Indian Journal of Animal Sciences **92** (9): 1116–1120, September 2022/Article
https://doi.org/10.56093/ijans.v92i9.117570

Prediction of first lactation 305 days milk yield using artificial neural network in Murrah buffalo

NRIPENDRA PRATAP SINGH[1]✉, TRIVENI DUTT[1], SHEIKH MOHD USMAN[1], MOHD BAQIR[1], RUPASI TIWARI[1] and AMIT KUMAR[1]

ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh 243 122 India

Received: 2 November 2021; Accepted: 7 July 2022

ABSTRACT

In the present study, first lactation test day and monthly milk records of 301 Murrah buffaloes were used for prediction of first lactation 305-day milk yield (FL305DMY) using artificial neural network (ANN) and was compared with multiple linear regression (MLR). Models were evaluated on the basis of coefficient of determination and root mean square error (RMSE). Two different input sets (Input set-1 and Input set-2) were used in the study. In input set-1, four test day milk yields (6^{th}, 36^{th}, 66^{th} and 96^{th} day of lactation) along with age at first calving (AFC) and peak yield (PY) were taken together and in input set-2, four monthly milk yields record (1^{st}, 2^{nd}, 3^{rd} and 4^{th} month yield) along with AFC and PY were taken together. The ANN was trained using back propagation (BP) algorithm which is also known as Bayesian regularization (BR). ANN achieved highest accuracy of 82% with lowest RMSE value of 16.46% for input set-1 while MLRs accuracy was 80.53% with RMSE value of 17.48%. Higher accuracy and lower RMSE value for ANN clearly showed its better performance than MLR model. Hence, ANN could be alternatively used as a tool for prediction of FL305DMY in Murrah buffaloes using input set-1 with more than 80% accuracy. So, 96^{th} day test day yield (TD4) can be used for prediction of FL305DMY and as a trait for early genetic evaluation of sires.

Keywords: Artificial neural network, FL305DMY, Murrah buffalo, Test day milk yields

The 305-day milk yield is an important basis for selection of cows as well as for genetic evaluation of sires. For timely culling, early identification of low-yielding animals is important with their related economic benefits (Kominakis *et al*. 2002, Njuibi *et al*. 2010). Sooner the bulls are identified for collection of semen, sooner insemination of buffaloes can begin (Sharma *et al*. 2007). Daily milk recording under field conditions is a costly and time-consuming process. So, it is better to use some early recorded test variables for predicting the 305 days lactation milk yield. The first option is to record a part of the lactation (called part-lactation milk yield-monthly and cumulative monthly milk yield) instead of recording a complete lactation and these partial milk yields can be used for predicting the lactation milk yield. The second option is to record milk yield once in a fortnight (fortnightly) or once in a month (monthly) or once in two months (bimonthly) and use these test-day milk yields for predicting the 305 days lactation milk yield (Kokate 2009).

Conventional models like multiple linear regression (MLR), partial least-squares regression, stepwise multiple linear regression, etc. have been extensively used for prediction of various real-life problems. However, these mathematical models have some intrinsic limitations (Kominakis *et al*. 2002) such as they impose restrictions on the number of input variables; over the whole dataset only one dependency function is assumed; other hypotheses enforced by their underlying theories (linearity, normality, data independence, etc.). For instance, a number of variables involved in modeling of various biological processes are not quantitative, and it is difficult to incorporate them into conventional empirical or statistical models; therefore, practical applications and prediction power of the models are limited (Fang *et al*. 2000).

The Artificial Neural Networks (ANNs) are one of the soft computing techniques used mainly for pattern recognition, modeling, and prediction. Neuro-computing paradigm, comparatively a new area of non-linear techniques is gaining popularity as plausible alternatives for solving real-life problems (Fang *et al*. 2000). ANNs are mathematical, algorithmic, software models inspired by biological neural networks (BNNs). ANN is an artificial representation of the human nervous system and is developed from the idea of mimicking the human brain. ANN has basic units called neurons, the design is proposed by its biological analogue. These artificial neurons have input pathways just like dendrites and output pathways just like axons of biological neurons

Present address: [1]ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh. ✉Corresponding author email: nripendrarawat@gmail.com

(Sharma *et al*. 2004). However, in dairying in India, there have been relatively fewer research about application of ANNs. Accordingly, the present investigation was undertaken to predict FL305DMY in Murrah buffaloes using early milk production traits.

