

Suggested method for prediction using semiparametric regression function

Aseel Sameer Mohamed

Family and Community Medicine Department, Al Kindy Medical College, University of Baghdad, aseelsameer@kmc.uobaghdad.edu.iq

Abstract:

Ferritin is a key organizer of protected deregulation, particularly below risky hyperferritinemia, by straight immune-suppressive and pro-inflammatory things. , We conclude that there is a significant association between levels of ferritin and the harshness of COVID-19.

In this paper we introduce a semi- parametric method for prediction by making a combination between NN and regression models. So, two methodologies are adopted, Neural Network (NN) and regression model in design the model; the data were collected from مستشفى دار التمريض الخاص for period 11/7/2021- 23/7/2021, we have 100 person, With COVID 12 Female & 38 Male out of 50, while 26 Female & 24 Male non COVID out of 50.

The input variables of the NN model are identified as the ferritin and a gender variable. The higher results precision was attained by the multilayer perceptron (MLP) networks when we applied the explanatory variables as the inputs with one hidden layer, which covers 3 neurons, as the planned many hidden layers are with one output of the fitting NN model which is use in stages of training and validation beside the actual data.

We used a portion of the actual data to verify the behaviour of the developed models, we find that only one observation is false prediction value. This mean that the estimation model has significant parameters to forecast the type of Covid cases (Covid or no Covid) .

Keywords: Semi- parametric method, Neural Network models (NN), regression, Ferritin level, COVID 19, multilayer perceptron (MLP).

1. Introduction

Ferritin is a key mediator of immune deregulation, especially under extreme hyperferritinemia, via direct immune-suppressive and pro-inflammatory effect. Many individuals with diabetes exhibit elevated serum ferritin levels³⁻⁵, and it is known that they face a higher probability to experience serious complications from COVID-19. On this basis, we briefly review evidence supporting the hypothesis that ferritin levels might be a crucial factor influencing the severity of COVID-19.[1]

Linlin Cheng et al , studies investigating ferritin in COVID-19 were collected from PubMed, EMBASE, CNKI, Sino Med, and WANFANG. A meta-analysis was performed to compare the ferritin level between different patient groups: non-survivors versus survivors; more severe versus less severe; with comorbidity versus without comorbidity; ICU versus non-ICU; with mechanical ventilation versus without mechanical ventilation, and conclude that Ferritin was associated with poor prognosis and could predict the worsening of COVID-19 patients. [2]

In this paper we introduce a semi- parametric method for prediction by making a combination between NN and regression models. So, two methodologies are adopted, Neural Network (NN) and regression model in design the model; the data were collected from مستشفى دار التمريض الخاص for period 11/7/2021- 23/7/2021, we have 100 person, With COVID 12 Female & 38 Male out of 50, while 26 Female & 24 Male non COVID out of 50, we used MATLAB to evaluate the NN result & MINITAB Student Release version 14 to evaluate the Regression model .

The input variables of the NN model are identified as the ferritin and a gender variable. The higher results precision was attained by the multilayer perceptron (MLP) networks when we applied the explanatory variables as the inputs with one hidden layer, which covers 3 neurons, as the planned many hidden layers are with one output of the fitting NN model which is use in stages of training and validation beside the actual data.

We used a portion of the actual data to verify the behaviour of the developed models, we find that only one observation is false prediction value. This mean that the estimation model has significant parameters to forecast the type of Covid cases (Covid or no Covid) .

2. Models Review

Nonlinear models are used to address the problem of observed data which does not fit standard structures in linear models, moreover, controlled to recover the predictable part to explain the performance procedure somewhat than to present certain models which are hard to grip or/ to enhance certain stochastic mechanisms.

The neural network used in fitting and predicting the incidence tendency of haemorrhagic fever with renal syndrome (HFRS) , in the mainland of China and the result showed better effects in fitting and predicting than using the traditional ARIMA model. Neural network models seemed to have strong application value in the prevention and control of HFRS [9].

