Websites:
http://www.sciencepub.net
http://www.sciencepub.net/report

Emails:
editor@sciencepub.net
reportopinion@gmail.com

Report and Opinion

MARSLAND PRESS
Multidisciplinary Academic Journal Publisher

A Comparative Study of Multinomial Logistic Regression and Artificial Neural Network Classifier with Application to Patient Data

Ejeh, Ominyi Samuel and Alabi, Olatayo Olusegun

dataspace1011@gmail.com

Abstract: Multinomial Logistic Regression (MLR) model and Artificial Neural Network (ANN) are widely used in predictive studies of data for various diseases. This research compares the performance of MLR models and ANN (Multilayer Perceptron, MLP) models using TB/HIV co-infected patients' data. The tools used are Goodness of fit, Pseudo R 2 measurement, Likelihood ratio test, Akaike Information Criterion (AIC), Bayesian Information Criteria, Wald test, and Classification comparison of all patients' variables. The effect of the three regimens does not have significant to TB preventive therapy for TB/HIV co-infected adults. The research established that ANN model classifies the dataset better than the MLR because overall prediction percentage was absolutely better. Moreover, after classifying both the MLR and ANN (MLP), it was observed that the training and testing process through Multilayer Perceptron of the dataset reveals that ANN performed better than the MLR.

[Ejeh, Ominyi Samuel and Alabi, Olatayo Olusegun. **A Comparative Study of Multinomial Logistic Regression and Artificial Neural Network Classifier with Application to Patient Data.** Rep Opinion 2021;13(11):37-49]. ISSN 1553-9873 (print);ISSN 2375-7205 (online). http://www.sciencepub.net/report. 7. doi:10.7537/marsroj131121.07.

Keywords: Multinomial logistic regression, Artificial neural network, machine learning

Introduction

Predictive models are used in a variety of medical domains for diagnostic and prognostic tasks. These models are built from ''experience'', which constitutes data acquired from actual cases. The data can be preprocessed and expressed in a set of rules, such as it is often the case in knowledge-based expert systems or serve as training data for statistical and machine learning models. This area of statistics deals with the use of data and machine learning algorithms, predicting the likelihood of future outcomes based on past data. Among the options in the latter category, the most popular models in medicine are logistic regression (LR), Multinomial Logistics Regression (MLR) and Artificial Neural Networks (ANN). These models have their origins in two different communities (statistics and computer science) but share many similarities. Predictive analytics is useful at every step in a patient's journey, including diagnosis, prognosis, and treatment. Predictive analytics can also inform remote patient monitoring and reduce adverse events. On a more macro level, predictive analytics can improve care quality while reducing costs. Today's healthcare organizations face increasing pressure to achieve better care coordination and improve patient care outcomes. To accomplish these results, organizations are turning to predictive analytics.

Classification is one of the important areas of research in the field of data mining and neural networks is one of the widely used techniques for classification. Neural networks have emerged as an important tool for classification. The recent vast research activities in neural classification have established that neural networks are a promising alternative to various conventional classification methods. The advantage of neural networks lies in the following theoretical aspects. First, neural networks are data driven self-adaptive methods in that they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal function approximators in that neural networks can approximate any function with arbitrary accuracy.

Multinomial logistic regression is also a classification algorithm like the logistic regression for binary classification. Logistic regression methods are common statistical tools for modelling discrete response variables such as binary, categorical, and ordinal responses. If the values of the response of these variables refer to different categories where a specific group of elements (called a sample), can be grouped into a class according to a series of characteristics which have previously been measured for each element, we are confronted with what is called a classification problem. Multinomial Logistic Regression is a classification technique that extends the logistic regression algorithm to solve multiclass possible outcome problems, given one or more independent variables. This model is used to predict the probabilities of a categorical dependent variable, which has two or more possible outcome classes. Whereas the logistic regression model is used when the dependent categorical variable has two outcome classes for example, students can either "Pass" or

"Fail" in an exam or bank manager can either "Grant" or "Reject" the loan for a person. Multinomial logistic regression is used to model nominal outcome variables, in which the log odds of the outcomes are modeled as a linear combination of the predictor variables. Therefore, the aim of this study is to compare the performance of the Multinomial Logistic regression models and the Artificial Neural Network in analyzing Tuberculosis/HIV co-infected patients.

Literature Review

Soo-Seon (2005), the study demonstrated the power of neural networks by comparing its predictive capability with that of a logit model in predicting hospitality bankruptcy. From empirical results of the two methodologies, it was shown that neural networks obtained a higher accuracy rate than did a logit model in an in-sample test as well as in holdout (testing) sample test. This result confirmed previous assertions made by many researchers stating the superiority of neural networks over logit models in classification and prediction tasks.

Khoshnevisan *et al* (2000) proposed a classification scheme to isolate truly benign tumors from those that initially start off as benign but subsequently show metastases. A non-parametric artificial neural network methodology has been chosen because of the analytical difficulties associated with extraction of closed-form stochastic-likelihood parameters given the extremely complicated and possibly non-linear behavior of the state variables. This was intended as the first of a three-part research output. They proposed and justified the computational schema. In the second part they set up a working model of our schema and pilot-test it with clinical data while in the concluding part we shall give an in-depth analysis of the numerical output and model findings and compare it to existing methods of tumor growth modeling and malignancy prediction.