MATERIALS AND METHODS

Data collection: The records on the first lactation traits of 301 Murrah buffaloes spread over a period of 18 years (2000-2018), maintained at IVRI, Izatnagar were collected. A total of 1204 test day milk yields, monthly milk yields, age at first calving and peak yields of first lactation records from 301 buffaloes were used in the study. The input variable for prediction of first lactation 305 days milk yield were TD1 (6^{th} day of lactation), TD2 (36^{th} day of lactation), TD3 (66^{th} day of lactation), TD4 (96^{th} day of lactation), M1 (1^{st} month yield), M2 (2^{nd} month milk yield), M3 (3^{rd} month milk yield), M4 (4^{th} milk yield), age at first calving (AFC) and peak yield (PY), respectively. The animals with milk yield of less than 500 kg were excluded from study, considering it abnormal lactation. Buffalo calves were permitted to suckle 1.5 kg of milk since birth to seven months of age. Thus, the amount of milk a buffalo calf consumed up to seven months of age and the pail yield were added for total lactation. Accordingly, correction of data was done in recorded peak yield, test days, monthly milk yield and FL305DMY or less milk yield to get original yields. Whole data was then divided into four main subsets according to training and test data division as subset-A, subset-B, subset-C and subset-D (Table 1). These input variables were further divided into two input sets that were input set-1 (TD1, TD2, TD3, TD4, AFC and PY) and input set-2 (M1, M2, M3, M4, AFC and PY) to predict FL305DMY. Descriptive analyses of data were done using SPSS 20.0 package.

Table 1. Distribution of data into four main subsets

Subset		No. of records	Division of data (%)
Subset-A	Training	200	66.67
	Test	101	33.33
Subset-B	Training	225	75
	Test	76	25
Subset-C	Training	240	80
	Test	61	20
Subset-D	Training	270	90
	Test	31	10

Statistical analysis: ANN model is an intelligent model which learns from the data presented to it while training and processes it to reach a better result with least error. Most commonly a multilayer feed forward network with back propagation is used. It is very efficient in error learning mechanism and shows better result in predictions. An input layer, hidden layer and an output layer are part of a basic fundamental neural network. Each layer in the model has its own important role in the network execution. The input vector and the corresponding target vector are used in back propagation procedure to train the network till it can estimate a prediction function (Fausett 1994).

In the current study, a multilayer feed forward neural network with back propagation of error learning mechanism was developed, using Neural Network Toolbox (NNT) of MATLAB 7.8.0 (MATLAB Users' Guide, R2009a) to predict FL305DMY. The network was trained and simulated using back propagation algorithm, viz. Bayesian regularization (BR) up to 6000 epochs or till the algorithm truly converged, where epoch means a single pass through the sequence of all input vectors. The input and target data were pre-processed so that mean is 0 and the standard deviation is 1 using the prestd feature available in NNT as per requirement of algorithm. Network parameters, learning rate (0.01), momentum (0.05) and error goal (0) were used as default setting of the algorithm. Rest, all the parameters were kept at their default values as enforced by the Neural Network Toolbox under MATLAB (Ghedira *et al*. 2004) used here. ANN plot was developed using RStudio Desktop 1.1.463. Alternatively, the multiple linear regression was used for prediction of 305 days milk yield.

