Indian Journal of Animal Sciences **84** (4): 427–430, April 2014/Article
https://doi.org/10.56093/ijans.v84i4.39848

Comparison of artificial neural network and multiple linear regression for prediction of first lactation milk yield using early body weights in Sahiwal cattle

MANOJ M[1], R S GANDHI[2], RAJA T V[3], A P RUHIL[4], A SINGH[5] and A K GUPTA[6]

National Dairy Research Institute, Karnal, Haryana 132 001 India

Received: 20 October 2013; Accepted: 15 December 2013

ABSTRACT

A comparative study of connectionist network (also known as artificial neural network, ANN)) and multiple regression is made to predict the first lactation 305 days or less milk yield (FL305DMY) from early body weights in Sahiwal cattle. A multilayer feed forward network with back propagation of error learning mechanism was used for prediction. Data collected from 221 Sahiwal heifers were partitioned into 2 data sets namely training data set comprising 75% data to build the neural network model and test data set comprising 25% to test the model. Early body weights, viz. birth weight, body weights at 6, 12, 18, 24, 30 months and weight at first calving were used as input variables and FL305DMY was considered as output variable. The same training and test data sets were used for multiple linear regression analysis (MLR). The prediction efficiency of both models was compared using the R^2 value and root mean square error (RMSE). The accuracy of prediction from both the models was observed to be very low. However, the accuracy of prediction was comparatively higher from ANN model (6.87%) than MLR model (3.32%) for test set of data. The comparatively low value of RMSE and high value of R^2 in case of connectionist network in comparison of MLR model shows that connectionist network model is a better alternative to the conventional regression model.

Key words: Artificial neural networks, Body weights, Milk yield prediction, Regression models, Sahiwal cattle

Multiple linear regression (MLR) models are generally used for prediction. However, this traditional regression analysis does not consider the dependency among explanatory variables and may lead to biased results. The connectionist models, artificial neural networks (ANNs), perform well in function approximation and pattern recognition and may constitute an effective alternative to these traditional models. The neural network consists of a set of processing elements, also known as neurons or nodes whose functionality is loosely based on biological neurons. These units are organized in layers that process the input information and pass it to the following layer. The processing ability of the network is stored in the inter unit connection strengths (or weights) that are obtained through a process of adaptation to a set of training pattern (Haykin 1999). Further, ANN approach is completely different from conventional statistical methods, which need a specified algorithm to be transformed by a computer programme (Grzesiak *et al.* 2003). Few workers used ANN model for prediction of milk yield in various breeds of cattle (Grzesiak *et al.* 2003, Grzesiak *et al.* 2006, Sharma *et al.* 2006, Gandhi *et al.* 2009) and reported that the performance of ANN model was slightly superior to that of conventional regression model. Therefore, the present investigation was undertaken to use the ANNs for prediction of first lactation milk yield of Sahiwal cattle on the basis of body weights at earlier ages and to compare their effectiveness with the traditional multiple regression analysis approach.

MATERIALS AND METHODS

The present investigation was undertaken with the data on body weights of 221 Sahiwal heifers maintained at NDRI Farm from 1970 to 2007. The traits considered were birth weight, body weights at 6, 12, 18, 24, 30 months and weight at first calving.

A multilayer feed forward network with back propagation of error learning mechanism is the most commonly used neural network architecture and have shown excellent results in dealing with functional approximation problems. In back-propagation technique, input vector and the corresponding target vectors are used

Present address: [1]Assistant Professor (manojanchuvilayil @yahoo.co.in), Department of Animal Breeding, Genetics and Biostatistics, College of Veterinary and Animal Sciences, Mannuthy. [2]ADG (ravindersinghgandhi@yahoo.co.in), Animal Science, ICAR, New Delhi. [3]Assistant Professor (sabitharaj2003 @yahoo.com),Department of Animal Genetics and Breeding, Kerala Veterinary and Animal Sciences University, Pookode. [4]Senior Scientist (ruhilap@gmail.com), ASRB, ICAR, New Delhi. [5,6]Principal Scientist (avtar54@gmail.com, guptaak2009 @gmail.com), Division of Animal Breeding.