Jaime and Javier (2012) results obtained in this thesis suggest the capability of ANN-based models to solve regression or classifications problems of the agricultural and industrial fields. Three main conclusions can be extracted from these results. The first one was that models proposed in scientific literature can be improved considering a set of variables larger or different than the proposed by other authors: for example, the tobacco drying model considers a measurement point inside the product to be dried, the switch grass drying model employs information about the rain events, the vibration model of a machine uses information from a measurement point placed far away from the rotary components considered, and the NDT model employs impedance data at different frequencies.

Divyansh *et al* (2015) this research does a comparative study of commonly used machine learning algorithms in predicting the prevalence of heart diseases. It uses the publicly available Cleveland Dataset and models the classification techniques on it. It brings up the differences between different models and evaluates their accuracy in predicting a heart disease. They have shown that less complex models such as logistic regression and support vector machines with linear kernels give more accurate results than their more complex counterparts. They have used F1 score and ROC curves as evaluative measures. Through this effort, we aim to provide a benchmark and improve earlier ones in the field of heart disease diagnostics using machine learning classification techniques.

Caliskan and Sevim (2018) said Artificial Neural Networks (ANN) was frequently used as a modelling tool in the analysis of complex problems that have been used to solve this issue. The aim of their study was to investigate the feasibility of ANN's including Multilayer Perceptron (MLP), Cascade Forward Back Propagation (CFBP) and compared the predictions for total time during log skidding operations stations in Eastern Black sea region (Giresun Forest District Directorates) of Turkey with those of the Multiple Regression Analysis (MRA). Moreover, standard times are calculated, and affective factors are determined, after which the effectiveness levels are evaluated at each working stage by way of timing determinations when skidder is used for timber extraction. The comparison of models was carried out by using the correlation coefficient, mean squared error, root mean square errors and mean absolute error. The comparison results indicate that MLP and CFBP models are better at predicting the total time during log skidding operations in comparison with the MRA model. These results have put forth that artificial neural networks have a greater prediction in comparison with multiple regression analysis for predicting the skidding time in timber extraction operations and that less erroneous results are obtained. It was observed that artificial neural networks can be preferred in cases for which the multiple regression analysis predictions have not been met and the analysis cannot be performed.

Paulo *et al* (2004) discussed Neural networks and logistic regression being among the most widely used AI techniques in applications of pattern classification. Much has been discussed about if there is any significant difference in between them but much less has been done with real-world applications data (large scale) to help settle this matter, with a few exceptions. The research presented a performance comparison between these two techniques on the market application of credit risk assessment, making use of a large database from outstanding credit bureau financial institutions (a sample of 180,000 examples). The comparison was carried out through a 30-fold stratified cross-validation process to define the confidence intervals for the performance evaluation. Several metrics were applied both on the optimal decision point and along the continuous output domain. The statistical tests showed that multilayer perceptions perform better than logistic regression at 95% confidence level, for all the metrics used.

Guoqiang (2000) summarized some of the most important developments in neural network classification research. Specifically, the issues of posterior probability estimation, the link between neural and conventional classifiers, learning and generalization tradeoff in classification, the feature variable selection, as well as the effect of misclassification costs are examined. Their purpose was to provide a synthesis of the published research in this area and stimulate further research interests and efforts in the identified topics.

Methodology

Study Population

The population target for this study comprised of all Patients with Tuberculosis related cases/issues in the DOTs Clinic of NIMR who had been registered between 2011 and 2020. The research design is a cross sectional design.

Study Site

The study was carried out at the DOTs Clinic of the Nigerian Institute of Medical Research (NIMR). A parastatal under the Federal Ministry of Health that has treated over 5000 TB patients in the last 10 years. The Institute has a Directly Observed Treatment Short Course (DOTS) Centre where it attends to patients infected with TB.

Sampling Procedure/Sample Size

All patients that were enrolled between 2011 and 2020 was included in the study; this will enable completion of the 8months treatment cycle for those enrolled in 2020.

Methods Used in the study

The two methods considered in this research work are as follows:

i. Artificial Neural Network Classifier and
ii. Multinomial Logistic Regression Mode

Artificial Neural Network

Artificial Neural Networks (ANNs) are processing models inspired by how the brain works, and they are one of the main data analysis techniques in the scientific literature. An ANN is composed by a set of neurons and the connection that joins these neurons, and it is characterized by having a good processing power despite the simplicity of the neurons that compose the ANN (Haykin, 1999).

Supervised learning Multilayer neural networks

A multilayer perceptron is a feedforward neural network with one or more hidden layers. network with one or more hidden layers. The network consists of an input layer of source neurons, at least one middle or hidden layer of computational neurons, and an output layer of computational neurons. The input signals are propagated in a forward direction on a layer direction on a layer-by-layer basis.

Multilayer neural networks are generally much more versatile than single neurons, with no linear separability requirement. Training is less obvious and potentially more time consuming.

The most common type of multilayer perceptron known as:

MLP (Multi-Level Perceptron) Backpropagation Network (alluding to a common method of training these networks; other training methods could conceivably be used.)

A hidden layer

A hidden layer “hides” its desired output. Neurons in the hidden layer cannot be observed through the input/output behavior of the network. There is no obvious way to know what the desired output of the hidden layer should be.

Commercial ANNs incorporate three and sometimes four layers, i one or two hidden layers. Each layer can contain from 10 to 1000 neurons. Experimental neural networks may have five or even six layers, including three or four hidden layers and utilize millions of neurons

Backpropagation

Werbos, in his Harvard PhD thesis in 1974 found a method while Rumelhart and McClelland, in 1985 discovered the method, presumably independently, and popularized it under the current name. In mathematics, such methods are in the category of “optimization”.

