AMERICAN JOURNALS
Open Access, Peer Reviewed Journals
American Journal of Business Management, Economics and Banking
ISSN (E): 2832-8078 Volume 13, | June, 2023

USING ARTIFICIAL NEURAL NETWORKS AND MULTIPLE LOGISTIC REGRESSION ANALYSIS TO STUDY SOME FACTORS RESULTING LEUKEMIA

Hayder Sami Alwan
College of Administration and Economics / University of Maysan-Iraq
haydersami.a87@uomisan.edu.iq

ABSTRACT

This study investigated some factors which lead to leukemia using, artificial neural networks (ANN) and multiple logistic regression analysis. The study also compared between these methods determine which method is better. The study included 150 observations and four variables: Chronic myeloid leukemia (CML) as a response variable, patient weight, blood platelets (PLT), and the bone marrow edema (EME) as explanatory variables. The model's goodness of fit was measured using the maximum likelihood ratio test, which followed a chi-square distribution. Results showed that the relationship between the explanatory variables and response variables is significant, indicating the importance of the explanatory variables in the model. The logistic regression model (LRM)'s parameters were estimated using the maximum likelihood ratio, and the neural network method was also used to analyze the data. The neural network method achieved a much higher accuracy rate than the multiple logistic regression method, with a difference of 20%, indicating the superiority of the neural network method over the multiple logistic regression method.

KEYWORDS

Analysis, artificial neural networks, cancer, factors, Leukemia, multiple logistic regression, Wald , MLE.

Introduction:

In recent years, there has been an increase in the use of modern statistical methods and applications for analyzing classified data, particularly in fields such as medical, agricultural, and social research (Al-Obeidi, 2013). This study discusses the use of two different statistical methods in analysis: artificial neural networks, which are an important product of artificial intelligence and simulate natural intelligence, specifically human intelligence. These two methods are known for their ability to interpret incomplete or noisy data, handle new types of problems, process data in parallel, and high flexibility compared to traditional mathematical methods. Additionally, neural networks have surpassed the limitations of traditional solutions because all cells within the network process a huge amount of data quickly at the same time (Davalo & Naim, 1991). Logistic regression analysis (LRA), the second method, is an important statistical method that deals with analyzing classified data, especially in cases where the response variable is nominal or ordinal and consists of two levels (two classifications) or more (Al-Afifi, 2010; Maiprasert & Kitbumrungrat, 2012). The analysis using logistic regression does not assume any conditions regarding the distribution of explanatory variables, making it relatively better and more robust than many other methods. It is characterized by flexibility and simplicity, as well as giving a clear and meaningful interpretation and implications for describing the relationship between the response variable and explanatory variables.

Research Objective

The current research aims is to study some factors which lead to leukemia using a multiple logistic regression (MLR) model and an intelligent method represented by artificial neural networks (ANN), and to compare between them.

Research Problem

Regression models in general, and binary and multiple response logistic regression models in particular, are characterized by their high ability to model causal relationships between explanatory variables on the one hand and the dependent variable on the other hand. However, they suffer from some assumptions that, if not met, will lead to misleading results. Therefore, more flexible methods with fewer assumptions compared to traditional regression models are used to avoid such problems.

Literature Review

Artificial neural networks (ANN)

The relationship between statistics and ANN has received significant and increasing attention from researchers and practitioners in both fields in recent years. Many ideas related to ANN are based on the fundamentals of statistics, and many statistical methods and techniques can be programmed using neural networks and building algorithms for them. The reason for this is due to the great similarity and congruence between the methods used in both fields (Jarrah, 2003). The general principle of building ANN is to develop a system for information processing to reach the closest point to human efficiency (Omar, 1999). Therefore, an ANN is a computer system for information processing, consisting of a very large number of interconnected processing elements with a dynamic nature. Its main function is the parallel division of network calculations. The neural network consists of a number of interconnected and homogeneous processing units, each of which is a computational tool that can model its behavior with simple mathematical equations (Rao & Raao, 1993). Neural networks are named after their resemblance to the internal communication systems of biological neurons, as these units are inspired by the study of biological neural systems , Figure 1 (Newbold, 1974).

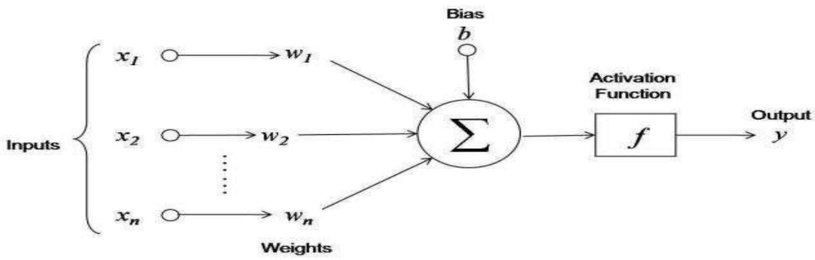

Figure 1. System of artificial neural networks

1- Components of an artificial neuron

Zurada (1997) listed the following components:

1. Synapses: The neuron receives input signals from other connected cells through input channels, and these incoming signals are called inputs and are denoted by (xi, i = 1, 2, 3, 4, ..., n).
2. Summation function: The function of this function is to unify the input signals into one signal.
3. Activation function: This function aims to propagate the input value according to the type of function used according to the scale of the output value limits.
4. Axon path's: The function of this channel is to send the output signal to other neurons, and it is considered an input signal to those cells, and the output signal is called outputs.

