Position Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 25–32
DOI: 10.15439/2024F8304
ISSN 2300-5963 ACSIS, Vol. 40

Relative Performance of Neural Networks and Binary Logistic Regression in a Variable Selection Framework

Castro Gbêmêmali Hounmenou 0000-0002-2306-6083*, Émile C. Agbangba 0000-0001-5280-6397†, Génevieve Amagbégnon 0009-0001-9809-7191‡ and Reine Marie Ndéla Marone 0000-0003-0787-1936 §
*Laboratoire de Biomathématiques et d'Estimations Forestières, University of Abomey-Calavi (LABEF/UAC), Benin
Centre de Recherche et de Formation en Infectiologie de Guinée, Université Gamal Abdel Nasser de Conakry, Guinea
University of Labé
Email: castrohounmenou@gmail.com
† Department of Environmental Engineering, Polytechnic School of Abomey-Calavi, Benin
Laboratoire de Biomathématiques et d'Estimations Forestières, University of Abomey-Calavi, Benin
Email: agbangbacodjoemile@gmail.com
‡Département de la Statistique et d'Economie Sectorielle, Ecole Nationale d'Economie Appliquée et de Management, University of Abomey-Calavi (ENEAM/UAC), Benin
Email: genevieveamagbegnon@gmail.com
§ Ecole des Bibliothécaires, Archivistes et Documentalistes, Cheikh Anta Diop University (UCAD), Sénégal
Email:reinemarie.marone@ucad.edu.sn

***Abstract*—This study evaluates the predictive capabilities of a binary response variable using Multilayer Perceptron neural networks (BLMLP) and binary logistic regression (BLR) in a variable selection context. The data used was related to the identification of prenatal factors linked to premature birth in women already in labor. The stepwise selection method on BLR and the Olden selection method based on the neural network approach were used to select the most relevant variables to predict the probability of premature birth by women. Then, the two selection methods were combined with BLR and BLMLP models. Using performance criteria such as sensitivity, precision, classification accuracy, F-score, and Area Under the Curve, the selection methods were compared to identify the best model. It appears from the analysis that the best procedure for selecting variables in a binary variable prediction is the use of the Stepwise procedure followed by multilayer perceptron neural networks.**

***Index Terms*—Binary logistic regression, neural network, multilayer perceptron, selection of variables, prediction**

I. Introduction

THE DURATION of a full-term pregnancy is 41 weeks of amenorrhoea. However, premature birth is defined as a baby born alive before 37 weeks of amenorrhea. [1] There are three levels of prematurity: (i) extreme prematurity (less than 28 weeks); (ii) great prematurity (between 28 and 32 weeks) and (iii) average or late prematurity (between 32 and 37 weeks). The World Health Organisation estimates that in 2018 there are 15 million premature babies each year, which represents more than one in 10 babies. Nearly one million children die each year from complications related to prematurity [2]. Many survivors suffer lifelong disabilities, including learning, visual and hearing impairments. Apart from the health problems and the number of lives lost as a result of premature birth , the consequences of premature birth for women in labour present enormous health, psychic and psychological risks [3], [4] that need to be mastered in order to develop better prevention solutions. Therefore, it is important not only to know the most significant factors responsible for preterm birth in women, especially in labour, but also to predict from a number of the most relevant prenatal factors whether women already in labour will conceive prematurely or not. For this purpose, binary logistic regression models are most commonly used.

Binary Logistic Regression (BLR) is one of the most widely used statistical modeling techniques in practice to predict or to explain a binary response variable [5], [6]. BLR models are more flexible than other techniques like parametric discriminant analysis, multi-channel frequency analysis, among other techniques [7]. The optimal conditions for good performance of BLR are: absence or very weak presence of multicollinearity between the explanatory variables, linearity of the independent variables and logarithm of odds ratios, a sufficient number of events per independent variable and absence of outliers having a strong influence [8], [9].

In real world situations, these conditions may not always met. Current models are more complex and often non-linear [10]–[13]. Among new tools to handle the complexity of the relationship between variables and possible noises in data are Multilayer Perceptron Neural Networks (MLP). MLP methods do not require verification of the assumptions and do not impose any restrictions on input variables. MLPs belong to a very rich family of continuous functions, the main characteristic of which is to allow great modeling flexibility. In addition, they have demonstrated their effectiveness in predicting empirical data compared to traditional methods and are applied in various fields [14]–[17]. Moreover, another important point for the establishment of any model is the selection of the variables to be included in the model in order to improve its explanatory and/or predictive power [18]. The selection of variables offers several advantages, such as: (i) facilitates the understanding or visualization of data, (ii) facilitate deployment, (iii) reduces physical storage and sizing requirements, (iv) improves the ratio of number of observations and dimension of representation , (v) reduces running time, (vi) improves knowledge of the phenomenon of causality between descriptors and the variable to be predicted and (vii) improves prediction performance [18].

There are several selection approaches (Manual, Backward, Forward, Stepwise, Olden, Garson, etc.) classified in two main categories: methods dependent on a model (wrapper methods), which allow the selection of a subset of variables resulting into construction of a good prediction model and the filter methods which ensure the search for relevant variables and then possibly their ordering [5], [6], [18]. The latter the user and who has the possibility of eliminating one of two variables which are significantly linked. However, the knowledge of the user is not sufficient to fully understand the underlying causalities, to discern the true links of simple artefacts, highlight the interactions, among others. Likewise, when the number of candidate variables is high, this knowledge-based approach or manual selection is not easy in practice. In this case, it is necessary to turn to automatic approaches (wrapper methods). However, there is a panoply of selection approaches and given the characteristics presented by the available data, it is up to the user to sort the method most suited to the available data and which leads to the lowest possible error rate. A method frequently used in classical logistic regression is the stepwise technique, which is more efficient compared to Backward and Forward selection methods since it is a combination of these two methods [19].