Learning in a multilayer network proceeds the same way as for a same way as for a perceptron. A training set of input patterns is presented to the network. The network computes its output pattern, and if there is an error there is an error – or in other words a difference between actual and desired output patterns − the weights are adjusted to reduce this error.

In a back-propagation neural network, the learning algorithm has two phases

First, a training input pattern is presented to the network input layer. The network propagates the input pattern from layer to layer until the output pattern is generated by the output layer.

If this pattern is different from the desired output, an error is calculated and then propagated backwards through the network from the output layer to the input layer. The weights are modified as the error is propagated.

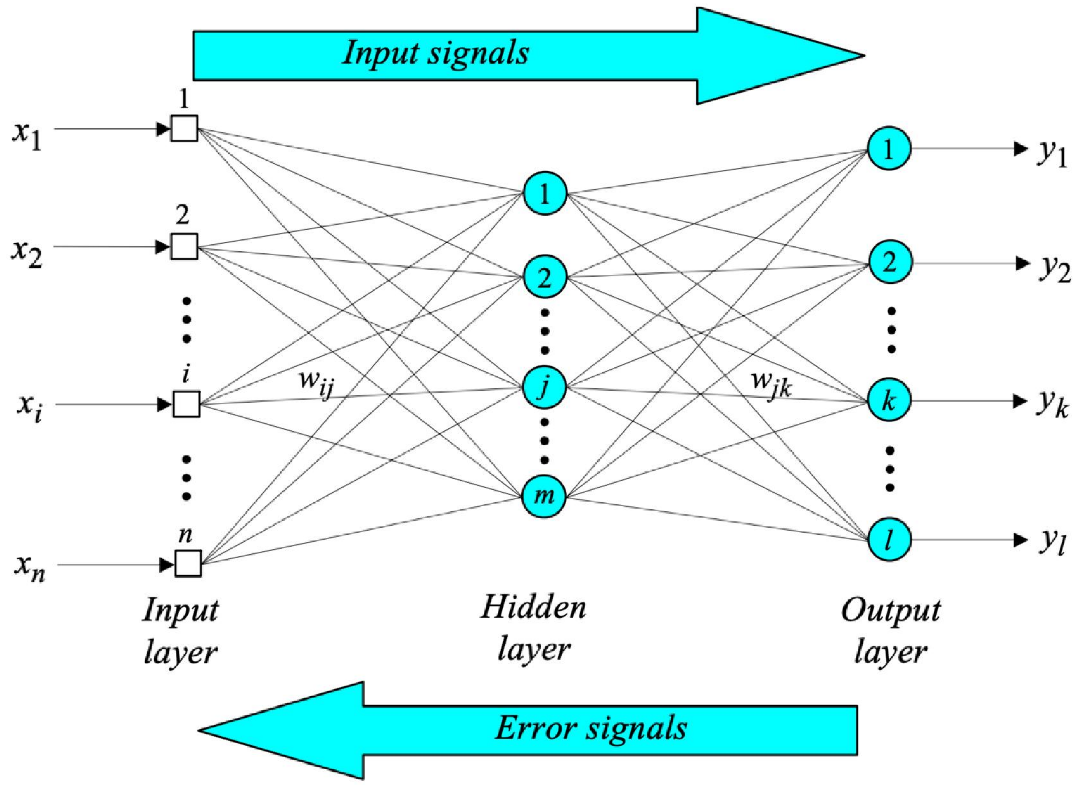

Fig. 1: *Backpropagation training cycle training cycle*

For forward propagation we derive the activation values (the inputs to the activation functions) at each neuron, and the final output. We then compute the error in the output and backpropagate the error through the network to get "sensitivities" at each neuron. (The gradient approximation is derivable from the sensitivities.) Using the sensitivities to derive weight changes, we then apply the weight changes.

Backpropagation is mathematically a lot like forward propagation. The sensitivities are used instead of signal values. The sensitivities are the partial derivatives of the MSE with respect to the activation values. Basically, both are iterated matrix multiplications.

Recall that the gradient consists of components where J is the mean-squared error and w is some weight (or bias) in the network.

For the Adeline, already derived:

$$\frac{\Delta J}{\Delta w} = -2\sum xi\ f'(n)$$

where xi is the input corresponding to weight wi, and n(net) is the weighted sum. This works for the multi-layer case at the output layer.

The back-propagation training algorithm

Step 1: Initialization: Set all the weights and threshold levels of the network to random numbers uniformly distributed inside a small range:

$$\left(-\frac{2.4}{F_i}, \quad +\frac{2.4}{F_i}\right)$$

where F_i is the total number of inputs of neuron in the network. The weight initialisation is done on a neuron-by-neuron basis

Step 2: Activation Activate the back-propagation neural network by applying inputs x_1(p), x_2 (p),…, x_n (p) and desired outputs $y_{d,1}$(p), $y_{d,2}$ (p),…, $y_{d,n}$ (p).

(a) Calculate the actual outputs of the neurons in the hidden layer:

where n is the number of inputs of neuron j in the hidden layer, and hidden layer, and sigmoid is the sigmoid activation function.

$$y_k(p) = sigmond\left[\sum_{i=1}^{m} x_{jk}(p).w_{jk}(p) - \theta_k\right]$$

where m is the number of inputs of neuron k in the output layer.

Step 3: Weight training Update the weights in the back-propagation network propagating backward the errors associated with output neurons.