1. 2 Types of artificial neural networks

1-2-1 Feedforward neural networks

These networks allow information to flow in only one direction, meaning that the data spreads from the back to the front (i.e., from the input layer to the output layer) because their structure lacks closed loops of connections between the layers (Stergiou & Siganos, 2006).

1-2-2 Backpropagation neural networks (NNN)

Backpropagation networks allow information to flow in both directions, and they are complex and powerful. They are dynamic networks that continuously change their stability until reaching a point of equilibrium, and they remain there until the input changes and a new equilibrium needs to be found (Stergiou & Siganos, 2006). BNN is based on the concept of training the network using the mean square error, by finding the minimum value of the total error square for the output values computed by the network. The weights between the layers are updated until the optimal weights that provide the best fit for the model are achieved (Al-Nasser & Al-Obaidi, 2003). The key to an ANN model is the structure of the information processing system that connects and organizes a large number of internally connected processing elements (neurons) that work together to solve specific problems. Signals pass between nodes (neurons) through connecting lines, and each line is associated with a certain weight. The incoming signals to the node (neuron) are multiplied by these weights, and the weighted inputs are summed up in the nodes or neurons. Then, the outputs of each node are processed by a non-linear function with a specific threshold called the activation function (Stergiou & Siganos, 2006).

1-3 Activation function

It is also known as conversion functions because it converts input values by mixing them with weights from one mathematical formula to another. It also has several forms, including linear and nonlinear, and is characterized by its ability to determine the type of relationship between inputs and outputs and forms of activation functions (Wasserman, 1989; Zurada, 1960).

1- Linear activation function

$$a=f(n)=n\ldots\ldots(1)$$

The derivative of this function is in this form

$$\bar{f}(n)=1\ldots\ldots\ldots(2)$$

2- Logistics activation function

This function is often used at the hidden layer of neural networks since the input values are basically real values (-∞, +∞) while the output values of the function for each node of the network are (0,1) and are according to the following formula:

$$f(n)=(1+e^{\frac{-n}{Q}})^{-1}\text{……..(3)}$$

This function can be derivated in this form

$$(n)=(f(n)(1-f(n)/Q\text{……..(4)}\bar{f}$$

3-The step activation function

This function is used in classification models, and its outputs depend entirely on the inputs. The output of the function equals (1) if the input is greater than or equal to (0), otherwise, the output equals (0). This function differs from the previous two functions in that it has no derivative. It can be expressed as:

$$a=f(n)=n\text{…….(5)}$$

$$a=\begin{cases}1, n\geq 0\\ 0, n<0\end{cases}$$

4-Threshold activation function

The threshold activation function is similar to the step function, as it has no derivative. The output values equal (+1) if the inputs are greater than or equal to (0), and (-1) if the inputs are less than (0). It can be expressed as:

$$a=f(n)\text{……(6)}$$

$$a=\begin{cases}1, n\geq 0\\ -1, n<0\end{cases}$$

5- Bipolar Sigmoid activation function

This function is commonly used in feature estimation. The input values range from (-∞ to +∞), while the output values range from (-1 to +1). It can be expressed as:

$$f(n)=\tanh(n/Q)=1-2/(1+e\text{^}(2n))\text{......(7)}$$

The derivative of the function is given by: $\bar{f}(n)=\frac{1-[f(n)]^2}{Q}$ …….(8)

1-4 Backpropagation Algorithm python

The backward backpropagation is one of the most important supervised training algorithms for neural networks. Its name comes from the fact that the error is propagated backward through the network from layer to layer. This algorithm relies on choosing a suitable error function, which is determined by the actual and desired output values, as well as the network's features such as weights and thresholds (Arabic Encyclopedia for Computer and the Internet, 2004)

Al-Sheikhly (2003) states that training an ANN using backpropagation involves three stages, as follows:

1. Forward propagation of error.
2. Backward propagation of error.
3. Weight tuning.

During the forward propagation stage, the input signal is transmitted to each node in the hidden layer, and the activation value for each node in the hidden layer is calculated based on this signal. Then, these nodes send their signals to each node in the output layer, and the activation value for each node in the output layer is calculated to form the network's response to the given input sample. During the training stage, each node in the output layer compares its calculated activation with the actual output value to determine the error value for that node. Based on this error value, the error correction factor δ_k is calculated. The error correction factor δ_k is used to distribute the error among the nodes in the output layer and is then sent back to each node in the previous layer. This factor is also used to update the weights in the output and hidden layers. Similarly, the error correction factor δ_j is calculated for each node in the hidden layer, and this factor is used to update the weights in the hidden and output layers. After determining all the error correction factors δ, the weights for all layers are adjusted simultaneously.

