane a far cry from the points of the two categories and the process of classification of observations are based on the following formulas:

w'xi + b +1 +1 for yi = +1

w'xi + b -1 -1 for yi = -1

3. Results and Discussion

The data under study were collected by studying the stools of patients with and without chronic lymphocytic leukemia from the Diwaniyah Teaching Hospital and the Oncology Consultation Unit. The sample size was 127 views and both sexes were divided into two groups described by following:-

1) Group I: People with a disease Chronic lymphocytic leukemia (77). The symbol (1) was the expression of the first group.
2) The second group: - Non- infected people Disease Chronic lymphocytic leukemia 50) View The symbol (2) was for the expression of the second set .

3.1 Description of the variables used in the research:

There are a number of factors that affect chronic lymphoblastic leukemia as mentioned in most medical studies in addition to the proposal of a set of variables affecting Y (variable response): A variable is a binary response that takes the value (1) if a person has lymphocytic leukemia and takes value (2) if the person is not infected with lymphocytic leukemia.

The independent variables are:
X1 - is a sex variable where (1) returns to males and (2) returns to females.
X2: - is a variable of age.
X3 - is a variant of red blood cells (RBC).
X4: - A variant of white blood cells (WBC).
X5: Hemoglobin Blood (HGB) is a special variable.
X6: - is a special variable of hematocrit HCT (Haematocrit
X7: - is a special variable in platelets (Blood Platelets).
X8: Lactate Dehydrogenates (LDH).
X9: - is a special variant of ESR (Erythrocyte Sedimentation Rate)
Applied

3.2 Analysis of the results using the Linear Discriminate function

The test is equal to the averages of the two groups

The results, which represent the F test obtained by analyzing the data, show that there is a significant significance for most of the research variables as shown in the following table**:**

Table 1: A test table is equal to the averages of the two groups

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
gen	.987	1.685	1	125	.197
age	.783	34.604	1	125	.000
WBC	.940	8.041	1	125	.005
RBC	.739	44.037	1	125	.000
HGB	945	7.284	1	125	.008
HCT	.935	8.717	1	125	.004
PLT	.992	1.005	1	125	318
LDH	.594	85.397	1	125	.000
ESR	.519	115.818	1	125	.000

Hence, the null hypothesis that there is no significant difference between the two groups and the acceptance of the alternative hypothesis is rejected. This means that there are significant differences between the two groups. The linear characteristic function has a high ability to excel.

3.3Analysis of the results using the Linear Discriminate function:

To obtain the results of the linear discriminate function, data on the variables that affect lymphocytic leukemia were introduced into the ready-made statistical program SPSS-v20 During the test results, X^2) chi-square , Shows that this function has a preference in analyzing the data of the phenomenon under study and this is evident by the value of X^2 Which amounted to $X^2 = 140.250$) As shown in the table below

Table 2: Displays the value of its test statistic for the linear discrimination function

Test of Function (s)	Chi-square	Df	Sig.
1	140.250	9	.000

Through the results listed in the table above we find that the value of very little moral (Sig = 0) Compared With the level of significance (0.01) , And through this result we will accept the alternative hypothesis H_1 , Wen reject the null hypothesis H_o , So we conclude that the function of excellence has a good ability to classify views to the original community. This function is very appropriate in analyzing the data of the phenomenon under study. After the conclusion of the preference of the function of linear excellence, it is necessary to achieve all the requirements and to find the requirements based on the following formula.
$Y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \ldots\ldots\ldots \alpha_r x_r$
Or using this formula: -

$$Y = \quad X \text{ whereas}$$

$$\acute{d} = (\bar{X}_1 - \bar{X}_2)' S^{-1}$$

Then subtract the first sum of the second group we get a set of transactions as shown in the following table:

Table 3: Shows the coefficients of the linear discriminant function

Variables	1.00	2.00	Total
Sex	6.321	7.989	-1.668
Age	0.228	0.168	0.06
RBC	-0.053	-0.112	0.059
WBC	2.118	3.133	-1.015
HGB	0.248	0.277	-0.009
HCT	0.553	0.561	-0.008
PLT	0.017	0.022	-0.005
LDH	0.069	0.024	0.045
ESR	0.209	0.084	0.125