(a) Calculate the error gradient for the neurons in the output layer:

$$\delta_k(p) = y_k(p).[1 - y_k(p)].e_k(p)$$

where $e_k(p) = y_{d,k}(p) - y_k(p)$

Calculate the weight corrections:

$$\Delta w_{jk}(p) = \alpha.y_j(p).\delta_k(p)$$

Update the weights at the output neurons:

$$w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p)$$

(b) Calculate the error gradient for the neurons in the hidden layer:

$$\delta_j(p) = y_j(p).\left[1 - y_j(p)\right].\sum_{k=1}^{l} \delta_k(p) w_{jk}(p)$$

Calculate the weight corrections:

$$\Delta w_{ij}(p) = \alpha . x_i(p) . \delta_j(p)$$

Update the weights at the output neurons:

$$w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p)$$

Step 4: Iteration: Increase iteration p by one, go back to Step 2 and repeat the process until the selected error criterion is satisfied. As an example, we may consider the three-layer back-propagation network. Suppose that the network is required to perform logical operation Exclusive-OR. Recall that a single-layer perceptron could not do this operation. Now we will apply the three-layer net.

Logistic Regression

Logistic regression analysis is a common method that has been increasingly used particularly in the social sciences. Logistic Regression is a traditional statistical inference that relates a binary output (such as good/bad, yes/no, l/O etc.) with a set of input variables$(x_1, x_2, \dots, x_k.)$. The dependent variable is the logarithm of the ratio of the probabilities of the two possible outcomes of the output variable, $(log[p/(1-p)])$ and can be turned into:

$$P(Y=1) = [1 + exp(-\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k]^{-1}$$

where Y is the dichotomic output variable, β_0 is the intersection and β_1 $(i = 1, \dots, k)$ are the coefficients corresponding to each explanatory variable x_j. The widespread maximum likelihood method is used here for estimating the binary response having its solution found through Newton-Raphson's iterative process.

Multinomial Logistic Regression (Mlr)

In brief, multinomial logistic regression model is an expanded version of the two-category model (binary model) for j category. Thereby, "$j-1$" multinomial logistic regression models occur (Long and Freese, 2006). It is allowed that the response probabilities to depend on nonlinear transformations of the linear function of equation:

$$x_i' \beta_i = \sum_{k=0}^{K} \beta_{jk} x_{ik}$$

where K is the number of the predictors, i represents ith individual, x represents independent variable and j expresses the category dependent variable. The multinomial logit model can be viewed as an extension of the binary logit model. For example, in case of three categories $(J = 3)$, the models can be written as below:

$$P_{i1} = P(y_i = 1|x_i) = \frac{1}{1+\exp(x_i'\beta_2)+\exp(x_i'\beta_3)}$$

$$P_{i2} = P(y_i = 2|x_i) = \frac{\exp(x_i'\beta_3)}{1+\exp(x_i'\beta_2)+\exp(x_i'\beta_3)}$$

$$P_{i3} = P(y_i = 3|x_i) = \frac{\exp(x_i'\beta_3)}{1+\exp(x_i'\beta_2)+\exp(x_i'\beta_3)}$$

In here, β_2 and β_3 denote the covariate effects specific to the second and third response categories with the first category as the reference. Besides, a reference category (baseline category) is determined at first to compare and analysis. At this stage, the researcher can select the reference category (j) optionally for instance, if there are categories such as 1, 2, and 3 in a dependent variable, 1 can be selected as the reference category. In this way, two different logistic models can be obtained for comparison of 1-2 and 1-3 (Hosmer and Lemeshow, 2000). On the other hand, the equation (3.6) for P_{i1} can be derived from the constraint that the three probabilities sum to 1

$$P_{i1} = 1 - (P_{i2} + P_{i3})$$

The sum of the probabilities of categories of dependent variable should be equal to 1 as in binary logit model. For instance, if the dependent variable has a category three-level structure, the sum of the probabilities for each category will be equal to 1 as follows (Hosmer *et al.* 2013):

$$[P_{i1} = P(y_i = 1|x_i)] + [P_{i2} = P(y_i = 2|x_i)] + [P_{i3} = P(y_i = 3|x_i)] = 1$$

In general, the probabilities of a dependent variable with j categories can be expressed in multinomial logit as below:

$$P_{ij} = P(y_i = j|x_i) = \frac{\exp(x_i'\beta_3)}{1+\sum_{j=2}^{J}\exp(x_i'\beta_j)} \qquad (3.11)$$

If is selected as the baseline category, the probability of the dependent variable to lie within the baseline category is defined as given in Equation 4 (Liao, 1994).

$$P_{i1} = P(y_i = 1|x_i) = \frac{1}{1+\sum_{j=1}^{J-1}\exp(x_i'\beta_j)} \quad j = 1, 2, \dots, J-1$$

Furthermore, the probability to lie within the baseline category can be computed with the help of other probabilities as given in Equation 5, if the other probabilities are known (Liao, 1994).

$$P_j = P(y = J) = 1 - [P(y=1) + P(y=2) + \cdots + P(y = J-1)]$$

In a multinomial logistic regression model, the logit transformation is obtained by taking the logarithms of the odds ratios after selecting the baseline category. For the four-category example, when 0 is selected as the baseline category, the logarithms of odds ratios can be obtained as given in Equation (3.11), Equation (3.12), and Equation (3.13) (Kienbaum and Klein, 2010).

$$In\left[\frac{P(y=1|x_1)}{P(y=0|x_1)}\right] = \beta_1 + \beta_{11} x_1$$

$$In\left[\frac{P(y=1|x_2)}{P(y=0|x_1)}\right] = \beta_2 + \beta_{21} x_1$$

$$In\left[\frac{P(y=1|x_3)}{P(y=0|x_1)}\right] = \beta_3 + \beta_{31}x_1$$

As its seen, the baseline category is taken as "□ = 0" in all three odds ratios. The notation of the model can be generalized as in Equation 9 with all these given (Liao, 1994).

$$In\left[\frac{P_j}{P_J}\right] = In\left[\frac{P(y=j)}{P(y=J)}\right] = (\sum_{k=1}^{K} \beta_{jk}x_k) \quad j = 1, \ldots J-1$$

As Equation (3.15) indicates, multinomial logistic regression model can be transformed into binary logit model for □ = 2.

