Journal of Statistics: Advances in Theory and Applications
Volume 16, Number 1, 2016, Pages 67-91
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/jsata_7100121706

ARTIFICIAL NEURAL NETWORKS AND LOGISTIC REGRESSION WITH AN ORDERED CATEGORICAL RESPONSE: STUDY CHRONIC KIDNEY DISEASE

MONA N. ABDEL BARY[1], DINA H. ABDEL HADY[2] and ZENAB M. EL GAMAL[3]

[1]Faculty of Commerce
Suez Canal University
Al Esmalia
Egypt
e-mail: mona_nazihali@yahoo.com

[2]Faculty of Commerce
Tanta University
Tanta
Egypt

[3]Faculty of Business Administration
Taibah University
Al Madinah
KSA

Abstract

The health research has become increasingly reliant on statistical modelling techniques to assess the effect of new health programs, the impact of risk factors on disease, the effects of health behaviours, and a host of other health concerns. Clinical researchers conduct studies about diagnostic tests mainly for

2010 Mathematics Subject Classification: 62Jxx, 62J20, 62Pxx, 62P10, 68Txx, 68T05.
Keywords and phrases: ordinal logistics regression model, artificial neural network, medical diagnosis, chronic kidney disease.
Received August 19, 2016

the purpose of either estimating the diagnostic accuracy of a test according to different patient or environmental characteristics or comparing diagnostic accuracy of different tests according to different patient or environmental characteristics. Studies are required to develop robust statistical methods to analyze data from diagnostic studies and assess the properties of available statistical methods. Multiple analytical statistical methods are available to analyze ordinal data. These methods can be a model based approach, such as models for cumulative response probabilities or a non-model based approach, such as a nonparametric method based on ranking. A common model based method used to analyze ordinal data is an ordinal logistic regression. In addition to statistical models, several machine learning algorithms are also available to analyze ordinal data, such as an artificial neural network model, a decision tree model, and a support vector machine model. The current study compares the performance of ordinal logistic regression model with artificial neural networks models for prediction of chronic kidney disease. The results of the current study show outstanding performance of artificial neural networks models for the prediction of the level of chronic kidney disease. In addition, the study illustrates the most affected variables are gender, surgical operations, the blood pressure and potassium ratio. There are many previous studies focus on demonstrated the ability of ANN models in applications binary classifications but through the current study we use neural networks models in multiple classifications. The results of the current study show the following: successful of ANN models in the process of separating and classifying accuracy rate of almost 100%, and the ordinal regression model succeeded in identifying risk factors for renal failure moral influence on the regression model.

1. Introduction

When the response variable for a regression model is categorical, linear models do not work. Logistic regression is one type of model that does, and it’s relatively straightforward for binary responses. When the response variable is not just categorical, but ordered categories, the model needs to be able to handle the multiple categories, and ideally, account for the ordering. We could run a multinomial regression model. The disadvantage is that we are throwing away information about the ordering. An ordinal logistic regression model preserves that information, but it is slightly more involved.

Many studies treat ordinal data as interval data (see [21], [22], [31]). Underlying this might be the fact that parametric tests with interval data are considered easier to interpret and provide more meaningful information than non-parametric tests (see [3], [5]). However, treating ordinal data as interval data may result in a misrepresentation of the results and lead to poor decision making since such treatment causes substantial bias by assuming equal intervals between points of the ordinal data and other assumptions related to the data distribution that are rarely fulfilled by ordinal data.

According to study of Hastie et al. [17] show that treating ordinal output data as interval data results in statistically significant interaction between independent variables. However, when this ordinal output data is analyzed as ordinal data, the interaction is not statistically significant. Therefore, many researchers recommend not analyzing ordinal data as interval data in order to achieve a higher capability of detecting meaningful trends of input variables on the response variable. Thus, analyzing ordinal data using methods that are able to maintain the rank-order of ordinal data without assuming equal distances between categories provide more valuable and useful results for further investigation and decision making (see [16], [19], [20]).

Dong [7] applied the models for ordinal response study, a self-efficacy in colorectal cancer screening. Adepoju and Adegbite [1] used ordinal logistic model to study the relationship between staff categories (as outcome variable) gender, indigenous status, educational qualification, previous experience, and age explanatory variables.

A common model-based method used to analyze ordinal data is an ordinal logistic regression (OLR) model. Several approaches are available to build the OLR model, such as the cumulative link model, the adjacent categories model, and the continuation ratio model. The most commonly used among these three approaches is the cumulative OLR model (see [2], [30]).

In addition to statistical models, several machine-learning algorithms are also available to analyze ordinal data, such as an artificial neural network (ANN) model, a decision tree model, and a support vector machine (SVM) model. The comparisons between the ANN model and the logistic regression model for classification or prediction problems of binary response data have been conducted extensively (see [6], [20], [27]).