MLP approach assesses the importance of a variable as the product of the raw input-hidden and output-hidden connections between each input and output neuron and adds the product across all neurons [20]. The variable selection approaches do not necessarily lead to the same types of explanatory variables selected or to the same number. Under these conditions, what are the best subset of explanatory variables to consider? And in which model to include them for prediction purposes? This paper aims to answer these questions by comparing the prediction of binary variables by multilayer perceptron neural networks and binary logistic regression in a variable selection framework for binary response prediction.

The rest of the document is structured as follows. Section 2 briefly describes the data source, provides the specifications of the models considered, offers a brief synthesis of variable selection approaches, outlines the statistical performance criteria used, and details the data analysis methodology. The results are presented in Section 3 and discussed in Section 4. Finally, Section 5 concludes the paper.

II. Methodology

A. Data source

The data used in this study focuses on prenatal factors (medical and personal) associated with preterm delivery in women already in preterm labor. They are recorded in an array of dimension 390×14 and can be accessed at [1]. They aim to get a better understanding and prediction of this threat to boost medical analysis. The summary of these data was generated by means of the calculation of some descriptive statistics parameters such as mean (the standard error) for the quantitative variables and the absolute frequency (the relative frequency) for the qualitative variables (Table I).

TABLE I
Description of variables, $n = 390$

Variable: Description	**Nature**	**Statistics**
Predictive variables		
GEST: Gestational age in weeks at the start of the study	Quantitative	30.30 (2.50)
DILATE: Cervical dilation in cm	Quantitative	1.24 (1.31)
EFFACE: Erasure of the collar in %	Quantitative	43.98 (34.86)
CONSIS: Consistency of the neck (1 = soft, 2 = medium, 3 = firm)	Ordinal qualitative	1: 55(14.10) 2: 127 (32.56) 3: 208 (53.33)
CONTR: Presence (= 1) or not (= 0) of contraction	Binary qualitative	1: 355 (91.03) 0: 35 (8.97)
MEMBRAN: Ruptured membranes(= 1) or not (= 0)	Binary qualitative	1: 91 (23.33) 0: 299 (76.67)
AGE: Patient's age	Quantitative	26.34 (5.31)
STRAT: Period of pregnancy	Quantitative	3.23 (0.83)
GRAVID: Gravidity (number of previous pregnancies including the current one)	Quantitative	2.30 (1.45)
PARIT: Parity (number of previous term pregnancies)	Quantitative	0.78 (1.01)
DIAB: Presence (=1) or non (=0) of a diabetes problem	Binary qualitative	1: 11 (2.56) 0: 380 (97.44)
TRANSF: Transfer (= 1) or not (= 0) transfer to a specialized care hospital	Binary qualitative	1: 188 (48.21) 0: 202 (51.79)
GEMEL: Simple pregnancy (= 1) or multiple (= 0)	Binary qualitative	1: 351 (90.00) 0: 39 (10.00)
Variable to predict		
PREMATURE: Premature delivery (= 1) or not (= 0)	Binary qualitative	1: 266 (68.21) 0: 124 (31.79)

B. Specification of models

1) Binary Logistic Regression (BLR): The relationship between the binary response variable, the premature by women in labor has two classes (premature delivery versus non-premature delivery) and various potential predictors (a collection of continuous, discrete and binary variables) is modeled by Binary logistic regression (BLR). If Y_i denotes the premature for the i^{th} woman in a sample of size $n = 390$ ($Y_i = 1$ if the woman in labor gives birth prematurely, and $Y_i = 0$ otherwise), and $\mathbf{X_i} = (X_{i1}, \cdots, X_{ia}) \in \mathbb{R}^a$ with $a \in \mathbb{N}^*$ denotes the corresponding predictors, the logistic regression model expresses the relationship between Y_i and

[1] http://eric.univ-lyon2.fr/~ricco/cours/slides/prematures.xls

$\mathbf{X_i}$ in term of the conditional probability $P(Y = 1|\mathbf{X_i} = \mathbf{x_i})$ of premature, as:

$$P(Y = 1|\mathbf{X_i} = \mathbf{x}) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x_i})}{1 + \exp(\boldsymbol{\beta}^\top \mathbf{x_i})} \quad (1)$$

where $\boldsymbol{\beta}^\top \mathbf{x_i} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_a x_{ia}$ is a linear combination between the vector $\mathbf{x_i}$ of predictor variables: $\mathbf{x_i} = (x_{i0}, x_{i1}, \cdots, x_{ia})' \in \mathbb{R}^{a+1}$ and the vector of logistic regression model coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_a)^\top \in \boldsymbol{B} \subset \mathbb{R}^{a+1}$; x_0 is an additional component of unit vector and β_0 is the intercept in the model.

By applying the logistic transformation and using the equation (Eq.1), we get the linear relation between the logarithm of the odds ratio (odds = $\exp(\boldsymbol{\beta}^\top \mathbf{x_i})$) and the independent variables (Eq.2).

$$\text{logit}\Big(P(Y = 1|\mathbf{X_i} = \mathbf{x_i})\Big) = \ln\left(\frac{P(Y = 1|\mathbf{X_i} = \mathbf{x_i})}{1 - P(Y = 1|\mathbf{X_i} = \mathbf{x_i})}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_a x_{ia}. \quad (2)$$

Assuming that we have n independent observations: $y_1, \cdots, y_n$, and that the i^{th} observation is a realization of the random response variable Y, the probability density function of Y is given by [21]:

$$f(y_i|\boldsymbol{\beta}) = P(Y = 1|\mathbf{X_i} = \mathbf{x_i})^{y_i}(1 - P(Y = 1|\mathbf{X_i} = \mathbf{x_i}))^{1-y_i} \quad (3)$$

and the conditional likelihood function is written:

$$\mathsf{L}(\boldsymbol{\beta}|y_i) = \prod_{i=1}^{n} P(Y_i = 1|X_i = x_i)^{y_i}(1 - P(Y_i = 1|X_i = x_i))^{1-y_i} \quad (4)$$