The multinomial logit model is estimated using maximum likelihood with the log-likelihood function for a sample of n-observations given by

$$1 + \sum_{j=1}^{J-1} \exp(x_i'\beta_j)$$

$$InL = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} log(P_{ij})$$

Where d_{ij} is a dummy variable that takes a value 1 if observation i takes the jth category and 0 otherwise because P_{ij} is a nonlinear function of parameters of the regression model.

Data Analysis and Results

Descriptive and Non-Parametric Analysis

First, descriptive statistics are used to give us information about the distributions of the variables. We get the baseline characteristics in 466 participants using the descriptive statistics (Table 4.1).

Table 4.1: Baseline characteristics in 466 participants

Variable	T	Age	BMI	CD4Count	Creatinine	Haemoglob
Total	3482	17528	162.89	94368.41	46812.18	24597.08
Mean	7.4721	37.6137	0.3495	202.5073	100.4553	52.7834
Std dev	2.2218	12.7125	0.1245	168.5325	68.4812	27.5678
C.V	29.7346	33.7975	35.6223	83.2229	68.1708	52.2282

Table 4.2: Description of Variables

Variable	**Description**	**Codes/values**
AGE	Age	year(s)
GENDER	Patient's sex	0 = Male, 1 = Female
STATUS	Marital status	0 = Single, 1 = Married 2 = Divorce, 3 = Widow
TYPE OF TB	Types of Tuberculosis Disease	Pulmonary = 0 Extra-pulmonary = 1
WEIGHT	Weight	Kg
ART STATUS	HIV positive ARV Status	HIV positive on ART = 0 HIV positive non-ART = 1
TREATMENT	Treatment Status	Completed and cured =0 Died = 1 Defaulted = 2
HAEMOGLOB	Hemoglobin	g/dL
CREAT	Creatinine level	Mg/Dl
CD4COUNT	CD4 cells level measurements	Cell/mm^3
GLUCOSE	Glucose Level	Mg/dL

Multinomial Logistic Regression Analysis (Mlr)

Goodness of Fit Test

Table 4.3: Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	796.997	906	.996
Deviance	231.289	906	1.000

The goodness of fit shows the significance of the model.

Decision

Since P_{-value} (0.996) > ($\alpha = 0.05$) for the Chi-square (Pearson), the effect of the three regimens does not have significant to TB preventive therapy for TB/HIV co-infected adults. Then the model is not different in the population which make the result not significance.

Table 4.4: Pseudo R^2 Measurements from Multinomial Logistics Regression

Measurement	R^2 values
Cox and Snell	.115
Nagelkerke	.165
McFadden	.103

Pseudo R^2 presented in the table are examined, explanation ratios of dependent variables upon independent variables are seen. From the above table, it is seen that dependent variable defined 11.5% of the variance of the independent variable according to Cox and Shell R^2 value, 16.5% according to Nagelkerke R^2 values and 10.3% according to McFadden R^2 value.
Nagelkerke R^2 values is the modified form of Cox and Shell coefficients and so it is always higher than Cox and Shell R^2 value (Garson, Hair *et al*, 2006).

Model Fitting Information

Model fitting information obtained from Multinomial Logistics Regression is presented in the table

Table 4.5: Model Table

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2Log Likelihood	Chi-Square	df	Sig.
Intercept Only	435.724	443.524	431.724			
Final	439.291	540.689	387.291	44.433	24	.007

By including the predictor variables and maximizing the log likelihood of the outcomes seen in the data, the "Final" model should improve upon the "Intercept Only" model. This can be seen in the differences in the -2(Log Likelihood) values associated with the models. The LR Chi-Square statistic can be calculated by -2*L(null model) – (-2*L(fitted model)) = 431.724 – 387.291 = 44.433, where *L(null model)* is from the log likelihood with just the response variable in the model (Intercept Only) and *L(fitted model)* is the log likelihood from the final iteration (assuming the model converged) with all the parameters. In other words, this is the probability of obtaining this chi-square statistic (44.433), or one more extreme, if there is, in fact no effect of the predictor variables. This p-value is compared to a specified alpha level, our willingness to accept a type I error, which is set at 0.05. The small p-value from the LR test, <0.007, would lead us to conclude that at least one of the regression coefficients in the model is not equal to zero.

Artificial Neural Network (Ann)

Table 4.6: The Case Processing Summary for Training and Testing of Network Data Exposure

Case Processing Summary

		N	Percent
Sample	Training	342	73.4%
	Testing	124	26.6%
Valid		466	100.0%
Excluded		0	
Total		466	

The case processing summary shows the number of samples that are used in training the model about 73.4% (342) and the 26.6% (124) which have not been seen by the model are used for testing

Model Summary for Ann (Multilayer Perceptron)

Table 4.7: Model Summary Table

Model Summary

Training	Sum of Squares Error	22.767
	Percent Incorrect Predictions	7.9%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.35
Testing	Sum of Squares Error	6.197
	Percent Incorrect Predictions	5.6%

Dependent Variable: Treatment outcome

a. Error computations are based on the testing sample.