According to study Amato et al. (2012); ANNs are used in chemical kinetics, optimization of electrophoretic methods [19], classification of agricultural products such as onion varieties [30], and even species determination (see [13], [23], [24]). In general, very diverse data such as classification of biological objects, chemical kinetic data, or even clinical parameters can be handled in essentially the same way. Lisboa and Taktak [28] had done a systematic review on artificial neural networks in decision support in cancer.

Advanced computational methods, including ANNs, utilize diverse types of input data that are processed in the context of previous training history on a defined sample database to produce a clinically relevant output, for example, the probability of a certain pathology or classification of biomedical objects. Due to the substantial plasticity of input data, ANNs have proven useful in the analysis of blood and urine samples of diabetic patients (see [4], [14]), diagnosis of tuberculosis (see [10], [11]), and leukemia classification [8].

Chronic kidney disease (CKD) is identified by a blood test for creatinine, which is a breakdown product of muscle metabolism. Higher levels of creatinine indicate a lower glomerular and as a result a decreased capability of the kidneys to excrete waste products. Creatinine levels may be normal in the early stages of CKD, and the condition is discovered if urinalysis shows the kidney is allowing the loss of protein or red blood cells into the urine. To fully investigate the underlying cause of kidney damage, various forms of medical imaging, blood tests, and sometimes a kidney biopsy are employed to find out if a reversible cause for the kidney malfunction is present [1].

Compared to the other regression methods is the literature, the ordinal regression method is the most suitable and practical technique for analyzing the effects of multiple explanatory variables on the ordinal outcome that cannot be assume to be a continuous measure with normal distribution. Artificial neural networks that can be used to perform nonlinear statistical modelling and to provide a new alternative to ordinal logistic regression methods. Neural networks have a number of advantages, including requiring less formal statistical training than other methods, the ability to implicitly detect complex relationships between dependent variables and explanatory variables, the ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms. In this study, ordinal logistic regression and artificial neural networks models are used to analyze and diagnostic of CKD.

2. Objective of the Study

In this paper, we address following problems:

(1) Identify the most important explanatory variables specific to the stage of renal failure.

(2) Suggest the best model for predicting the level of renal failure patient by using the following steps:

(2.1) Using the artificial neural networks models for forecasting with stage renal failure.

(2.2) Comparing the performance of different models of neural networks in the prediction stage renal failure and determine the best model of them.

(3) Comparing the performance of artificial neural networks models and ordinal logistic regression models to predict the stage of renal failure.

3. Importance of the Study

The study is important because of the following reasons:

(1) Chronic kidney disease (CKD) is a growing health problem worldwide. Predicting the renal failure progression with reasonable accuracy is necessary because of the dynamic nature of renal disease, its covert nature in the early stages, and heterogeneity of patients. So that, early diagnosis of the disease, especially determining the appropriate time to apply medical treatments for CKD is prime significance to control and manage the consequences.

(2) There are many previous studies focus on demonstrated the ability of ANN models in applications binary classifications but through the current study we use neural networks in multiple classifications.

(3) The comparison between artificial neural networks models and ordinal logistic regression model for prediction is important for get the best model.

4. Assumptions of the Study

The current study assumes that:

(1) The relationship among the variables is linear.

(2) Y be ordinal response variable with possible values $1, \ldots, n$.

(3) $X = x_1, x_2, \ldots, x_k$ be independent predictor variables.

(4) $\propto_1, \ldots, \propto_{n-1}$ and $\beta = \beta_1, \ldots, \beta_k$ be unknown regression coefficients.

5. Materials and Methods

5.1. Data collection

CKD is mainly determined by a decrease in glomerular filtration rate, the rate at which blood is filtered in the glomeruli of the kidney. This is detected by a decrease in or absence of urine production or determination of waste products (creatinine or urea) in the blood. Depending on the cause, hematuria (blood loss in the urine) and proteinuria (protein loss in the urine) may be noted. Chronic kidney failure is measured in five stages, which are calculated using a patient's GFR, or glomerular filtration rate, see Table 1.

Table 1. Chronic kidney failure stages

Stage	Description	GFR(ml, min, 1.73m^2)
1	Kidney damage with normal or increase GFR	≥ 90
2	Kidney damage with mild decrease GFR	60-90
3	Moderate decrease GFR	30-59
4	Severe decrease GFR	15-29
5	Kidney failure	< 15

Cockcroft and Gault equation

$$\text{GFR for males} = \frac{(140 - age(years)) \times weight(kg)}{72 \times serum\ creatinine(mgldl)}. \tag{1}$$

$$\text{GFR for females} = \frac{(140 - age(years)) \times weight(kg) \times 0.85}{72 \times serum\ creatinine(mgldl)}. \tag{2}$$

The study used data available to the kidneys and urinary tract Center at Mansoura University, during the period from January 2004 to January 2008. The study was selected a random sample of 126 patients.

Table 2. Risk factors for renal failure

P	Patient	Y	Creatinine
X1	Age	X8	Treatment
X2	Sex	X9	Condition
X3	Operation	X10	Diagnosis
X4	Systolic	X11	Blood sugar
X5	Diastolic	X12	NA
X6	Pulse	X13	K
X7	Weight	X14	Calcium

5.2. Statistical software

This study is used the following software:

(1) MINITAB 16 software is used for fitting the ordinary logistic model through the procedure of maximum likelihood estimation.