To simplify the maximization of the equation (4), which allows to obtain the values of $\boldsymbol{\beta}$, its logarithm is used:

$$\begin{aligned} \ln \mathsf{L}(\boldsymbol{\beta}|y_i) &= \sum_{Y_i=1} \ln P(Y = 1|\mathbf{X_i} = \mathbf{x_i}) \quad (5) \\ &+ \sum_{Y_i=0} ln(1 - P(Y = 1|\mathbf{X_i} = \mathbf{x_i})) \end{aligned}$$

And replacing the expression $P(Y = 1|\mathbf{X_i} = \mathbf{x_i})$ (see equation (1)) in equation (5), we obtain:

$$\ln \mathsf{L}(\boldsymbol{\beta}|y_i) = \sum_{i=1}^{n} \Big(y_i(\mathbf{x_i}\boldsymbol{\beta}) - \ln\big(1 + \exp(\mathbf{x_i}\boldsymbol{\beta})\big)\Big) \quad (6)$$

The maximization of the relation ((6)) gives the estimation of $\boldsymbol{\beta}$ and this includes partial differentiation using iterative procedures as Newton-Raphson algorithm, Fisher scoring method, etc. [22], [23].

2) Formalism of Binary Logistic Multilayer Perceptron Neural Network (BLMLP): Binary Logistic Multilayer Perceptron Neural Networks are mathematical models inspired by human brain function and represented as directed graph (Fig.1). They are made up of neurons organized in successive layers. The first layer is called the input layer, the last output layer, and the middle layers are called the hidden layers. Neurons are interconnected with each other by synaptic weights (model parameters) and on the same layer, neurons cannot

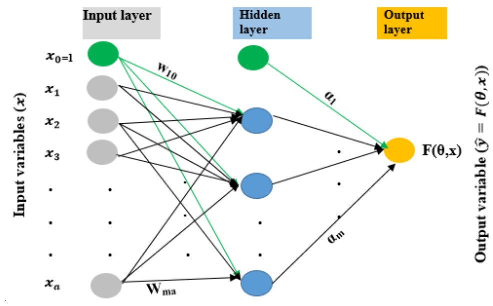

Fig. 1. An example of binary multilayer perceptron neural network model, BLMLP(a,m,1)

interconnect. Considering $n \in \mathbb{N}^*$, the number of women to give birth in the sample where $i\,(i = 1, \ldots, n)$ represents any women in the sample, after through passing the examples $(x_i, y_i)_{1 \le i \le n}$ in the network, the output F (the likelihood of a woman delivering a premature baby or not) is calculated using the following equation [24]:

$$F(\theta, x) = f(\sum_{k=1}^{m} \alpha_k f(\sum_{l=1}^{a} w_{kl} x_l + w_{k0}) + \alpha_0) \quad (7)$$

where $F(.,.) : \Theta \times \mathbb{R}^{a+1} \longrightarrow [0,1]$; $\theta = (w_{10}, \ldots, w_{m0}; w_{11}, \ldots, w_{1a}, \ldots, w_{m1}, \ldots w_{ma}; \alpha_0; \alpha_1, \ldots, \alpha_m) \in \Theta \subset \mathbb{R}^{m(a+2)+1}$ et $f(.) : \mathbb{R} \longrightarrow [0,1]$ (real value function) are respectively the output of the model, the vector of parameters of the model and the activation function of the output unit and each hidden unit $(f(z) = \frac{1}{1+e^{-z}})$. $w_k = (w_{k0}, \ldots, w_{ka})' \in \mathbb{R}^{a+1}$ is a vector of parameters of a hidden unit k with $k \in [\![1, m]\!]$; et $\alpha = (\alpha_0, \ldots, \alpha_m)' \in \mathbb{R}^{m+1}$ a vector of parameters for the single output unit.

The parameter θ is estimated by minimizing the cross-entropy error function defined by :

$$E(\theta) = -\frac{1}{n} \sum_{i=1}^{n} [y_i \ \log(F(\theta, x_i) + (1 - y_i)\log(1 - F(\theta, x_i))] \quad (8)$$

For this purpose, different algorithms are used and based on the gradient descent procedure. The basic idea is to calculate the partial derivatives $\partial(\theta)/\partial w_k$ et $\partial E(\theta)/\partial \alpha_k$ using the chain rule. There are two steps: The first is propagation learning, which calculates the error and partial derivatives, and the second is reverse propagation learning, which calculates the update of the resulting weight. From one algorithm to another, only the second step changes. We briefly present the one used in this study which is the resilient backpropagation algorithm (Rprop) as well as a local adaptive learning program [25].

$$\theta(k+1) = \theta(k) + \triangle\theta(k) \quad (9)$$

$$\begin{cases} \eta^+ \times \triangle(k-1) \ \ if \dfrac{\partial E(\theta)}{\partial \theta}(k-1) \times \dfrac{\partial E(\theta)}{\partial \theta} > 0 \\ \eta^- \times \triangle(k-1) \ \ if \dfrac{\partial E(\theta)}{\partial \theta}(k-1) \times \dfrac{\partial E(\theta)}{\partial \theta} < 0 \\ \triangle\theta(k-1) \ \ else \end{cases} \quad (10)$$

where k = number of iterations; η^- et η^+ are reduction and increase factors, $0 < \eta^- < \eta^+$. These factors are fixed at $\eta^+ = 1,2$ et $\eta^- = 0,5$ based on theoretical considerations and empirical assessments. This reduces the number of free parameters to two, namely $\triangle_0$ and $\triangle_{max}$. The computation is slightly more expensive than the ordinary back-propagation but is an answer to the problems of convergence and over-adjustment.