By compensating the coefficients described above in the linear function equation , we obtain the following:
$Y = -1.668X_1 + 0.06X_2 + 0.059X_3 - 1.015X_4 - 0.009X_5 - 0.008X_6 - 0.005X_7 + 0.045X_8 + 0.125X_9$

After finding the coefficients of the linear function , the separation point is calculated as a second step , which is used to classify the observations into their original collections as shown in the formula below
$\mathbf{Z} = \frac{\bar{y}_1 + \bar{y}_2}{2}$
In the above formula , we need a (Y1,Y2) Which are calculated by the following formulas:
$\bar{y}_1 = (\bar{X}_1 - \bar{X}_2)' S^{-1} \bar{X}_1$
$\bar{y}_2 = (\bar{X}_1 - \bar{X}_2)' S^{-1} \bar{X}_2$
After finding the values of($\bar{y}_2$،$\bar{y}_1$) Which was equal to$\bar{y}_1 = -12.66$*and* $\bar{y}_2 = -16.498$*and compensate them in chapter point equation getZ= -14.579After getting this point may be classified according to the following rule: new views:-*
New viewing x Return Of the group the first if It was
New viewing x Returns to the second group if it is

Table 4: Shows the wrong and correct classification of the discriminatory function

Yhat	set 1	set 2	Sum
set 1	75	2	77
set 2	4	46	50
Sum	Games	48	127

Preparation of the researcher based on the results of the program (R- language)

Through the results listed in the table above, which are related to the computation of the calculated classification Through the way the Linear Discriminate function (LDA) And for (127) of the patients with lymphoma and non – infected leukemia, The first group (people with lymphocytic leukemia) was classified as (77) patients, according to the following classification base:

Classification (75) to the first patient group (people infected with lymphatic leukemia) out of 77 patients, with the proportion of discrimination. Correct up to 97% machine mode according to the vector booster (LDA) (2) patients in the first group (people with lymphocytic leukemia) They were classified into the second group (people who did not have lymphocytic leukemia).

As for the second group (people who do not have lymphatic leukemia) 50) Were classified according to the following classification:

Classification (46) to the second patient group (non- people were infected by lymphatic leukemia) out of (50) patients, with the proportion of discrimination Correct up to 92%, according to linear function either discriminatory (4) patients remaining from the second group (non - persons infected with disease lymphatic leukemia) They were classified into the first group (people with lymphocytic leukemia) .

The percentage of total classification by method Special linear and special function The two groups reached (96%).

The importance of each variable

An important step in the qualitative analysis is to determine the relative importance of each variable by using the formula shown below.

$$*X_i = \alpha \sqrt{Vii}$$

And compensation for the values of $*$ *Multiplied by the values of the root of the variation we will get the following:*

$$X_1^* = -1.668\sqrt{0.249} = -0.83233$$
$$X_2^* = 0.06\sqrt{149.622} = 0.73392$$
$$X_3^* = 0.059\sqrt{28.58} = 0.31542$$
$$X_4^* = -1.015\ \sqrt{2.46} = -1.591965$$
$$X_5^* = -0.009\ \sqrt{1132.861} = -0.302922$$
$$X_6^* = -0.008\sqrt{239.881} = -0.123905$$
$$X_7^* = -0.005\sqrt{11025.95} = -0.525023$$
$$X_8^* = 0.045\sqrt{2744.801} = 2.35759$$
$$X_9^* = 0.125\sqrt{297.388} = 2.15562$$

By calculating the relative importance of each variable can be arranged regardless of the signal

Table 5: Shows the degree of importance of each variable

The degree of importance of each variable	Variable	Description of the variable
1	X_8^*	Rate LDH
2	X_9^*	Rate ESR
3	X_4^*	Rate WBC
4	X_1^*	Sex
5	X_2^*	New
6	X_7^*	Rate PLT
7	X_3^*	Rate RBC
8	X_5^*	Rate HGB
9	X_6^*	Rate HCT

From the results shown in the table above, we find the following:

That the variable (X8: Lactate Dehydrogenates) (LDH) ranked first in the impact on the diagnosis of lymphoid leukemia with an effective value of () The second rank belongs to variable (X9: speed of blood deposition) (Erythrocyte Sedimentation Rate) (ESR) in influencing the diagnosis of lymphocytic leukemia with an effective value in the language (), and the third rank is related to the variable (White Blood Cells) (WBC), in influencing the diagnosis of lymphocytic leukemia (X1: Sex) is ranked fourth in influencing the diagnosis of lymphoid leukemia with an effective value. (X2: age) in influencing the diagnosis of lymphoid leukemia with a value of (), while the sixth rank is due to the variable (X7: blood platelets) in PLT. (X3: Red Blood Cells) RBC ranked seventh in influencing the diagnosis of lymphocytic leukemia with an influence value of (). The variable (X5: Hemoglobin Blood (HGB) is ranked eighth in influencing the diagnosis of lymphocytic leukemia with an (B) It grew variable (X6: - represents the percentage of hematocrit ((Haematocrit HCT) is ranked ninth and last in influencing the diagnosis of lymphatic leukemia in a language worth of splash (). The order of relative importance of the variables affecting the diagnosis of lymphatic leukemia can be summarized by the diagram below.

3.4 Support Vector Machine (SVM) results

To obtain the results of the SVM method, an algorithm was constructed from the code within the statistical program (R) to obtain the following results.

3.5 Results of vector support machine Support Vector Machine (SVM)

For method results SVM , An algorithm was built from the codes within the statistical program (language R) For the following results.

Table 6: Showing the wrong and correct classification of the mode of the support vector machine "SVM"

Yhat	set 1	set 2	Sum
set 1	75	2	77
set 2	2	48	50
Sum	77	50	127

Preparation of the researcher based on the results of the program (R- language)

Through the results listed in the table above, which are related to the computation of the calculated classification Through the way the vector machine supports Support Vector Machine (SVM) And for (127) of the patients with lymphoma and non - infected leukemia, The first group (people with lymphocytic leukemia) was classified as (77) patients, according to the following classification base:

Classification (75) to the first patient group (people infected with lymphatic leukemia) out of 77 patients, with the proportion of discrimination · Correct up to 97% machine mode according to the vector booster (SVM) (2) patients in the first group (people with lymphocytic leukemia) They were classified into the second group (people who did not have lymphocytic leukemia) .

As for the second group (people who do not have lymphatic leukemia) 50) Were classified according to the following classification:

Classification (48) to the second patient group (non infected people disease lymphatic leukemia) out of (50) patients, with the proportion of discrimination. Correct up to 96% machine mode according to the vector booster (SVM) (2) patients who remained in the second group (people who do not havelymphatic leukemia) They were classified into the first group (people with lymphocytic leukemia).

The overall rating ratio is according to the method of the supporting vector machine (SVM) And special The two groups reached (96%).

3.6 Calculate the correct classification ratios:

The correct classification of the first group (people with lymphatic leukemia) can be found in the following:
75 / (75 + 2) * 100 = 97%

But the correct classification rate for the second group (people not suffering from lymphatic leukemia) is calculated by the following:
48 / (48 + 2) * 100 = 96%

The correct overall classification ratio for the two groups can be found together
(77 + 48) / 127 * 100 = 98%
In contrast to the correct classification is the wrong classification which is defined as follows: is the classification of the new viewing to a particular group, but in fact the return of this view to the other type.

3.7 Finding supporting vectors Support Vectors

3.7.1Supporting vectors in group I Persons with lymphocytic leukemia

Through the following target function Y = w * x + b variable values can be determined which are the supporting vectors, as shown below.