(2) STATISTICA7 software is used for classifying the data using artificial neural networks model.

5.3. Methods

5.3.1. Ordinal logistic regression

Many variables of interest are ordinal. That is, you can rank the values, but the real distance between categories is unknown. The ordinal logistic method is a generalization of the linear regression method. The ordinal regression method is used to model the relationship between response variables and set of explanatory variables, which can be either categorical or numerical [12]. When a dependent variable is ordinal, we face a quandary. The model is:

$$y_i = x_i\beta + \varepsilon_i. \tag{3}$$

However, since the dependent variable is categorized, we must instead use:

$$c_x(x) = \ln\left[\frac{P(y \leq j|x)}{P(y > j|x)}\right], \tag{4}$$

and

$$\ln\left(\frac{\sum pr(event)}{1 - \sum pr(event)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k. \tag{5}$$

An ordinal logistic model for k predictors with $p-1$ levels response variable.

$$\ln\left(\frac{\sum pr(y \leq j|x)}{1 - \sum pr(y \leq j|x)}\right) = \propto_j + \beta_I x_{i,I},$$

$$i = 1, \ldots, k \quad \text{and} \quad j = 1, 2, \ldots, p-1, \tag{6}$$

where

$\propto_j$ or β_0 = called threshold,

β_I = parameter,

$x_{i,I}$ = sets of factors or predictors,

5.3.2. Artificial neural networks

Artificial neural network models were inspired by the biological sciences which study how the neuroanatomies of living animals have developed in solving problems. According to Nelson and Illingworth [26], ANNs are also called parallel distributed processing models, connectivity models, adaptive systems, self-organizing systems, neuro computing, and neuromorphic systems.

There are usually many inputs. These inputs are commonly called input neurons. There are only hypothetical neurons that produce an output equal to their supposed input. No processing is required by an input neuron. Similar to inputs, there are also one or more output neurons, typically very few.

Signal propagation: The input layer comprises n neurons that code for the n pieces of input signal $(X1 \ldots Xn)$ of the network (independent variables). The number of neurons of the hidden layer is chosen empirically by the user. Finally, the output layer comprises k neurons for the k classes (dependent variables). Each connection between two neurons is associated with a weight factor (random value between -0.3 and $+0.3$ at first); this weight is modified by successive iterations during the training of the network according to input and output data. In the input layer, the state of each neuron is determined by the input variable; the other neurons (hidden layer and output layer) evaluate the state of the signal from the previous layer as:

$$a_j = \sum_{i=1}^{I} x_i w_{ji},$$

where a_j is the net input of neuron j; x_i is the output value of neuron i of the previous layer; w_{ji} is the weight factor of the connection between neuron i and neuron j. The activity of neurons is usually determined via a sigmoid function:

$$f(a_j) = \frac{1}{1 + \exp^{-a_j}}.$$

Thus, weight factors represent the response of the NN to the problem being faced.

Training the network: The back propagation technique is akin to supervised learning as the network is trained with the expected reply: replies. Each iteration modifies the connection weights in order to minimize the error of the reply (expected value-estimated value). Adjustment of the weights, layer by layer, is calculated from the output layer back to the input layer. This correction is made by

$$\Delta w_{ji} = \eta \delta_j f(a_j),$$

where Δw_{ji} is the adjustment of weight between neuron j and neuron I from the previous layer; $f(a_j)$ is the output of neuron i, η is the learning rate, and δ_j depends on the layer. For the output layer, δ_j is:

$$\delta_j = (y_j - \hat{y}_j) f'_j(a_j),$$

where y_j is the expected value (observed value) and $\hat{y}_j$ is the current output value (estimated value) of neuron j. For the hidden layer, δ_j is:

$$\delta_j = f'_j(a_j) \sum_{k=1}^{K} \delta_k w_{kj},$$

where K is the number of neurons in the next layer. The learning rate plays an important role in training. When this rate is low, the convergence of the weight to an optimum is very slow, when the rate is too high, the network can oscillate, or more seriously it can get stuck in a local minimum (Gallant [15]). To reduce these problems, a momentum term α is used and Δw_{ji} becomes:

$$\Delta w_{ji} = \eta \delta_j f(a_j) + \propto \Delta w_{ji}^{prev},$$

where Δw_{ji}^{prev} denotes the correction in the previous iteration.

The training, performed on a representative data set, runs until the sum squared of errors (SSE) is minimized:

$$SSE = \frac{1}{2} \sum_{p=1}^{P} \sum_{j=1}^{N} (y_{pj} - \hat{y}_{pj})^2,$$

where y_{pj} is the expected output value, $\hat{y}_{pj}$ is the estimated value by the network, $j = 1, \ldots, N$ is the number of records and $p = 1, \ldots, P$ is the number of neurons in the output layer.

