Indian J Dairy Sci 76(1): 64-68
https://doi.org/10.33785/IJDS.2023.v76i01.008

RESEARCH ARTICLE

Comparative study of multiple linear regression and artificial neural network for prediction of first lactation 305-days milk yield in Tharparkar cattle

Subhita[1] (✉), M Nehara[2], U Pannu[3], M Bairwa[4] and Rashmi[5]

Received: 22 July 2022 / Accepted: 05 October 2022 / Published online: 20 February 2023

Abstract: The present investigation was undertaken on 3266 weekly test day milk yield records of first lactation Tharparkar cows spread over a period of 8 years (2012-2020) maintained at Livestock Research Station, Beechwal, Bikaner. The weekly test day milk yields (WTD) were used to develop best multiple linear regressions (MLR) and artificial neural network (ANN) model for prediction of first lactation 305-days milk yield (FL305DMY). Further, the comparison was made between MLR and ANN model based on coefficient of determination (R^2) and root mean square error (RMSE). Artificial Neural Network was trained using back propagation algorithms viz. Scaled conjugate gradient (SCG). It has been observed that the coefficient of determination of the models was increased with the addition of test day milk yields as input variables. It was inferred from the study that artificial neural network was better than the multiple linear regression to predict FL305DMY with more than 70% accuracy by almost all the input sets at early as 117th day of the lactation with lesser value of RMSE in comparison to MLR. Therefore, it is concluded that ANN is a potential tool for the prediction of the first lactation305-days milk yield in Tharparkar cattle than multiple linear regression.

Keywords: Artificial neural network, First lactation 305-days milk yield, Multiple linear regressions, Tharparkar cattle, Weekly test day milk yields

[1,4] Department of Animal Husbandry, Rajasthan, India
[2,3] Department of Animal Genetics and Breeding, College of Veterinary and Animal Science, RAJUVAS, Bikaner (Rajasthan) 334 001
[5]Department of Veterinary Pathology, College of Veterinary and Animal Science, RAJUVAS, Bikaner (Rajasthan) 334 001

Subhita (✉)
Department of Animal Husbandry, Rajasthan, India
E-mail: subhitapilania95@gmail.com

Introduction

The ability to predict first lactation 305-days milk yield (FL305DMY) of a cow from its test-day milk yields would determine the success of dairy herd culling programs. In dairy cattle, a high rate of genetic improvement is only possible through the early culling of low-producing cows. The selection of dairy cattle based on test day milk yields is advantageous to the dairy farmer as it cuts down the cost of progeny testing, genetic evaluations studies specially, sire evaluation at an early age and also saves time in decision-making for selection. However, in developing countries such as India, there is a limited level of milk recording, and the use of test day models would result in lower recording costs because we could have longer intervals between milk recording and less frequent collection of milk samples. As a result, use of test day milk yields intervals is receiving more importance for prediction of milk yield in dairy cattle. Prediction of FL305DMY using test day milk yields in an early stage of lactation with maximum accuracy is one of the criteria of selection for lifetime profitability of dairy cows (Kannan and Gandhi, 2006). Currently, multiple linear regressions are using for predict lactation milk yield or first lactation 305 days milk yield at most of the dairy evaluation programmes. In multiple linear regressions, several explanatory variables are used to predict the outcome of the response variable. It estimates the coefficients of the linear relationship between input and output variable. The neural network is a set of non-linear data modeling tools and consisting of three layers i.e. input layer, output layer, and one or two hidden layers. The weights associated with the connections between neurons in each layer are iteratively adjusted by the training. The knowledge is acquired by the network through a learning process and synaptic weights are used to store the knowledge so it resembles the human brain.

Mostly, the prediction of 305-days milk yield is done on the basis of prediction equations constructed by multiple regression analysis, which does not consider co-linearity among explanatory variables and may lead to biased results (Pindyick and Rubinfeld, 1991). On the other hand, the connectionist models also known as artificial neural networks (ANNs) are algorithmic and mathematical models that mimic the learning process of the human brain and can be applied to non-linear and complex data, even if

it is imprecise and noisy, therefore nowadays it is receiving more importance for the prediction of milk yield.

It has found wide applications viz., prediction of second parity milk yield and fat percentage of dairy cows based on first parity information using neural network system (Edriss et al. 2008), prediction of first lactation 305-day milk yield in Sahiwal cattle (Dongre et al. 2012), comparisons of artificial neural network and multiple linear regression for prediction of first lactation 305-day milk yield in Murrah buffaloes (Rana et al. 2021). Hence, the present investigation was carried out to compare the relative efficiency of multiple linear regression and artificial neural network for prediction of first lactation 305-days milk yield in Tharparkar cows based on weekly test day milk yields.