The algorithm or methodology for this network can be summarized in the following steps:

1. Generate initial weight values from a statistical distribution.
2. Each node in the input layer receives input signals and sends them to all nodes in the hidden layer.
3. Each node in the hidden layer aggregates weighted input signals according to the equation:

$$hj = 2 / (1 + \exp(- \ xiwij)) - 1 \ \ldots\ldots\ (9)$$

4. Applying the activation function to estimate the outputs of the hidden layer, and sending the activation values to all nodes in the output layer.
5. Each node in the output layer aggregates weighted input signals according to the equation:

$$y_k = 2 / (1 + \exp(\sum_j h_i w_{ik})) - 1 \ldots\ldots\ldots\ldots(10)$$

6. Calculating the error for each output node by finding the difference between the activation value y_k and the target value t_k, as follows:

$$E_k = t_k - y_k \ \ldots\ldots\ \ (11)$$

The output of the neural network is compared with the real values to estimate the error according to the equation:

$$\delta_k = (t_k - y_k) \ .\ f(v) \ \ldots\ldots\ (12)$$

Where f(v) represents the logistic or tansig function when the output node is nonlinear and equals 1 when the function is linear. Then, calculate the change in weight Δw_{jk} using the following equation:

$$\Delta w_{jk} = \alpha \ .\ \delta_k \ .\ h_j \ \ldots\ldots\ (13)$$

7. Each node in the hidden layer aggregates weighted error signals from the output layer according to the equation:

$$\Delta_j = \sum_j \ \delta_k w_{jk} \ \ldots\ldots\ldots\ (14)$$

Then this value is multiplied by the activation function to calculate δ_j' after which the change in the magnitude of the error Δv_{ij} is calculated and according to the equation:

$$\Delta v_{jj} = \alpha \ .\ \delta_j \ .\ x_i \ \ldots\ldots\ (15)$$

8. Update the weights of each node in the output layer, according to the equation:

$$w_{jk}\text{ (new)} = w_{jk}\text{ (old)} + \Delta w_{jk} \text{ .... } (16)$$

Then the weights for each node is updated in the hidden layer, according to the equation:

$$v_{ij}\text{ (new)} = v_{ij}\text{ (old)} + \Delta v_{ij} \text{ .... } (17)$$

9. The network continues to update the weights "i.e. the process of learning and training" until the optimal weights are obtained. Thus, the desired outputs are obtained, i.e. reaching the best fit for the model under study.

Where:

x_i: represents the network inputs, w: represents the weights between levels.

2. Logistic Regression

Logistic regression is used to find the best fit model that describes the relationship between a response variable (dependent) and one or more explanatory variables (independent). Many natural phenomena exhibit non-linear behavior, and to analyze such phenomena, non-linear models are used. However, due to the difficulty in using these models, linear regression models are often used for this purpose. Logistic regression models are one such statistical model used to describe and analyze such phenomena. The logistic model is used to describe the relationship between a response variable (y) and one independent variable (x) or several independent variables (x_1, ..., x_n) (Al-Afifi, 2010; Ghanem & Al-Jaouni, 2011; Johnson, 1998). When there is a binary response variable and one independent variable, the general formula for the logistic model is:

$$\text{.......}(18)\, p(x) = \frac{1}{1+e^{-\alpha-\beta x_i}}$$

Since: i=1,2,3,4,……,n,

B and α: are the two parameters of the model to be estimated.

β>0

P(x): the probability of the response

x_i: the explanatory variable

Where:

$-\infty < x_i < \infty \;\&\; -\infty \leq a \leq \infty$

The above formula is known as the logistical response function and is characterized by the fact that P (x) is defined between (1,0) and that the two parameters (B, α) are unrestricted. However, in the case of each of the nominally dependent variables that have multiple responses consisting of two levels or two classifications or more and there are several independent variables. Therefore we have the multiple logistic model; it is formulated as:

$$\text{.......}(19)\, p(x) = \frac{1}{1+e^{-Z}}$$

Where, $Z=a+B_1X_1+B_2X_2+\text{.........}+B_pX_p\text{.........}(20)$

There are two types of LRM, the first model is the two-response LRM and this model is used in the case if the response variable consists of two levels (two classifications) only. The second model is the multiple-response LRM and is used in the case if the response variable consists of three levels and more.

1-Two-response logistic regression model

The LRM is based on a basic assumption, which is that the dependent variable (y), the response variable that we are interested in studying, is a binary variable and follows the Bernoulli distribution, and takes the value (1) with a probability of (π) and the value (0) with a probability of (1-π), that is, the occurrence of the response and not occurring (Al-Afifi, 2010; Timm, 2002).

2. Multiple response logistic regression model

The multiple-response LRM is considered one of the important statistical models in analyzing classified data, and it is generally used when the response variable belongs to variables of nominal or ordinal type consisting of two levels or two classifications or more. Likewise, the multiple-response LRM is a development and expansion of the two-response LRM, and it depends mainly on the multinomial distribution (Al-Afifi, 2010; Maiprasert & Kitbumrungrat, 2012).