The sum of squares Error for Training and Testing (22.767 and 6.197) respectively is relatively low since the percentage incorrect prediction of 7.9% in Training and 5.6% in testing shows that the model achieved 92.1% accuracy in training and 94.4% in testing respectively.

Table 4.8: Independent variable importance

Independent Variable Importance

	Importance	Normalized Importance
Age	.056	17.0%
Gender	.036	10.9%
Marital Status	.014	4.2%
Type of TB	.121	36.9%
ART Status	.012	3.6%
Hemoglob	.023	7.1%
Creatinine	.327	100.0%
CD4Count	.083	25.5%
Glucose	.266	81.4%
Weight	.062	19.0%

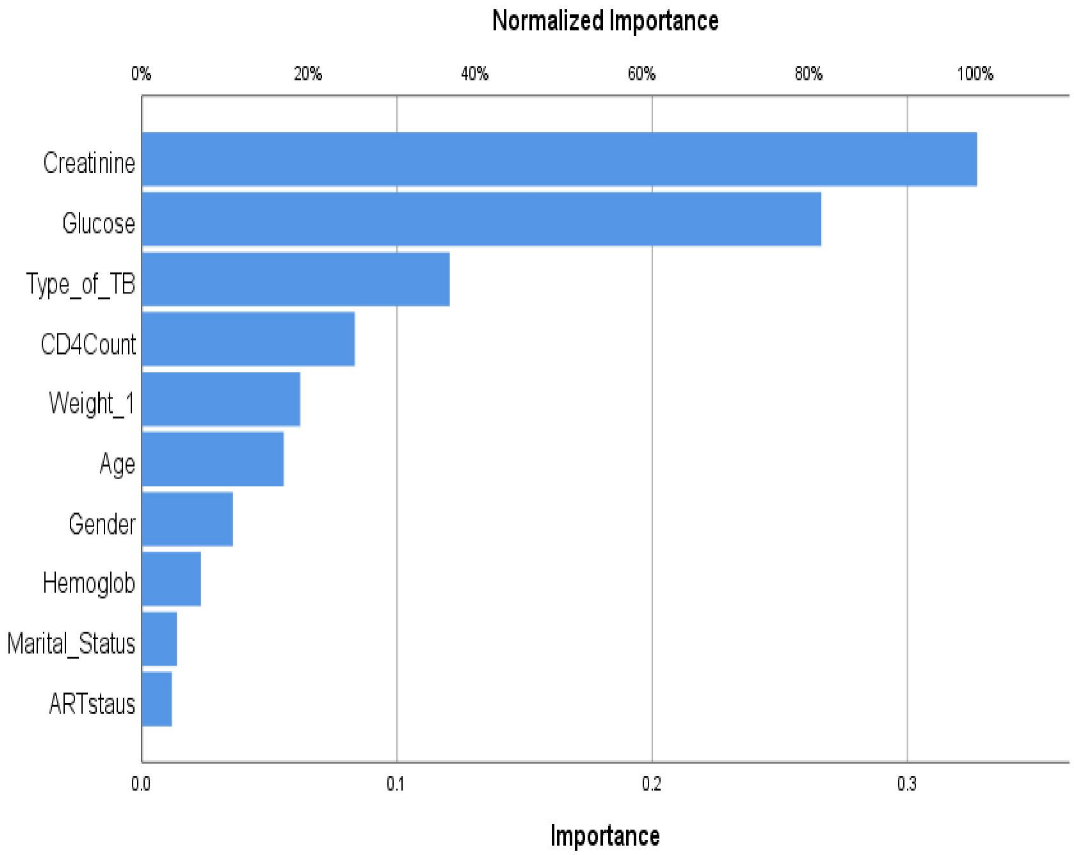

Fig. 4.1: *Independent Variable Chart*

The independent Variable Importance shows the contribution of each independent variable to predicting the Outcome (dependent). The Normalized importance shows that creatinine level (100%) follows by Glucose level (81.4%), Type of Tuberculosis (36.9%) and CD4 count (25.5%) has major contribution to predicting the output of the dependent variables. A simple bar chart for showing individual contribution is given below.