Table 7: Determine the values of the supporting vectors from the first set according to the supporting vector machine "SVM"

The supporting vectors of the first set									
obvs	gen	Age	WBC	RBC	HGB	HCT	PLT	LDH	ESR
1	1.1038117	1.81980036	-0.39607091	-0.8311213	-0.440994	0.33033378	-0.75582046	-0.5614973	0.647000516
2	1.1038117	2.22856344	-0.22772125	-0.3593359	2.3191225	-2.1277462	0.38698745	-0.5614973	-0.10684412
10	1.1038117	0.18474804	-0.24455622	0.25270988	-0.259759	1.18260075	-0.90819485	1.02274886	-0.22282022
12	1.1038117	0.10299542	-0.128582019	-0.02143566	-0.4172261	0.38198632	0.26318326	-0.37062435	0.647000516
13	1.1038117	0.51175850	-0.111747053	-0.75461561	-0.4974448	-0.21201793	1.55836556	0.18290747	0.415048320
16	1.1038117	0.18474804	2.389180577	-0.36571146	2.7647818	-2.12406597	2.45356509	2.75969355	0.647000516
17	1.1038117	0.67526374	1.248144036	-0.28283025	-0.4766473	0.12372360	2.73926707	2.01528868	1.284869056
19	1.1038117	0.34825327	-0.614925462	-1.00963472	2.2002800	-2.12994145	-2.03005128	1.69080451	-0.048856073
25	1.1038117	-0.14226243	-0.375494844	-0.43584172	-0.3875154	0.49820454	-0.56535247	-0.40879896	0.241084173
30	1.1038117	-0.79628336	-0.263261742	-0.25732834	-0.3696891	0.61442277	0.14890247	-0.37062435	-0.106844122
32	-0.8988181	-0.63277812	-0.766440150	-1.00325924	-0.5099232	-0.21201793	-1.27008402	-0.56149739	-0.164832171
33	-0.8988181	-0.46927289	-0.347436568	-0.58885319	2.4676756	-2.12806904	-0.09870591	-0.56149739	0.009131976
37	-0.8988181	-0.06050981	0.337185356	0.01681721	-0.4261393	0.55631366	-0.38440789	-0.48514817	0.183096123
39	-0.8988181	-0.63277812	0.380208045	0.08694746	-0.3815733	0.65316218	-0.31774409	-0.19883861	0.009131976
40	-0.8988181	-0.38752028	-0.324989948	-0.14256974	-0.3964286	0.58859650	-0.42250148	-0.08431478	-0.106844122
41	-0.8988181	-0.79628336	-0.592478842	-0.76736656	-0.4915026	-0.08288657	1.35837418	-0.27518783	0.878952712
43	-0.8988181	1.41103728	0.892739212	-1.09889141	2.0487558	-2.13039341	-0.56535247	2.93147929	3.778355167
49	-0.8988181	-0.30576766	-0.272614500	0.22720797	-0.3875154	0.59505306	1.39646777	0.08747095	-0.164832171
50	-0.8988181	1.00227420	0.082790324	-0.04693757	-0.3399784	0.70481472	-0.18441650	-0.56149739	-0.164832171
51	-0.8988181	0.43000589	0.387690252	0.47585161	-0.2983836	-2.11573700	-2.20909119	-0.48514817	-0.222820220
52	-0.8988181	-0.30576766	-0.020090020	-1.14989523	1.9923056	-2.13000601	0.47269805	-0.46606087	0.067120025
53	-0.8988181	1.90155298	8.468473619	-1.09889141	2.2002800	-2.12910209	-1.77291950	-0.21792591	1.980725645

54	-0.8988181	2.22856344	0.372725838	-0.53147389	-0.4499078	0.27222467	1.48217837	0.88913773	2.328653940
64	-0.8988181	1.81980036	0.069696462	0.20808154	-0.3191810	0.93079460	-1.35579461	-0.19883861	3.198474676
67	-0.8988181	2.22856344	0.174447357	0.63523855	-0.2538176	1.47960288	0.14890247	0.43104243	0.820964663
68	-0.8988181	1.32928466	-0.180957467	0.45672517	-0.5339888	0.20765899	1.02505520	0.24016939	0.820964663
69	-0.8988181	1.90155298	0.080919772	1.09427295	-0.3578048	0.94370773	-1.46055200	0.79370121	-0.048856073
74	-0.8988181	-0.71453074	0.133295220	-0.97775733	2.1438298	-2.12981232	0.80601702	0.14473287	0.009131976
75	-0.8988181	-0.38752028	-0.826297805	-1.25190288	-0.4974448	-0.14745225	-1.40341161	0.37378052	0.415048320
127	1.1038117	-0.30576766	-0.255779535	-0.32745860	-0.5687503	-0.58004231	-1.10818623	-0.84780695	-0.744712662