BR minimizes a combination of squared errors and weights, for determining correct combination to produce a network that generalizes well (Foresee and Hagan 1997). In 'BR algorithm', there is no need of regularization for weight-decay method as it possesses inbuilt regularization feature. Weight and bias values in BR network function are updated according to Levenberg-Marquardt optimization. BR uses Jacobian for calculations which assumes that performance is a mean or sum of squared errors. Adjustments of variables are done on the basis of Levenberg–Marquardt algorithm.

Performance evaluation of network: Accuracy of each input set and further its subsets were done using the coefficient of determination (R^2 value) and RMSE technique of the test data. Further, the network was tested with 1 hidden layer keeping 3, 5, 7, and 9 neurons and with 2 hidden layers keeping 2:5, 2:10, 3:5, 3:6, 3:7, 4:7, 4:10, 5:5, 5:7 and 10:5 neurons. Initial weights and bias matrix were randomly initialized between -1 and 1. A non-linear transformation function tangent sigmoid was used to determine the output from the summation of weighed neuron inputs in each hidden layer. For network response, a pure linear transformation function was used on the output layer. The designed network was trained in supervisory mode with Bayesian regularization (BR) variant of back propagation of error learning algorithm.

RESULTS AND DISCUSSION

The mean (± standard error) monthly milk yields for M1, M2, M3 and M4 were 119.45±4.39, 227.31±3.35, 238.13±3.28 and 232.63±52.05 kg, respectively whereas mean (±standard error) test day milk yields for TD1, TD2, TD3 and TD4 were 5.39±0.10, 7.64±0.12, 7.94±0.12 and 7.66±0.11 kg, respectively. It was observed that the milk

Table 2. Optimum equation along with R^2 values developed using MLR and ANN on test data of set-1 and set-2

Training-test data	MLR equations	R^2 value (%)		RMSE (%)	
		MLR	ANN	MLR	ANN
Set-1					
Subset-A (66.67:33.33)	Y = 76.64 - 0.17AFC + 37.71PY + 16.57TD1 + 34.76TD2 + 69.69TD3 + 75.51TD4	72.92	74.90	16.80	16.76
Subset-B (75:25)	Y =1.7.46 - 0.10AFC + 19.38PY + 3.99TD1 + 55.76TD2 + 71.02TD3 + 73.08TD4	76.93	77.57	19.63	18.35
Subset-C (80:20)	Y = 110.09 - 0.13AFC + 27.67PY + 14.36TD1 + 44.56TD2 + 66.17TD3 + 76.67TD4	80.53	82.00	17.48	16.46
Subset-D (90:10)	Y = 99.75 - 0.13AFC + 28.63PY + 18.28TD1 + 45.51TD2 + 64.98TD3 + 75.95TD4	76.30	77.32	16.68	16.86
Set-2					
Subset-A (66.67:33.33)	Y = (-24.69) - 0.11AFC + 21.1PY - 0.02M1 + 2.08M2 + 1.27M3 + 3.94M4	74.71	74.32	17.80	17.92
Subset-B (75:25)	Y = (-2.17) - 0.07AFC + 8.45PY - 0.25M1 + 2.39M2 + 1.19M3 + 4.15M4	76.92	76.97	21.90	21.23
Subset-C (80:20)	Y = (-6.49) - 0.10AFC + 17.06PY - 0.03M1 + 2.02M2 + 0.69M3 + 4.68M4	78.07	78.49	21.40	19.58
Subset-D (90:10)	Y = (-28.21) - 0.09AFC + 16.96PY - 0.01M1 + 2.17M2 + 0.71M3 + 4.56M4	71.42	74.43	17.87	16.44

production increased from first test day till 3rd test day and then declined on 4th test day. Similarly, for monthly yield, milk yield increased till 3rd month and then declined in 4th month. Thus the highest peak yield (10.35±0.124 kg) was witnessed in 3rd month of lactation and mean first lactation milk yield was 1757±27.79 kg in Murrah buffaloes.

Prediction using MLR: The test day milk yields, monthly milk yields, AFC and PY were used to predict first lactation 305-day milk yield (FL305DMY). Prediction equations were developed for both the input sets (set-1 and set-2).