2-1 Logistic model assumptions

Zelen (1983) stated that there are many hypotheses for the logistics model, the most important of which is:

1. If the value of the dependent variable is confined or restricted to the period [1,0]
2. The values of the dependent variable are either (0) or (1) and the remainders are generally large and the value of (R^2) is small.
3. The logistic model does not assume a specific distribution of the dependent variable, but it is assumed that its distribution is within the distributions of the exponential family (i.e., binomial distribution).
4. To ensure a sufficient number of items for the response variable, the sample sizes must be large in the logistic regression.
5. The logistic model does not assume a certain distribution of the explanatory variables, and there should not be a high correlation between the explanatory variables, because this causes problems in estimation.
6. There shouldn't be a linear relationship between the independent variable and the dependent variable, but it assumes linearity between the independent variables and the logarithm of probability (Cook et al., 2001).

2-2 Model estimation

2-2-1 Maximum likelihood estimation

The parameters of the multiple-response LRM will be estimated in the manner of greatest possibility (Agresti, 1990; Al-Bayati, 2005; Al-Razami, 2000) which can be described as follows:
The probability density function for a polynomial distribution is as follows:

$$\ldots\ldots.(21) P_r\left(Y_{i1} = y_{i1},\ .\ .\ .\ .\ .\ , Y_{ij} = y_{iJ}\right) = \binom{ni}{y_{ij\ ,\ldots\ldots,} y_{iJ}} \pi_{i1}^{y_{i1}}\ .\ .\ .\ \pi_{iJ}^{y_{iJ}}$$

Hence,

$\pi_{i1}^{y_{i1}} \ldots.\ \pi_{iJ}^{y_{iJ}}$ represents the percentage of response groups in the community.

$$L(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{J} \left[\pi_j\ (\underline{x})\right]^{y_{ij}} \quad \ldots\ldots..(22)$$

Calculating the logarithm of the possible function, we get:

$$\ldots\ldots..(23) LN\ L(\beta) = \sum_i \sum_j y_{ij}\ log\ \pi_j(\underline{x})$$

Substituting π_j (x) into equation (23), we get:

$$\ldots\ldots.(24) LN\ L\ (\beta) = \sum_i \sum_j y_{ij}\ log \frac{e^{\alpha_j + \underline{\beta}_j'\ \underline{x}}}{1+\sum_{j=1}^{J} e^{\alpha_j + \underline{\beta}_j'\ \underline{x}}}$$

$$\ldots\ldots.(25) = \sum_i \sum_j y_{ij}(\alpha_j + \underline{\beta}_j'\ \underline{x}\ ) - \sum_i \sum_j y_{ij}\ Log(1 + \sum_{j=1}^{J} e^{\alpha_j + \underline{\beta}_j'\ \underline{x}})$$

By deriving the maximum possibility function with respect to the parameter vector (β)

$$\ldots\ldots\ldots(26) \frac{\partial \mathrm{LnL}(\beta)}{\partial(\beta)} = \frac{\sum_i \sum_j y_{ij}((\alpha_j + \underline{\beta}_j'\ \underline{x}\ )}{\partial(\beta)} - \frac{\sum_i \sum_j y_{ij}\ \mathrm{Log}(1+\sum_{j=1}^{J} e^{\alpha_j + \underline{\beta}_j'\ \underline{x}})}{\partial(\beta)}$$

When equating the derivative to zero ($\frac{\partial \mathrm{LnL}(\beta)}{\partial(\beta)} = 0$), we get natural equations, and the roots of these equations represent the values of the estimated parameters resulting from the process of maximization and because the equations are non-linear, we resort to using one of the iterative methods (Newton Raphson), which produces appropriate estimates according to the following formula:

$$\ldots\ldots.(27) \mathrm{B}(t+1) = \mathrm{B}(t) + \{\acute{X}\ v(t) X\}^{-1} \acute{X} r(t)$$

Where,

t : represents the number of cascades

r : represents the remainder vector with dimension [N(g-1),1]

v: is $\mathrm{Diag}(v_1 \ldots \ldots v_n)$and each v_i is a square array written in the following form:

$$\ldots\ldots(28) v_i = n_i\ \{P_s\ (X_i)(\delta_{st} - P_t(X_i))\} S_t$$

And that $\delta_{st}$is the Kronecker delta and is defined by the following:

$$= \begin{bmatrix} 1 & S = t \\ 0 & S \neq t \end{bmatrix} \delta_{st}$$

Accordingly, we can start from initial values by equating the feature vector to zero, and in the binary case, the graph or the usual least squares method can be used, or the estimates of the linear discrimination function can be used as initial values in estimating the features, as their use will reduce the number of successive cycles, and when obtaining the required convergence between sessions, these optimal estimates are required.

2-3 Tests

2-3-1 Wald

One of the important tests used to test the significance of the effect of the independent variable on the dependent variable in non-linear models is the Wald test. This test corresponds to the t-test used in linear models (Garson, 2013). The formula for the Wald test is as follows:

$$T_W = \left(\frac{\hat{\beta}}{Se\left(\hat{\beta}\right)}\right)^2 \qquad \ldots\ldots..(29)$$

In statistical analysis, ($\hat{\beta}$) represents the parameter estimate and)$Se\left(\hat{\beta}\right)$(represents the standard error of the parameter. The test value is compared to χ^2 a chi-squared distribution with one degree of freedom. However, the accuracy of this test is questionable in small samples because the standard error may be large. Menard (Zelen, 1983) pointed out that large logit coefficients lead to large standard errors, which can reduce the accuracy of the Wald statistic and lead to type II errors. This is one of the drawbacks of the Wald statistic, as very large effects can lead to large standard errors. In most cases, the results of this test are very similar to the results of the odds ratio test (i.e., they rarely differ). Therefore, it is best to use the odds ratio test to test for the significance of variables with and without parameters. Additionally, logistic regression analysis is sensitive to violations of assumptions about large sample sizes. For these reasons, the odds ratio test is generally preferred (Cook et al., 2001).