C. Variable selection

Variable selection eliminates irrelevant variables from the model to improve its accuracy and also reduce the risk of overfitting [26]. For logistic regression models, it is possible to test the statistic of the significance of the coefficients associated with the variables in the model [27]. These tests can be used to build models step by step. The three most common approaches are to start with an empty model and successively add variables (forward selection), to start with the complete model and remove variables (backward selection) or by adding and removing covariates (stepwise selection). Due to the nonlinear nature of multilayer perceptron neural networks, the statistical tests for the significance of the coefficients that are used in classical logistic regression cannot be applied here. We can use the automatic relevance determination [28] or the sensitivity analysis [20], [29] to heuristically evaluate the importance of the input variables on the target variable. One method used for the selection of variables is the Olden method. This method is similar to Garson's [30] algorithm modified by [31] in that the connection weights between layers of a neural network form the basis for determining varying importance. This Olden method calculates the importance of a variable as the product of the raw input-cached and output hidden connections between each input and output neuron and adds the product across all the hidden neurons. An advantage of this approach is that the relative contributions of each connection weight are maintained in terms of amplitude and sign with respect to Garson algorithm which only takes into account the absolute amplitude. Moreover, the need to reduce the number of input variables was not linked only to the performance of neural network models. Indeed, before the work of [32], neural networks were treated as a " black box " because they provided little information to explain the influence of independent variables in prediction process. Thus, [32] have proposed and demonstrated a randomization approach to statistically assess the importance of axon connection weights and input variables contribution to the neural network. Researchers have the possibility of eliminating null connections between neurons whose weights do not significantly influence the output of the network thus facilitating the interpretation of the individual and interactive contributions of the input variables in the network. By using this randomization procedure, the mechanism of the " black box" is clarified and improves the predictive ability of neural networks.

Variable selection methods, particularly the Olden procedure and the stepwise procedure, are favored in this work due to their numerous advantages. The Olden method stands out for its interpretability, allowing for easy analysis of variable importance and their contributions to predictions, its ability to account for correlations between variables, and its flexibility in application to various types of complex and even nonlinear models. On the other hand, the stepwise procedure offers an automatic selection process that simplifies modeling, strikes a balance between complexity and performance to avoid overfitting, while producing simpler and more generalizable models, as well as a solid statistical foundation to justify variable choices. It is particularly more utilized in the health field, where understanding the impact of each variable is crucial for clinical decision-making. In comparison to other methods, filtering methods evaluate variables independently of the model, risking the neglect of interactions and potentially leading to less relevant selection, whereas Olden and stepwise analyze the effect of variables within the model, promoting better selection. Clustering selection methods, while effective in reducing the number of variables, may omit crucial information by grouping features without considering their individual importance; in contrast, Olden and stepwise assign a distinct value to each variable. Finally, wrapper methods can produce excellent results but are often computationally expensive, while Olden and stepwise prove to be more efficient and suitable for large datasets while maintaining good performance.

D. Statistical performance criterion

To evaluate and select the best performing model, goodness of fit statistics such as sensitivity, precision, F-score, classification accuracy (Accuracy) and the area under the curve (AUC) are used. The closer the values of these criteria are to 1, the better the model. They are calculated from a confusion matrix (Table II). The notations in this table are as follows: all true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN) [33]. In the table above, True Positives are observations that have been rated positive and actually are. False Positives are individuals classified as positive and who are in fact negative. Likewise, False negatives are individuals classified as negative but who are actually positives and True negatives are observations that have been classified as negative and are actually negative.

TABLE II
CONFUSION MATRIX

	Predict: No (0)	Predict: Yes(1)
Actual: No (0)	True negatives (TN)	False positives (FP)
Actual: Yes (1)	False positives (FN)	True positives (TP)

Sensitivity: It measures the proportion of current positives that are correctly identified. The formula is as follows :

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

Accuracy: It is the proportion of the total number of predictions that are correct.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

Precision: This is the proportion of positives that are correctly identified.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

F-score: It is the combination of sensitivity and positive predictive value, which can be further called precision.

$$F - score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \tag{14}$$

The AUC criterion of ROC: It expresses the probability of placing a positive individual in front of a negative individual

$$AUC = \frac{W_1 - \frac{n_1.(n_1+1)}{2}}{n_1.n_0} \tag{15}$$

where :W_1: the sum of the ranks of mis-classified individuals; n_1: the number of misclassified individuals ; n_0: the number of well-ranked individuals

Akaike Information Criterion (AIC): is a measure of the quality of a statistical model. It applies to models estimated by the maximum likelihood approach such as logistic regressions. It is defined by :

$$AIC = -2log(L) + 2a \tag{16}$$

where L: the likelihood of the model and a the number of parameters in the model. It is a criterion for penalizing the log likelihood taking into account the number of explanatory variables. The best model is the one with the lowest AIC.

E. Data analysis methods

The data analysis was done in 5 steps :

1^{st} *Step* : Data processing

Initial data (x_{ij}, y_i) (1≤i≤390 and 1≤j≤13) are normalized using the formula (Eq.17). Therefore, they are partitioned into training data (70%) and test data (30%). The training data is used to establish models and test data is used to assess the model generalization abilities.

$$new_v = \frac{v - min_z}{max_z - min_z} \tag{17}$$

2^{nd} *Step* : Establishment of models

Two different models were considered for the prediction of preterm delivery. First, the binary logistic regression (BLR) model using the regression (Eq.2) with the function “glm” from the default package “stat” of R software [34] and based on binomial distribution. Second, multilayer perceptron neural networks, MLP (see Eq. 7) were used by varying the number of hidden neurons (2, 5, 8, 11, 15 and 20). The Rprop algorithm is applied. The function ”neuralnet” from the package ”neuralnet” of R software [34] is used [35]. The best MLP architecture is obtained based on the performance criteria value close to 1.