Preparation of the researcher based on the results of the program (R- language) Of the results listed in the table above, we find that the number of vectors supporting special Ba first group (people infected with lymphatic leukemia) are (30) heading supportive. The first column of the table above shows the set of views for supporting vectors Support Vectors When using method Support Vector Machine (SVM) Which takes views (1,2,10, ..., 127) , The sequence of columns from the second to the ninth are the values of the variables for those views, for example, the value (10), which returns to the first column represents the third view , and find that the value in the second column was equal (1.1038117) It dates back to the first variable (gen) , And their value in the third column was equal (-0.24455622) It dates back to the third variable(WBC) So for the rest of the variables.

According to the codes used in the statistical program R For the implementation of the vector method, the original values are converted to the standard or standard formula Stander) Which is described by subtracting all of the available values from their arithmetic mean and divide the deviation Ha standard

Supporting vectors in the second group (persons who do not have lymphocytic leukemia)

The table below represents the observations of the supporting vectors in the second group as follows:

Table 8: Views supporting the second set of vectors according to a for method of supporting vector machine "SVM"

The supporting vectors of the second group									
obvs	**Gen**	**age**	**WBC**	**RBC**	**HGB**	**HCT**	**PLT**	**LDH**	**ESR**
77	-0.8988181	-0.30576766	-0.644854289	-0.69086083	-0.4499078	0.04624479	-0.52725888	0.28925669	0.183096123
78	1.1038117	-0.38752028	-0.637372083	-1.31565765	-0.5212133	-0.17327852	-1.07961603	0.14473287	0.125108074
Games	-0.8988181	1.00227420	3.324456429	0.06782103	-0.3696891	0.61442277	1.92977813	0.06838365	0.531024418
80	-0.8988181	0.67526374	-0.659818703	-0.23182643	-0.3340363	0.75646726	-0.49868868	-0.84780695	-0.280808269
81	-0.8988181	0.26650065	-0.715935254	-0.14894522	-0.4588209	0.20120242	0.88220421	-0.84780695	-0.280808269
82	-0.8988181	-0.63277812	-0.852485529	-0.76099109	-0.4796184	0.02687508	-1.33674781	-0.84780695	-0.454772416
86	1.1038117	-1.04154120	-0.826297805	-0.79924395	2.3785437	-2.12858557	-2.30146816	-0.84780695	-1.672521447
87	1.1038117	-0.30576766	-0.946013114	-1.09889141	-0.4974448	-0.14745225	-0.39393129	-0.84780695	-0.628736564
88	-0.8988181	1.90155298	-0.783275116	-0.39121337	-0.4796184	0.18828928	2.30119070	-0.84780695	-0.570748515
89	1.1038117	-0.14226243	-0.882414356	-0.44859267	-0.4974448	0.11726703	-1.42245841	-0.84780695	-0.744712662
90	-0.8988181	-0.55102551	-0.678524220	-0.01506018	-0.3786023	0.58213993	-1.07961603	-0.84780695	-0.918676809
92	-0.8988181	-0.79628336	-0.657948151	-0.13619426	2.9430456	-2.12503446	0.53936184	-0.84780695	-1.672521447
100	-0.8988181	-0.79628336	0.668273008	4.16725325	-0.3340363	0.87914205	0.50126824	-0.84780695	-0.860688760
101	-0.8988181	-1.61380952	0.138906875	2.25460991	-0.3667180	0.77583697	-0.49868868	-0.84780695	-0.280808269
102	1.1038117	-0.30576766	-0.497080705	0.08057198	-0.3934576	0.38844289	-1.30817762	-0.84780695	-0.744712662
108	-0.8988181	-0.95978859	-0.006996158	1.75094717	-0.4023708	0.51111768	-0.61296947	-0.84780695	-0.280808269
113	-0.8988181	0.18474804	-0.220239052	1.04964461	-0.3340363	0.84040265	3.09163284	-0.84780695	-0.860688760
115	-0.8988181	-1.12329382	-0.676653668	-0.23820191	-0.5185393	-0.30886645	2.35833110	-0.84780695	-1.672521447
118	1.1038117	-0.30576766	-0.218368501	1.05602009	-0.3072968	1.29236241	-0.62249287	-0.84780695	-0.280808269
119	-0.8988181	1.57454251	-0.502692360	0.08694746	-0.3459206	0.63379247	-0.35583769	-0.84780695	-0.280808269
120	1.1038117	-0.38752028	-0.222109604	1.04326913	-0.5137856	0.10435390	0.04414508	-0.84780695	-0.280808269
121	1.1038117	1.00227420	-0.534491739	-0.04056209	-0.4231682	0.15600644	0.72982983	-0.84780695	-0.744712662
124	1.1038117	-1.53205690	-0.629889876	-0.34658503	-0.5687503	-0.65752112	2.10119932	-0.84780695	-0.280808269
125	1.1038117	-1.12329382	-0.719676358	-0.67173440	-0.3904865	0.40135603	1.19647639	-0.84780695	-0.280808269
126	1.1038117	-1.04154120	-0.753346288	-1.30928218	-0.4974448	0.51111768	-0.79391405	-0.84780695	-0.860688760