2-3-2 Likelihood ratio test

If we want to test two models, the first () includes all the variables, and the second (LM_2) includes a part of those variables, then in this case we use the Likelihood-Ratio test to show which of the two models has better significance than the other(Dessens & Janson, 2005).

The formula for the probability test is:

$$D = -2 \log (LM_1 / LM_2) = -2 (\log LM_1 - \log LM_2) \quad ......(30)$$

whereas :-

LM_1: represents the possibility function under the null hypothesis. lm1

LM_2: represents the possibility function under the alternative hypothesis.LM2

The D value is close to the distribution of and compared with the tabular value of the distribution with a degree of freedom that represents the difference between the parameters of the two models.

2-4 Model quality

The quality of relevance means how appropriate the statistical model is to the data of the study sample. The quality of relevance measures measure the closeness between the observed and expected values of the model. Here are some important tests for relevance quality.

2-4-1 Chi square test χ^2

It is one of the important tests used to fit or define the quality of the model, and it shows how the model describes the response variable, which is equal to the sum of squares. The differences between the observed and expected values divided by the expected values, and estimating the suitability of the model includes the convergence of the expected values from the observed values, as in the following formula (Cowan, 1994):

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected} \quad ....(31)$$

Where (observed) means the observed values and (expected) means the expected values, and compares the value of the test statistic with the $\chi 2$ distribution with a degree of freedom equal to the number of classes minus the number of parameters in the logistic model.

$$d.f = n - p \quad ....(32$$

whereas:-

p: represents the number of parameters,

n: represents the number of items

2-4-2 Hosmer–Lemeshow

This test is widely used in estimating the quality of fit of the model and allows any number of independent variables, which may be continuous or discrete, and this test is somewhat similar to the $\chi 2$ test for the quality of fit. This test is as in the $\chi 2$ method for the quality of fit and approaches the $\chi 2$ distribution with a degree of freedom equal to the number of classes minus the number of parameters in the logistic regression model. With a small sample size, this test is weak and this test requires a sample size of more than 400 observations (Cowan, 1994), and it is possible using equation (31).

Where the (observed) means the observed values and (expected) means the expected values

2-5 The coefficient of determination R^2

R^2 does not explain the quality of fit of the model as in linear regression, but it is an indicator of the importance of explanatory variables to predict the response variable and then measure the effect size (Agresti, 2002).

We can mention the Cox & Snell R^2 formula as follows:

$$R^2cs = 1 - \left[\frac{\hat{L}(B_{(0)})}{\hat{L}(B)} \right]^{2/N} \quad ..... \ (33)$$

The Nagelkerke R^2 formula can be explained by the following formula:

$$R^2N = \left[\frac{R^2cs}{1 - \{\hat{L}(B_{(0)})\}^{2/N}} \right] \quad ..... \ (34)$$

The $L(\beta(0))$ represents the possibility function of the model under the null hypothesis, and $L(\beta)$ represents the possibility function of the model under the alternative hypothesis. The large sample size gives power to the test analysis of the logistic model, as the larger the sample size, the greater the test power of the model

Results

This section includes an applied study of the multiple LRM using the maximum likelihood method, as well as an analysis of ANN for data on CML obtained from a master's thesis by the researcher (Mirdas, 2023). The data represent 150 observations and four variables, which are as follows:

- Type of CML infection: as a response variable with three categories (Chronic phase CP, Accelerated phase AP, and Blast phase BP), represented by the code (1) for CP patients, (2) for AP patients, and (3) for BP patients.
- Patient weight: The first explanatory variable.
- Blood platelets (PLT): The second explanatory variable.
- Number of cancer cells in bone marrow (BME): The third explanatory variable.

Statistical data analysis using multiple logistic regression

Model fitness quality

To measure the goodness of fit of the model, in the case of the multiple LRM, the likelihood ratio test is used, which follows a chi-square distribution. The outputs of the SPSS program are as follows:

Table 1 Goodness of fit of the multiple LRM

<table>
<tr><th>Model</th><th>-2 Log Likelihood</th><th>Chi-Square</th><th>df</th><th>Sig.</th></tr>
<tr><td>Fixed</td><td>323.294</td><td rowspan="2">42.240</td><td rowspan="2">6</td><td rowspan="2">0.000</td></tr>
<tr><td>Final model</td><td>281.054</td></tr>
</table>

According to the values of Sig in Table 1, the relationship between the explanatory variables and the response variable is significant. This means that there is a difference in the model when the explanatory variables are present compared to when they are absent. Other tests also support this conclusion, as shown in Table 2:

Table 2 Goodness of fit of the multiple LRM using non-parametric tests

Model	Chi-Square	df	Sig.
Pearson	284.938	290	0.573
Deviance	279.667	290	0.658

It can be observed from Table 2 that the Sig values for both tests are greater than the significance level (0.05), indicating the importance of the explanatory variables in the model.