3^{rd} *Step* : Variables selection (identification of the determinants of preterm birth)

The variable selection methods used for an effective prediction of preterm delivery in women are : the Stepwise procedure applied on the BLR model with the ”stepAIC” function from ”MASS” package of R software [34] and the AIC fit statistic is used to measure the fit of the model during the variable selection process. The best model is the one with the lowest value of AIC.

The Olden procedure applied on the MLP identified in step 2 as best. The ”olden” function of the ”NeuralNetTools” package [36] is used and the higher the value of importance of an explanatory variable, the more this variable affects the response variable and the better the results . Since the number of input variables has decreased, a new MLP architecture has been chosen again taking into account the variables selected by the Olden procedure.

4^{th} *Step* : Effective prediction of premature baby delivery with selected variables and identification of the best model.

Four types of models have been developed but with regard to the use of MLPs, the number of hidden neurons has always been varied. These models are: (i) MLP on the variables selected from the Olden procedure, (ii) MLP on the variables selected from the Stepwise procedure, (iii) BLR on the variables selected from the Olden procedure and (iv) BLR on the variables selected from the Stepwise procedure. Based on the performance criteria value near 1, the best model is identified.

5^{th} *Step* : Analysis of the variables of preterm delivery according to the best approach.

III. 3. RESULTS

A. Determination of the best architecture of multilayer perceptron neural networks and establishment of classical binary logistic regression

The performance of BLMLPs models varies depending on the number of neurons in the hidden layer (Table III). The BLMLP model $(13, 20, 1)$ provided the best architecture with 13 input variables, 20 hidden neurons and an output variable (value closed to 1 for all performance criteria: $TBC = 0.99$, $Sensitivity = 0.99$, $Precision = 1$, $F - score = 0.99$ and $AUC = 0.99$).

Regarding the binary logistic regression model, the residual deviance (246.11) is deviated from the degrees of freedom (274) and their ratio is equal to 0.90, $AIC = 274.11$, $TBC = 0.76$, $Sensitivity = 0.88$, $Precision = 0.80$, $F - score = 0.84$, $AUC = 0.84$.

B. Identification of selected variables according to Olden and Stepwise procedures

Fig. 2 provides information on the importance of the explanatory variables compared to the variable explained by the

TABLE III
IDENTIFYING THE BEST NEURAL NETWORK

Architecture	Sensitivity	Precision	F-score	Accuracy	AUC
BLMLP(13,2,1)	0.75	0.85	0.80	0.72	0.81
BLMLP(13,5,1)	0.91	0.90	0.90	0.87	0.88
BLMLP(13,8,1)	0.95	1.00	0.97	0.96	0.95
BLMLP(13,11,1)	0.96	0.99	0.97	0.96	0.95
BLMLP(13,15,1)	0.96	0.99	0.97	0.96	0.95
BLMLP(13,20,1)	0.99	1.00	0.99	0.99	0.99

Olden procedure. It reveals that a subset of 5 explanatory variables are retained among the initial 13 . These are: GEMEL, TRANSF, GRAVID, PARIT and DILATE (importance value greater than 0). With the Stepwise procedure, 8 explanatory variables are selected (AIC = 266.91, lower than that of the full model, AIC = 274.11): CLEAR, MEMBRAN, STRAT, DIAB, in addition to the 4 variables obtained by the Olden procedure except by GRAVID.

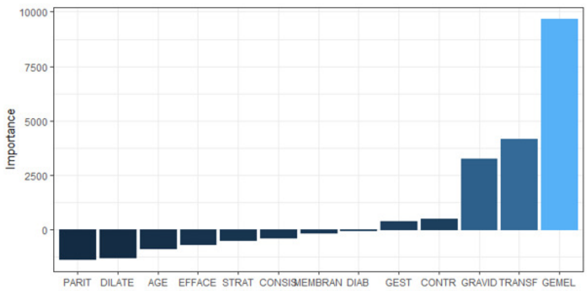

Fig. 2. ”Importance diagram of explanatory variables derived from the Olden procedure

C. Comparative analysis of modeling approaches for an efficient prediction of premature

The five variables selected with the Olden procedure and the 8 resulting from the Stepwise procedure were used as input for the BLMLPs with variation in the number of hidden neurons (Tables IV and V). The analysis of the performance criteria reveals that the best architectures of the binary logistic multilayer perceptron neural network are respectively $BLMLP(5,2,1)$ and $BLMLP(8,20,1)$ for a good prediction of the PREMATURE.

So for comparisons, the pure neural network approach ($BLMLP_{Olden}$, $BLMLP(5,2,1)$) and the approach BLMLP and Stepwise ($BLMLP_{Olden}$, $BLMLP(8,20,1)$) are retained. Added to this are the binary logistic regression models with the 8 variables retained by stepwise procedure ($BLR_{stepwise}$) and the one with the 5 variables retained by Olden procedure (BLR_{Olden}).