Preparation of the researcher based on the results of the program (R- language). Through the results shown in the table above, we find that the number of support vectors for the second set to be (25) heading supportive , the first column shown in the table above due to the views of the special e configured supporting Vectors Support Vectors Through the use of method Support Vector Machine (SVM) Which starts from viewing (77,78,79, ..., 126) The rest of the columns starting from the second to the ninth are the values of the variables for those views .

According to the data, the total number of supporting vectors was 55.

3.8 Rating Views

Classification of group 1 observations (people with lymphocytic leukemia)

Through the following function or function y = w * x + b It can be classified me in views to their original collections, since the results of the target function described above represents the patient classification to any group

belongs. The patient may be classified into a group of people with lymphocytic leukemia or is classified into the group of people who do not have lymphocytic leukemia as shown in the table below:

Table 9: Classification of the first group views according to the method of the vector machine SVM

Observations	Value	observations	Value
1	1.00053966	**40**	0.67254893
2	1.00016826	**41**	1.00000013
3	1.37995132	**42**	1.71447915
4	1.06435987	**43**	1.00008752
5	1.37985413	**44**	1.35662500
6	1.10890650	**45**	1.68702170
7	1.65758168	**46**	1.27162155
8	1.61864687	**47**	1.04134907
9	1.68205203	**48**	1.43779617
10	1.00001253	**49**	0.40084152
11	1.87057001	**50**	0.39379417
12	0.90729845	**51**	1.00011148
13	0.97166174	**52**	0.99971686
14	1.36233255	**53**	1.00010634
15th	1.27823659	**54**	0.99960069
16	0.99963109	**55**	1.29949124
17	0.99968854	**56**	1.03791483
18	1.33671569	**57**	1.54634198
19	1.00006744	**58**	1.69518191
20	1.20221930	**59**	1.34357025
21	1.17188073	**60**	1.50382606
22	1.09301987	**61**	1.58764758
23	1.48684328	**62**	1.43666738
24	1.44736059	**63**	1.33396827
25	0.42860135	**64**	0.99991495
26	1.18998768	**65**	1.62632486
27	1.15819186	**66**	1.72349874
28	1.07429687	**67**	1.00040922
29	1.19642720	**68**	1.00009159
30	**-0.06145279**	**69**	0.99961799
31	1.18590617	**70**	2.03416244
32	0.11734183	**71**	1.40448619
33	0.81558884	**72**	1.33622020
34	1.56732619	**73**	1.19248038
35	1.38645318	**74**	1.00048356
36	1.14847226	**75**	1.00023865
37	0.98802905	**76**	1.10559794
38	1.02632864	**127**	**-0.62226483**
39	0.99979973		

Preparation of the researcher based on the results of the program (R- language) Of the results listed in the above table for group I group ratings (people with lymphocytic leukemia). We find that the method of vector support machine succeeded in the classification of a large range of views correctly and at the same time did not succeed in the classification of some observations. Returning to the table above, the first and third columns include the current study views, and the second and fourth columns include the values of those observations calculated according to the SVM method.