Testing the network: After training, the performance of the network has to be tested. As in discriminate analysis, a first indication is given by the percentage of correct classifications of the training set records. Nevertheless, the performance of the network with a test set is more relevant. In the test step, the input data are fed into the network and the desired values are compared to the network's output values. The agreement or disagreement of the results thus gives an indication of the performance of the trained network. We have the following types of networks:

Multi-layer perceptron network: According to Eberhart & Dobbins [9], multi-layer perceptron (MLP) is feed forward networks with one or more hidden layers. A MLP becomes a nonparametric sieve if the number of hidden neurons is allowed to increase with the sample size.

Linear neural networks implement a basic linear model, used principally for regression. Linear models are equivalent to simple forms of neural network, with no hidden layer.

Radial basis function (RBF) networks combine a single radial hidden layer with a dot product output layer. The hidden layer neurons act as cluster centers, grouping similar training cases, and the output layer forms a discriminate function or regression. Since the clustering transformation is non-linear, a linear output layer is sufficient to perform an overall non-linear function. RBF networks use a two stage training process -first, assignment of the radial centers and their deviations; second, optimization of the output layer.

Generalized regression neural network (GRNN) is a type of neural network using kernel based approximation to perform regression. GRNN work in a similar fashion to probabilistic neural networks (PNNs). The GRNN copies the training cases into the network to be used to estimate the response on new points.

6. Results

6.1. Ordinal logistic regression model

6.1.1. Interpreting the results

● Response information displays the number of observations that fall into each of the response categories, and the number of missing observations. The ordered response values, from lowest to highest, are shown. Here, we use the default coding scheme which orders the values from lowest to highest: 1 is $\geq$ 90 (kidney damage with normal or increase glomerular filtration rate) $(\text{GFR})(\text{ml/min}/1.73\text{m}^2)$, 2 = 60 to 89 $(\text{GFR})(\text{ml/min}/1.73\text{m}^2)$, 3 = 30 to 59 $(\text{GFR})(\text{ml/min}/1.73\text{m}^2)$, 4 = 15 to 29 $(\text{GFR})(\text{ml/min}/1.73\text{m}^2)$, and 5 < 15 $(\text{GFR})(\text{ml/min}/1.73\text{m}^2)$, see Table 3. Factor information displays all the factors in the model, the number of levels for each factor, and the factor level values. The factor level that has been designated as the reference level is first entry under values.

Table 3. Ordinal logistic regression: ckd (dep.) versus X2; X3; ...

Link Function: Logit		
Response Information		
Variable	Value	Count
ckd (dep.)	1	1
	2	6
	3	15
	4	37
	5	67
	Total	126

● Logistic regression table shows the estimated coefficients, standard error of the coefficients, *z*-values, and *p*-values*,* see Table 4.

Table 4. Logistic regression output

Predictor	Coef	SE Coef	Z	P	Odds Ratio	Lower	Upper
Const(1)	– 12.0224	6.23902	– 1.93	0.054			
Const(2)	– 9.98981	6.16332	– 1.62	0.105			
Const(3)	– 8.54479	6.14201	– 1.39	0.164			
Const(4)	– 6.68638	6.12513	– 1.09	0.275			
X2	– 0.444429	0.378263	– 1.17	0.240	0.64	0.31	1.35
X3	0.0612419	0.461787	0.13	0.894	1.06	0.43	2.63
X4	0.0113170	0.0136855	0.83	0.408	1.01	0.98	1.04
X5	– 0.0341705	0.0212322	– 1.61	0.108	0.97	0.93	1.01
X6	– 0.0018833	0.0168451	– 0.11	0.911	1.00	0.97	1.03
Z1	– 0.0010064	0.0053928	– 0.19	0.852	1.00	0.99	1.01
Z2	0.0803270	0.0421898	1.90	0.057	1.08	1.00	1.18
Z3	– 0.761116	0.229597	– 3.32	0.001	0.47	0.30	0.73
Z4	– 0.0455957	0.194901	– 0.23	0.815	0.96	0.65	1.40
X8	1.68173	0.409384	4.11	0.000	5.37	2.41	11.99
X9	– 0.950128	0.540038	– 1.76	0.079	0.39	0.13	1.11
Log-Likelihood = – 122.172							
Test that all slopes are zero: G = 41.019				DF = 11		*P*-Value = 0.000	

When we use the logit link function, we see the calculated odds ratio, and a 95% confidence interval for the odds ratio, see Table 4.

• The values labelled Const(1), Const(2), Const(3), and Const(4) are estimated intercepts for the logits of the cumulative probabilities of survival for ≥ 90 (GFR)(ml/min/1.73m^2), and for 60 to 89 (GFR)(ml/min/1.73m^2), and for 30 to 59 (GFR)(ml/min/1.73m^2), and for 4 = 15 to 29 (GFR) (ml/min/1.73m^2), respectively. Because the cumulative probability for the last response value is 1, there is no need to estimate an intercept for < 15 (GFR) (ml/min/1.73m^2).