The comparison of predictive performances for these four models (Table VI): Sensitivity, precision, F-score, rate of good classification and AUC showed that the model $BLMLP_{stepwise}(8,20,1)$ is the best model (Table VI). Therefore, stepwise selection gives the neural network better performance in terms of prediction. Fig. 3 presents this network. We can therefore retain that stepwise procedure is better compared to Olden procedure. Thus, the relevant variables to better predict the premature delivery of a baby are :

- DIAB: presence or absence of a diabetes problem
- GEMEL: single or multiple pregnancy
- STRAT: period of pregnancy
- TRANSF: transfer or not to a hospital for specialized care
- DILATE: cervical dilation
- PARIT: parity (number of previous term pregnancies)
- EFFACE: the erasure of the collar
- MEMBRAN: rupture of membranes

TABLE IV
IDENTIFICATION OF THE BEST NETWORK WITH THE SELECTED INPUT VARIABLES WITH OLDEN PROCEDURE

Architecture	Sensitivity	Precision	F-score	Accuracy	AUC
BLMLP(5,2,1)	1	0.82	0.90	0.71	0.73
BLMLP(5,5,1)	1	0.81	0.90	0.68	0.72
BLMLP(5,8,1)	1	0.79	0.88	0.70	0.72
BLMLP(5,11,1)	1	0.80	0.89	0.64	0.68
BLMLP(5,15,1)	1	0.80	0.89	0.70	0.69
BLMLP(5,20,1)	0	0.78	0.00	0.54	0.58

TABLE V
IDENTIFICATION OF THE BEST NETWORK WITH THE VARIABLES SELECTED WITH THE STEPWISE PROCEDURE

Architecture	Sensitivity	Precision	F-score	Accuracy	AUC
BLMLP(8,2,1)	1	0.86	0.92	0.75	0.82
BLMLP(8,5,1)	1	0.85	0.92	0.80	0.88
BLMLP(8,8,1)	1	0.85	0.93	0.82	0.88
BLMLP(8,11,1)	1	0.88	0.88	0.83	0.89
BLMLP(8,15,1)	1	0.90	0.95	0.86	0.94
BLMLP(8,20,1)	1	0.91	0.95	0.88	0.97

TABLE VI
COMPARISON OF ANALYTICAL APPROACHES IN THE CONTEXT OF VARIABLE SELECTION

Models	Sensitivity	Precision	F-score	Accuracy	AUC
$BLR_{stepwise}$	1	0.81	0.90	0.77	0.82
BLR_{Olden}	1	0.77	0.87	0.73	0.73
$BLMLP_{stepwise}$	1	0.91	0.95	0.88	0.97
$BLMLP_{Olden}$	1	0.82	0.90	0.71	0.73

IV. DISCUSSION

The predictive performance of empirical data based on binary logistic multilayer perceptron neural network (BLMLP) model is better than that of classical logistic regression (BLR), taking into account all the starting independent variables (full model). Likewise, for the same subset of variables resulting from the same variable selection procedure and serving as input for the BLMLP and BLR models, BLMLPs give the best prediction performance. These results could be justified by the fact that BLMLPs are semi-parametric classifiers and are more flexible than parametric models. They

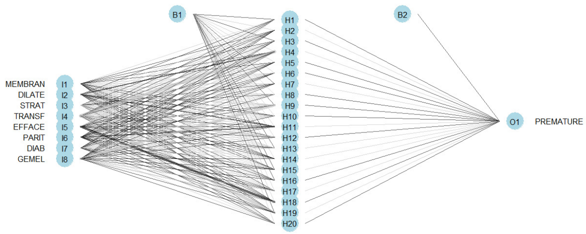

Fig. 3. Best model for PREMATURE prediction

use learning by example which makes them more powerful in pattern recognition and have more ability to mimic complicated patterns than classical logistic regression [37], [38]. In addition, they do not require a hypothesis [39] and are able to find models despite the presence of noisy data or missing data and even in the presence of multi-collinearities between the descriptors [8], [17], [40]. Furthermore, the use of BLR models requires satisfaction of many assumptions which may not be true in some real cases. This is probably the case with the PREMATURE data on which the study is focused. Failure to respect these assumptions can affect the predictive performance of BLR models and consequently lead to errors in predictions [8], [9], [18] Likewise, several studies have shown that multilayer perceptron neural networks have better prediction skills compared to classical binary logistic regression. [41]–[45]. But since the sample size of our data is not large, this result is contrary to those obtained by [44], [46] who worked on a large sample size where BLMLPs and logistic regression classic have almost similar performance although PCMs are powerful in concept. However, logistic regression requires large sample sizes to make maximum likelihood estimates powerful. [9]. The independent variables selected vary according to the selection procedure and as well as their numbers. This observation is certainly due to the approaches used which are related to the estimation criterion (AIC for the Stepwise procedure and Importance for Olden procedure). With the Olden procedure, we can know the order of importance and the direction of influence of each identified descriptor, which is not the case with Stepwise where we can only know the group of significant descriptors. Considering the selected variables by the Stepwise procedure as input variables for the BLMLP model ($BLMLP_{Stepwise}$) has good predictive power than the models $BLMLP_{Olden}$, $BLR_{Stepwise}$ and BLR_{Olden}. This could be justified by the fact that Stepwise procedure got rid of all the explanatory variables not relevant than Olden procedure. These irrelevant variables could make the estimates numerically unstable and negatively affect the predictive capacity of the BLMLP and BLR models [47]. This approach seems to give a result contrary to the principle of Occam's Razor, which in favor of selecting, for the same number of observations, a model with few variables with a better chance of being more robust in generalization. However, the number of descriptors identified with the Stepwise procedure is higher than that obtained with Olden by considering the same number of observations. This contraction could be explained by the complicity of the data or of the possible interactions existing between them, that the MLPs models have the capacity to manage [44], [45]. Although the selected variable prediction approach $BLMLP_{Stepwise}$ gives better predictive performance, it would be advantageous for a study to compare Olden procedure to stepwise one depending on the complexity of the relationship between variables. Another advantage may be to vary the sample size and the dimension of the variables to see how the four models will behave as the sample size increases. Another important aspect of networks is the choice of hyper-parameters (activation functions in hidden layers, number of layers and hidden neurons, learning rate, learning algorithm, etc.). The latter influence the performance of neural networks and would be useful to explore them for the selection of variables with Olden procedure. Moreover, the comparisons were based on empirical data and it would be important to repeat them on several databases through a simulation in order to generalize the conclusions.