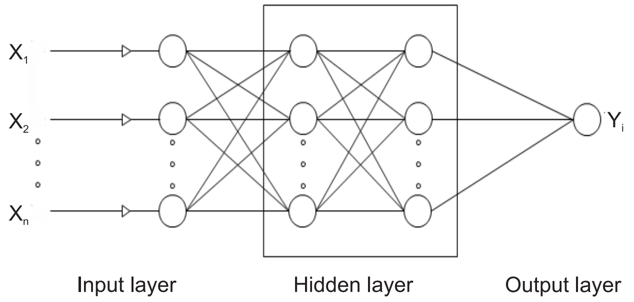

Fig. 1. Neuron connections in a systematic ANN.

to train a network until it can approximate a prediction function. A properly trained network is likely to give reasonable output when presented with new inputs. A general schematic diagram of multilayer feed forward network with input layer, two hidden layers and output layer is shown in Fig. 1.

In the present study, a multilayer feed forward neural network with back propagation of error learning mechanism was developed using Neural Network Toolbox (NNT) of MATLAB 7.0 to predict the body weight. The network had 7 nodes at input layer and one node at output layer for producing the network response. The input and output layer of the network included the variables as shown in Fig. 2.

The whole data set (221) was separated at random in to two subsets viz., the training set consisting of 75% and testing subset comprising 25% of data. The training sets were used to train the neural network models and the testing sets were used to validate the models.

The network was tested with 1 and 2 hidden layers with 3 to 25 neurons in each hidden layer. Initial weights and bias matrix was randomly initialized between -1 and 1. A non-linear transformation (or activation) function *tangent sigmoid* was used to compute the output from summation of weighted inputs of neurons in each hidden layer. A *pure linear* transformation function was used at output layer for getting network response. The designed network was trained in supervisory mode with several variant of back propagation of error learning algorithms like Gradient descent, Widrow Hoff learning rule, conjugate gradient, Quasi-Newton, Levenberg-Marquardt, resilient back propagation and Bayesian regularization to evaluate the performance of ANN models. Among all these algorithms, Bayesian regularization algorithm gave stable and consistent results for the given data. This algorithm has the ability to adjust the network parameters like learning rate, momentum rate, performance goal etc. automatically as per the situation to avoid the problem of over prediction. Therefore, only the results of this algorithm are discussed.

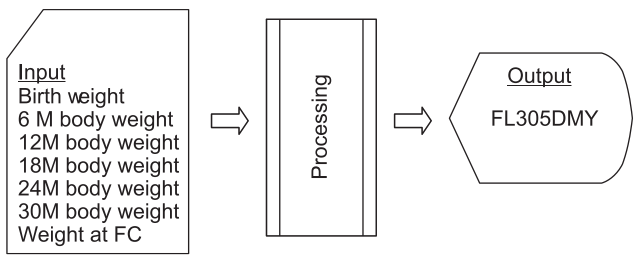

Fig. 2. Input/output variables used in ANN and MLR models.

The network was trained using above mentioned learning algorithms for up to 2000 epochs or till the algorithms truly converged. The input and target data was preprocessed so that the mean is 0 and the standard deviation is 1 using the feature *prestd* available in NNT as per the requirement of algorithms. The parameters like learning rate, performance/ error goal, learning rate increment etc. were used at the default setting of the algorithms in the MATLAB. Most of the time it was observed that algorithms were truly converged which means that performance/error goal was achieved. The network was trained with training data set for a number of times to get consistent results. The prediction performance was tested using a new data set (test data).

To compare the effectiveness of the ANNs for prediction of FL305DMY; the MLR model was developed using the seven body weights as input variables to predict the FL305DMY. The same training data set was used to develop the regression equations and the effectiveness of prediction from MLR model was tested using test data set. The performance of both ANN model and the MLR model were compared using R^2 value and root mean square error (RMSE) as mentioned in Grzesiak *et al.* (2003).