Material and Methods

The data on 3266 weekly test day milk yields records of first lactation of Tharparkar cows maintained at Livestock Research Station, Beechwal, Rajasthan University of Veterinary and Animal Sciences, Bikaner, India over a period of 8 years (2012–2020). Data were analysed to predict first lactation 305-days milk yield (FL305DMY) using weekly test day milk yield records. The farm is located at 28° 1' N Latitude and 73° 1 9' E Longitude. It has an average elevation of 234.84 meters above mean sea level. The soil is sandy. The maximum temperature goes as high as 50 °C during the summer months and falls down to the level of 0 °C during the winter months. Low and erratic rainfall is also a common feature in this area. Thus, it is obvious that Tharparkar cattle maintained at this farm have been exposed to extreme climatic conditions.

The records on animals that dried up before 100 days of lactation were not included in the study. The statistical analysis was performed using Statistical Package for the Social Sciences version 20 software. The whole data was divided into four main training data -test data sets (%) as SET-A (66.67-33.33), SET-B (75-25), SET-C (80-20) and SET-D (90-10).

From each animal, a total of 43 test day milk yield records were collected at weekly interval starting from the 6th day of lactation onwards. Out of all test days, a total of three test day records were selected using backward elimination method of multiple linear regressions (MLR) to use as input variables for artificial neural network. MLR starts with all explanatory variables included in the model. The least significant explanatory variable, that is, the one with the highest p-value, is then removed at each step until all variables have been added. When the overall fit of the model is considered, variables were automatically removed until the optimum model was found. The optimum model has three test-day milk yield records i.e. 3th, 14th and 24th which were recorded on 20th, 97th, and 167th day of lactation.

Further, a total of three input sub-sets have been prepared with a total of three test days which were used as input variables. The first input set included two test day milk yields records viz. WTD 3 and WTD24, second input set included two test days namely WTD14 and WTD24, and finally, the last input set included three test days as, WTD3, WTD 14 and WTD24.

Multiple linear regression

$$v_i = a + b_i \text{ “ } X_i$$

Where,

v_i = Estimated first lactation 305-day or less milk yield of the ith animal

a = Intercept

X_i = First lactation weekly test day milk yield record of ith animal

b_i = Regression coefficient of first lactation 305-day or less milk yield on weekly test day milk yields

The accuracy of fitting the regression model was calculated by coefficient of determination (R^2) using the following formula:

Coefficient of determination R^2=

$$\frac{\text{Sum of squares due to regression}}{\text{Total sum of square}} X100$$

Artificial neural network (ANN)

Artificial neural network was used to predict the first lactation 305-days or less milk yield from weekly test day milk yield records, using SPSS (Statistical Package for the Social Sciences) software version 20.0. ANN model is basically an intelligent data processing system which learns the predictive ability automatically from the data set presented while training the network. The artificial neural network consists of input layer, hidden layer and output layer. Each layer has a specific role in execution of the neural network. In back propagation technique, input vector and the corresponding target vectors are used to train a network until it can approximate a prediction function. The network was trained using back propagation algorithm i.e. scaled conjugate gradient. Network parameters such as learning rate, momentum, and error goal was used as the default setting of the algorithms.

The performance efficiency of the artificial neural network model was calculated using the value of coefficient of determination (R^2-value) and Root Mean Square Error (RMSE) value.

Coefficient of Determination (R^2)

The proportion of the variance in the dependent variable that is predictable from the independent variable(s).

Calculated as

$R^2 = (\text{“}_i (Y_i - \acute{e})^2 \text{“}_i (Y_i - v_i)^2 / \text{“}_i (Y_i - \acute{e})^2)$

Where

Y_i = Observed value

é = Mean of observed values

v_i = Estimated value

Root Mean Square Error (RMSE)

The root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample or population values) predicted by a model and the values observed.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(\hat{Y}_i - Y_i\right)^2}{n}}$$

Where,

v_i = Values predicted by the model.

Y_i = Actual values

n = Number of observations

Comparison of Multiple Linear Regression (MLR) and Artificial Neural Network (ANN)

The comparison of multiple linear regression and artificial neural network for prediction of first lactation 305-days or less milk yield was done based on R^2 value (coefficient of determination) and RMSE (Root Mean Square Error) value. The higher R^2 value and smaller RMSE value denoted a better fit model of prediction.