Prediction equation of whole data set for input set-1

Y = 74.16 - 0.11AFC + 25.90PY + 14.48TD1 + 42.84TD2 + 72.68TD3 + 75.90TD4

where Y, Predicted milk yield.

Prediction equation of whole data set for input set-2

Y = (-62.35) - 0.07AFC + 18.66PY + 0.08M1 + 1.91M2 + 1.19M3 + 4.36M4

Further on testing the training equation with test data (Table 2), subset-C (80:20; training-test data) explained highest coefficient of determination (R^2=80.53%) and RMSE (17.48%) in case of input set-1 and also for input set-2, the highest coefficient of determination (78.07%) was obtained in subset-C (80:20; training-test data) with RMSE value 21.40%.

Prediction using ANN: In the present study, it was observed that coefficient of determination (R^2) and RMSE had no regular trend with the decrease in the percentage of test data set (Table 2) for both the input sets. But the study done by Dongre *et al*. (2012) in Sahiwal cows revealed that R^2 increased with decreased test set. The difference may be due to less or no test data in subset-D. In subset-C (training-test data: 80:20) with one hidden layer and 3 neurons, ANN explained highest, coefficient of determination (R^2=82.00%) and lowest root mean square error (RMSE=16.46%) for input set-1 (Table 2). Further, in input set-2, subset-C (training-test data: 80:20) with one hidden layer and 9 neurons, explained highest coefficient of determination (R^2=78.49%) and lowest root mean square error value (RMSE=19.58%) (Table 2). Further, in all the subsets, performance of ANN was better than MLR except in the subset-A of input set-2 where MLR showed slightly better results than ANN. Hence it is clear from the present investigation that prediction could be done using early lactation test day records of up to 96th test day of lactation with better accuracy. The ANN plots were developed for best subset, i.e. subset-C for both input sets (Fig. 1 and Fig. 2).

Comparison of MLR and ANN on test data: In the

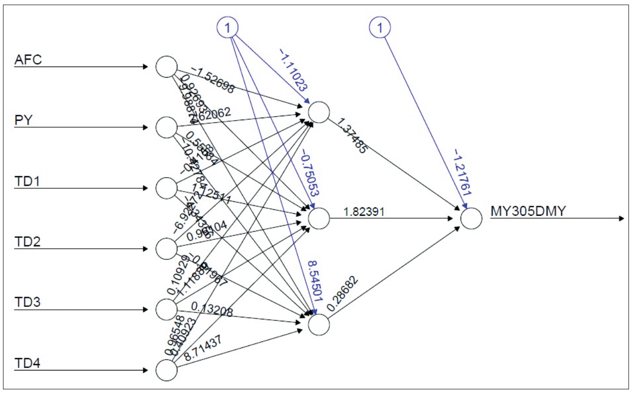

Fig. 1. ANN plot for input set-1 with best accuracy (Subset-C).

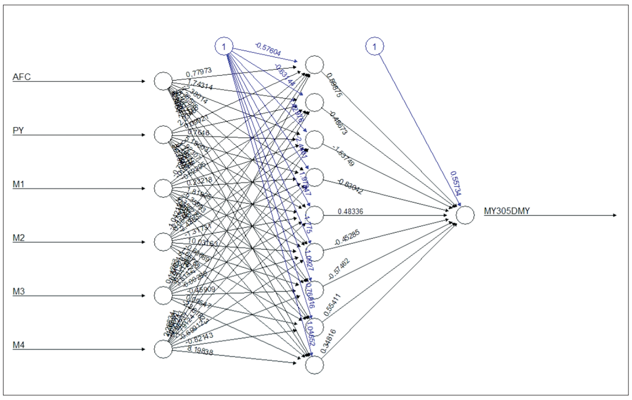

Fig. 2. ANN plot for input set-2 with best accuracy (Subset-C).