Parameters estimation of the multiple LRM

The parameters of the multiple LRM were estimated using the maximum likelihood method. Table 3 summarizes the estimates of these parameters, along with their standard errors, Wald statistics, and accompanying Sig values:

Table 3. Estimates of parameters of the multiple LRM

CML [1]		B	Std. Error	Wald	df	Sig.
CML=1.00	Fix	0.155	2.450	0.004	1	0.950
	Weight	-0.007	0.034	0.047	1	0.829
	PLT	0.007	0.004	2.867	1	0.090
	BME	-23.410	6.054	14.954	1	0.000
CML=2.00	Fix	5.151	2.341	4.843	1	0.028
	Weight	-0.069	0.034	4.222	1	0.040
	PLT	-0.001	0.004	0.098	1	0.754
	BME	-5.665	3.025	3.506	1	0.061

It can be observed that the effect of the constant term and the explanatory variables on the response variable varies from category to category. For example, the effect of the constant term on the logit of the Chronic phase is 0.155, while it is 5.151 for the AP. Additionally, the weight variable has a negative

[1] The value (3) was used as a reference category.

effect on the logit of the Chronic phase with a regression coefficient of -0.007, and -0.069 for the Accelerated phase of the illness.

The results of the classification are exhibited in Table 4 as follows:

Table 4. Confusion matrix for the multiple LRM

Classification		Predictor				Correct rating ratio
		Chronic phase	Accelerated phase	Blast phase	Total	
Observation	Chronic phase	50	11	1	62	80.6%
	Accelerated phase	22	19	7	48	39.6%
	Blast phase	22	5	13	40	32.5%
Total		94	35	21	150	54.7%

Table 4 shows that the correct classification rate (CCR) for the CP is 80.6%, but it is lower for both the PB and BP. The overall CCR is only 54.7%, which is relatively low.

Statistical data analysis of using ANN

The results of the ANN were obtained from the SPSS program after applying it to the study data. The network for the study model is shown in Figure 1.

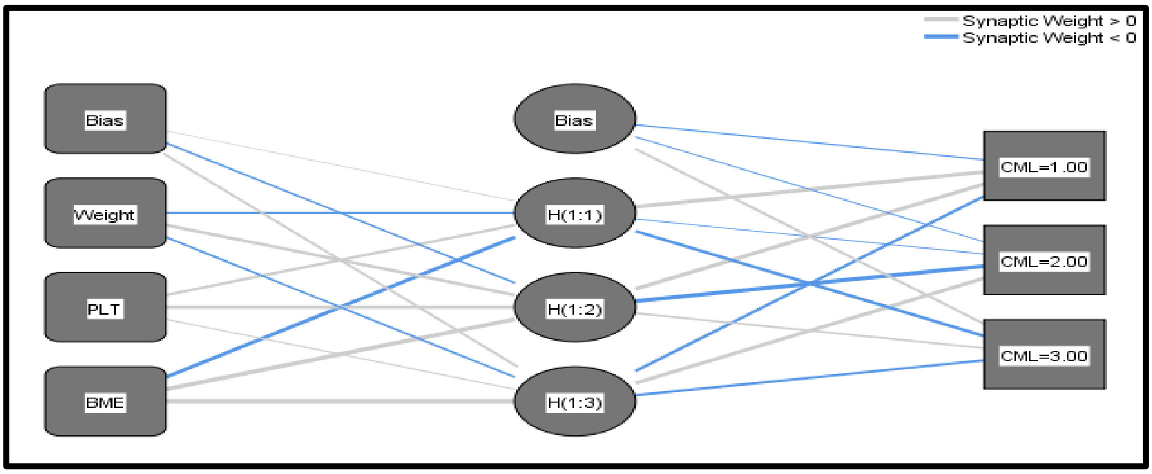

Figure 1. Artificial neural networks

The estimated parameters were as following:

Table 5. Estimated parameters

Predictor		Predicted					
		Hidden Layer 1			Output Layer		
		H(1:1)	H(1:2)	H(1:3)	CML=1	CML=2	CML=3
Input Layer	(Bias)	0.062	-0.772	1.069			
	Weight	-0.655	1.823	-0.914			
	PLT	1.428	1.84	0.543			
	BME	-1.889	2.864	3.01			
Hidden Layer 1	(Bias)				-0.551	-0.469	1.204
	H(1:1)				2.318	-0.402	-1.773
	H(1:2)				2.053	-2.714	0.982
	H(1:3)				-1.422	2.045	-1.309

This method achieved a much higher correct classification rate than the multiple LRM, as shown in Table 6 below:

Table 6. Confusion matrix for the artificial neural network

Classification		Predictor				Correct rating ratio
		Chronic phase	Accelerated phase	Blast phase	Total	
Observation	Chronic phase	57	3	2	62	91.9%
	Accelerated phase	15	30	3	48	62.5%
	Blast phase	10	4	26	40	65.0%
Total		82	37	31	150	75.3%

It can be seen from Table 6 that the CCR for the Chronic phase is 91.9%, which is higher than the CCR in the multiple logistic regression model (80.6%). The correct classification rate for the Accelerated phase is 62.5%, which is higher than in the LRM (39.6%), and the CCR for the Blast phase is 65%, which is higher than in the LRM (32.5%). Therefore, the overall CCR for the ANN method is 75.3%, which is higher than the overall CCR in the multiple LRM (54.7%).