Results and Discussion

Multiple linear regression

The weekly test day milk yields were used to predict first lactation 305-day milk yield (FL305DMY) by using the multiple linear regression. The backward elimination method was used to find the optimum equation (Table 1). When all the weekly test day records (WTD1 to WTD43) were incorporated in an equation, the R^2-value was 98.80% and RMSE value was 145.25. The reported estimate of R^2-value was very close to the estimate reported by Garcha and Dev (1994) for prediction of lactation milk yield from all the ten monthly test day records in crossbred cattle with 99% R^2-value. On the contrary, Joshi et al. 1996 (93.07%) in Hariana cattle, reported lower estimate of R^2-value for prediction of lactation milk yield from all the ten monthly test day records.

Table 1 Prediction equations for first lactation 305-day or less milk yield on the basis of weekly test day milk yields by backward elimination method

S. NO.	Prediction equations	R^2-Value (%)	RMSE
1	Y= 43.148 + 140.809 WTD1 + 11.819 WTD2 - 25.901 WTD3 + 118.937 WTD4 + 9.356 WTD5 - 142.577 WTD6 + 17.683 WTD7 + 117.676 WTD8 - 145.657 WTD9 + .002 WTD10 - 117.109 WTD11 + 16.023 WTD12 + 52.899 WTD13 + 372.996 WTD14 - 317.096 WTD15 + 181.363 WTD16 + 19.019 WTD17 - 86.921 WTD18 + 34.447 WTD19 - 263.533 WTD20 + 150.524 WTD21 + 47.449 WTD22 + 154.210 WTD23 - 53.880 WTD24 -21.888 WTD25 - 27.284 WTD26 + 93.759 WTD27 + 134.732 WTD28 - 291.026 WTD29 + 325.472 WTD30 + 69.655 WTD31 -85.348 WTD32 + 308.736 WTD33 -90.686 WTD34 - 173.402 WTD35 - 529.390 WTD36 + 529.167 WTD37 - 64.459 WTD38 - 66.411 WTD39 - 49.851 WTD40 - 5.274 WTD41 + 28.088 WTD42 +40.344 WTD43	98.80	145.25
2	Y = 304.194 + 71.944 WTD3 - 56.944 WTD6 + 37.150 WTD7 + 86.935WTD14 - 69.248 WTD15 + 52.926 WTD20 -38.251 WTD23 + 206.351 WTD24	73.20	239.05
3	Y = 382.155+ 67.912 WTD3 - 55.416 WTD6 + 43.092 WTD7 + 87.076 WTD14 - 65.36015 + 202.462 WTD24	72.50	238.64
4	Y = 412.027+ 69.774 WTD3 - 42.371 WTD6 + 67.878 WTD14 + 178.606 WTD24	71.10	241.14
5	Y = 391.959 + 46.932WTD3 + 53.605 WTD14 + 171.223 WTD24	70.10	243.79

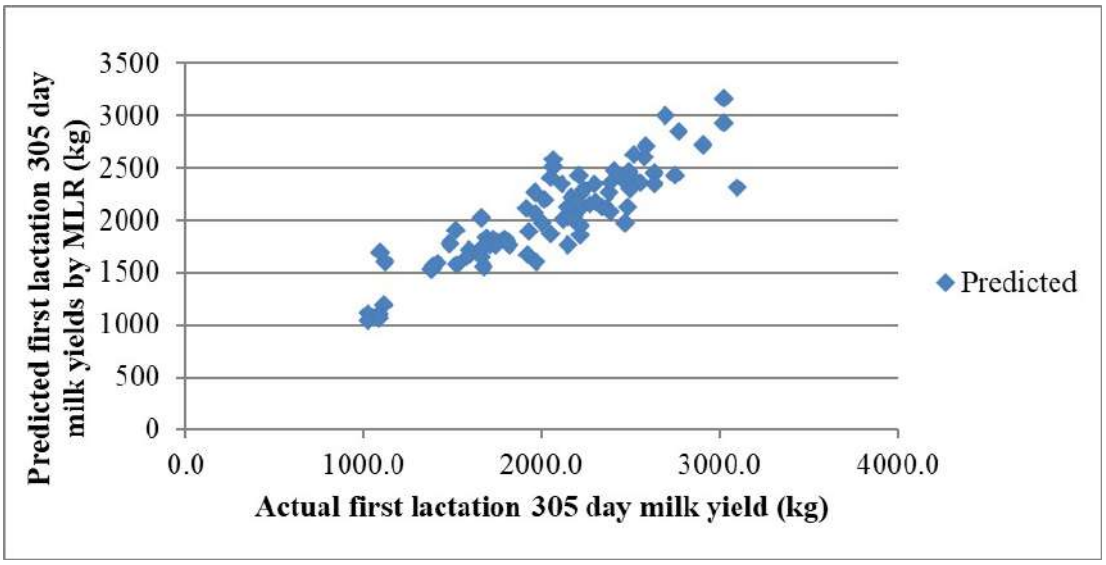

Fig 1. Predicted versus actual FL305DMY by Multiple linear regression in Tharparkar cattle