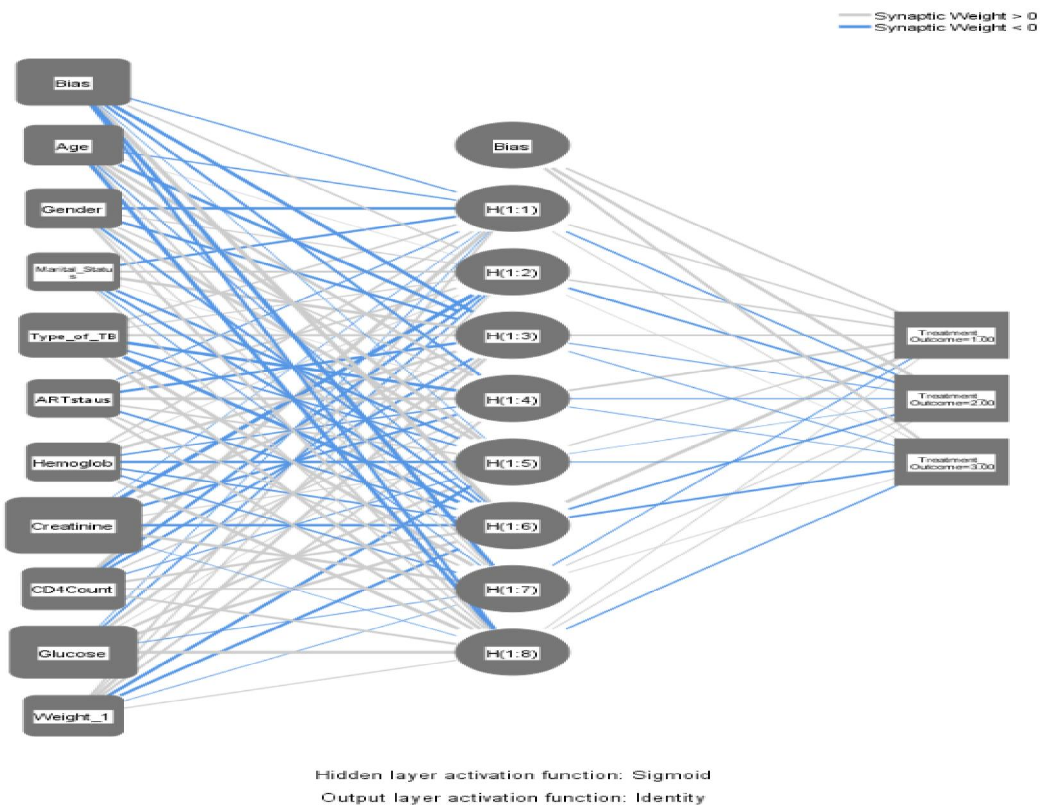

Fig 4.2: *Network relationship between the dependent and independent variable*

As the relationship network presented in Figure 2 is examined and the effect of 10 dependent variables in the model on independent variable is searched, it was observed that dependent variables changed the parameter coefficients (Wi) for 224 times during the interaction in "treatment" and "perceptron" processes to improve the model, and in this sense, it was observed that a function with 224 pieces was produced. Afterwards, this polynomial number was decreased into 3, and eventually, dependent variable was attempted to be estimated with the common effect of this three-piece function formed.

Relative Operating Characteristics Curve

Table 4.9: Area under the Curve Table

Area Under the Curve

		Area
Treatment outcome	COMPLETED AND CURED	0.817
	DIED	0.742
	DEFAULTED	0.896

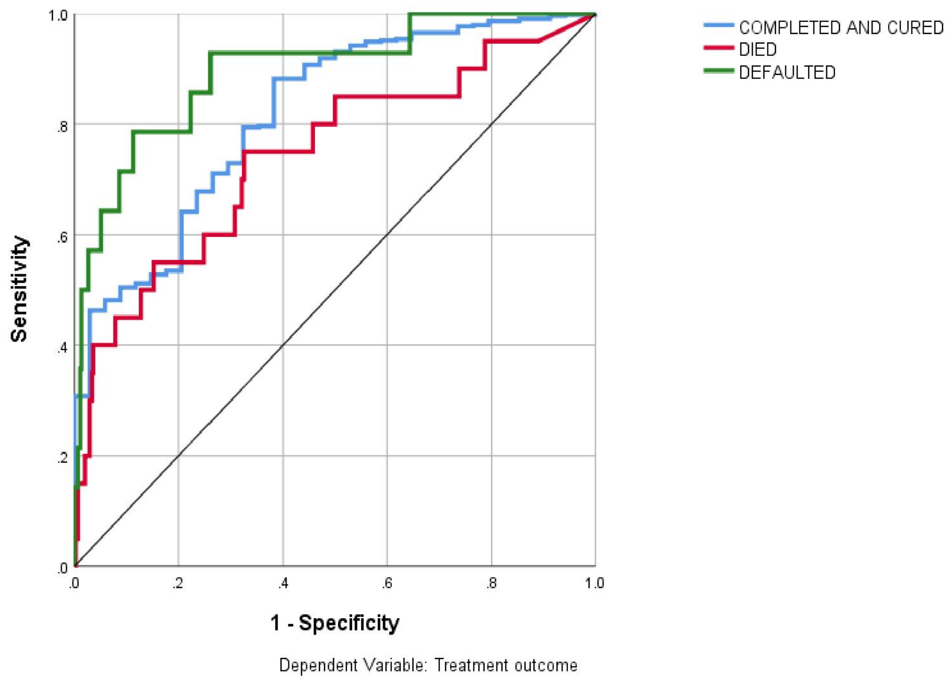

Fig 4.3: *Relative Operating Characteristics Curve*

The ROC curve shows the Sensitivity (the proportion of true positive that are classified as Positive i.e. the proportion of COMPLETED AND CURED that are actually Classified as COMPLETED AND CURE) while Specificity (the proportion of true negative that are actually classified as negative i.e. the proportion of not COMPLETED AND CURED that are correctly classified as not COMPLETED and CURED). The ROC can be better explained by the area cover under the cure give in the table above.

Comparison Between Mlr and Mlp (Ann) Classification

Multinomial Logistic Regression

Table 4.10: Multinomial Logistic Regression Classification

Observed	Predicted			
	COMPLETED AND CURED	DIED	DEFAULTED	Percent Correct
COMPLETED AND CURED	429	1	2	99.3%
DIED	19	1	0	5.0%
DEFAULTED	12	0	2	14.3%
Overall Percentage	98.7%	0.4%	0.9%	92.7%

These are classification statistics used to determine which group membership were best predicted by the model.