It is known that the number of observations of the first group (people with lymphocytic leukemia) was (77) views, we note that (75) of this group were classified into the first group (people with lymphocytic leukemia), that is, method SVM Successfully rated 75 views, and these observations have positive values as shown in the table above. At the same time we note that (2) observations of this group were classified into the second group (people who do not have lymphocytic leukemia) SVM I mistakenly rated 2 views incorrectly and these views have negative values which are both views (30,127) as shown in the table above.

Classification of observations of the second group (persons with leukemia)

According to the equation of the decision function and according to the method "SVM" Group 2 observations (people not affected by lymphocytic leukemia) can be classified as shown in the table below:

Table 10: Classification of the second group views according to the vector machine method "SVM"

Observations	Value	Observations	Value
77	**1.09864704**	102	-0.99969460
78	**0.45817109**	103	-1.34066294
Games	-0.24587592	104	-1.22344137
80	-0.39184774	105	-1.39074047
81	-0.63318858	106	-1.35002540
82	-0.43074978	107	-1.19173217
83	-1.20565016	108	-0.90060013
84	-1.29306489	109	-1.37002358
85	-1.61713498	110	-1.25858241
86	-0.70405421	111	-1.48252476
87	-0.99935710	112	-1.27002006
88	-0.99991129	113	-1.00041936
89	-0.99984064	114	-1.02210764
90	-0.97071613	115	-0.99983502
91	-1.32863529	116	-1.10114439
92	-0.39589707	117	-1.28956880
93	-1.46994878	118	-0.87690257
94	-1.62687785	119	-0.32251555
95	-1.45749304	120	-11.00037289
96	-1.37059343	121	-0.93486574
97	-1.09678567	122	-1.48453937
98	-1.35119355	123	-1.20552832
99	-1.31526376	124	-1.00008992
100	-1.00009486	125	-0.92227593
101	-1,00020032	126	-0.99982713

Table by the researcher using the program (R- language)

From the results shown in the above table and special B classification group views the second (the people of others infected with lymphatic leukemia) . We find that the method of vector support machine succeeded in categorizing a large range of observations correctly and at the same time did not succeed in the classification of some observations. To return to the table above, the vertical n the first and the third includes the current views of our study, the vertical n second and fourth includes the values of those views which have been calculated according to the way SVM .

It is known that the number of observations of the second group (people who did not have lymphatic leukemia) was (50) observations , we note that (48) views of this group were classified into the second group (people without lymphoma), SVM You successfully rated 48 views, and these views were negative and as shown in the table

above. At the same time, we note that (2) observations of this group were classified into the first group (people with lymphocytic leukemia) SVM I mistakenly rated 2 views correctly and these observations have positive values and are all views (77,78) As shown in the table above.

4. Conclusions and Recommendations

4.1 Conclusions

1) The method of the Discriminate function and the Support Vector Machine (SVM) method was used to analyze the data of the phenomenon under study, and both methods had a high capacity for good classification of the data.
2) The vector-propulsion technique has a comparative advantage over the Discriminate function method because the fault of the vector-machine method is much less than the typo error of the Discriminate analysis function.
3) Through the results of the extraction of both methods we find that they have a high classification capacity in the classification of new views to the original group to which it belongs.
4) The method of the support vector machine has a high superiority in the classification of the first group (people with lymphatic leukemia) with a correct classification of up to 86% compared to 75% for the function of the analysis of the distinction, and the correct classification of the second group (people who do not have lymphatic leukemia)) In the language of 98% and both methods. Through the above results we note that the vector support machine outperforms the qualitative analysis.
5) The results have been concluded that the hormone LDH has a high potential in the diagnosis of chronic lymphocytic leukemia, this variable comes first, and then variable rate of blood deposition ESR comes second, and the variable ratio of white blood cells This variable comes in fourth place and the variable age. This variable comes in fifth place, and the variable is the ratio of blood vessels PLT. This variable comes in sixth place and the red blood cell variable (RBC). This variable comes in seventh place. The rest of the variables were of slight effect.