Conclusions

From the results, we conclude that the correct classification rate (CCR) when modeling using the multiple logistic regression method is approximately 55%, while CCR when using the artificial neural network method was approximately 75%. This indicates a superiority of the artificial neural network (ANN) method by about 20%.

Recommendations

1. Adopting the use of ANN models in studies and research due to their high accuracy, low error rates, and adaptability in dealing with data as they are considered adaptive methods.
2. It is important to measure the quality of fit of the model in the case of MLR to determine the significance of the relationship between response and explanatory variables before starting the analysis process.
3. Developing the healthcare sector by the relevant authorities and create a plan to increase health awareness through educating and encouraging citizens to have regular check-ups for early detection of blood cancer diseases.
4. It is necessary to record complete information and data for each patient in their own medical records to benefit from them as data is the basis for conducting research and studies.

References

1. Agresti, A. (2002). Categorical Data analysis , (2 nd edition). Hoboken: Wiey.
2. Agresti, A. (1990). Categorical data analysis . New York: John Wiley Sons , Inc.
3. Al-Afifi, R .M. (2010). The Use of Multinomial Logistic Regression Model on Physical Violence Data. (Master thesis). Gaza, Al- Azhar University.

4. Al-Bayati, H. I. S. (2005). Path analysis in the logistic regression model with a practical application, (Unpublished master's thesis). Al-Mustansiriya University.
5. Al-Nasser, A. H. & Al-Obaidi, M. A. (2003). Using neural networks to predict future values of irregular time series in seasonal length. Iraqi Journal of Statistical Sciences, 3, (6), 1-16
6. Al-Obeidi, M. (2013). Artificial neural networks. Computer Research Journal, 4(1)
7. Al-Razami, A. Y. (2000). Employing the logistics model and severity measures to study the epidemiology of surgical nervous system diseases at martyr Adnan Khairallah Specialist Hospital (Unpublished Master's Thesis). Al-Mustansiriya University
8. Al-Sheikhly, E. A. L. (2003). Designing a nonlinear dynamic mathematical system using neural networks (NARMAX) for analytical and predictive purposes of sales activity in the baghdad electricity company, (Unpublished Master Thesis). University of Baghdad.
9. Cook, D., Dixon,P., Duckworth,W. M., Kaiser,M. S., Koehler,K., Meeker, W. Q. & Stephenson, W. R. (2001). Binary response and logistic regression analysis. in project beyond traditional statistical methods (pp. 1-23).USA: Iowa State University NSF/ ILI
10. Cowan, R. (1994). Technological forecasting and social change. Journal Technological Forecasting and Social Change, 47, 89-102.
11. Davalo, E. & Naim, P. (1991). Neural Networks. Macmillan, Hong kong.
12. Dessens, J., & Janson, W. (2005). Logistic regression analysis. New : York Wiley.
13. Garson, G. D. (2013). Hierarchical linear modeling: Guide and applications. Sage.
14. Ghanem, A. M. & Al-Jaouni, F. K. (2011). The use of the two-response logistic regression technique in studying the most important economic and social determinants of family income adequacy, an applied study on a random sample of families in Damascus Governorate. Damascus Journal of Economic and Legal Sciences, 27(1), 113 – 132
15. Jarrah, R. A. (2003). Principal component analysis using artificial neural networks with application to breast cancer patients. (Unpublished PhD thesis). Iraq, University of Mosul.
16. Johnson, R. W. (1998). Statistical analysis of multiple variables from an applied point of view. King Saud University.
17. Maiprasert, D. & Kitbumrungrat, K. (2012). Comparison logistic regression and discriminant analysis in classification groups for breast cancer. International Journal of Computer Science and Network Security (IJCSNS), 12(5), 44-52
18. Mirdas, M. A. (2023). A comparison between the parametric and non-parametric methods using multiple logistic regression with a practical application (Unpublished master's thesis). Iraq: Wasit University.
19. Newbold, P. (1974). The exact likelihood function for a mixed autoregressive-moving average process. Biometrika, 61(3), 423-426.
20. Omar, M. M. A.(1999). Artificial neural networks: a comparative study, (Higher Diploma Thesis). The National Center for Computing and Informatics.
21. Rao, V. B., & Raao, H. V. (1993). C++ Neural network and fuzzy logic, management information. Inc., New York, NY, 408.
22. Stergiou, C. & Siganos, D. (2006). Neural Networks. Surveys and Presentations in Information Systems Engineering. SURPRISE 96 Journal.
23. The Arabic Encyclopedia for Computer and the Internet. (2004).Neural Networks, retrieved from: http://www.c4arab.com.mejhar.php

24. Timm, N. H. (Ed.). (2002). Applied multivariate analysis. New York, NY: Springer
25. Wasserman, P. D. (1989). Neural computing: theory and practice. Van Nostrand Reinhold Co.
26. Zelen, M. (1983). Biostatistical Science as a discipline: A look into the future. Biometrics 39 827-837.
27. Zurada, J. M. (1960). introduction to artificial neural system. Jaico Publishing House.
28. Zurada. (1997). Introduction to artificial neural network. Jaico Publishing House.