Table 3. R^2 and RMSE value at different hidden layers and neurons with BR algorithm

BR Algorithm							
		Input set-1			Input set-2		
Subset	Hidden layer	Neurons	R^2 (%)	RMSE (%)	Neurons	R^2 (%)	RMSE (%)
Subset-A (66.67:33.33)	1	3	74.80	17.16	3	73.02	18.38
	1	5	74.92	16.91	5	71.77	19.35
	1	7	74.93	16.81	7	70.40	19.61
	1	9	74.90	16.76	9	74.32	17.92
	2	3:5	74.45	17.23	2:5	74.56	18.30
	2	5:7	74.34	17.05	3:5	71.22	18.89
	2	5:5	74.22	17.14	5:7	67.56	20.62
	2	10:5	74.47	17.03	5:5	67.33	20.66
Subset-B (75:25)	1	3	77.57	18.35	3	76.60	20.13
	1	5	76.48	17.98	5	76.81	21.20
	1	7	76.42	17.96	7	76.97	21.23
	1	9	72.23	18.66	9	76.96	21.25
	2	2:10	77.02	18.97	3:5	75.64	19.41
	2	3:5	75.68	18.33	5:5	73.57	21.16
	2	3:6	77.24	18.28	4:7	76.12	20.01
	2	3:7	77.27	18.26	4:10	76.18	19.97
Subset-C (80:20)	1	3	82.00	16.46	3	76.88	21.86
	1	5	81.83	16.30	5	78.20	21.90
	1	7	81.75	16.27	7	77.90	21.99
	1	9	81.69	16.26	9	78.49	19.58
	2	3:5	80.20	16.16	3:5	76.87	19.83
	2	5:7	77.63	16.70	5:7	73.48	20.52
	2	5:5	81.15	15.98	5:5	75.15	20.80
	2	10:5	80.34	16.84	10:5	74.99	20.67
Subset-D (90:10)	1	3	77.32	16.86	3	75.04	20.16
	1	5	76.61	15.86	5	74.43	16.44
	1	7	77.29	15.91	7	73.03	17.23
	1	9	77.21	15.93	9	72.90	17.25
	2	3:5	75.84	16.07	3:5	70.26	17.72
	2	5:7	71.33	17.31	5:7	70.45	17.69
	2	5:5	71.30	17.32	5:5	72.75	17.09
	2	10:5	71.81	15.88	10:5	70.38	17.31

present study, ANN model explained higher value of coefficient of determination (R^2) and lower RMSE value as compared to MLR model for predicting FL305DMY in Murrah buffaloes. For input set-1, best artificial neural network algorithm achieved an accuracy of 82.00% and a RMSE value of 16.46%, whereas the multiple linear regression model achieved 80.53% accuracy and a RMSE value of 17.48% (Table 3). Further, for input set-2, best artificial neural network algorithm had an accuracy of 78.49% and a RMSE value of 19.58%, whereas the multiple linear regression model had 78.07% accuracy and a RMSE value of 21.40% (Table 3). All the four subsets ANN performed better with training tests data of 80:20. The FL305DMY predictions for input set-1 and input set-2 by the best ANN model and MLR models were developed. The ANN performs better than MLR in both input sets though the difference is not very high. Similar findings were reported in Sahiwal cattle where the ANN had better accuracy and lower RMSE than MLR models (Gandhi *et al*. 2009, 2010; Mundhe *et al*. 2015). Further, similar findings were reported in crossbreds by Gorgulu (2012) in Brown Swiss, Chaturvedi *et al*. (2013) in Karan Fries and Bhosale and Singh (2015) in Frieswal cattle, respectively. On the contrary, Kumar *et al*. (2019) reported multiple regression model had higher accuracy of prediction for lactation milk yield than artificial neural network in Murrah buffalo.

The artificial neural network performed better than multiple linear regression for prediction of FL305DMY in early stage of lactation. The input data sets of first four test days, AFC and PY had more than 80% accuracy of prediction for FL305DMY as compared to input set of monthly milk yield with AFC and PY. Further, the accuracy can be increased by adding some more early lactation non-linear traits for prediction.