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\left[\sum_{1}^{N}\left(Q_{exp}-Q_{cal}\right)^{2}\right]}$$

where, Q_{exp}, observed value; Q_{cal}, predicted value; N, number of observations.

RESULTS AND DISCUSSION

The multiple linear regression (MLR) and artificial neural network (ANN) models were developed using body weights at different ages as independent variables and the first lactation 305-day or less milk yield as dependent variable. For this purpose the training set was used to build the MLR and ANN models and the test set was used to test the prediction efficiency of the models. The training data set was subjected to MLR and the linear regression equation developed is given below.

$$\mathrm{FL305DMY} = 843.8685+30.3496X_1-7.8723X_2+13.2199\,X_3-6.6147\,X_4+1.1309X_5-0.3574X_6+0.8522X_7$$

The accuracy of prediction from the above model was 11.22% in training data set. However, when the model was applied on test data set it gave an accuracy of 3.32% only (Table 1). The R^2-value of prediction in both the data sets was very low suggesting that the relationship between the predictors and response variable is not linear. Moreover, the correlation of FL305DMY and body weights at various stages were also low which revealed that the early body weights were not good indicators of the FL305DMY.

Aduli *et al.* (1996) reported similar R^2 value (0.141) when used multiple linear regression to predict milk yield from body weights at birth, three, six, nine and twelve months of age in Friesian × Bunaji half breds in Nigeria.

Table 1. Various criteria of judging the effectiveness of MLR and ANN analyses

Number of records (221)	Training data set 166		Test data set 55	
	MLR	ANN	MLR	ANN
RMSE	621.85	623.67	628.79	617.15
R^2- value (%)	11.22	10.70	3.32	6.87
SD ratio	0.9423	0.9450	0.9832	0.9650
Pearson' s correlation coefficient	0.3349**	0.3382**	0.2432	0.2623

** Highly significant (P<0.01).

SD ratio and RMSE value were lower in ANN model than in MLR model in test data set. The regressions of predicted FL305DMY on actual yield predicted by ANN and MLR were plotted and are presented in Fig. 3.

The results of the paired 't' test for comparison amongst actual and predicted FL305DMY using ANN and MLR models revealed statistically non-significant difference (Table 2). However, the predictions based on ANN model gave higher R^2 values with lower RMSE and SD ratio in comparison to MLR in test data set indicating that the former method was comparatively more efficient to predict FL305DMY using body weights in Sahiwal cattle. It was observed that the predicted FL305DMY from ANN model gave the lower standard deviation and CV (%) indicating that fitting of this model smoothened the predicted values to the maximum extent in comparison to MLR model.

Table 2. Actual and predicted milk yield (kg) in test data set

	No.	Maximum	Minimum	Mean	SD	CV%
Actual milk yield	55	2975.5	510.20	1744.28 ^{a}	645.39	37.00
MLR predicted	55	2206.92	744.67	1735.59 ^{a}	260.64	15.02
ANN predicted	55	2123.29	1392.82	1751.83^a	171.90	9.81

Means bearing same superscript did not differ significantly.