V. Conclusion

In this study, two models of prediction of a binary variable (binary logistic regression and multilayer perceptron neural networks) were combined with two variable selection procedures (stepwise and Olden) in order to propose a new prediction approach.Starting from the example of predicting the premature or non-premature delivery of a baby, binary logistic multilayer perceptron neural network (BLMLP) models best predict these data compared to classical logistic regression (BLR) models with all the starting independent variables (full model). Also, for the same group of variables resulting from the same variable selection procedure and serving as input for the PCM and BLR models, the BLMLPs give the best prediction performance. Moreover, the use of the variables selected by the stepwise selection procedure as input variables to neural networks has a good predictive power that the models $BLRMLP_{Olden}$, $BLR_{Stepwise}$ and BLR_{Olden}.

References

[1] D. Wolke, "Preterm birth: high vulnerability and no resiliency?" *Journal of Child Psychology and Psychiatry*, vol. 59, no. 11, pp. 1201–1204, 2018.

[2] L. Liu, S. Oza, D. Hogan, Y. Chu, J. Perin, J. Zhu, J. E. Lawn, S. Cousens, C. Mathers, and R. E. Black, "Global, regional, and national causes of under-5 mortality in 2000–15: An updated systematic analysis with implications for the sustainable development goals," *The Lancet*, vol. 388, no. 10063, pp. 3027–3035, 2016.

[3] R. Alijahan, S. Hazrati, M. Mirzarahimi, F. Pourfarzi, and P. A. Hadi, "Prevalence and risk factors associated with preterm birth in ardabil, iran," *Iranian journal of reproductive medicine*, vol. 12, no. 1, p. 47, 2014.

[4] A. T. Deressa, A. Cherie, T. M. Belihu, and G. G. Tasisa, "Factors associated with spontaneous preterm birth in addis ababa public hospitals, ethiopia: cross sectional study," *BMC pregnancy and childbirth*, vol. 18, no. 1, pp. 1–5, 2018.

[5] C. J. Peng and T. H. So, *Logistic regression analysis and reporting: A Primer*, 2002.