4.2 Recommendations

1) Recent medical studies indicate the risk of chronic lymphocytic leukemia. Therefore, we recommend that people be sensitized to conducting laboratory tests periodically for the early detection of this disease, and urged people to do these tests through the media such as radio, internet, television or the establishment of conferences and seminars for this purpose
2) You should see a doctor specializing in cancer and tumors when feeling any symptom of the presence of lymphoma, lymphoma, in order to accelerate the treatment of this disease in its initial stages.
3) Focus on SVM technology and its use in various fields, especially medical fields, for the efficiency and flexibility of this method compared to other classification methods.
4) Focus on the use of the method of SVM instead of the classical method of classical analysis to prove the efficiency of this method and ability to classify views to its original collections. And its high flexibility in dealing with data linear and non-linear
5) Given the spread of cancer diseases significantly recommends the researcher to carry out advanced studies to know the trends of the spread of lymphatic cancer and develop good strategies to address this disease.

References

[1] Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., & Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. Ecological Informatics, 4(4),pp 206-214.
[2] Byrd JC, Furman RR, Coutre SE, et al. Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. New England Journal of Medicine. 2013;369(1): 32-42.
[3] Cai, C. Z., Wang, W. L., Sun, L. Z., & Chen, Y. Z. (2003). Protein function classification via support vector machine approach. Mathematical biosciences, 185(2),pp. 111-122.
[4] W. Deng; J. Hu; and J. Guo, "Robust Fisher Linear Discriminant Model for Dimensionality Reduction," IEEE, pp. 1-4, 2006.
[5] Omar, Ibrahim Bashi (2015), "Face Recognation Based On PCA , LBP and SVM Techniques", Eng .&Tech.Journal , Vol(33), part(B),no.(3),pp 384-396
[6] P. Viszlay, M. Lojka and J. Juhár,(2014)Class-dependent two dimensional linear discriminant analysis using two-pass recognition strategy, in: Proceedings of the 22nd European SignalProcessing Conference (EUSIPCO), IEEE, , pp. 1796– 1800.
[7] Priyabalasubramanian ,(2014)," Automated classification of EEG signals using component analysis and support vector machine", A Thesis of master of Science in Engineering , Padnos College of Engineering and Computing, Grand valley state university .
[8] Qi, Z., Tian, Y., & Shi, Y. (2012). Laplacian twin support vector machine for semi-supervised classification. Neural Networks, 35, pp.46-53.
[9] QinghuaGuo et al,(2004)," Support vector machines for predicting distribution of Sudden Oak Death in California"EcologicalModelling vol.182 (2005)pp(75–90).
[10] Rancher , A . C . (2002) " Methods of Multivariate Analysis " second Edition John Wily and sons .
[11] Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., &Plümer, L. (2010). Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. Computers and Electronics in Agriculture, 74(1),pp. 91-99.
[12] Thair A saleh, Mustafa Zuhaernayef,(2012),"Discrimination Analysis and Support vector Machine ",Eng.&Tech.Journal , Vol(31), part(A),no.(12) , pp.2261-2272
[13] Takeuchi, K., & Collier, N. (2005). Bio-medical entity extraction using support vector machines. Artificial Intelligence in Medicine, 33(2),pp. 125-137.

[14] T. Li, S. Zhu and M. Ogihara, (2006), Using discriminant analysis for multi-class classification: An experimental investigation, Knowledge and Information Systems 10(4), 453–472. doi:10.1007/s10115-006-0013-y.
[15] Yao, X. J., Panaye, A., Doucet, J. P., Zhang, R. S., Chen, H. F., Liu, M. C., ... & Fan, B. T. (2004). Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression. Journal of chemical information and computer sciences, 44(4),pp. 1257-1266.
[16] Zhiquan Qi et al,(2012)," Laplacian Twin Support Vector Machine For Semi-Supervised Classification " Preprint submitted to Neural Networks October 28, 2012
[17] http//ar.wikipedia.org Wikipedia free encyclopedia
[18] https// . ebteb.com › Medical Web Site
[19] https://www.mayoclinic.orgMayo Clinic