The whole of nonlinear statistical demonstrating methods is large and can categorize to parametric method which resources that the number of the connected parameters are known and the construction of the purpose to estimate is firm; nonparametric models do not oblige the function to any exact procedure; and semi parametric models as a mixture of parametric and nonparametric parts [2]. The NN model has to be stated counting the number of parameters - earlier it is estimated, so it can categorize as parametric models [4]. Then the NN models can be mentioned as parametric model in the arithmetical logic. It is established by the contrast of estimating linear functions neural networks to traditional linear methods are obviously the achieve of linear neural function little effective since of the unnecessary extra exertion, but results in [1] showed that NN models are the finest in the situation of a nonlinear in statistics patron.

The model design using Regression and including neural model in this equation:

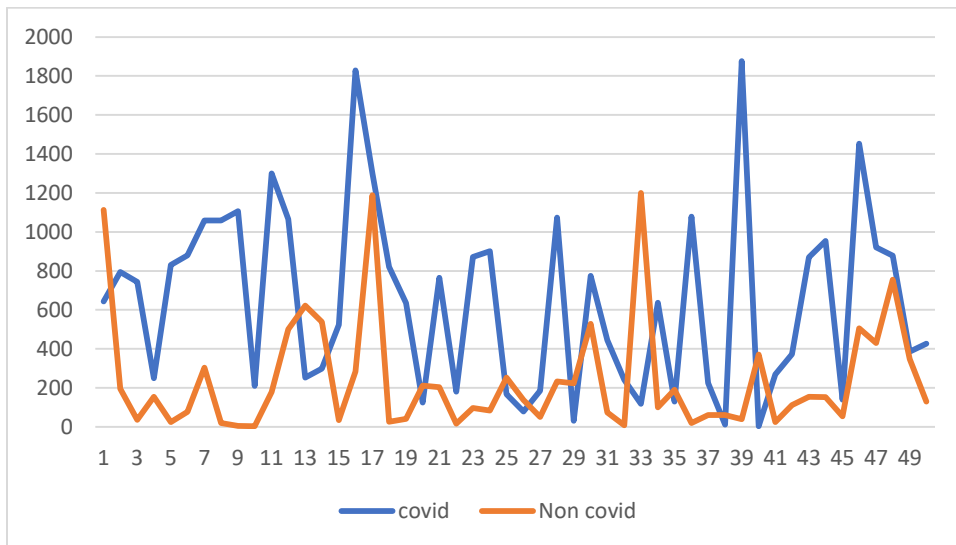
$$x_t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \theta_1 F(x_1, x_2, \dots, x_p) \text{ -----(1)}$$

Where F is function of the nonlinear part means a multilayer perceptron (MLP) by input vector x_p and weight (parameter) vector θ_1 .

3. Data Analysis

The correlation coefficient used to figure out the relationship between the level of ferritin in the blood and the case of covid-19 infection and no infections. The result demonstrated that Pearson correlation of two variables (with and without covid) = -0.052 at P-Value = 0.721. This mean significant difference between the two cases of ferritin level.

Fig 1 the comparison between ferritin level in two cases patient covid and control (non covid)



It is clear from this fig the case with covid has high level than case no covid. Therefore, when construction prediction model must include this impact factor.

The another important factor that impacting on the output is the gender (male, female). It is clear the significant influence of this variable when computed the frequencies of the female numbers in the infection cases which is =12 out of 50, while in cases no infection is =26 out of 50. Due to this analysis dependent these two variables in building the prediction models.

4. Model building of Covid Prediction

Based on data analysis; two methodologies (NN and regression model) are adopted when built the semi-model prediction. Then using the combination methodologies in this estimation equation:

$$P_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p + \hat{\theta}_1 \hat{F}(x_1, x_2, \dots, x_p) \text{ -----(2)}$$

Then, the multiple regression model uses where P_{t+1} is the prediction value and $\hat{F}(x_1, x_2, \dots, x_p)$ is the result of NN prediction model.

The input variables of the NN model are identified as the ferritin and a gender variable. The higher results precision was attained by the multilayer perceptron (MLP) networks when we applied the explanatory variables as the inputs with one hidden layer, which covers 3 neurons, as the planned many hidden layers are with one output of the fitting NN model which is use in stages of training and validation beside the actual data.