ACKNOWLEDGEMENTS

The authors are thankful to Director, ICAR-IVRI Izatnagar, Bareilly, Uttar Pradesh and CAAST-ACLH project for providing necessary facilities. Authors are also thankful to Dr. A P Ruhil, Principal Scientist, ICAR-National Dairy Research Institute for providing guidance on data analysis.

REFERENCES

Bhosale M D and Singh T P. 2015. Comparative study of feed-forward neuro-computing with multiple linear regression model for milk yield prediction in dairy cattle. *Current Science* **108**(12): 2257–60.

Chaturvedi S, Gupta A, Yadav R and Sharma A K. 2013. Life time milk amount prediction in dairy cows using artificial neural networks. *International Journal of Recent Research and Review* **5**: 1–6.

Dongre V B, Gandhi R S, Singh A and Ruhil A P. 2012. Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal cattle. *Livestock Science* **147**: 192–97.

Fang Q, Hanna M A, Haque E and Spillman C K. 2000. Neural network modeling of energy requirements for size reduction of wheat. *American Society of Agriculture and Biological Engineers* **43**: 947–52.

Fausett L V. 1994. *Fundamentals of Neural Networks: Architectures, Algorithms, and applications*. Prentice Hall, Englewood Cliffs, New Jersey, USA.

Foresee F D and Hagan M T. 1997. *Gauss–Newton approximation to Bayesian regularization*. Proceedings of the IEEE International Joint Conference on Neural Networks, Houston, TX, USA, June 12.

Gandhi R S, Raja T V, Ruhil A P and Kumar A. 2010. Artificial neural network versus multiple regression analysis for prediction of lifetime milk production in Sahiwal cattle. *Journal of Applied Animal Research* **38**: 233–37.

Gandhi R S, Raja Thiruvothur, Ruhil A P and Kumar Amit. 2009. Prediction of lifetime milk production using artificial neural network in Sahiwal cattle. *Indian Journal of Animal Sciences* **79**: 1038–40.

Ghedira H and Bernier M. 2004. *The effect of some internal neural network parameters on SAR texture classification performance*. Proceedings of the IEEE International Geosciences and Remote Sensing Symposium, 6, Anchorage, Alaska, USA, September 20-24.

Gorgulu O. 2012. Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. *South African Journal of Animal Science* **42**(3): 280–87.

Kokate L S. 2009. 'Genetic evaluation Karan Fries sires based on test day milk yield records.' MVSc Thesis, Deemed University, NDRI, Karnal, India.

Kominakis A P, Abas Z, Maltaris I and Rogdakis E. 2002. A preliminary study of the application of artificial neural networks to prediction of milk yield in dairy sheep. *Computer and Electronics in Agriculture* **35**: 35–48.

Kumar V, Chakravarty A K, Magotra A, Patil C S and, Shivahre P R. 2019. Comparative study of ANN and conventional methods in forecasting first lactation milk yield in Murrah buffalo. *Indian Journal of Animal Sciences* **89**(11): 1262–68.

Mundhe U T, Gandhi R S, Das D N, Dongre V B and Gupta A. 2015. Genetic and non-genetic factors affecting monthly part lactation milk yields in Sahiwal cattle. *Indian Journal of Animal Sciences* **85**(5): 517–18.

Njubi D M, Wakhungu J W and Badamana M S. 2010. Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein–Friesian dairy cows. *Tropical Animal Health Production* **42**: 639–44.

Sharma A K, Sharma R K and Kasana H S. 2007. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Applied Soft Computing* **7**: 1112–20.

Sharma R K and Sharma A K. 2004. *Neuro-computing paradigms with application to dairy production,* Lecture Compendium, National Training Programme on Information Resources on Genetics and Documentation Techniques for Livestock Improvement, Centre for Advanced Studies AG&B, Dairy Cattle Breeding Division, NDRI, Karnal, India. 173–178.