• The coefficient of – 0.444429 for X2 is the estimated change in the logit of the cumulative ckd probability When the X2 is 0 compared to X2 being 1, with the covariate other variables held constant. Because the *p*-value for estimated coefficient is 0.240, there is insufficient evidence to conclude that X2 has an effect upon ckd.

• There is one estimated coefficient for each covariate, which gives parallel lines for the factor levels. Here, the estimated coefficient for the covariate, Z3, X8, are 0.229597, 0.409384, with a *p*-value of < 0.05. The *p*-value indicates that for most α-levels, there is sufficient evidence to conclude that Z3, X8 affect ckd. The positive coefficient, and an odds ratio that is greater than one indicates that higher Z3, X8 levels tend to be associated with higher values of ckd.

• Next displayed is the last log-likelihood from the maximum likelihood iterations along with the statistic G. This statistic tests the null hypothesis that all the coefficients associated with predictors equal zero versus at least one coefficient is not zero. In our results, $G = 41.019$ with a *p*-value of 0.000, indicating that there is sufficient evidence to conclude that at least one of the estimated coefficients is different from zero.

● Goodness-of-fit tests displays both Pearson and deviance good ness of-fit tests. In our results, the p-value for the Pearson test is 0.968, and the p-value for the deviance test is 1.000, indicating that there is insufficient evidence to claim that the model does not fit the data adequately. If the p-value is less than selected α-level = 0.05, the test rejects the null hypothesis that the model fits the data adequately. If the p-value is less than the accepted α-level, the test would reject the null hypothesis of an adequate fit, see Table 5.

Table 5. Goodness-of-fit tests

Method	Chi-Square	DF	P
Pearson	432.804	489	0.968
Deviance	244.345	489	1.000

● Measures of association display a table of the number and percentage of concordant, discordant and tied pairs, and common rank correlation statistics. These values measure the association between the observed responses and the predicted probabilities. Based on the model, a pair is concordant if the individual with a low pulse rate has a higher probability of having a low pulse, discordant if the opposite is true, and tied if the probabilities are equal. In this study, 76.7% of pairs are concordant and 22.8% are discordant. We can use these values as a comparative measure of prediction, for example in comparing fits with different sets of predictors or with different link functions. Somers' D, Goodman-Kruskal Gamma, and Kendall's Tau-a are summaries of the table of concordant and discordant pairs. These measures most likely lie between 0 and 1 where larger values indicate that the model has a better predictive ability. The measure range from 0.33 to 0.54 which implies less than desirable predictive ability, see Table 6.

Table 6. Measures of association (between the response variable and predicted probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	3740	76.7	Somers' D = 0.54
Discordant	1112	22.8	Goodman-Kruskal Gamma = 0.54
Ties	26	0.5	Kendall's Tau-a 0.33
Total	4878	100.0	

6.2. Artificial neural networks models

6.2.1. Interpreting the results

We have investigated the accuracy of the neural network classification for the MLP, PNN, RBF, linear neural network methods is compared, for knows witch neural networks gives the high performance of classifying the data imports. The training performance rate and the error rate for each neural network show in Table 7. The PNN and RBF give the highest train performance result followed by MLP and Linear with 98.5% train performance. The PNN results give a lower classification error during the implantation than the RBF. The MLP take a longest time to train the neural network and classify the data with 3 minute and 22 second, while the rest of neural networks take a 4 second in this task.

Table 7. Neural network classification performance and error

Profile	Train Perf.	Select Perf.	Test Perf.	Train Error	Select Error	Test Error	Input	Hidden (1)	Hidden (2)
MLP 12:12-14-14-5:5	0.984127	1.000000	1.000000	0.173449	0.318376	0.356457	12	14	14
PNN 12:12-63-6-5:5	1.000000	0.967742	1.000000	0.249716	0.377968	0.372309	12	66	0
RBF 12:12-1-5:5	1.000000	1.000000	0.967742	0.345865	0.309624	0.367240	12	1	0
Linear 12:12-5:5	0.984127	1.000000	1.000000	0.287949	0.358013	0.363730	12	0	0

● *Multilayer perceptron* (*MLP*): The MLP model consist 4 layer; 14 inputs, 1 output, and 14 hidden neurons in layer 1 and 14 hidden neurons also in layer 2, determined the 0.98 training performance, and 0.17 training error show in Figure 1 and Table 8.