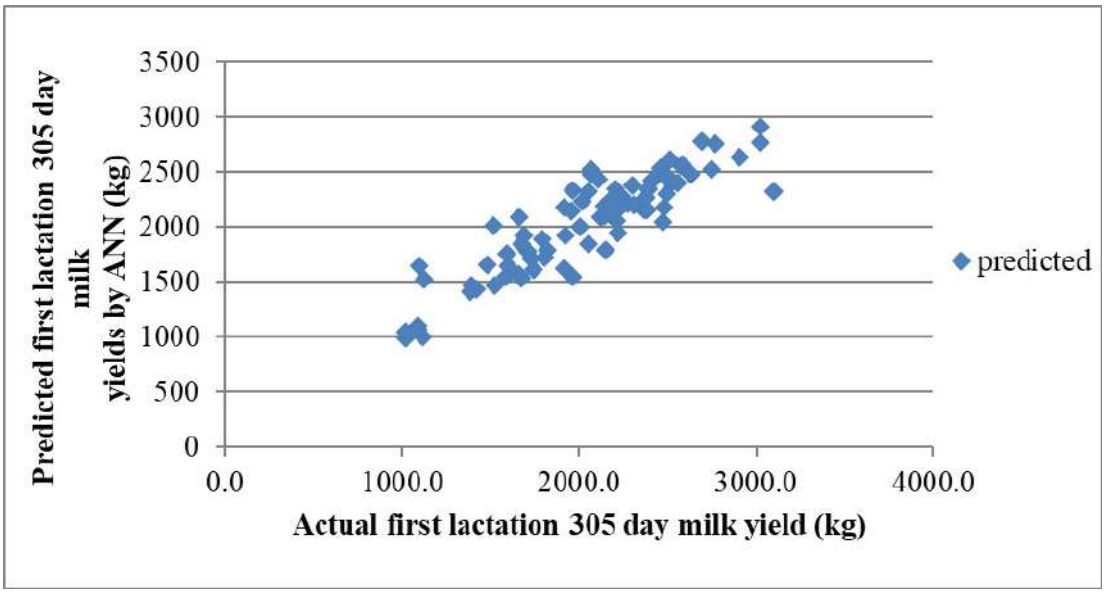

Fig 2. Predicted and actual FL305DMY by Artificial neural network in Tharparkar cattle

The optimum equation had total three variables (test days) viz. WTD3, WTD14 and WTD24. This equation gave an accuracy of prediction of 70.10% and 243.79 RMSE value. Kokate et al. (2014) reported, regression equation with 3 variables viz. BTDY-2, BTDY-3 and BTDY-5 was considered more appropriate for prediction of first lactation 305-day milk yield with 83 % accuracy and 7.54 percent error of prediction in 305-day milk yield in Karan Fries cattle.

Artificial neural network

The ANN was trained by a training data set with three test days, which was incorporated in the best equation from regression analysis. It was observed that the coefficient of determination (R^2) was increasing while the percentage of test data set was decreasing (Table 2). In SET-D (training data–test data: 90–10%), the artificial neural network explained highest (76.70 %) coefficient of determination and lower RMSE value 217.66. Bhosale MD & Singh TP (2015) repoted 85.07% accuracy as early as 126th day of lactation with Bayesian regularization neural network model and it has been also found that R^2 value of the models increases with increase in the number of test-day milk yield records in dairy cattle. Akilli & Hulya (2020) reported 79.18% accuracy with Scaled Conjugant Gradient (SCG) algorithm with decimal scaling normalization technique.The coefficient of determination and RMSE values for different input sets on test data for SET-A, SET-B, SET-C and SET-D have been presented in Table 2.