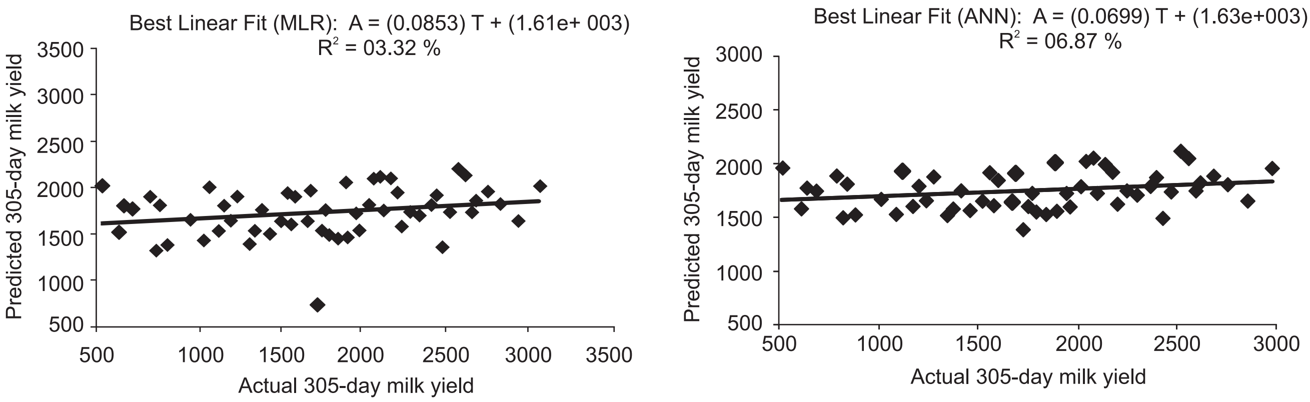

Fig. 3. Regressions of predicted (A) on the actual (T) 305-day milk yield (kg) for test set using MLR (left) and ANN (right) models in Sahiwal cattle.

However, Haile *et al.* (2008) observed higher R^2 value (21.76%) when predicted FLMY from birth weight, weaning weight, body weights at 6 months, 12 months, 18 months, 24 months and age at first calving in Ethiopian Boran breed of cattle and their crosses with Holstein Friesian.

The neural network model was trained using the training data set to predict the FL305DMY and a maximum goal of 99% accuracy was set to be achieved in 2,000 epochs (cycles). Several combinations of hidden layers (1–2 layers) with varying number of neurons (1–25 neurons) were experimented to train the network and the best results was obtained with the combination of 1 hidden layer and 1 neuron in that hidden layer in 51 epochs. Various criteria of judging the effectiveness of MLR and ANN analyses are given in Table 1. The R^2-value of prediction of FL305DMY was 10.70%. However, when the network was trained and simulated with test data set it gave an accuracy of 6.87%, which was higher than that from MLR model (3.32%). The

The accuracy of prediction of FL305DMY using early body weights from both the models was observed to be very low. However, the accuracy of prediction was comparatively higher from ANN model (6.87%) than MLR model (3.32%) for test set of data. There was no significant difference amongst actual and predicted FL305DMY using ANN and MLR models. As the ANN predictions gave higher R^2 values with lower RMSE and SD ratio in comparison to MLR in test data set, it can be interpreted that ANN is comparatively more accurate to predict FL305DMY using body weights in Sahiwal cattle.

REFERENCES

Aduli A E O, Abubakar B Y and Dim N I. 1996. Phenotypic relationship between body weights and subsequent milk production of Friesian × Bunaji crosses. *American Journal of*

Animal Sciences **21**: 274–77.

Gandhi R S, Raja T V, Ruhi A P, Kumar A. 2009. Evolving prediction equations for lifetime milk production using artificial neural network methodology in Sahiwal cattle. *Proceedings of XXXVII Dairy Industry Conference*. 7–9 February, Goa. pp 103.

Grzesiak W, Blaszczyk P and Lacroix R. 2006. Methods of predicting milk yield in dairy cows – predictive capabilities of Wood's lactation curve and Artificial neural networks (ANN). *Computer and Electronics in Agriculture* **54**: 69–83.

Grzesiak W, Lacroixr R, Wojcik and Blaszczyk P. 2003. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Sciences* **83**: 307 – 10.

Haile A, Joshi B K, Ayaleq W, Tegegne A, Singh A and Chakravarty A K. 2008. Prediction of first lactation milk yield of Boran cattle and their crosses with Holstein-Friesian in Central Ethiopia using multiple regression and principal components analysis. *Indian Journal of Animal Sciences* **78** (1): 66–69.

Haykin S. 1999. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Englewood Cliffs, NJ

Sharma A K, Sharma R K and Kasana H S. 2006. Empirical comparisons of feed forward connectionist and conventional regression models for prediction of first lactation 305-day milk yield in Karan Fries dairy cows. *Neural Computer and Application* **15**: 359 – 65.