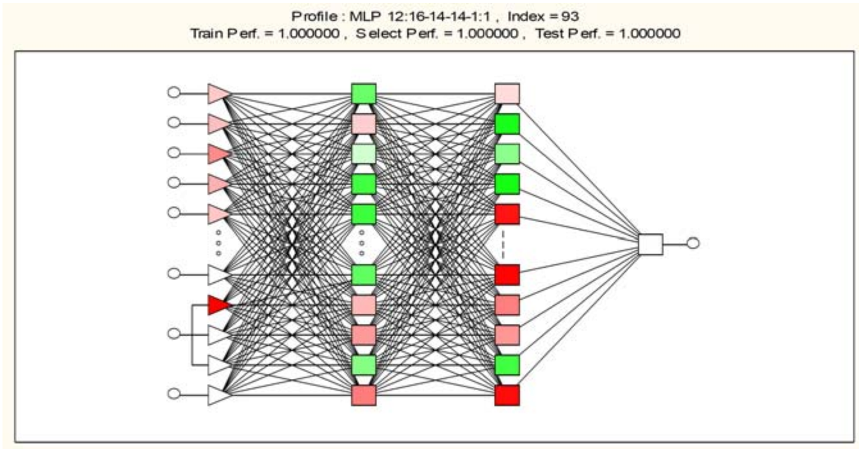

Figure 1. The MLP model.

Table 8. The MLP model results

	Profile	**Train Perf.**	**Select Perf.**	**Test Perf.**	**Train Error**
93	MLP 12:16-14-14-1:1	1.000000	1.00000	1.00000	0.00000
Select Error	Test Error	Training/Members	Inputs	Hidden (1)	Hidden (2)
0.00000	**0.00000**	**BP100,CG6b**		**12**	**14**

● *Linear neural network* (*LNN*): LNN institute with 2 layer; 12 inputs, 1 output, without hidden layer, shows in Figure 2. That will resolute the full training performance, selecting performance and testing performance and 0 training error, shows in Table 9.

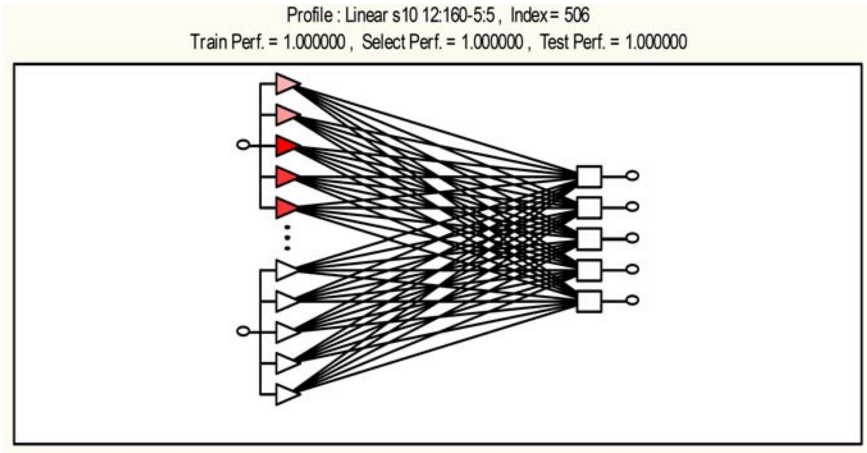

Figure 2. Linear neural network model.

Table 9. The LNN model result

	Profile	**Train Perf.**	**Select Perf.**	**Test Perf.**	**Train Error**
506	Linears 10:12:160-5:5	1.000000	1.000000	1.000000	0.287949
Select Error	Test Error	Training/Members	Inputs	Hidden (1)	Hidden (2)
5.067857E-01	**5.254003E-01**	**PI**	**12**	**0**	**0**

● *Radial basis function* (*RBF*): RBF neural network established with 3 layers; 12 inputs, 1 output, and one hidden layer with 11 neurons, shows in Figure 3 and Table 10.

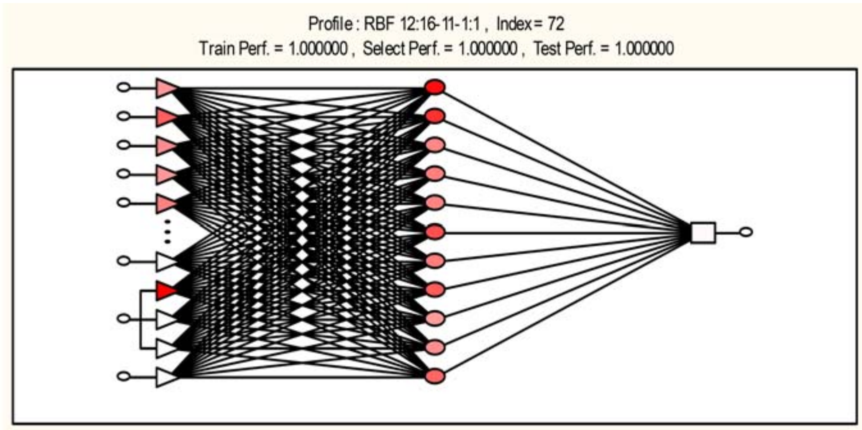

Figure 3. The RBF model.

Table 10. The RBF model result

	Profile	Train Perf.	Select Perf.	Test Perf.	Train Error
72	RBF 12:16-11-1:1	1.000000	1.000000	1.000000	0.082040
Select error	Test error	Training/members	Inputs	Hidden (1)	Hidden (2)
0.202727	0.138416	KM, KN, PI	12	11	0

● *Generalized regression neural network* (*probabilistic neural networks*): The PNN establish with 3 layer; 12 inputs, 1 output, and one hidden layer with 66 neurons in hidden layer, shows in Figure 4 and Table 11.