Comparison of Multiple linear regression and artificial neural network

In the present investigation, significant difference found between multiple linear regression and artificial neural network for predicting FL305DMY in Tharparkar cows. The artificial neural

Table 2 Comparison of R^2 values of different input sets with test data

Training-Test data (%)	Input Sets	R^2-value (%)	RMSE
SETA (66.67- 33.33)	1	73.70	233.12
	2	73.40	232.89
	3	70.20	261.90
SETB (75-25)	1	73.80	232.90
	2	74.10	230.07
	3	74.90	229.21
SETC (80-20)	1	72.60	236.43
	2	70.40	246.00
	3	75.70	222.79
SETD (90-10)	1	73.30	233.26
	2	72.80	239.84
	3	76.70	217.66

network model was found better than multiple linear regression for prediction of FL305DMY in Tharparkar cattle. Similar results were also reported by Sharma et al. (2007) in Karan Fries cattle, Njubi et al. (2010) in Kenyan Holstein- Friesian cattle, Dongre et al. (2012) in Sahiwal cattle, Gorgulu (2012) in Brown Swiss cattle, Bhosale MD & Singh TP (2015), Atil& Akilli (2016) in dairy cattle, Norouzian et al. (2021) in dairy cow and Singh et al. (2022) in Murrah buffalo. The FL305DMY predictions made by the best ANN model and the MLR model developed here are graphically depicted in Figs. 1 and 2, respectively.

Conclusions

The comparison was made between MLR and ANN on the basis of R^2 value and RMSE value. The arterial neural network using scaled conjugant gradient (SCG) algorithm achieved 76.70 % R^2 value and 217.66 RMSE value while in MLR it was found 70.10 % R^2 value and 243.79 RMSE value. Finally, it is concluded that artificial neural networks is better method for prediction of FL305DMY in Tharparkar cows.

Acknowledgement

We gratefully acknowledge the help offered by Dean, College of Veterinary and Animal Science, Bikaner for providing infrastructure and necessary facilities to conduct the research.

References

Akilli A, Hulya A (2020) Evaluation of normalization techniques on neural networks for the prediction of 305-day milk yield. Turk J Agric Eng Res 1: 354-367

Atýl H, Akýllý A (2016) Comparison of artificial neural network and K-means for clustering dairy cattle. Int J Sustain Agric Manag Inform 2 : 40-52

Bhosale MD, Singh TP (2015) Comparative study of feed-forward neuro-computing with multiple linear regression model for milk yield prediction in dairy cattle. Curr Sci 108: 2257-2261

Dongre VB, Gandhi RS, Singh A, Ruhil AP (2012) Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal cattle. Livest Sci 147 : 192-197

Edriss MA, Hosseinnia P, Edrisi M, Rahmani HR, Nilforooshan MA (2008) Prediction of second parity milk performance of dairy cows from first parity information using artificial neural network and multiple linear regression methods. Asian J Anim Vet Adv 3: 222–229

Garcha DS, Dev DS (1994) Number of daughters required to progeny test dairy sires under different sampling schemes. J Dairy Foods Home Sci 13: 113-118

Gorgulu O (2012) Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. S Afr J Anim Sci 42: 280-287

Joshi BK, Tantia MS, Vij PK, Kumar P, Gupta N (1996) Performance of Hariana cows under farmer's herd condition. Indian J Anim Sci 66**:** 383-397

Kannan DS, Gandhi RS (2006) Prediction of lifetime production in Sahiwal cattle. Indian J Anim Sci 9 : 768–769

Kokate LS, Singh A, Banu R, Gandhi RS, Chakravarty AK, Gupta AK, & Sachdeva GK (2014) Prediction of 305-day lactation milk yield based on bimonthly test day values in Karan Fries cattle. Indian J Anim Res 48 : 103-105

Njubi DM, Wakhungu JW, Badamana MS (2010) Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein–Friesian dairy cows. Trop Anim Health Prod 42: 639-644

Norouzian MA, Bayatani H, Vakili Alavijeh M (2021) Comparison of artificial neural networks and multiple linear regression forý prediction of dairy cow locomotion score. Vet Res Forum 12 (1): 33-37

Pindyick RS, Rubinfeld (1991) In: Econometric Models and Economic Forecasts Mc Graw Hill Inc., New York

Rana E, Gupta AK, Singh A, Ruhil AP, Malhotra R, Yousuf S, Ete G (2021) Prediction of first lactation 305-day milk yield based on bimonthly test day milk yield records in Murrah buffaloes. Indian J Anim Res 55 : 486-490

Sharma AK, Sharma RK, Kasana HS (2007) Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. Appl Soft Comput 7: 1112-1120

Singh NP, Dutt T, Usman SM, Baqir M, Tiwari R, & Kumar A (2022) Prediction of first lactation 305 days milk yield using artificial neural network in Murrah buffalo. Indian J Anim Sci 92 : 1116-1120