[6] P. K. Josepha and A. Ame, "Effect of testing logistic regression assumptions on the improvement of the propensity scores," *International Journal of Statistics and Applications*, vol. 8, no. 1, pp. 9–17, 2018. [Online]. Available: http://doi:10.5923/j.statistics.20180801.02
[7] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics*, 2007.
[8] J. C. Stoltzfus, "Logistic regression: A brief primer," *ACADEMIC EMERGENCY MEDICINE*, vol. 18, pp. 1099–1104, 2011. [Online]. Available: http://doi:10.1111/j.1553-2712.2011.01185.x
[9] H. Park, "An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain," *J Korean Acad Nurs*, vol. 43, no. 2, pp. 154–164, 2013. [Online]. Available: https://doi.org/10.4040/jkan.2013.43.2.154
[10] D. M. Bates and D. G. Watts, *Nonlinear regression analysis and its applications*. Wiley New York, 1988, vol. 2.
[11] J. K. Lindsey, *Nonlinear models in medical statistics*. Oxford University Press on Demand, 2001.
[12] S. Wang, L. Zheng, J. Dai *et al.*, "Empirical likelihood diagnosis of modal linear regression models," *Journal of Applied Mathematics and Physics*, vol. 2, no. 10, p. 948, 2014.
[13] J. Hagenauer, H. Omrani, and M. Helbich, "Assessing the performance of 38 machine learning models: The case of land consumption rates in bavaria, germany," *International Journal of Geographical Information Science*, vol. 33, no. 7, pp. 1399–1419, 2019. [Online]. Available: https://doi.org/10.1080/13658816.2019.1579333
[14] M. Cottrell, M. Olteanu, F. Rossi, J. Rynkiewicz, and N. Villa-Vialaneix, "Neural networks for complex data," *KI-Künstliche Intelligenz*, vol. 26, no. 4, pp. 373–380, 2012.
[15] G. Daniel, *Principles of artificial neural networks*. World Scientific, 2013, vol. 7.
[16] O. Asogwa and A. Oladugba, "On the comparison of artificial neural network (ann) and multinomial logistic regression (mlr)," *West African Journal of Industrial and Academic Research*, vol. 13, no. 1, pp. 3–9, 2015.
[17] A. Mollalol, M. K. Rivera, and B. Vahedi, "Artificial network modeling of novel coronavirus (covid 19) incidence rates across the continental united states," *Int. J. Environ. Res. Public Health*, vol. 17, no. 4204, pp. 1–13, 2020. [Online]. Available: http://doi:10.3390/ijerph17124204
[18] S. Sperandei, "Understanding logistic regression analysis," *Biochemia Medica*, vol. 24, no. 1, pp. 12–18, 2014. [Online]. Available: http://dx.doi.org/10.11613/BM.2014.003
[19] P. Du Jardin, "Prevision de la defaillance et reseaux de neurones : L'apport des methodes numeriques de selection de variables," in *These de Doctorat, Universite Nice Sophia Antipolis, France*. HAL, 2007.
[20] J. D. Olden, M. K. Joy, and R. G. Death, "An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data," *Ecological modelling*, vol. 178, no. 3-4, pp. 389–397, 2004.
[21] J. D. Jobson, *Applied multivariate data analysis: volume II: Categorical and Multivariate Methods*. Springer Science & Business Media, 2012.
[22] S. A. Czepiel, "Maximum likelihood estimation of logistic regression models: theory and implementation," *Available at czep. net/stat/mlelr. pdf*, pp. 1 825 252 548–1 564 645 290, 2002.
[23] A. Diop, A. Diop, and J.-F. Dupuy, "Maximum likelihood estimation in the logistic regression model with a cure fraction," *Electronic journal of statistics*, vol. 5, pp. 460–483, 2011.
[24] C. G. Hounmenou, R. Tohoun, K. E. Gneyou, and R. GLèLè Kakaï, "Empirical determination of optimal configuration for characteristics of a multilayer perceptron neural network in nonlinear regression," *Afrika Statistika*, vol. 15, no. 3, pp. 2413–2429, 2020.
[25] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The rprop algorithm," in *IEEE international conference on neural networks*. IEEE, 1993, pp. 586–591.
[26] C.-K. Chen and H. John Jr, "Using ordinal regression model to analyze student satisfaction questionnaires. ir applications." *Association for Institutional Research (NJ1)*, vol. 1, 2004.
[27] D. Hosmer and S. Lemeshow, "Applied logistic regression 2nd edition wiley," *New York*, 2000.
[28] C. M. Bishop *et al.*, *Neural networks for pattern recognition*. Oxford university press, 1995.
[29] J. M. Zurada, A. Malinowski, and I. Cloete, "Sensitivity analysis for minimization of input data dimension for feedforward neural network," in *Proceedings of IEEE International Symposium on Circuits and Systems-ISCAS'94*, vol. 6. IEEE, 1994, pp. 447–450.
[30] D. G. Garson, "Interpreting neural network connection weights," *Artificial Intelligence Expert*, vol. 9, no. 3, pp. 46–51, 1991.
[31] A. T. Goh, "Back-propagation neural networks for modeling complex systems," *Artificial Intelligence in Engineering*, vol. 9, no. 3, pp. 143–151, 1995.
[32] J. Olden and D. Jackson, "Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks," *Ecological Modelling*, vol. 154, pp. 135–150, 2002. [Online]. Available: https://doi.org/10.1016/S0304-3800(02)00064-9
[33] J. Zarifis, V. Grammatikou, M. Kallistratos, A. Katsivas, and I. of the Prospective Noninterventional Observational Study of the Antianginal Efficacy of Ivabradine During a 4-Month Treatment of a Greek Population With Coronary Artery Disease, "Treatment of stable angina pectoris with ivabradine in everyday practice: a pan-hellenic, prospective, non-interventional study," *Clinical cardiology*, vol. 38, no. 12, pp. 725–732, 2015.
[34] R. R Core Team, "A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. version 3.3.6," *https://www.R-project.org/*, 2019.
[35] S. Fritsch, F. Guenther, and M. F. Guenther, "Package ?neuralnet?" *Training of Neural Networks. Recuperado de https://cran. r-project. org/web/packages/neuralnet/neuralnet. pdf*, 2019.
[36] M. W. Beck, "Neuralnettools: Visualization and analysis tools for neural networks," *Journal of statistical software*, vol. 85, no. 11, p. 1, 2018.
[37] J. Song, S. Venkatesh, E. Conant, P. Arger, and C. Sehgal, "Comparative analysis of logistic regression and artificial neural network for computer-aided diagnosis of breast masse," *Acad Radiol*, vol. 12, pp. 487–495, 2005. [Online]. Available: https://doi.org/10.1016/j.acra.2004.12.016
[38] S. R. D. N. C. Barwar, "A comparative study of multilayer perceptron, radial basis function networks and logistic regression for healthcare data classification," 2016.
[39] P. M. West, P. L. Brockett, and L. L. Golden, "A comparative analysis of neural networks and statistical methods for predicting consumer choice," *Marketing Science*, vol. 16, no. 4, pp. 370–391, 1997.
[40] B. Eftekhar, K. Mohammad, H. Ardebili, M. Ghodsi, and E. Ketabchi, "Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data," *BMC Med Inform Decis Mak*, vol. 5, no. 3, p. 20, 2005. [Online]. Available: https://doi.org/10.1186/1472-6947-5-3
[41] P. J. Adeodato, G. C. Vasconcelos, A. L. Arnaud, R. A. Santos, R. C. Cunha, and D. S. Monteiro, "Neural networks versus logistic regression: A comparative study on a large data set." in *ICPR (3)*, 2004, pp. 355–358.
[42] V. Bourdès, S. Bonnevay, P. Lisboa, R. Defrance, D. Pérol, S. Chabaud, T. Bachelot, T. Gargi, and S. Négrier, "Comparison of artificial neural network with logistic regression as classification models for variable selection for prediction of breast cancer patient outcomes," *Advances in Artificial Neural Systems*, vol. 2010, 2010.
[43] C.-p. LI, X.-y. Zhi, M. Jun, C. Zhuang, Z.-l. Zhu, C. Zhang, and L.-p. HU, "Performance comparison between logistic regression, decision trees, and multilayer perceptron in predicting peripheral neuropathy in type 2 diabetes mellitus," *Chinese medical journal*, vol. 125, no. 5, pp. 851–857, 2012.
[44] M. Parsaeian, K. Mohammad, M. Mahmoudi, and H. Zeraati, "Comparison of logistic regression and artificial neural network in low back pain prediction: second national health survey," *Iranian journal of public health*, vol. 41, no. 6, p. 86, 2012.
[45] M. García, C. Valverde, M. I. López, J. Poza, and R. Hornero, "Comparison of logistic regression and neural network classifiers in the detection of hard exudates in retinal images," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 5891–5894.
[46] A. Kazemnejad, Z. Batvandi, and J. Faradmal, "Comparison of artificial neural network and binary logistic regression for determination of impaired glucose tolerance/diabetes," *EMHJ-Eastern Mediterranean Health Journal*, vol. 16, no. 6, pp. 615–620, 2010.
[47] S. Menard, *Applied logistic regression analysis (Second Edition)*. Sage Publications, 2002, vol. 106.