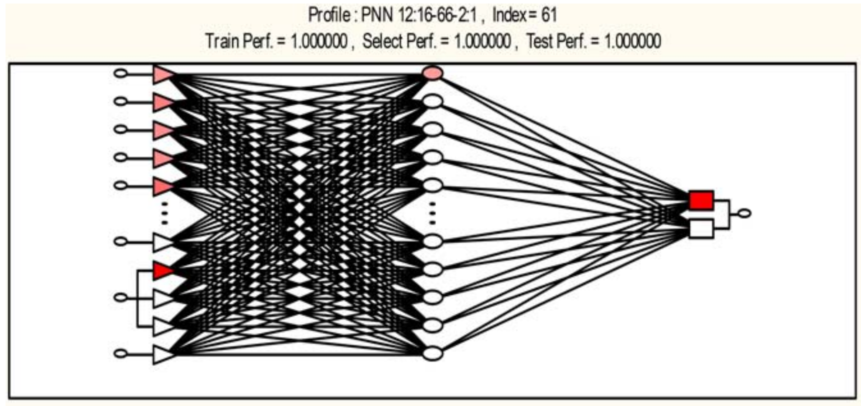

Figure 4. The PNN model.

Table 11. The PNN result

	Profile	Train Perf.	Select Perf.	Test Perf.	Train Error
61	PNN 12:16-66-2:1	1.000000	1.000000	1.000000	0.000158
Select Error	Test Error	Training/Members	Inputs	Hidden (1)	Hidden (2)
0.026999	**0.053660**		**12**	**66**	**0**

● *Sensitivity analysis*: The sensitivity analysis indicates which input variable is most important by the neural network models. The sensitivity is the ratio between the error with omission and the baseline error and ranks the variables in the order of importance, show in Table 12.

Table 12. Sensitivity analysis

Variables	Gender	Is there other operations	Systolic blood pressure	Diastolic blood pressure	Pulse	Blood sugar	Sodium	Potassium	Calcium	Summary of treatment	The patient's condition when you go out	Final diagnosis
MLP	1.106586	1.168614	1.06442	1.149585	1.07814	1.01046	1.100221	1.200451	1.04871	1.129099	1.090193	1.117668
	6	2	10	3	9	12	7	1	11	4	8	5
PNN	1.008271	1.009980	1.00650	1.011080	1.00644	1.00260	1.007148	1.010176	1.007808	1.012093	1.024715	1.031362
	7	6	10	4	11	12	9	5	8	3	2	1
RBF	0.997655	0.98317	0.998774	1.000275	1.001142	0.99700	1.000646	0.998507	0.998688	0.99387	1.006572	1.005676
	9	12	6	5	3	10	4	8	7	11	1	2
LNN	1.009155	0.99617	1.009635	1.037911	1.003760	0.99945	0.99456	1.013583	1.004372	1.016807	1.003043	1.036626
	6	11	5	1	8	10	12	4	7	3	9	2

7. Conclusion

The current study is focused on diagnosis of chronic kidney disease (CKD). Predicting the renal failure progression with reasonable accuracy is necessary because early diagnosis of the disease is prime significance to control and manage the consequences. The previous studies focused on demonstrated the ability of ANN models in applications binary classifications, but the current study focused on demonstrated the ability of ANN models in applications ordinal data with multiple classifications. We have investigated the accuracy of the neural network classification for the following types of networks; multilayer perceptron, linear neural network, radial basis function, generalized regression neural network.

The result shows the following:

● The most affected variables are gender, surgical operations, the blood pressure and potassium ratio.

● Successful of ANN models in the process of separating and classifying accuracy rate of almost 100%, and the ordinal regression model succeeded in identifying risk factors for renal failure moral influence on the regression model.

References

[1] A. A. Adepoju and M. Adegbite, Application of ordinal logistic regression model to occupational data, J. Sci. Ind. Stud. 7 (2009), 39-49.

[2] A. Agresti, Analysis of Ordinal Categorical Data, Wiley, 2nd Edition, 2010.

[3] I. E. Allen and C. A. Seaman, Likert scales and data analysis, Quality Progress 40(7) (2007), 64-65.

[4] M. Catalogna, E. Cohen, S. Fishman, Z. Halpern, U. Nevo and E. Ben-Jacob, Artificial neural networks based controller for glucose monitoring during clamp test, PloS One. 7: e44587, 2012.

[5] J. R. Chimka and H. Wolfe, History of ordinal variables before 1980, Scientific Research and Essay 4(9) (2009), 853-860.

[6] W. Deng, W. Chen and W. Pei, Back-propagation neural network based importance-performance analysis for determining critical service attributes, Expert Systems with Applications 34(2) (2008), 1115-1125.