It is found the series correlation between the ferritin and gender when estimation the parameters of the above equation. Therefore, it was used this formal to get optimal Covid prediction:

$$P_{t+1} = \begin{cases} 0.000415 * ferritin + 0.547 NNoutput & \text{Male} \\ 0.000563 * ferritin & \text{Female} \end{cases} \text{ ----- (3)}$$

The Table below represent the DW statistic and F-statistic for the two estimating models

Table 1. The statistical attributes of prediction models

t-test	constant	ferritin	NN output	DF	DW	F	P	MS	R ²
Model (male)	NA	0.0002	0.2	6	1.40726	54.02	0.000	9.8	0.91167
Model (Female)	NA	0.0005632	0.0001782	1	1.20855	9.99	0.003	2.5571	0.505812

From Table 1, we see that the Durbin Watson Statistic (DW) equal to 1.40726 for male and 1.20855 for female indicate there is a positive autocorrelation between residuals .

In this section we used a portion of the actual data to verify the behaviour of the developed models.

4. Results discussion

It is clear from the Table 2 just one observation is false prediction value. This is mean the estimation model has significant parameters to forecast the type of Covid cases (Covid or no Covid).

Table 2. Expected values of the two Covid cases of Female gender.

Ferritin level	Actual (Case of Covid)	Prediction (Case of Covid)	Accurate
385.03	1	0	False
151.64	0	0	True
505.8	0	0	True
754.84	0	0	True

Table 3. Expected values of the two Covid cases of male gender.

NN-output	Ferritin level	case of Covid	Prediction	Accurate
1	1453.3	1	1	True
1	920.41	1	1	True
1	878.78	1	1	True
1	427.14	1	1	True
0	53.4	0	0	True
1	430.38	0	1	False
0	128.69	0	0	True

The result in above Table is represented the optimal prediction depend on two variables NN-output and Ferritin level.

5. Conclusion:

- 1- This suggested model showing the association between the serum ferritin level and clinical features of COVID-19 patients with disease
- 2- We recommended to depend on the ferritin test in an emergency to detect COVID -19 patients
- 3- Future clinical educations should be did to additional clarify its predictive and pathogenic part in COVID-19.

6. References:

- 1- A.S.Mohamed & N.A.Mohamed , “ Comparison between regression and artificial Neural Network in Forecasting “ published in the 4th international Scientific Conference of Arab Statisticians , 20-21, November , 2013.
- 2- James H. Stock & Mark W. Watson , “Business Cycles, Indicators and Forecasting “ ,January 1993. <http://www.nber.org/books/stoc93-1>
- 3- Jonathan Feld, Douglas Tremblay, Santiago Thibaud & Alaina Kessler, “ Ferritin levels in patients with COVID -19 : A poor predictor of mortality and hem phagocytic lymphohisticocytosis “ International Journal of Laboratory Hematology , Vol.42, Issue 6, December 2020, Pages 773-779.
- 4- Kin –Wai Lee & Cheng – Few Lee, “Cash holdings and corporate governance in family – controlled firms 1992.
- 5- Linlin Cheng, Haolong Li, Liubing Li, Chenxi Liu, Songxin Yan & Haizhen Chen, , “ Ferritin in the coronavirus disease 2019 (COVID -19) : A systematic review and meta – analysis “, Journal of Clinical Laboratory Analysis , Vol.34, Issue 10,October 2020 .
- 6- Lomax, Richard G. (2007) “ Statistical Concepts : A Second Course “, P.10, ISBN0-8058-8550-4.
- 7- Pan American Health Organization (WHO/PAHO) “ Ferritin levels and COVID-19”, International Repository for Information Sharing , <https://iris.paho.org/handle/10665.2/3021>.
- 8- Sawilowsky, S.(2002). “ Fermat Suhubert Einstein , and Behrens- Fisher . The Probable Difference Between Two Means When $\delta_1^2 = \delta_2^2$, Journal of Modern Applied Statistical Methods , 1(2) ,461-472.
- 9- Guoju Wu, Xiaohong Liu, Tuo Chen ,Guobao Xu & Wenzhi Wang , “ Elevation – dependent variations of tree growth and intrinsic water- use efficiency in Schrenk spruce (Picea schrenkiana) in the western Tianshan Mountains , China “ frontiers in Plant Science , 2015