Appendix

Leukemia patient data

No.	CML	Weight	PLT	(BME)	No.	CML	Weight	PLT	(BME)
1	1	53	147	0.03	76	1	66	331	0.03
2	1	60	139	0.01	77	1	68	213	0.02
3	2	52	311	0.09	78	3	63	194	0.24
4	1	51	130	0.04	79	2	61	217	0.18
5	1	61	186	0.03	80	1	63	251	0.02
6	2	63	194	0.18	81	1	63	217	0.03
7	3	70	315	0.24	82	1	66	321	0.04
8	1	57	215	0.03	83	3	51	312	0.23
9	1	55	197	0.04	84	1	67	215	0.01
10	2	57	311	0.14	85	1	71	314	0.01
11	1	63	219	0.03	86	1	61	215	0.01
12	1	67	312	0.04	87	1	53	313	0.02
13	1	71	221	0.02	88	2	62	211	0.09
14	3	59	332	0.25	89	1	63	281	0.04
15	1	61	212	0.03	90	1	67	223	0.04
16	1	66	211	0.04	91	1	71	277	0.03
17	1	70	311	0.02	92	1	61	217	0.04
18	1	68	286	0.01	93	2	59	319	0.14
19	1	65	147	0.02	94	1	52	271	0.03
20	2	61	137	0.17	95	1	51	233	0.04
21	3	67	360	0.25	96	3	53	247	0.23
22	1	60	294	0.03	97	1	71	313	0.01
23	1	70	148	0.01	98	1	68	306	0.03
24	1	65	393	0.04	99	1	65	217	0.02
25	3	70	325	0.23	100	2	53	333	0.18
26	2	75	312	0.18	101	1	71	193	0.01
27	2	63	330	0.09	102	1	66	268	0.04
28	1	65	314	0.01	103	1	75	256	0.03
29	1	60	299	0.03	104	3	53	321	0.24
30	1	61	260	0.01	105	1	72	222	0.04
31	1	52	321	0.02	106	1	59	187	0.03
32	1	50	360	0.03	107	1	73	212	0.01
33	1	61	315	0.04	108	1	68	325	0.02
34	1	66	217	0.01	109	1	63	198	0.03

35	1	52	314	0.02	110	1	59	219	0.04
36	1	60	187	0.03	111	2	52	227	0.12
37	2	61	220	0.11	112	1	57	300	0.04
38	1	58	265	0.01	113	1	50	311	0.01
39	1	66	255	0.01	114	1	49	343	0.03
40	1	52	255	0.01	115	3	51	288	0.23
41	3	62	315	0.22	116	1	63	216	0.03
42	1	51	180	0.01	117	1	67	325	0.01
43	1	61	257	0.02	118	1	73	213	0.03
44	1	65	251	0.03	119	1	64	313	0.02
45	2	70	341	0.13	120	2	51	279	0.15
46	1	60	266	0.02	121	1	67	317	0.04
47	1	67	317	0.02	122	1	66	316	0.01
48	3	59	218	0.24	123	1	72	213	0.02
49	1	71	219	0.01	124	2	52	331	0.17
50	1	62	307	0.01	125	1	62	227	0.03
51	1	52	222	0.03	126	1	48	232	0.02
52	1	55	237	0.04	127	1	49	331	0.03
53	1	58	315	0.01	128	1	69	234	0.04
54	1	51	357	0.02	129	2	51	331	0.16
55	1	71	313	0.01	130	1	67	239	0.04
56	1	65	310	0.02	131	1	62	277	0.02
57	1	70	218	0.01	132	3	51	312	0.24
58	1	60	317	0.01	133	1	62	212	0.04
59	2	52	218	0.07	134	1	67	214	0.01
60	1	55	218	0.01	135	1	71	341	0.02
61	1	72	211	0.02	136	1	62	279	0.03
62	3	58	351	0.23	137	2	53	377	0.16
63	1	51	217	0.02	138	1	62	288	0.02
64	1	71	315	0.01	139	1	61	361	0.03
65	2	61	217	0.11	140	1	72	241	0.04
66	1	68	317	0.03	141	1	67	342	0.01
67	1	73	217	0.01	142	1	57	271	0.03
68	1	61	219	0.03	143	3	49	197	0.03
69	1	61	317	0.04	144	1	71	313	0.03
70	1	62	211	0.01	145	2	48	201	0.04
71	1	71	306	0.03	146	1	67	319	0.02
72	1	61	211	0.01	147	1	56	287	0.01
73	1	65	312	0.03	148	1	64	267	0.03
74	1	62	213	0.03	149	1	72	312	0.01
75	2	61	222	0.13	150	1	67	2.79	0.03