COMPLETED AND CURED were correctly predicted by the model 99.3% of the time [as 429 of the 432 patient who are COMPLETED AND CURED actually were predicted to do so by the model; 429/(429+1+2) = .993]. The model particularly did a not very good job at predicting DIED and DEFAULTED (at a rate of 5.0% and 14.3% respectively) but the overall prediction percentage of 92.7% was excellent.

Multilayer Perceptron
Table 4.11: Multilayer Perceptron Classification
Classification

Sample	Observed	Predicted			
		COMPLETED AND CURED	DIED	DEFAULTED	Percent Correct
Training	COMPLETED AND CURED	315	0	0	100.0%
	DIED	18	0	0	0.0%
	DEFAULTED	9	0	0	0.0%
	Overall Percent	100.0%	0.0%	0.0%	92.1%
Testing	COMPLETED AND CURED	117	0	0	100.0%
	DIED	2	0	0	0.0%
	DEFAULTED	5	0	0	0.0%
	Overall Percent	100.0%	0.0%	0.0%	94.4%

Dependent Variable: Treatment outcome

These are classification statistics used to determine which group memberships were best predicted by the model.

The multilayer perceptron assigns all test observations from the three groups in one group (Completed and cured). In terms of classification rate, it performs well. However, the performance is poor in the sense that all the test observations in other groups were assigned to the group "Completed and cured" in both training and testing experiments.

After the classification task done by both the Multinomial Logistic Regression and the Multilayer Perceptron of (ANN), it was observed that the training and the testing process of the dataset reveals that the Multilayer Perceptron (ANN) perform better than the Multinomial Logistic Regression.

Conclusion

This study is based on many participants from Lagos residents in Nigeria, where the prevalence of TB infection and HIV are very high. In this study, the MLR model and the MLP (ANN) model have been compared using TB/HIV co-infected data. Association of the TB/HIV preventive therapies with the three-regimen effect was examined through the linkage of the signs and symptoms to replication of the virus.

After the classification task done by both the Multinomial Logistic Regression and the Multilayer Perceptron (MLP-ANN), it was observed that the training and the testing process of the dataset reveals that the Multilayer Perceptron (ANN) perform better than the Multinomial Logistic Regression.

Recommendation

In contrast, the Artificial Neural Network Classifier provides an adequate description of the data. The family of the ANN containing the Multilayer Perceptron, Feed Forward Neural Networks, Convolutional Neural Network, Modular Neural Network, are applied to this dataset.

We select the model that best describes the data. In addition, the example illustrates that the MLP model have a more realistic interpretation and provides more informative results as compared to MLR model for the available data.

Therefore,

a) We suggest that using the multinomial logistic regression model may not be the optimum approach. The Artificial Neural Network may provide an alternative method to fit Patient data.
b) Determining the effect of the three regimens may be additional values to research.

The results from this model could then be compared with the standard MLP and MLR models. In addition, further study can be carried out to evaluate the effects of practical cases such as large censoring.

References

[1]. Abdalla, M.E., (2012): "An Application on Multinomial Logistic Regression Model". *Pakistan Journal of Statistics and Operation Research Vol.:8 No.2. pp 271-291.*

[2]. Abdolmaleki, P., Buadu, L.D., Murayaama, S., Murakami, J., Hashiguchi, N., Yabuuchi, H., Masuda, K., (1997). Neural Network Analysis of Breast Cancer from MRI Findings. *Radiation Medicine, **15:** 283-293.*

[3]. Abdolmaleki, P., Buadu, L.D., Naderimansh, H. (2001). Feature extraction and classification of breast cancer on dynamic magnetic resonance imaging using

artificial neural network. *Cancer Lett.,* ***10:*** *171-183.*

[4]. Agresti, A. (1990). *Categorical data analysis. New York, NY: J.VV Iley & Sons.*

[5]. Asogwa, O.C., and Oladugba, A.V., (2015): On The Comparison of Artificial Neural Network (ANN) and Multinomial Logistic Regression (MLR). *West African Journal of Industrial and Academic Research April 2015 Vol.13 No. 1: 3-9.*

[6]. Benito, U., Hanns, F., Claudio, E., and Marisela, F. (2020): "Multinomial Logistic Regression to Estimate and Predict the Perceptions of Individuals and Companies in the Face of the COVID-19 Pandemic in the Ñuble Region, Chile". *Sustainability 12, 9553; doi:10.3390/su12229553* www.mdpi.com/journal/sustainability.

[7]. Broomhead, DS and Lowe, D. (1988) Radial basis functions, multi-variable functional interpolation and adaptive networks. DTIC Document.

[8]. Caliskan, E. and Sevin, Y (2018) "A Comparative Study of Artificial Neural Networks and Multiple Regression Analysis for Modelling Skidding Time. *Applied Ecology and Environmental Research Journal 17(2): 1741-1756*

[9]. Divyansh, K., Rohan, S., Vecky, B and Bharat, D. (2015) "Comparative Study of Classification Technique (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease*". International Journal of Machine Learning and Computing Vol.5, No.5*

[10]. Elif, B. (2016) "Using Neural Network and Logistic Regression Analysis to Predict Prospective Mathematics Teachers Academis Success Upon Entering Graduate Education. *Education Sciences: Theory and Practice. 16, 943-964.*

[11]. Erkan, A. (2016): "Using Multinomial Logistic Regression to Examine the Relationship Between Children's Work Status and Demographic Characteristics". *Research Journal of Politics, Economics and Management, 2016, Year:4, Volume:4, Issue:1 77-93.*

[12]. Erkan, A. and Noyan, A. (2016): "Examination by Multinomial Logistic Regression Model of the Factors Affecting.

11/19/2021