[7] Y. Dong, Logistic Regression Models for Ordinal Response, Texas Medical Center Dissertations, 2007,

http://digitalcommons.library.tmc.edu/dissertations/AA11445931.

[8] P. Dey, A. Lamba, S. Kumari and N. Marwaha, Application of an artificial neural network in the prognosis of chronic myeloid leukemia, Anal. Quant. Cytol. Histol. 33 (2012), 335-339.

[9] R. C. Eberhart and R. W. Dobbins, Neural Network PC Tools: A Practical Guide, Academic Press, San Diego, California, USA, 1990.

[10] E. Elveren and N. Yumuşak, Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm, J. Med. Syst. 35 (2011), 329-332.

[11] O. Er, F. Temurtas and A. Tanrýkulu, Tuberculosis disease diagnosis using artificial neural networks, J. Med. Syst. 34 (2008), 299-302.

[12] Kevin L. Fox, Rhonda R. Henning and Jonathan H. Reed, A Neural Network Approach Towards Intrusion Detection, In Proceedings of the 13th National Computer Security Conference (1990).

[13] P. Fedor, I. Malenovsky, J. Vanhara, W. Sierka and J. Havel, Thrips (Thysanoptera) identification using artificial neural networks, Bull. Entomol. Res. 98 (2008), 437-447.

[14] J. Fernandez de Canete, S. Gonzalez-Perez and J. C. Ramos-Diaz, Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes, Comput. Meth. Progr. Biomed. 106 (2012), 55-66.

[15] S. I. Gallant, Neural Network Learning and Expert Systems, MIT Press, Cambridge, MA, 1993.

[16] T. G. Gregoire and B. L. Driver, Analysis of ordinal data to detect population differences, Psychological Bulletin 101(1) (1987), 159-165.

[17] T. J. Hastie, J. L. Botha and C. M. Schnitzler, Regression with an ordered categorical response, Statistics in Medicine 8(7) (1989), 785-794.

[18] J. Havel, E. Peña, A. Rojas-Hernández, J. Doucet and A. Panaye, Neural networks for optimization of highper formance capillary zone electrophoresis methods, J. Chromatogr. A 793 (1998), 317-329.

[19] S. Jamieson, Likert scales: how to (ab)use them, Medical Education 38(12) (2004), 1217-1218.

[20] M. G. Karlaftis and E. I. Vlahogianni, Statistics versus neural networks in transportation research: Differences, similarities and some insights, Submitted to Transportation Research Part C: Emerging Technologies 19(3) (2011), 387-399.

[21] T. R. Knapp, Treating ordinal scales as interval scales: An attempt to resolve the controversy, Nurs. Res. 39(2) (1990), 121-123.

[22] L. Mayer, A note on treating ordinal data as interval data, American Sociological Review 36(3) (1971), 519-520.

[23] V. Michalkova, A. Valigurova, M. Dindo and J. Vanhara, Larval morphology and anatomy of the parasitoid Exorista larvarum (Diptera: Tachinidae), with an emphasis on cephalopharyngeal skeleton and digestive tract, J. Parasitol. 95 (2009), 544-554.

[24] N. Murarikova, J. Vanhara, A. Tothova and J. Havel, Polyphasic approach applying artificial neural networks, molecular analysis and post abdomen morphology to West Palaearctic Tachina spp. (Diptera, Tachinidae), Bull. Entomol. Res. 101 (2010), 165-175.

[25] National Kidney Foundation, K/DOQI clinical practice guidelines for chronic kidney disease, Retrieved 2008-06-29 (2002).

[26] M. Nelson and W. T. Illingworth, A Practical Guide to Neural Nets, Addison-Wesley, Reading, MA, 1990.

[27] M. Paliwal and U. A. Kumar, Neural networks and statistical techniques: A review of applications, Expert Systems with Applications 36(1) (2009), 2-17.

[28] Paulo J. Lisboa and Azzam F. G. Taktak, The use of artificial neural networks in decision support in cancer: A systematic review- School of Computing and Mathematical Science, Liverpool John Moores University, Liverpool, UK, Department of Clinical Engineering, Royal Liverpool University Hospital, Liverpool, UK, Neural Networks-Volume 19(4) (2006); M. G. Penedo, M. J. Carreira, A. Mosquera and D. Cabello, Computer-aided diagnosis: a neural-network-based approach to lung nodule detection, IEEE Trans. Medical Imaging 17(6) (1998).

[29] Galdón B. Rodríguez, E. Pená-Méndez, J. Havel, Rodríguez E. Rodríguez and Romero C. Díaz, Cluster analysis and artificial neural networks multivariate classification of onion varieties, J. Agric. Food. Chem. (2010), 11435-11440.

[30] Tutz, Regression for Categorical Data, Volume 34, Cambridge University Press, 2012.

[31] P. F. Velleman and W. Leland, Nominal, ordinal, interval, and ratio typologies are misleading, The American Statistician 47(1) (1993), 65-72.

[32] H. White, Some asymptotic results for learning in single hidden layer feed forward network models, Journal of American Statistical Association 84 (1989).

■